EGEE User Forum

Abstracts book

http://indico.cern.ch/conferenceDisplay.py?confId=7247

Table of contents

Abstract ID : 5

Supporting MPI applications on the EGEE Grid

Description of the activity :

This work has been undertaken to improve support for parallel applications using MPI. Such applications are common in many fields of computational science including earth sciences, computational chemistry, astrophysics and climate modelling. Existing users of high-performance clusters are accustomed to MPI support and in many cases it is a pre-requisite for migrating their applications to the Grid. Better MPI support would greatly increase the potential user-base of the EGEE Grid.

Grid added value :

The first phase of the Grid has focussed on very loosely-coupled applications such as those common in high-energy physics. However, there is a significant community of computational scientists who need to run more tightly-coupled applications using standards such as MPI. Most of these scientists currently run their jobs directly on clusters. If they were to migrate to the Grid, they would be able to access multiple clusters with a single sign-on. This would be of particular benefit to existing multi-site collaborations. To date, the lack of good MPI support has been one of the factors preventing greater adoption of the Grid by such scientists. Improvements would open up a range of application areas.

Experience or proposed activity :

It has been technically possible to run MPI jobs on the Grid for some time. However, support has been lacking in these areas: i) standards for advertising availability of MPI libraries; ii) recipes for configuring sites for MPI; iii) clear user instructions for locating MPI sites and submitting jobs; iv) middleware limitations that assume (and force) inflexible methods for submitting MPI jobs that are not acceptable to many sites.

Based on discussion within EGEE and with the int.eu.grid (I2G) project, we have formulated simple solutions that should greatly ease the use of MPI code on the Grid. The approach is for users to submit their jobs wrapped in a script that performs any necessary setup (e.g. compilation), using site-defined environment variables and I2G's MPI-start package to execute their binary using the desired version of MPI. Script templates will be made available for users to customise. We have also produced guidelines for configuring sites to support MPI.

Future evolution :

The solutions we propose are workable and can be rapidly implemented by sites. Even without changes to the core middleware, we will be able to provide users with a sensible methodology for submitting MPI jobs and site admins with a practical recipe for configuring their sites. There are issues that will need to be addressed in the future including methods for selecting custom interconnects and compilers and flexible support for parallel jobs in the WMS.

Primary authors : Dr. CHILDS, Stephen (Trinity College Dublin) ; Dr. LOOMIS, Charles (LAL)

Co-authors : Dr. COGHLAN, Brian (Trinity College Dublin) ; DIJKSTRA, Fokke (University of Groningen) ; KELLER, Rainer (HLRS) ; STORK, Sven (HLRS) ; BORGES, Goncalo (LIP) ; HEYNMANN, Elisa (UAB) ; CAMPOS, Isabel (CSIC) ; GERVASI, Osvaldo (University of Perugia) ; COSTANTINI, Alessandro (University of Perugia) ; STERZEL, Mariusz (Cyfronet)

Presenter : Dr. CHILDS, Stephen (Trinity College Dublin)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : CHILDS, Stephen

Submitted on Thursday 04 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 6

Interactive European Grid: an interoperable infrastructure targetting interactivity, visualization and parallelism

Description of the activity :

The int.eu.grid project (http://www.interactive-grid.eu) aims to deploy and operate a production-level e-Infrastructure for demanding interactive applications with advanced visualization capabilities, and quality support for parallel jobs. The infrastructure is inter-operable with EGEE resources. We shall present in this forum the progress of the project in its first year to address the necessities of users interested in these features

Grid added value :

Deploying and operating an interoperable production-level e-Infrastructure for demanding interactive applications will impact the daily work of researchers. The main features of this scientific initiative are:

* Distributed Parallel (MPI) Interactive Computing and Storage at the Tera level
* User Friendly Access through a Grid Interactive Desktop with powerful visualization
* Supporting Virtual Organizations at all levels: setup, collaborative environment, grid enhancement of applications, execution and monitoring tools, discussion of results.

Experience or proposed activity :

Concerning interactivity, our modified Broker is able to start inmediately short jobs requiring a quick answer on the screen, and in combination with the Migrating Desktop, provides true application steering of the simulation. For this purpose the middleware developpers have enhanced the Migrating Desktop from the Crossgrid project to support glogin and advanced visualization tools like GVid.

From the point of view of parallel jobs the project has developped a self-consistent schema to support OpenMPI jobs through our clusters, and works towards the inter-cluster MPI support using PACX-MPI. The results in this respects will be presented together with the EGEE MPI working group in another presentation of this Forum.

Future evolution :

Our project is complementary to the EGEE project because it complements the batch oriented approximation of the EGEE grid, and at the same time maintaining compatibility with the EGEE infrastructure. There are many applications in different fields which relay on the possibility of having a fast answer from the simulation point of view. Such are for example medical image analysis, which are only foreseeble as a hospital true tool if waiting time does not exceed 30 minutes.

Primary authors : Dr. CAMPOS PLASENCIA, Isabel (Instituto de Fisica de Cantabria CSIC) ; ON BEHALF OF THE INTERACTIVE EUROPEAN GRID, Coordinator Jesus Marco (www.interactive-grid.eu)

Co-authors :

Presenter : Dr. CAMPOS PLASENCIA, Isabel (Instituto de Fisica de Cantabria CSIC)

Track classification : Interactivity and Portals

Contribution type : oral presentation

Submitted by : CAMPOS PLASENCIA, Isabel

Submitted on Wednesday 10 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Interactivity and Portals

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 7

ETICS User feedback and Usability Workshop

Description of the activity :

As ETICS is gaining momentum, its user community is growing. We now have over 15 projects (mainly Grid related) from different grid user communities; for example: gLite developers and integrators, high-level services developers (e.g. GridWay and DILIGENT).

This workshop will provide an opportunity for exchange between the growing ETICS user community and ETICS developers.

Grid added value :

ETICS is powered by Grid middleware, via the NMI/Condor. This is why ETICS can claim that "We're using the Grid to build the Grid".

The Grid allows ETICS users to transparently build and test complex software on a wide range of architecture, operating system and environments, in turn accelerating the ability for software to properly execute in heterogeneous environment like Grids.

Experience or proposed activity :

We are not yet able to run on the EGEE infrastructure, since we're missing an important element: virtualisation. As soon as we are able to ship with a build and test job a VM of the environment required for the job, we'll be able to exploit the full power of the EGEE infrastructure. This is work in progress and an important focus for ETICS work plan for 2007.

Future evolution :

As mentioned above, virtualisation is an important enabler for ETICS's usage of a public Grid such as EGEE. Software built and tested with ETICS require a highly controlled environment in order to ensure build and test reproducibility over time. This is why virtualisation is key for ETICS, such that it provides the required controlled environment no matter what the state of the host computer is in, as long as it can execute the VM.

Primary authors : BEGIN, Marc-elian (CERN)

Co-authors :

Presenter : BEGIN, Marc-elian (CERN)

Track classification : ETICS Users Workshop

Contribution type : --not specified--

Submitted by : BEGIN, Marc-elian

Submitted on Thursday 11 January 2007

Last modified on : Sunday 18 February 2007

Comments :

If possible, we would like to have a double 1.5 hours sessions (3 hours in total).

Thanks,

Meb

Status : SUBMITTED

Track judgments :

Track : ETICS Users Workshop

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 8

Structural Biology in the context of EGEE

Description of the activity :

Electron microscopy (EM) is a crucial technique which allows Structural Biology researchers to characterize macromolecular assemblies in distinct functional states. Image processing in three dimensional EM (3D-EM) is used by a flourishing community (exemplarized by the EU funded 3D-EM NoE) and is characterized by voluminous data and large computing requirements, making this a problem well suited for Grid computing and the EGEE infrastructure.

Grid added value :

There are various steps in the 3D-EM refinement process that may benefit from Grid computing. To start with, large numbers of experimental images need to be averaged. Nowadays, typically tens of thousands of images are used, while future studies may routinely employ millions of images.
Our group has been developing Xmipp, a package for single-particle 3D-EM image processing. Using Xmipp, the classification of 91,000 ribosome projections into 4 classes took more than 2500 CPU hours using the resources of the MareNostrum supercomputer at the Barcelona Supercomputing Center. As few groups will have access to such resources, we propose to use the EGEE infrastructure for Xmipp (ML2D/ML3D), in collaboration with the Network of Excellence in 3D-EM. Enabling widespread adoption of 3D-EM will have a long-term profound impact in our understanding complex biological structures (such as viruses, organelles and macromolecular assemblies) to exploit their biomedical applications.

Experience or proposed activity :

We have adapted our Structural Biology applications for production use over EGEE with the help of the DIANE framework for resource and job management.
To spread knowledge of our solution, CNB is organizing a seminar with wet-lab users (Structural Biology researchers) and developers where we will introduce it and collect their response and feedback to its implementation.
We think that the success of our activity within VO Biomed and NA4 depends on:
- Production level quality of the services running on the Grid: for this reason we need an efficient port of our applications to EGEE and a correct adaptation of the software to this environment. Interaction with the EGEE infrastructure cannot pose any added handicap to users, as this would dissuade potential users from using EGEE.
- We must rely in the availability of the resources of EGEE. Dealing with infrastructure problems external to the code being ported add a heavy burden on application developers.

Future evolution :

Currently, our main concerns with support on the EGEE infrastructure are:
- Better and widespread support for parallel processing (MPI) is needed to improve response time in our applications.
- Data management needs improvements for usability and transparency (e.g. a Grid file system like ELFI).
- EGEE needs to go past its current transition to gLite: we still need to use LCG commands to interact with the Information system to avoid gLite shortcomings detecting resource availability.

Primary authors : Mr. GARCíA, David (CNB/CSIC) ; Mr. CARRERA, Germán (CNB/CSIC) ; Dr. CARAZO, Jose María (CNB/CSIC) ; Dr. VALVERDE, José Ramón (CNB/CSIC)

Co-authors :

Presenter : Mr. CARRERA, Germán (CNB/CSIC)

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : DAVID, García

Submitted on Tuesday 16 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 9

EasyGrid: a job submission system for distributed analysis using grid

Description of the activity :

We automate job submission system to grid farms with EasyGrid, an intermediate layer between grid middleware and user software that provides functionality to perform data and functional parallelism. Users without grid skills were able to use grid. The case studies were hadronic tau decays distributed analysis and neutral pion discrimination using genetic programming algorithms.

Grid added value :

The added value was a user's transparent framework for reliable data gridification (support execution and results' recovery of many copies of the same binary code running independently and at same time in many computers with different data files) and functional gridification (one binary code running distributed in many grid computers at same time).
Data gridification can be used to run Monte Carlo Events generation, raw data analysis, any Root application, or any other generic software.
Functional parallelism is done through a library with several functions to run conventional software on the grid with minor changes in the source code. It provides an efficient and secure communication mechanism to allow data transfer between jobs in different worker nodes. If any node goes down, the master program re-submits the task to another server. For more information see http://www.geocities.com/jamwer2002/gridgeral.pdf

Experience or proposed activity :

Data gridification: The first benchmark was eta(540) reconstruction to test what is the best approach to data distribution. The second benchmark was tau decays to neutral pions. This benchmark selected events over 482 million real events and generated 5 million MC events. The third benchmark was search for anti-deuterons in all events available in BaBar Run 3 (1,500 million events in one week using 250 computers in parallel).
Functional gridification: Genetic programming was used to discriminate reconstruction of real neutral pions from background evolving a mathematical model with maps the variables hyperspace to a real value, an algebraic function of pions kinematics variables. Applying the discriminator to a given pair of gammas, if the discriminate value is bigger than zero, the pair of gammas is deemed to come from pion decay. Otherwise, the pair is deemed to come from another (background) source.
More information see http://www.hep.man.ac.uk/u/jamwer/

Future evolution :

Easygrid has performed its tasks submitting, recovering results, and providing listings for further analysis when something goes wrong. However, data services (store, recover and transfer data) still a bottleneck. Most worker nodes were running distributed analysis (an IO bond application) with 50% IOWAIT, and consequently 50% CPULOAD. This is a severe efficiency problem that needs to be tackled before CERN distributed analysis production starts next year.

Primary authors : Dr. WERNER, James (University of Manchester)

Co-authors :

Presenter : Dr. WERNER, James (University of Manchester)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Dr. WERNER, James

Submitted on Tuesday 16 January 2007

Last modified on : Saturday 10 February 2007

Comments :

The author thanks GridPP and PPARC for funding this project, and the □ BaBar collaboration for granting access to their data.

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 10

Security in a Replicated Metadata Catalogue

Description of the activity :

The gLite-AMGA metadata has been developed by NA4 to provide simple relational metadata access for the EGEE user community. As advanced features, which will be the focus of this presentation, AMGA provides very fine-grained security also in connection with the built-in support for replication and federation of metadata. AMGA is extensively used by the biomedical community to store medical images metadata, digital libraries, in HEP for logging and bookkeeping data and in the climate community.

Grid added value :

The biomedical community intends to deploy a distributed metadata system for medical images consisting of various sites, which range from hospitals to computing centres. Only safe sharing of the highly sensitive metadata as provided in AMGA makes such a scenario possible. Another scenario are digital libraries, which federate copyright protected (meta-)data into a common catalogue.

Experience or proposed activity :

The biomedical and digital libraries have been deployed using a centralized structure already for some time. They now intend to decentralize their activity to increase reliability and scalability without cutting back on security. The deployed systems make
use of the EGEE workload management system in case of the biomed use case to schedule jobs to run over the stored data and both the biomed use case and the digital library one use a file-catalogue and the storage elements.

Future evolution :

Security in a replicate environment is a very complex problem, because it requires the nodes to establish some sort of trusted relationship. We will show how these problems have been tackled, which may be of interest also for other services in a Grid environment.

Primary authors : Dr. KOBLITZ, Birger (CERN-IT) ; Mr. SANTOS, Nuno (EPFL Lausanne)

Co-authors :

Presenter : Mr. SANTOS, Nuno (EPFL Lausanne)

Track classification : Data Management

Contribution type : oral presentation

Submitted by : Dr. KOBLITZ, Birger

Submitted on Tuesday 23 January 2007

Last modified on : Sunday 18 February 2007

Comments :

This is a presentation about the secure replication features of the glite-AMGA metadata catalogue including a short presentation of some use cases. We expect the biomedical and digital library communities to elaborate futher on the way they are currently using AMGA or intend to use this system in the future.

Status : SUBMITTED

Track judgments :

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 11

Experiment Dashboard - monitoring system for the LHC experiments

Description of the activity :

LHC experiments are depending on the distributed EGEE infrastructure for their core activities.
The Experiment Dashboard is a monitoring framework aiming to provide for the LHC experiments the overview of their activities on the EGEE infrastructure with a special emphasis in support for users community.

Grid added value :

Existing monitoring tools are usually focusing onto a specific usage like specific Grid middleware/infrastructures, specific submission tool, etc. The Experiment Dashboard has been built to aggregate the existing monitor infrastructure (from experiment specific software, infrastructure itself, monitoring tools) and provide unified views and information correlation.
Experiment Dashboard is covering different areas of the LHC activities - job processing, data transfer, data publishing. It is deployed for four LHC experiments (CMS, ATLAS, LHCb, ALICE).
Some of the core functionality of the Experiment Dashboard like job monitoring can be used for other virtual organizations.

Experience or proposed activity :

Experiment Dashboard is currently in production and is used by LHC users with different roles for their everyday work.

Future evolution :

The whole monitoring infrastructure can be considerably improved. Very often the error messages
indicating various failures are not clear and do not point to a real problem.
The variety of the local fabrics monitoring systems used by local sites complicates the task of creation of the common framework for aggregation of the monitoring data in the central repository.
Transparent navigation of the monitoring data provided by different monitoring systems is often not possible.

Primary authors : ANDREEVA, Julia (CERN) ; GAIDIOZ, Benjamin (CERN) ; BRITO DA ROCHA, Ricardo (CERN) ; HERRALA, Juha (CERN) ; SAIZ, Pablo (CERN)

Co-authors : CATALIN, Cirstoiu (CERN) ; TSAI, Fu-ming (ASGC) ; CHEN, Tao-sheng (ASGC) ; CHIU, Shih-chun (ASGC)

Presenter : ANDREEVA, Julia (CERN)

Track classification : Grid Monitoring and Accounting

Contribution type : oral presentation

Submitted by : ANDREEVA, Julia

Submitted on Wednesday 24 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Grid Monitoring and Accounting

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 12

A Grid application to ensemble ocean forecasting

Description of the activity :

The Istituto Nazionale Geofisica e Vulcanologia runs an operational oceanographic system called Mediterranean Forecasting System.
The long term goal of this project is to collect high quality data and provide reliable environmental predictions for the Mediterranean Sea. End user applications of ocean forecast span from contaminant drift predictions to search and rescue operations.
An ensemble forecast system is currently under investigation to provide estimates of forecast uncertainties.

Grid added value :

The subject of this study is the understanding and the demonstration of the usage of the Grid system in ensemble ocean forecasting.
Short term ocean forecast are externally driven by atmospheric forecasts. The MFS system collects ECMWF atmospheric forecast daily. To be effective an ocean ensemble forecast must be run within the time constraints imposed by the availability of the ECMWF products. To test the potential of the Grid for this purpose, the time window allowed for a 10 days ocean forecast with an ensemble of 1000 forecasts was set to 6 hours.
This experiment was run in collaboration with the Istituto Nazionale di Fisica Nucleare on the INFN Grid. This work demonstrated that an extremely large ocean ensemble forecast, that would be unfeasible on most of the other computer infrastructures, could be run on a Grid system under operational forecasting working conditions and normal Grid configuration.

Experience or proposed activity :

Two techniques for generating an ocean ensembles were successfully tested on the Grid.
The first approach is based on a random perturbation of the initial conditions of the ocean forecast. This is highly efficient because all the members are initialized using the same input files, only the seed numbers needed to be transferred to the Grid Working Nodes. The results show that a minimum number of 450 members were successfully accomplished in 5 hours. This result was achieved using 15 Computing Elements and imposing a simple requirement policy to the Grid software manager.
The second approach relies on the perturbation of the wind forcing. All the ocean members start from the same initial condition but are driven by different wind realizations. This significantly increase the amount of data that must be transferred through the network. Preliminary results showed that the Grid system was able to sustain this working load for a 100 members ensemble set up.

Future evolution :

So far every member of the ocean ensemble was run on a single Working Node. This represents a limitation to the size of the problem that can be treated.
While a natural evolution to this problem will be apply a Message Passage Interface a second strategy will also be tested. Since many events that would benefit from an ocean ensemble forecast are localized in a small portions of the model domain, an high-resolution limited-domain ocean model might be the best compromise for the Grid environment.

Primary authors : Mr. BONAZZI, Alessandro (INGV - Italy)

Co-authors : Prof. PINARDI, Nadia (Universita degli Studi di Bolgna) ; Dr. NAVARRA, Antonio (INGV - Italy) ; Dr. GHISELLI, Antonia (INFN - Italy) ; Mr. SCOCCIMARRO, Enrico (INGV - Italy) ; Mr. VERONESI, Paolo (INFN - Italy)

Presenter : Mr. BONAZZI, Alessandro (INGV - Italy)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Mr. BONAZZI, Alessandro

Submitted on Wednesday 24 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 13

Integration of the ATLAS VOMS system with the ATLAS Metadata Interface

Description of the activity :

The ATLAS collaboration has 1700 members from 144 institutions and 33 countries. Authorisation plays a key role in the process of gaining access to resources in a computational grid. AMI is an application which stores and allows access to dataset metadata for the ATLAS experiment. This abstract will summarise how the relationship between the ATLAS VOMS system and AMI was defined and implemented.

Grid added value :

The key to success of Grid computing is the development of the 'middleware', the software that organises and integrates the disparate computational facilities belonging to the Grid. Its main role is to automate all the machine-to-machine negotiations required to interlace the computing and storage resources and the network into a single, seamless computational fabric.

A suite of tools are available to assist with authorisation on the grid and VOMS authorisation in particular. The AMI-VOMS solution which is the focus of this abstract provides a generic solution for the delegation of a VOMS proxy certificate from any VOMS server to a web-based Grid application. It solves the current problem that browsers cannot handle proxy certificates and therefore provides a mechanism for any Grid application that uses a browser client to be capable of VOMS authorisation. This delegation mechanism also allows the Grid application to access other external Grid applications on behalf of the user.

Experience or proposed activity :

ATLAS uses the EGEE infrastructure. The wider ATLAS grid activity with regards authorisation is
towards the use of VOMS. Currently all ATLAS VO users have the same rights and capabilities. The fully deployed VOMS system will give additional capabilities to the management of VOs and users. This transition depends on users being divided in groups inside the VO and holding different roles. In this way, some users will be able to access resources and perform actions that others will not. The success of the AMI-VOMS solution therefore depends on the VOMS system being fully deployed within ATLAS.
In the future the method used within the solution may become obsolete if either VOMS servers are redesigned to administer VOMS credentials in a different manner, or browser support for proxy certificates improves; both of which are quite likely.

Future evolution :

The AMI-VOMS solution, solves two problems for Grid Applications that are browser based, namely delegation and authorisation of VOMS proxy certificates. This is not currently handled in the EGEE infrastructure. It is natural therefore for this solution to, due to its generic nature, extend its applicability to the field of industrial (browser-based) Grid applications whenever the access to resources is governed by VOMS authorisation.

Primary authors : Mr. DOHERTY, Thomas (University of Glasgow)

Co-authors : Dr. ALBRAND, Solveig (IN2P3) ; Mr. FULCAHIER, Jerome (IN2P3) ; Mr. LAMBERT, Fabian (In2P3)

Presenter : Mr. DOHERTY, Thomas (University of Glasgow)

Track classification : Data Management

Contribution type : oral presentation

Submitted by : Mr. DOHERTY, Thomas

Submitted on Thursday 25 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 14

Secure Storage

Description of the activity :

The scientific and technical community using the EGEE infrastructure and involved in the Secure Storage project is composed by a public research centre, INFN, and a private company, UNICO S.R.L. (http://www.unicosrl.it/).
The aim of the activity is to design a secure storage service. This means to create a mechanism to store in a secure way and in an encrypted format the data deployed on the grid storage elements. This stored data will be accessible and readable only by their owners.

Grid added value :

One of the main benefit of the Grid Infrastructure is the possibility to use distributed storage space. A community could want to use storage elements owned by an external organization to delegate the management of this machines and to avoid to buy specialized hardware. In this way the community could rent the storage space as needed and minimize human and hardware costs.
In the case of confidential data this scenario is not feasible. Indeed, the community should satisfy strongly privacy requirements, as in the case, for example, it have to manage medical or financial data. To store the confidential data in a storage element managed by an external organization a mechanism to prevent the administrator of the machine accessing the data is required. This is the "insider abuse" problem and the Secure Storage project provides a solution to this problem.

Experience or proposed activity :

A secure version of some lcg-utils commands and a keystore service has been developed:
lcg-scr: encrypts a file and uploads it on a storage element, registering its Logical File Name in a LFC catalog. Moreover, it stores the key used to encrypt the file in the keystore. An ACL will be associated to each key on the repository. This ACL will contain all users authorized to access the file.
lcg-scp: downloads an encrypted file, gets the key to decrypt the file from the keystore, decrypts the file and then store it on a local file-system.
The keystore service stores the key and the associated ACL received by the lcg-scr commands on its repository and provides the key to the lcg-scp command. The communications between the commands and the keystore is established on a secure GSI authenticated channel. The keystore provides the key to the lcg-scp command only if the request is coming from an authorized user (thanks to the GSI authentication, it knows the distinguished name of the user).

Future evolution :

The main issues encountered in using the EGEE infrastructure during the development of the secure storage service are the following:
We cannot use the last version (and then more secure) of the OpenSSL library for a library conflict with OpenSSL version used by Globus.
The development of a GSI Client in C language was been hard. The Globus GSI API are not very intuitive.

Primary authors : Dr. SCARDACI, Diego (INFN Catania) ; Mr. SCUDERI, Giordano (UNICO S.R.L.) ; Dr. CALANDUCCI, Tony (INFN Catania)

Co-authors :

Presenter : Dr. SCARDACI, Diego (INFN Catania)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : Dr. SCARDACI, Diego

Submitted on Friday 26 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 15

Grid systems' capacity metrics from a user's point of view

Description of the activity :

The activity we propose involves benchmarking of grid resources, as they are effectively available for all users. We do this in order to measure real grid characteristics, which serve as a proof that metrics-guided resource selection is nearly imperative, if not to optimally select resources, at least to specifically avoid ones which are known a priori that they don't perform as good as necessary. Our results hint in favour of a more intelligent matchmaking process which involves metrics.

Grid added value :

The potential community that benefits is in effect all grid users, since optimization of the system as a whole, can lead to direct and indirect advantages for everyone, in terms of total AND individual job throughput. What we want specifically to demonstrate is, that by skipping benchmarking and resource characterization, enormous amount of grid resources can be waisted or sub-optimally exploited. For example, the systems that are best on floating point of a given algorithm, say 64bit operations, are not the ones that are optimal on memory transfers, and vice versa. The results are conclusive in demonstrating that the current GIIS-based scheme is, at best, incomplete.
We started our activity within SEE VO, then verified the situation also within ATLAS VO, and expect that if operations' teams (dteam, ops) align and perform similar metrics, the same will be proven for the system as a whole. Currently this is not possible from our side, because it requires operators' and/or VO approval.

Experience or proposed activity :

As explained earlier, the metrics service can be deployed within the context of a VO, in order to benefit the users of that particular realm directly and in their own discretion or, it could be provided as an integrated servide within the RB/WMS mechanisms that can now make resource selection with more detailed and accurate algorithms. The basic framework for doing the first part is already available as a python code package, which is able to submit a self-compiling lmbench source and some related scripts that gather other system information -software & hardware-, and collect their reports. In order to make the results technically correct some statistical validation is necessary.

What is very important to specifically clarify, is that the benefits of applying the benchmarking technique and resource characterization can greatly outnumber the measurement system's overhead in itself (typically less than 0.5% of site's capacity).

Future evolution :

To begin with, we have been able to make our first experiments with the technology, but not to implement it as a regular service, say in the ATLAS VO, because we suspect that we might be colliding with the VO's AUP. We do think that the service is useful though and should be expanded across multiple VOs.
It is imperative to note that resource characterization makes sense, as long as sysadmins keep the type of resources within a single queue homogeneous (ie. similar systems across one queue).

Primary authors : Mr. GEORGATOS, Fotis (University of Cyprus) ; Mr. KOUVAKIS, Ioannis (University of the Aegean) ; Mr. KOYRETIS, Ioannis (National Technical University of Athens)

Co-authors :

Presenter : Mr. GEORGATOS, Fotis (University of Cyprus) ; Mr. KOUVAKIS, Ioannis (University of the Aegean) ; Mr. KOYRETIS, Ioannis (National Technical University of Athens)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : GEORGATOS, Fotis

Submitted on Sunday 28 January 2007

Last modified on : Saturday 10 February 2007

Comments :

We have multiple measurement tables that would be of interest to the audience, with real-data taken by lmbench at various sites (memory bandwith, file open/close times, integer and floating point speed etc) and other system related information. User forum is a great opportunity to demonstrate to people that metrics are very important, because the related community is present there.

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 16

Interconnection & Interoperability of Grids between Europe and China the EUChinaGRID Project

Description of the activity :

EUChinaGRID provides specific support actions to foster the integration and interoperability of the Grid infrastructures in Europe (EGEE) and China (CNGrid) for the benefit of eScience applications and worldwide Grid initiatives. The project studies and supports the extension of a pilot intercontinental infrastructure using the EGEE-supported applications and promotes the migration of new application in Astroparticle Physics (ARGO-YBJ) and Biology (Never Born Proteins).

Grid added value :

Funded by European Union, EUChinaGRID Project officially started on 1st January 2006 with the aim to interconnect the existing European and Chinese Grid Infrastructures and enable their interoperability, thus creating a network of collaboration between Europe and China.

Experience or proposed activity :

EUChinaGRID has started two relevant technical activities:
- Interoperability of gLite with GOS Middleware
- Compatibility of Grid Middleware with IPv6 and interoperability with IPv4.

Future evolution :

The studies on IPv6 compatibility have shown that most of the gLite code is not compliant and needs to be rewritten following the rules of independency from the IP stack.
Comparison of gLite and GOS services lead to the proposal of a gateway that will allow job submission across the two infrastructures of EGEE and CNGRID.

Primary authors : Dr. RUGGIERI, Federico (INFN)

Co-authors :

Presenter : Dr. RUGGIERI, Federico (INFN)

Track classification : Related Projects

Contribution type : oral presentation

Submitted by : Dr. RUGGIERI, Federico

Submitted on Sunday 28 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Related Projects
Judgment :

Judged by :
Date :
Comments : ""

Abstract ID : 17

A GRID infrastructure to promote cooperation across the Mediterranean - The EUMEDGRID Project

Description of the activity :

The interest of the EUMEDGRID experience does not restrict to scientific issues although the opportunity to port applications of regional importance, such as the hydro-geological and medical ones, on the pilot infrastructure is not less important than fostering Grid awareness. The growth of new competences among our Neighbors' scientific community is a concrete initiative towards bridging digital divide and, moreover, to promote a peaceful and effective collaboration among all Partners.

Grid added value :

Funded by EU within the 6th FP for Research and Development and Coordinated by INFN, EUMEDGRID aims to support the development of EGEE-compliant Grid eInfrastructure in the Mediterranean Area and promote porting of new applications on the Grid platform, thus allowing Mediterranean scientist to collaborate closely with their European colleagues.
Conceived as the extension of EGEE in the region, EUMEDGRID is disseminating Grid awareness and competences across the Mediterranean and, in the meanwhile, identifying new research groups to be involved in the project, helping them to exploit Grids' enormous potential to improve their own applications.

Experience or proposed activity :

The results of the first year of activity are mainly related to:
- the the progess made by many countries in the deployment of Grid Sites which joined the pilot infrastructure;
- the advanced process of creation Certification Authorities and National Grid Initiatives;
- the deployment of existing EGEE supported applications in the fields of Physics (ATLAS), Biomedical (WISDOM);
- the testing and porting of new regional applications in Hydrology.

Future evolution :

EUMEDGRID is expanding the EGEE infrastructure to the Mediterranean Area, and will need in the future a ROC and related services dedicated to the specific geographical environment.

Primary authors : Dr. RUGGIERI, Federico (INFN)

Co-authors :

Presenter : Dr. RUGGIERI, Federico (INFN)

Track classification : Related Projects

Contribution type : oral presentation

Submitted by : Dr. RUGGIERI, Federico

Submitted on Sunday 28 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Related Projects

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 18

Fault Detection and Diagnosis from the Logging and Bookkeping Data

Description of the activity :

Autonomic Computing (AC) is defined as "computing systems that manage themselves in accordance with high-level objectives from humans" [1]. AC is now a well-established scientific domain, and a priority for industry. Automated detection, diagnosis, and ultimately management, of software/hardware problems define autonomic dependability. The paper reports on applying state of the art autonomic dependability methods to the Logging and Bookeeping data, with promising results on detection.

Grid added value :

The production status and integration level reached by the EGEE middleware and monitoring provide immense datasets. These are challenging targets for the Machine Learning (ML) community, whose techniques are at the base of AC. The fundamental motivation for this interest is the complexity of the hardware/software components, and the intricacy of their interactions, which defeat attempts to build models only from a-priori knowledge. Furthermore, EGEE is not a steady-state system, not only because it is yet ramping-up, but more profoundly because of the externally-driven collective behaviour of its users.

EGEE monitoring data exemplify to the extreme two classical issues in ML: 1) curse of dimensionality (state space exponential in the number of variables); and 2) data sparsity, most of the state-action space being actually unexplored. EGEE data offer an extra complexity, not addressed in this paper: integration of heterogeneous sources of information (ontology building).

Experience or proposed activity :

L&B records are oriented towards operational semantics: each service logs its own vision of the job information and status. The result is a very large amount of highly redundant data, with in many cases no a-priori syntax or semantics (blobs in the long_fields table). We have first developed a software suite that segments the data in order to discover the basic attributes, and cautiously filter out redundant information. The software also allows to convert the categorical data into a boolean description, convenient for many off-the-shelf mining and learning software. The next step was analysis. Elementary methods (independently scoring attributes) provided little information. The ROGER algorithm [2] developed in our lab provides a good predictor, which can be interpreted through sensitivity analysis. On-going work deals with intelligent clustering in order to reduce dimensionality (frequent itemsets) and learning non-linear models, which can detect compound failure conditions.

Future evolution :

This paper is an attempt to explore the performance and limits of this purely passive approach of detection and diagnosis: a minimally invasive failure analyser would be based solely on analysis of the production (L&B) records of the production jobs. A more intrusive approach would use SAM (or equivalent software) and active learning methods to (approximately) design an optimal probe set and infer the system state [3]. Support for gathering and interpreting data in this area would be required.

Primary authors : Prof. GERMAIN-RENAUD, Cecile (LRI and LAL (CNRS - University Paris-Sud 11)) ; Ms. ZHANG, Xiangliang (LRI) ; Dr. SEBAG, Michèle (LRI (CNRS) and INRIA) ; Dr. LOOMIS, Charles (LAL)

Co-authors :

Presenter : Prof. GERMAIN-RENAUD, Cecile (LRI and LAL (CNRS - University Paris-Sud 11))

Track classification : Grid Monitoring and Accounting

Contribution type : oral presentation

Submitted by : Prof. GERMAIN-RENAUD, Cecile

Submitted on Sunday 28 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Comment 1: There is no field for the references
[1] Kephart, J. O. and Chess, D. M. 2003. The Vision of Autonomic
Computing. Computer 36 :1. 2001
[2] Sebag M., Lucas N., Azé J. ROC-based Evolutionary Learning:
Application to Medical Data Mining
Procs the 6th Int. Conf. on Artificial Evolution. Springer Verlag LNCS 2936
pp 384-396. 2003
[3] Rish I. et al. Adaptive Diagnosis in Distributed Systems. IEEE
Transactions on Neural Networks (special issue on Adaptive Learning
Systems in Communication Networks). 16:5. 2005.

Comment 2:
The order of the author list should be Zhang, Sebag, Germain, Loomis

Status : SUBMITTED

Track judgments :

Track : Grid Monitoring and Accounting

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 19

Using the Grid - the BaBar Perspective.

Description of the activity :

The BaBar experiment is a high-energy physics collaboration based at SLAC, California, USA, with 570 members from 10 countries investigating the properties of anti-matter. The Grid is used for simulating the experiment, reconstructing and processing the experimental results, and analysing the results for publication.

Grid added value :

In 2006, the experiment required 1.4 Petabytes of storage and 5500 KSI2K of computing power to simulate, process, reconstruct and analyse our large dataset with low latency and high efficiency. For the last two years we have investigated the use of the Grid as an alternative resource to achieve our current and future goals. The main goals have been a need to double are processing requirements while keeping manpower at current or reduced levels.

Experience or proposed activity :

In this contribution, we discuss the experience of converting pre-Grid computing models to using the Grid; the relative costs and benefits of the Grid to the three core tasks of simulation, central reconstruction and user analysis; and the ability of the Grid to meet the time-critical needs of an experiment that runs 24/7. Key to the success of the project is high throughput, high reliability, high efficiency, low latency, good monitoring and reporting, and guaranteed future resource use. We compare the Grid and classic batch systems against these key metrics.

Future evolution :

The Egee infrastructure is not as reliable as current batch systems. The instablility of key elements such as resource brokers and VOMS, the incompatibilities between software releases, constantly changing data access policies, poor error reporting, lack of user documentation, the disconnect betwen the user and the support, and the general high learning curve all contribute to difficulties for the experts and failure to motivate users.

Primary authors : Dr. WILSON, Fergus (CCLRC) ; Prof. BARLOW, Roger (Manchester University) ; Dr. BREW, Chris (CCLRC) ; Dr. BAILEY, David (Manchester University)

Co-authors :

Presenter : Dr. WILSON, Fergus (CCLRC)

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : Dr. WILSON, Fergus

Submitted on Monday 29 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 20

Archaeological Applications on e-Infrastructures by ArchaeoGRID

Description of the activity :

ArchaeoGRID enables the possibility to exploit advanced grid computational and storage technologies in archaeology for the analysis of the emerging large region archaeology. The advantage of multidisciplinary research,"specialists work alone using appropriate techniques", and of interdisciplinary research, "specialists cooperate and discover new aspects of their
data", are combined and enabled by ArchaeoGRID. ArchaeoGRID community uses methods of
analysis and data from other grids.

Grid added value :

ArchaeoGRID is not only grid applications. Some of ArchaeoGRID applications could be deployed on an e- Infrastructure only if this structure will satisfy specific needs of the archaeological research and of the use of the archaeological knowledge.
In fact, the middleware DILIGENT (or other Digital Library middleware) must be present on the e- Infrastructures, integrated at some level with g- Lite, for the final narration by production of a digital document; a Visualization Laboratory, with the most advanced Virtual Reality and Augmented Reality technologies available, must be accessible; the archaeological geospatial -
temporal data must be accessible and manageable by means of some GRID GIS; the grid interoperability must be ensured to operate with the other grid; last but not least, through the e- Infrastructure must be possible design and access to large repositories of the archaeological
information .

Experience or proposed activity :

The research project on the origin of the City in the Mediterranean Region between the XVIII Centuries B.C., has been approved by EGEE- II Committee and a paleoclimatic application is running on GILDA t-Infrastructure. Paleoclimate simulation is made using the MM5 Mesoscale Model. MM5 package is installed and is running on GILDA t-Infrastructure. This topic can be considered as a natural application of ArchaeoGRID.
In fact the protohistorical urbanisation process can be considered as the starting point of the Mediterranena and European Civilisations
Paleoclimate evaluation is relevant for evaluating:agricultural and pastoral activities; forests amount and distribution ; exchange and trade; environment ; habitation and dressing;
epidemiological situation, etc.
With MM5 seasonal weather will run to derive information about the climate of that period and how that compares to the present climate. Series of daily weather are produced by Stochastic
Weather Generator and used for hydrology

Future evolution :

Future evolution of ArchaeoGRID include the prosecution of programs testing on t-Infrastructure GILDA and the integration of the ArchaeoGRID system on e-Infrastructure EUMEDGrid for running archaeological applications.

Primary authors : Dr. PELFER, Giuliano (University of Florence - CSDC) ; Dr. CECCHINI, Roberto (INFN) ; Prof. PELFER, Pier Giovanni (Dept. Physics, University of Florence and INFN) ; Dr. POLITI, Antonio (ISC-CNR, Firenze)

Co-authors :

Presenter : Prof. PELFER, Pier Giovanni (Dept. Physics, University of Florence and INFN)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Dr. PELFER, Giuliano

Submitted on Monday 29 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 21

Porting ATLAS Trigger & Data Acquisition System to Run on the Grid

Description of the activity :

The ATLAS Trigger & Data Acquisition System has been designed to use more than 2000 CPUs.

Grid added value :

During the current development stage it is crucial to test the system on a number of CPUs of similar scale. A dedicated farm of this size is difficult to find, and can only be made available for short periods. On the other hand many large farms have become available recently as part of computing grids, leading to the idea of using them to test the TDAQ.

Experience or proposed activity :

However the task of adapting the TDAQ to run on the Grid is not trivial, as the TDAQ system requires full access to the computing resources it runs on and real-time interaction. Moreover the Grid virtualises the resources to present a common interface to the user. We will describe the implementation and first tests of a scheme that resolves these issues using a pilot job mechanism. The Tier2 cluster in Manchester was successfully used to run a full TDAQ system on 400 nodes using this implementation.

Future evolution :

Apart from the tests described above, this scheme also has great potential for other applications, like running Grid remote farms to perform detector calibration and monitoring in real-time, and automatic nightly testing of the TDAQ.

Primary authors : Mr. GARITAONANDIA ELEJABARRIETA, Hegoi (Instituto de Fisica de Altas Energias (IFAE)) ; Dr. FORTI, Alessandra (University of Manchester) ; Dr. WHEELER, Sarah (CERN) ; Dr. THORSTEN, Wengler (University of Manchester)

Co-authors :

Presenter : Mr. GARITAONANDIA ELEJABARRIETA, Hegoi (Instituto de Fisica de Altas Energias (IFAE))

Track classification : Interactivity and Portals

Contribution type : oral presentation

Submitted by : GARITAONANDIA ELEJABARRIETA, Hegoi

Submitted on Monday 29 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Interactivity and Portals

Judgment :

Judged by :

Date :
Comments : ""

Abstract ID : 22

Accessing Grid-based Data Via a Web Browser

Description of the activity :

Researchers in the life sciences need easy but secure access to grid-based data from their work stations. Their working environments are highly variable (different OSes, tools, etc) and include platforms not typically supported by gLite. Having secure access to grid-based data via a web browser facilitates access to data without the need to install grid client software. Equally, other communities could benefit such a solution to browse data, log files, or reports residing in the grid.

Grid added value :

The primary benefit of this development is making grid-based data directly and universally accessible, hence lowering the barriers for grid technology adoption. Adding the https protocol to grid storage makes access to data easy from a variety of existing languages and tools (web browsers included) without needing to install any grid client software. The direct access to the storage avoids needing to copy the data to an intermediate server as is typically done for web portals.

Experience or proposed activity :

This project developed a system to access the grid-based data with a web-browser. The system consists of two components. First we developed a trivial web interface to the LHC File Catalog (LFC). This allows users to browse logical file names and recover the transport URL of files to be accessed. Second and most important we have added the https protocol to the DPM (Disk Pool Manager) implementation of SRM (Storage Resource Manager). We adapted the GridSite Apache module (http://www.gridsite.org/) to perform authorization callouts to DPM. DPM facilitates the addition of new protocols and requires only an http server for each DPM disk server running the modified GridSite module and a minor configuration change on the DPM server.

Future evolution :

The current prototype allows https access to grid-based data and provides tangible benefits to grid users. The prototype uses standard grid certificates allowing authorization based only on the user's identity (DN). It will be extended to use VOMS proxies to allow full group and role authorization. One could imagine also a more complete front-end that incorporates, for example, file metadata and the ability to push files into the grid.

Primary authors : Mr. JOUVENOT, Daniel (NA4)

Co-authors : Mr. LOOMIS, Charles (NA4)

Presenter : Mr. JOUVENOT, Daniel (NA4)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : Mr. JOUVENOT, Daniel

Submitted on Monday 29 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 23

Proceedings on collaboration between EGEE and Platform Computing - User impact and experience

Description of the activity :

At EGEE'06 conference, Platform Computing was announced as one of the first EGEE Business Associates (EBA). This talk will report on the achievements of this collaboration and partnership and on further proceedings. We will discuss, how extended exploitation of readily available scheduler options started to change and improve usage and user experience. One key result is improved handling of different workload types in the EGEE infrastructure.

Grid added value :

Impact and potential: As several Platform Computing customers participate in EGEE, already more than 30% of the compute capacity is provided by Platform LSF sites. Making the already available advanced LSF features working for gLite users and EGEE resource providers is impacting the effective throughput and responsiveness situation significantly, considered it potentially be used on more than a third of EGEE infrastructure. First results and details are presented. LSF integration: In the past, LSF integration into gLite and prior EGEE software stacks has proven to be improvable: Robustness and performance, ease of use, well sorted documentation, support for extended features are targeted. A side effect of the LSF exploitation is a freshly reviewed and debugged LSF integration, updated documentation and certification by the EGEE labs. Examples will be shown.

Experience or proposed activity :

Exploitation of advanced LSF features: The exploitation of LSF is based on the recently introduced gLite mechanism for forwarding arbitrary information from the grid user to the batch system. This allows users to request the functionalities of the batch system to greater detail. The gLite team in Bologna together with Platform Computing currently explores how this feature can be used to better exploit the capabilities of LSF in order to build complex algorithms, in particular for parallel jobs (MPI) and service-level agreement (SLA) scheduling and enforcement. Target is to define a generic way of using grid-user provided information for interfacing to local batch systems. The exploitation of LSF's data scheduling features will also be
explored. The potential impact on users and resource providers is tremendous. Again, we will show first results and details, further plans. Questions from the audience are encouraged, discussion expected

Future evolution :

Future collaboration items: Some potential topics have been identified and will be presented. Contributions and ideas on collaborations are welcome.

Primary authors : Mr. SCHOTT, Bernhard (Platform Computing)

Co-authors :

Presenter : Mr. SCHOTT, Bernhard (Platform Computing)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Mr. SCHOTT, Bernhard

Submitted on Monday 29 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 24

Job prioritization in LHCb

Description of the activity :

LHCb is one of the four high-energy experiments running in the near future at the Large Hadron Collider (LHC) at CERN. LHCb will try to answer some fundamental questions about the asymmetry between matter and anti-matter. The experiment is expected to produce about 2PB of data per year. Those will be distributed to several laboratories all over Europe and then analyzed by the Physics community. To achieve this target LHCb fully uses the Grid to reprocess, replicate and analyze data.

Grid added value :

The access to the Grid happens through LHCb's own distributed production and analysis system, DIRAC (Distributed Infrastructure with Remote Agent Control). Dirac implements
the “pull” job scheduling paradigm, where all the jobs are stored in a central task
queues and then pulled via generic grid jobs called Pilot Agents. The whole LHCb community (about 600 people) is divided in sets of physicists, developers, production and software managers that have different needs about their jobs on the Grid. While a Monte Carlo simulation jobs need several days of intensive CPU time, the analysis jobs just need to start immediately. The current state of affairs, where all the users access the Grid through a single entry point, does not prevent to have sub-communities running most of the jobs and then monopolizing the use of Grid resources. The way to avoid this is to implement a system that ensures job priority and
fair share of the resources among all the community users.

Experience or proposed activity :

There are two possible approaches to encompass it: a site-wise approach where the VO just takes care of filling up its queues and leaves the site-specific software to redistribute the jobs accordingly to early negotiations; a VO-wise approach, best tailored to the LHCb computing model, where the site just allocates the quota of resources competing to the VO and the VO decides how to share it across its users sub-communities.
A rough priority algorithm based on the VO-wise approach has already been implemented. The introduction of a "Priority'' flag in the specification of the job and some changes in the resource-job matching mechanism already proved to guarantee the
right precedence to short analysis jobs or to Reconstruction jobs with respect of cumbersome Monte Carlo jobs. Our Priority algorithm must be considered as a work-in-progress development. Accounting information based on both the user, job length and community past CPU consumption will also be considered.

Future evolution :

The job priority mechanism needs to be extensively tested. An aging system will also be
introduced to avoid that some jobs to stay too long in the central queues before being picked-up at the first suitable resource available. The mechanism relies on the assumption that DIRAC is the only access to the Grid but does not prevent
users to bypass it and access the Grid somehow else. A tool to enforce VO policy at site
level is then highly desired.

Primary authors : CASTELLANI, Gianluca (European Organization for Nuclear Research (CERN))

Co-authors :

Presenter : CASTELLANI, Gianluca (European Organization for Nuclear Research (CERN))

Track classification : Workflow

Contribution type : --not specified--

Submitted by : CASTELLANI, Gianluca

Submitted on Monday 29 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 25

Intelligent, distributed climate data management

Description of the activity :

Climate research is generally data-intensive. Observation, analysis and output data of climate simulations are traditionally stored in large archives and central databases. They vary highly in quality and in accessibility. Thus searching, finding and retrieving the data is often highly inefficient. To enable and ease effective data retrieval as well as collaborative work in the often interdisciplinary projects, an intelligent and sustainable (meta-)data management infrastructure is needed.

Grid added value :

Currently, to find, retrieve and process climate data mostly complex individual solutions are used. Processed data is commonly stored locally and undocumented. Thus, identical analysis are redone by various scientists.

Enabling searching and browsing of the various data in a central catalog according to content, quality and processing history would ease the discovery of data. An intelligent, transparent data access would simplify the data retrieval. Selectable basic processing options and an automatic republishing of the processed data would support the daily workflows of climate scientists and facilitate further processing or usage of the results.

To realize this, besides conceptual agreements also common protocols and logging facilities, common authorization and authentication standards and finally also common resources to effectively share tools and data are required. EGEE offers solutions for most of these challenges and is thus a logical choice as basis for such an infrastructure.

Experience or proposed activity :

As a first step, we used adapted tools and concepts developed in the German community driven Grid initiative C3-Grid (http://www.c3-grid.de/) to port the prototype of a typical climate data analysis workflow to the EGEE infrastructure (demonstrated at EGEE'06 conference).

Based on these experiences we now improve the structure and tighten the integration of the two systems. Of the C3-Grid, the central Web portal and uniform access interfaces, implemented at German Earthsystem science data centers, are used to initially search, find and retrieve the data and also to publish resulting metadata of already processed data. Processing can be done on EGEE. The Amga catalog is used as a central instance to receive and update the necessary runtime information of data produced or altered in the processing jobs. An interface is implemented to the Amga catalog to enable the automatic harvesting and hence republishing of the resulting metadata into the central C3 Web portal.

Future evolution :

The outlined approach does not only advance the interoperability of the two grid projects, but also stimulates synergy effects by combining the strengths of the two systems. Using standards along with international agreements allows for interoperability with international partners, such as the British NERC DataGrid and the US-American Earth System Grid, and with other communities, such as the GIS community. The next issue that needs to be solved is the match of the different security systems.

Primary authors : Dr. RONNEBERGER, Kerstin (DKRZ, Hamburg) ; Dr. KINDERMANN, Stephan (DKRZ, Hamburg) ; Dr. STOCKHAUSE, Martina (MPI-M, Hamburg)

Co-authors :

Presenter : Dr. RONNEBERGER, Kerstin (DKRZ, Hamburg) ; Dr. KINDERMANN, Stephan (DKRZ, Hamburg) ; Dr. STOCKHAUSE, Martina (MPI-M, Hamburg)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : Mrs. RONNEBERGER, Kerstin

Submitted on Monday 29 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 26

Intelligent, distributed climate data management

Description of the activity :

Climate research is generally data-intensive. Observation, analysis and output data of climate simulations are traditionally stored in large archives and central databases. They vary highly in quality and in accessibility. Thus searching, finding and retrieving the data is often highly inefficient. To enable and ease effective data retrieval as well as collaborative work in the often interdisciplinary projects, an intelligent and sustainable (meta-)data management infrastructure is needed.

Grid added value :

Currently, to find, retrieve, and process climate data mostly complex individual solutions are used. Processed data is commonly stored locally and undocumented. Thus, identical analysis are redone by various scientists.

Enabling searching and browsing of the various data in a central catalog according to content, quality and processing history would ease the discovery of data. An intelligent, transparent data access would simplify the data retrieval. Selectable basic processing options and an automatic republishing of the processed data would support the daily workflows of climate scientists and facilitate further processing or usage of the results.

To realize this, besides conceptual agreements also common protocols and logging facilities, common authorization and authentication standards and finally also common resources to effectively share tools and data are required. EGEE offers solutions for most of these challenges and is thus a logical choice as basis for such an infrastructure

Experience or proposed activity :

As a first step, we used adapted tools and concepts developed in the German community driven Grid initiative C3-Grid (http://www.c3-grid.de/) to port the prototype of a typical climate data analysis workflow to the EGEE infrastructure (demonstrated at EGEE'06 conference).

Based on these experiences we now improve the structure and tighten the integration of the two systems. Of the C3-Grid, the central Web portal and uniform access interfaces, implemented at German Earthsystem science data centers, are used to initially search, find and retrieve the data and also to publish resulting metadata of already processed data. Processing can be done on EGEE. The Amga catalog is used as a central instance to receive and update the necessary runtime information of data produced or altered in the processing jobs. An interface is implemented to the Amga catalog to enable the automatic harvesting and hence republishing of the resulting metadata into the central C3 Web portal.

Future evolution :

The outlined approach does not only advance the interoperability of the two grid projects, but also stimulates synergy effects by combining the strengths of the two systems. Using standards along with international agreements allows for interoperability with international partners, such as the British NERC DataGrid and the US-American Earth System Grid, and with other communities, such as the GIS community. The next issue that needs to be solved is the match of the different security systems.

Primary authors : Dr. RONNEBERGER, Kerstin (DKRZ) ; Dr. KINDERMANN, Stephan (DKRZ, Hamburg) ; Dr. STOCKHAUSE, Martina (MPI-M, Hamburg)

Co-authors :

Presenter : Dr. KINDERMANN, Stephan (DKRZ, Hamburg)

Track classification : Data Management

Contribution type : oral presentation

Submitted by : Mrs. RONNEBERGER, Kerstin

Submitted on Monday 29 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 27

Experiment Dashboard - monitoring system for the LHC experiments

Description of the activity :

LHC experiments are depending on the distributed EGEE infrastructure for their core activities. The Experiment Dashboard is a monitoring framework aiming to provide for the LHC experiments the overview of their activities on the EGEE infrastructure with a special emphasis in support for users community.

Grid added value :

Existing monitoring tools are usually focusing onto a specific usage like specific Grid middleware/infrastructures, specific submission tool, etc. The Experiment Dashboard has been built to aggregate the existing monitor infrastructure (from experiment specific software, infrastructure itself, monitoring tools) and provide unified views and information correlation. Experiment Dashboard is covering different areas of the LHC activities - job processing, data transfer, data publishing. It is deployed for four LHC experiments (CMS, ATLAS, LHCb, ALICE). Some of the core functionality of the Experiment Dashboard like job monitoring can be used for other virtual organizations.

Experience or proposed activity :

Experiment Dashboard is currently in production and is used by LHC users with different roles for their everyday work.

Future evolution :

The whole EGEE monitoring infrastructure can be considerably improved. Very often the error messages indicating various failures are not clear and do not point to a real problem. The variety of the local fabrics monitoring systems used by local sites complicates the task of creation of the common framework for aggregation of the monitoring data in the central repository. Transparent navigation of the monitoring data provided by different monitoring systems is often not possible.

Primary authors : ANDREEVA, Julia (CERN) ; GAIDIOZ, Benjamin (CERN) ; BRITO DA ROCHA, Ricardo (CERN) ; HERRALA, Juha (CERN) ; SAIZ, Pablo (CERN)

Co-authors : CATALIN, Cirstoiu (CERN) ; TSAI, Fu-ming (ASGC) ; CHEN, Tao-sheng (ASGC) ; CHIU, Shih-chun (ASGC)

Presenter : ANDREEVA, Julia (CERN)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : ANDREEVA, Julia

Submitted on Monday 29 January 2007

Last modified on : Friday 16 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 28

An Adaptive Framework for WCDMA System Analysis in the EGEE Grid Infrastructure

Description of the activity :

We plan to use the EGEE infrastructure in order to evaluate and analyze various techniques and scenarios for the design of mobile communication systems. Thus far, we have performed simulations of various scenarios for Wideband Code Division Multiple Access (WCDMA) systems. In the future, we intend to use EGEE resources for performing further simulations on other mobile communication techniques, such as OFDM, smart antennas and multiple-input multiple-output (MIMO) techniques.

Grid added value :

The systems we are analyzing and evaluating are stochastic; Monte Carlo methods are used in order to perform system simulations. These Monte Carlo simulations may become extremely computationally demanding as the physical dimensions and the loading of the simulated system increase. Indicatively, a single simulation may last a day or even more in a standard desktop computer and for the convergence of the Monte Carlo to be satisfied, hundreds of simulations may be eventually needed. The EGEE infrastructure provides vast resources that may be exploited in order to efficiently reduce execution time. Use of these resources will enable the simulation and analysis of more complex systems that would be infeasible to perform in a desktop computer or a conventional small cluster. Monte Carlo methods are widely used in research fields (ray-tracing, molecular dynamics, finance, etc) and our proposed framework may be used under these contexts with minimal modifications.

Experience or proposed activity :

In order to speedup the execution of the simulations, we have exploited the parallelizability of Monte Carlo methods and have developed a task-farming framework for their parallel execution. The framework is tailored to the gLite middleware and implements and manages a task-farming workflow for parallel Monte Carlo simulations. Briefly, once the initialization parameters are given, the application dispatches several jobs to the grid, assigning a number of single simulations to each job. Due to the heterogeneity of the infrastructure and the maximum CPU time available for each job, the number of simulations assigned to each job is dynamically determined by a simple heuristics that takes into consideration execution time and MaxCPUTime and that is incorporated to the job itself. Until now, we have executed several scenarios regarding a WCDMA system, for various base station configurations, physical dimensions and system loading.

Future evolution :

A restraining factor to our approach is that in certain scenarios a single simulation may exceed the MaxCPUTime limit mentioned above. In order to surpass this problem, we plan to exploit any parallelization pattern may be applicable to a single simulation and develop MPI programs for concurrent execution at simulation level. This, however, restricts usable resources to those of an MPI supporting site, as the support for MPI applications across different sites is still missing.

Primary authors : Mr. ATHANAILEAS, Theodoros (National Technical University of Athens) ; Mr. GKONIS, Panagiotis (National Technical University of Athens) ; Dr. TSOULOS, Georgios (National Technical University of Athens) ; Prof. KAKLAMANI, Dimitra (National Technical University of Athens)

Co-authors :

Presenter : Mr. ATHANAILEAS, Theodoros (National Technical University of Athens)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : ATHANAILEAS, Theodoros

Submitted on Monday 29 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 29

BiG: A Grid Service to Distribute Large BLAST Runs

Description of the activity :

Bioinformatics community use habitually computers for their large computationally-intensive genomic and proteomic analysis. Bioinformatics is a consolidated community in the frame of EGEE since phase I, who is using the resources through applications such as the GPS@ portal.
The main challenge of this community (in the frame of the present work) is the identification and annotation of the function of protein and nucleotide sequences through their comparison with other well-known databases.

Grid added value :

BLAST-enabled Grid services do exist. However, most of them use sequential BLAST as their processing engine. However, MPI-enabled versions of BLAST do exist, and performance of large runs could be extensively improved. Then, the objective has been to develop an off-line, robust service to BLAST, based on MPI-Blast and delivering higher performance. Currently, this services are used daily by hundreds of users for short runs.
The service can be integrated in portals or as a part of an application, since it exports an interface based on a statefull web service that enables launching the execution, disconnecting and reconnecting later on. The service splits the input sequences and the target database in multiple jobs and MPI processes, being possible to launch several simultaneous analyses on different protein and nucleotides databases.

Experience or proposed activity :

The BiG service is already operational in the EELA infrastructure. It is being used through the Ibero-American Bioinformatics Portal, (http://portal-bio.ula.ve/) and runs on the resources of the Technical University of Valencia and Federal University of Rio de Janeiro, being currently other resource providers under negotiation. Registered users can access the Grid service through the portal. Other source of users is the community of biologists using B2GO tool (http://www.blast2go.de/), mainly interested on agricultural biology. The average usage has been of 1.5 CPU/days per day since its release in June, but jobs take sequentially more than 3 CPU/days, and results are obtained in a few hours. This peak demand of resources has fitted very well the usage fashion of Grids. This service has been used for analysis of segments of pathogens causing diseases in humans and agriculture.

Future evolution :

The main problems we have experimented have been related to the reliability of the services and the proxy renewal. Automatic proxy renewal and VOMS credential through myproxy repositories seems to be still not completely solved. Other problems are not expected, since the system has extensively run on a similar production infrastructure.

Primary authors : Dr. BLANQUER, Ignacio (UPV) ; Mr. APARICIO, Gabriel (UPV) ; Prof. HERNáNDEZ, Vicente (UPV)

Co-authors :

Presenter : Dr. BLANQUER, Ignacio (UPV)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Dr. BLANQUER, Ignacio

Submitted on Monday 29 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 30

Gridification: Porting New Communities onto the WLCG/EGEE Infrastructure

Description of the activity :

The computational and storage capability of the Grid are attracting several research communities and we will discuss the general patterns observed in supporting new applications, porting them on the the EGEE environment.
In this talk we present the general infrastructure we have developed inside the application and support team at CERN (PSS and GD groups) to merge in a fast and feasible way all these applications inside the Grid, as for example Geant4, HARP, Garfield, UNOSAT or ITU.

Grid added value :

All these communities have different goals and requirements and the main challenge is the creation of a standard and general software infrastructure for the immersion of these communities onto the Grid. This general infrastructure is effectively "shield" the applications from the details of the Grid (the emphasis here is to run applications developed independently from the Grid middleware). On the other hand, it is stable enough to require few control and support by the members of the Grid team and also of the members of the users communities. Finally, it is flexible and general enough to match the requirements of the different productions without including mayor changes in the design of the tool.

Experience or proposed activity :

As general submission, tracking and monitoring tool we have chosen the Ganga/DIANE infrastructure as official tool for all new gridifications. This infrastructure is adapted to the requirements of each production with a minimum impact in the general tool. It also includes a layer to MonALISA to monitor the status of the jobs at each site and keep processing history information.
We have also created a whole VOMS infrastructure for a new VO called “gear” (applied for generic applications). This VO hosts a whole EGEE infrastructure, which enables the full and immediate immersion of new communities inside the Grid. At any moment the production of these communities and also the behavior and use of the VO policy is strictly observed by the ARDA team.
The main value of the work presented in this talk is the large level abstraction of the Grid that we have achieved with the infrastructure that we have created for new communities.

Future evolution :

The successful execution of all productions requires a minimum amount of dedicated computational resources and robust and fully supported services. This includes RBs and CEs being in an unconditional stable shape throughout the entire period of each production. Particular attention has to be given to a stable and fully supported computing and storage systems which do not compromise the basic services.

Primary authors : Dr. MENDEZ LORENZO, Patricia (CERN IT/PSS) ; Mr. MOSCICKI, Jakub (CERN IT/PSS) ; Dr. LAMANNA, Massimo (CERN IT/PSS) ; Mr. MURARU, Adrian (CERN IT/PSS)

Co-authors :

Presenter : Dr. MENDEZ LORENZO, Patricia (CERN IT/PSS)

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : Dr. MENDEZ LORENZO, Patricia

Submitted on Monday 29 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 31

Experiences on Grid Production for GEANT4

Description of the activity :

Geant4 is a world-wide collaboration of scientists and software engineers whose goal is to develop, maintain and provide support for the Geant4 package, representing a general purpose Monte Carlo toolkit for simulating the propagation of particles through matter. Geant4 is currently employed in a row of particle physics experiments (BaBar, HARP, ATLAS, CMS, LHCb), and is also applied in other areas like space science, medical applications, and radiation studies.

Grid added value :

Due to the complexity of the Geant4 code and due to the broad spectrum of possible configurations, involving a variety of physical processes for different source particles, particle energies and target or shielding materials, an intensive testing of new release candidates is mandatory to carefully and thoroughly test each of its components, especially before major releases, generally twice a year. Regression tests are required to be performed by the Geant4 team within a short period of time (2 to 3 weeks), basically demanding vast computational resources (equivalent to approximately 4 CPU-years). In order to cope with this abrupt rise in numbers of required CPU's, EGEE Grid resources are utilized, where in general several EGEE sites provide CPU's primarily dedicated for these Geant4 production series.

Experience or proposed activity :

The first Geant4 production on the Grid was performed in December 2004, and since December 2005 Geant4 is officially recognized as WLCG/EGEE VO. Basically, the test series compare a reference and candidate version of the Geant4 toolkit to detect any significant difference between them, by looking at calorimeter observables. In order to guarantee the successful and complete testing of the release candidates within the required period of time, both, a minimum amount of dedicated computational resources (of the order of 100 CPU's) and dedicated, robust and fully supported services, are vital. This includes in particular RB's and batch systems being in an unconditional stable shape throughout the entire period. Involved sites are required to provide a shared file system to the Geant4 VO, to prevent a time-consuming multiple reinstallation of the Geant4 reference (and other required software) at each worker node, since only the candidate version changes throughout the testing period.

Future evolution :

Geant4 production series encompass the need to submit a vast number of Grid jobs within 2 to 3 weeks, where delays are not acceptable due to fixed release dates. Thus, particular attention has to be given to provide a stable computing environment and to ensure an uncompromising performance of basic services. This includes an instantaneous service support, since a missing response to performance problems might cause significant interruptions of the regression testing.

Primary authors : Mr. LECHNER, Anton (Atominst. der Oest. Universitaeten - Technische Universitaet Wie)

Co-authors : Dr. LAMANNA, Massimo (CERN) ; Dr. MENDEZ LORENZO, Patricia (CERN) ; Mr. MOSCICKI, Jakub (CERN) ; Dr. RIBON, Alberto (CERN)

Presenter : Mr. LECHNER, Anton (Atominst. der Oest. Universitaeten - Technische Universitaet Wie)

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : ANTON, Lechner

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 32

Protein structure prediction of "never born proteins". An experience within the EUChinaGRID framework

Description of the activity :

In nature there exists only a tiny fraction of all the theoretically possible protein sequences. It is thus of interest for the biologists to study the properties of proteins not present in nature (the "never born proteins") as a way to improve our knowledge on the fundamental properties that make existing protein sequences so unique. Protein structure prediction tools combined with the use of large computing resources allow to tackle this problem.

Grid added value :

The study of never born proteins requires the generation of a large library of protein sequences not present in nature and the prediction of their three-dimensional structure. This is not trivial when facing 10^5-10^7 protein sequences. Indeed, on a single CPU it would require years to predict the structure of such a large library of protein sequences. On the other hand, this is a trivial parallelism problem in which the same computation (i.e. the prediction of the 3D structure of a protein sequence) must be repeated several times (i.e. on a large number of protein sequences). The use of the GRID infrastructure makes feasible to approach this problem in an acceptable time frame.
In addition, once the simulation in a grid environment has been set up, the same approach can be used to tackle problems of immediate biomedical relevance such as the prediction of the structure of the entire set of proteins of a virus or a bacterial pathogen.

Experience or proposed activity :

The biological community is increasingly taking advantage of the informatics tools available. However informatics training is still scarce in biology courses. In our experience in porting protein structure prediction applications in a grid environment, the main difficulty we encountered was in being able to formalize job description and submission as well as output retrieval through the use of grid middleware. Essential to our experience was the integration of our applications within the GENIUS portal which allows job submission and management through the use of a user friendly grid interface. This allows non grid-trained users to profit from the advantages provided by the grid infrastructure.

Future evolution :

According to our experience, a simplified approach to the use of grid resources will be the key point to attract new communities, especially as far as the biomedical community is concerned. From this viewpoint, the further development of web-based, user friendly services will be critical to allow the access of non informatics trained scientists to grid resources and for the success of grid computing approach.

Primary authors : Prof. POLTICELLI, Fabio (Department of Biology, University Roma Tre, Italy)

Co-authors : Dr. MINERVINI, Giovanni (Department of Biology, University Roma Tre, Italy) ; Dr. LA ROCCA, Giuseppe (INFN Catania, Italy) ; Prof. LUISI, Pier Luigi (Department of Biology, University Roma Tre, Italy)

Presenter : Dr. MINERVINI, Giovanni (Department of Biology, University Roma Tre, Italy)

Track classification : Poster session

Contribution type : poster

Submitted by : Dr. POLTICELLI, Fabio

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 33

"High throughput" protein structure prediction application in EUChinaGRID

Description of the activity :

In nature there exists only a tiny fraction of all the theoretically possible protein sequences. It is thus of interest for the biologists to study the properties of proteins not present in nature as a way to improve our knowledge on the unique properties of the existing protein sequences. The demo will deal with the use of the protein structure prediction software Rosetta in a grid enviornment as a tool to study the structural properties of large protein sequence libraries.

Grid added value :

The demo will show how, through the use of grid resources, the problem of studying the structural properties of a huge number of protein sequences can be tackled. Through the use of a parametric job submission procedure, the user can submit multiple structure prediction jobs and perform a large number of predictions in a reasonable time frame. This is essential when dealing with a library made up of 10^5 10^7 protein sequences. In fact, the size of the library is essential to derive general rules that allow us to pinpoint the particular properties of natural proteins that make them unique.

Experience or proposed activity :

The demo will show how the application and related services for job submission and retrieval of the output have been integrated within the GENIUS portal. In addition it will be also demonstrated how a non grid trained user can submit jobs, retrieve the output and obtain a graphical representation of the output also through the use of molecular graphics plugins. This is a critical point when dealing with the biology and biomedical community as proper informatics training is still scarce in these disciplines, while the use of informatics tools in general and of the grid infrastructure in particular is of unvaluable help in the research in these fields.

Future evolution :

The demo presentation will also deal with aspects that still need development/improvement. In particular, a simplified, web-based access to the grid infrastructure allows easy and fast submission of jobs and further developments along this path are critical for the access of novel communities to the grid resources.

Primary authors : Prof. POLTICELLI, Fabio (Department of Biology, University Roma Tre)

Co-authors : Dr. MINERVINI, Giovanni (Department of Biology, University Roma Tre, Italy) ; Dr. LA ROCCA, Giuseppe (INFN Catania, Italy) ; Prof. LUISI, Pier Luigi (Department of Biology, University Roma Tre, Italy)

Presenter : Dr. MINERVINI, Giovanni (Department of Biology, University Roma Tre, Italy)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : Dr. POLTICELLI, Fabio

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 34

Grid2Win: gLite on Microsoft Windows platform.

Description of the activity :

Grid2Win is a joint development of the EGEE GILDA Team and the TriGrid VL Project (http://www.trigrid.it) funded by the Sicilian Regional Government.

Grid added value :

The goal of the Grid2Win project is to port gLite services to MS-Windows based platforms to be integrated into the Sicilian grid e-Infrastructure of TriGrid. Main services on which we are focusing are the User Interface, to let also Windows users access the grid services, and the Computing Element, to add Windows applications to the set of the ones already deployed on the infrastructure as well as to exploit the large computing power provided by existing Windows clusters.
Currently, we have already ported the gLite User Interface and developed our first Grid2Win Graphical UI (Grid2Win GUI) running on Windows desktops. Further, we have already built a Torque/MAUI based Computing Element on Windows XP and published it into an Information System as a grid resource.
Next steps are the deployment of an entire Windows XP cluster on our infrastructure and the integration of a new type of CE built on top of the Compute Cluster Server (CCS) local resource manager recently released by Microsoft.

Experience or proposed activity :

At the moment, our experience is based on a set of tests we made to verify the validity of the system and to evaluate performances. We have just put in production a Windows-based site in the GILDA t-Infrastructure. Tests made executing simple Windows applications were successfull. We are now preparing a more complex set of tests to be done in the next weeks.

Future evolution :

During the process of porting, we encountered many problems in terms of portability of the code. There is no a unique way to build packages and software. For each package, a different solution has been found and adopted. There are no documents describing the source from a developer point of view and we were obliged to make a very deep analysis of the code. Moreover, we were obliged to adjust some portions of the code to be able to execute on Cygwin POSIX environment.

Primary authors : Dr. RUSSO, Dario (INFN Catania)

Co-authors : Dr. SCIBILIA, Fabio (INFN Catania)

Presenter : Dr. RUSSO, Dario (INFN Catania)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : RUSSO, Dario

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 35

CAVIAR, CArdio Vascular Image Analysis on gRid

Description of the activity :

The progress of image acquisition devices makes possible to better explore the dynamics of moving organs such as the heart. The CAVIAR project objective is to exploit the huge computing power of grid architectures in time consuming cardio-vascular 4D image analysis applications (huge amount of data typically around 0.5GB per patient and per examination) such as cardio-vascular 3D+time structure segmentation and motion tracking.

Grid added value :

Grid architectures offer the opportunity to run the image analysis programs on large sets of patient data, allowing for large scale inter-patient cardio-vascular (CV) studies and comparisons. The complex analysis programs can not be run on standard PC architectures with the great amount of data and require high computing power.
Developments using grids can thus help for :
- construction of CV atlases for normal and patholohical cases
- analysis of patient specific data with prior atlases and providing faster diagnostic
results
- epidemiologic studies on populations suffering from CV pathologies, still one of the leading cause of mortality in Europe.

Such a principle can be envisaged in other applications involving image analysis tasks like in image sequence indexing and retrieval or geopraphical and satellite image analysis.

Experience or proposed activity :

Parallel versions of spatio-temporal motion estimation and segmentation methods from Magnetic Resonance Image sequences have been developped (multithreaded and MPI versions) and first deployed on PC clusters. In a second step, they are currently ported on the EGEE grid.
Security and confidentiality of processed data must be insured.
Also grid services should enable the parallel execution of application softwares through MPI, in particular when processing one specific case. However the full interest of the grid would be evident when processing several cases in parallel.

Future evolution :

Our plan is to develop a web portal to remotely access tools developped in the context of the CAVIAR Project
- to conduct first tests on clinical data
- to evaluate the feasibility of massive experimental data processing

It is essential however to be able to execute MPI based parallel programs which would require the development of specific node/cluster selection mechanisms.

Primary authors : Mr. DELHAY, Bertrand (CREATIS)

Co-authors : Mr. SCHAERER, Joel (CREATIS) ; Mr. PERA, Christophe (CREATIS) ; Mr. CLARYSSE, Patrick (CREATIS)

Presenter : Mr. DELHAY, Bertrand (CREATIS)

Track classification : Poster session

Contribution type : poster

Submitted by : Mr. DELHAY, Bertrand

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 36

Ganga - an Optimiser and Front-End for Grid Job Submission

Description of the activity :

The presentation will introduce the Ganga job-management system (http://cern.ch/ganga), developed as an ATLAS-LHCb common project. The main goal of Ganga is to provide a simple and consistent way of preparing, organising and executing analysis tasks, allowing physicists to concentrate on the algorithmic part without having to worry about techical details.

Grid added value :

Ganga provides a clean Python API that reduces and simplifies the work involved in preparing an application, organizing the submission, and gathering results. Technical details of submitting a job to the Grid, for example the preparation of a job-description file, are factored out and taken care of transparently by the systems. By changing the parameter that identifies the execution back-end, a user can trivially switch between running an application on a portable PC, running higher-statistics tests on a local batch system, and analysing all available statistics on the Grid.

Experience or proposed activity :

Although Ganga is being developed for LHCb and ATLAS, it is not limited to use with HEP applications, and already has several non-HEP users. These include users on projects in bio-medicine, engineering, and (Grid) software testing.

Future evolution :

Ganga is a higher-level Grid tool and therefore tries to circumvent typical problems when submitting jobs to the Grid, easing the user experience. Ganga has a plug-in mechanism, so that it can be highly customised to suit the needs of a given user community.

Primary authors : Dr. MAIER, Andrew (CERN)

Co-authors : Mr. BROCHU, Frederic (University of Cambridge) ; Dr. EGEDE, Ulrik (Imperial College London) ; Dr. ELMSHEUSER, Johannes (Ludwig-Maximilians-Universität, München) ; Dr. GAIDIOZ, Benjamin (CERN) ; Mr. HARRISON, Karl (High Energy Physics Group, Cavendish Laboratory) ; Mr. LEE, Hurng-chun (ASGC, Taipei and CERN) ; Dr. LIKO, Dietrich (CERN) ; Mr. MOSCKICKI, Jakub (CERN) ; Mr. MURARU, Adrian (CERN) ; Mr. ROMANOVSKY, Vladimir (State Res.Center of Russian Feder. Inst.f.High Energy Phys. (IFVE)) ; Dr. SOROKO, Alexander (University of Oxford)) ; Mr. TAN, Chun Lik (University of Birmingham) ; KOBLITZ, Birger (CERN)

Presenter : Dr. MAIER, Andrew (CERN)

Track classification : Interactivity and Portals

Contribution type : oral presentation

Submitted by : MAIER, Andrew

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Interactivity and Portals

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 37

VO VOCE -- Numerical Trip to the Galactic Centre

Description of the activity :

The young massive stars in the Galactic Centre represent one of the challenges of modern astrophysics. We can study such stellar system either by near infra red (NIR) observations or by numerical simulations. We study dynamics of young massive stars in the central parsec region around the galactic super-massive black hole (SMBH) surrounded by several stellar structures which act as a perturbing component to the central SMBH potential.

Grid added value :

The study of young stellar cluster in the Galactic Centre helps us to better understand the conditions in vicinity of SMBHs in the galactic centres in general. We study motion of test particles (single young massive stars) in the composed gravitational potential -- potential of the SMBH plus the perturbation potential from those surrounding stellar structures, therefore we can study the effect of the Kozai resonance on the orbits of young massive stars (test particles). This study includes scanning the huge parameter-space of the system by means of numerical integration of trajectories of testing particles. The lower limit on basic computations CPU time consumption is of the order of months (normalised to 2 GHz CPU), the added value of the Grid is in its computational power and in the possibility to "paralelize" the computation. The Grid provides us with the possibility scan larger volume of the parameter space in sufficiently shorter time.

Experience or proposed activity :

We performed test of availability and stability of VOCE resources with bunches of ultra-short jobs (running few seconds of CPU time) in order to get percentage of successfully computed jobs and its distribution over the VOCE resources. We performed short jobs computations (few minutes to few hours of CPU time) in order to test the Grid suitability to our activity. From this testing we got an impression that it is possible to perform such computations in the Grid environment. The most important Grid tools we need to solve our problem on the Grid are successful submission of the job and job flow control in general, gcc compiler at our disposal on each Computing Element (CE), access to the Storage Elements (SE) and accessible data on the SEs.

Future evolution :

One of the possible limitations to our computations should be a given upper limit on CPU time consumption per job. The stability of the Grid environment is the most crucial part of our long-lasting computations.

Primary authors : Ms. SCHOVANCOVA, Jaroslava (CESNET, Prague, Czech Republic)

Co-authors : Dr. SUBR, Ladislav (Astronomical Institute, Charles University in Prague, Czech Republic)

Presenter : Ms. SCHOVANCOVA, Jaroslava (CESNET, Prague, Czech Republic)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Ms. SCHOVANCOVA, Jaroslava

Submitted on Tuesday 30 January 2007

Last modified on : Saturday 10 February 2007

Comments :

We would like to present our contribution either as the oral presentation or as
the poster presentation.

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 38

Using GridWay for the Management of Virtual Machines in the EGEE Infrastructure

Description of the activity :

Virtual machines (VMs) are a promising technology to overcome some of the problems found in current Grids, like heterogeneity, performance partitioning or application isolation. This presentation describes how GridWay can be used for the management of VMs in Grid infrastructures. The novel approach provides end users with a custom environment for the execution of legacy numerical codes, without requiring additional grid middleware to be installed and so easily deployable within EGEE.

Grid added value :

Since the late 1990s, we have witnessed an extraordinary development of Grid technologies. Nowadays, different Grids are being deployed within the context of a growing number of national and transnational research projects (e.g. EGEE, TeraGrid or OSG). These projects have achieved unseen levels of resource sharing, offering a dramatic increase in the number of processing and storage resources that can be delivered to applications. However, a growing heterogeneity on the organizations that joins these projects hinders the development of large scale Grid infrastructures. Grid resources do not only differ in their hardware but also in their software configurations (OS, libraries, and applications). This heterogeneity increases the cost and length of the application development cycle, as they have to be tested in a great variety of environments where the developers have limited configuration capabilities. Therefore, some of the users are only able to use a small fraction of the Grid.

Experience or proposed activity :

The proposed approach consists in encapsulating a virtual machine in a grid job, incorporating the functionality offered by a general purpose meta-scheduling system. So, the genuine characteristics of a Grid infrastructure (i.e. dynamism, high fault rate, heterogeneity) are naturally considered in the proposed solution. This strategy does not require either additional middleware to be deployed or execution of VMs in the Computing Elements, as it is based on well-tested procedures and standard services. The only requirement is to have Worker Nodes running an hypervisor (Xen, UML…). The presentation also assesses the suitability of this deployment in the execution of a high throughput scientific application, the XMM-Newton Scientific Analysis System.

Future evolution :

Although the solution is fully functional, it presents several drawbacks, namely: limited use of the potential benefits offered by the virtualization technology, as the underlying LRM is not aware of the nature of the job, and limited support for interactive jobs, as only batch jobs are supported. We are working on the integration of GridWay with the Globus Virtual Workspace service which allows a remote client to securely negotiate and manage a virtual resource.

Primary authors : Prof. MARTIN LLORENTE, Ignacio (Universidad Complutense de Madrid) ; Prof. SANTIAGO MONTERO, Ruben (Universidad Complutense de Madrid) ; Mr. RUBIO MONTERO, Antonio Juan (CIEMAT) ; Mr. TAPIADOR DE PEDRO, Daniel (European Space Astronomy Centre, ESAC-ESA) ; Prof. HUEDO CUESTA, Eduardo (Universidad

Complutense de Madrid)

Co-authors :

Presenter : Prof. SANTIAGO MONTERO, Ruben (Universidad Complutense de Madrid)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : MARTIN LLORENTE, Ignacio

Submitted on Tuesday 30 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 39

GridWay as a Tool for Porting Applications to the EGEE Infrastructure

Description of the activity :

The GridWay Metascheduler is a Globus project that performs job execution management and resource brokering, allowing unattended, reliable, and efficient execution of jobs, job arrays, and workflows on heterogeneous and dynamic Grids. GridWay is completely functional on EGEE, being able to interface with its computing, file transferring and information services. The demonstration will mainly show the functionality provided by GridWay to port scientific production codes to EGEE.

Grid added value :

GridWay provides the following benefits to the different stakeholders involved in a Grid environment. For project and infrastructure directors, GridWay is a community project, adhering to Globus philosophy and guidelines for collaborative development. For system integrators, GridWay is open-source software, released under Apache license v2.0. For system managers, GridWay gives a scheduling framework similar to that found on local DRM systems, supporting resource accounting and the definition of scheduling policies, with minimum installation requirements, being compatible with a wide variety of platforms. For application developers, GridWay implements the OGF standard DRMAA API (C and JAVA bindings), assuring compatibility of applications with LRM systems that implement the standard, such as SGE, Condor, Torque… For end users, GridWay provides a LRM-like CLI for submitting, monitoring, synchronizing and controlling jobs.

Experience or proposed activity :

The demonstration will show the main functionality provided by GridWay, focusing on its support for the execution of typical execution profiles, namely: embarrassingly distributed, master-worker and workflow. GridWay is being successfully used in the NA4 Fusion activities and is being evaluated in other VOs, such as BioMed.

Future evolution :

GridWay provides a lighter-weigh alternative to the resource brokers available in EGEE (gLite-WMS and LCG-RB), offering additional application porting functionality and higher performance for given execution profiles. GridWay exhibits shorter scheduling latencies as it reduces the number of submission stages and provides mechanisms, such as opportunistic migration and performance slowdown detection, that improve the usage of the underlying resources.

Primary authors : Prof. MARTIN LLORENTE, Ignacio (Universidad Complutense de Madrid) ; Prof. SANTIAGO MONTERO, Ruben (Universidad Complutense de Madrid) ; Prof. HUEDO CUESTA, Eduardo (Universidad Complutense de Madrid) ; Mr. VAZQUEZ BLANCO, Constantino (Universidad Complutense de Madrid) ; Mr. VAZQUEZ POLETTI, Jose Luis (Universidad Complutense de Madrid) ; Mr. HERRERA SANZ, Jose (Universidad Complutense de Madrid) ; Mrs. LEAL ALGARA, Katia (Universidad Rey Juan Carlos)

Co-authors :

Presenter : Mr. VAZQUEZ BLANCO, Constantino (Universidad Complutense de Madrid)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : MARTIN LLORENTE, Ignacio

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 40

FOOTPRINT@work, A COMPUTING FRAMEWORK FOR LARGE SCALE PARAMETRIC SIMULATIONS : APPLICATION TO PESTICIDE RISK ASSESSMENT AND MANAGEMENT

Description of the activity :

FOOTPRINT is an EU-funded FP6 project which aims at developing risk assessment and management tools for different end-user communities in Europe to limit the contamination of water resources by pesticides. The project is relevant to everyone involved in pesticide management and use (researchers,policy-makers,ministries,local authorities,water managers,farmers). It relies on the definition of a large number of agro-environmental scenarios and their modeling using validated pesticide fate models.

Grid added value :

Results from the very numerous (1-10,000,000 runs, several Petabytes of data) precomputed runs will be used to build a series of meta-models. The huge requirement for computing power has already led to collaboration with a national grid initiative (IGGI) and more than 500 desktop PCs of BRGM's internal grid will be used during their idle period for the modeling. The pesticide community has praised FOOTPRINT for its innovative and ground-breaking approach and numerous proposals for improvement of the methodology used have been expressed, resulting in demands for computing power far exceeding the capabilities of desktop PCs. Hence, the success of this European research project clearly calls for the use of high performance computing facilities such as those offered by the EGEE infrastructure. Key individuals in research institutes, registration authorities, consultancy and the agrochemical industry have already expressed their support to see the FOOTPRINT concept developed further.

Experience or proposed activity :

Based on the principles of the PEST/SENSAN package which is widely used in hydrology for model calibration and sensitivity/uncertainty analyses, a set of tools have been developed to mimic its principal functionalities and to handle inputs and outputs of large parametric simulations. Although the pesticide fate models used for pesticide registration are typically locked, access to the code of the MACRO model has been granted as part of FOOTPRINT, which allows its porting under various OS, including Linux. Windows-based pesticide fate models have already been deployed on BRGM PCs running Linux using dedicated emulators. A parameter to the success of the deployment of MACRO onto the EGEE infrastructure would be the availability/development of a parametric jobs manager although alternatives can be used to avoid this aspect.

Future evolution :

We have not made use of the EGEE infrastructure at this stage. From a

prospective point of view, capacities and limitations with regard to storage resources on the EGEE architecture will be of particular interest. Also, swift and rapid access to the modelling results is essential.

Primary authors : Mrs. BOULAHYA, Faïza (BRGM) ; Mr. DUBUS, Igor (BRGM) ; Mr. DUPROS, Fabrice (BRGM) ; Mr. LOMBARD, Pierre (Icatis)

Co-authors :

Presenter : Mrs. BOULAHYA, Faïza (BRGM) ; Mr. DUBUS, Igor (BRGM)

Track classification : Related Projects

Contribution type : oral presentation

Submitted by : Mrs. BOULAHYA, Faïza

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Related Projects

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 41

On porting low-energy nuclear reaction codes on Grid

Description of the activity :

There was always a stringent need for high accuracy nuclear data due to nuclear fusion and fission projects, and not only, that have intensified in the last years due to new foreseen facilities. The experiments are becoming more complex, expensive and lack the power to measure all the required quantities. The comparison of the measured and calculated cross sections for nuclear reactions showed in several occasions noticeable differences between nuclear computer codes predictions above the energy regions with experimental data and also a significant deviation from unity of calculated-to-experimental ratio [1], moreover these evaluations being reported without estimates of uncertainties. Recent efforts have begun to address this uncertainty issue through the generation of covariance files for nuclear model calculations [2], a task that requires large scale computations. It is thus justified the effort made to improve the actual status of the nuclear computer codes which in turn translates in reliable, safer and cost effective modern facilities [3].

Grid added value :

Continuous improvement of a nuclear code is essential for its predicting power and reliability, obviously the most important aspect being the underlying physical models and parameters but another key issue is the computational time performance. This performance is affected by the complexity of the actual physical model which in turn dictates the level of approximations employed to the algorithm used. The simplifications are most of the time satisfactory but with time they become too general for user requirements and tend to restrict the applicability of the algorithm, thus new techniques or models are needed.

Therefore it is desirable to exploit modern hardware and software platforms in order to perform faster and with better precision the necessary calculations and thus improving overall the process of evaluating the nuclear data.

The paper presents the techniques used to improve the performance of low-energy nuclear reaction codes with respect to computational time, namely the porting process of EMPIRE-II and SCAT2MIN nuclear computer codes to grid environment [4], method that provides the highest achievable performance at a reduced cost and a higher security.

Experience or proposed activity :

Porting the EMPIRE-II [5] nuclear reaction code to computational GRIDs was done by integrating it on the GILDA [6] Grid testbed [7] during ICTP/INFM-Democritos workshop on “Porting Scientific Applications on Computational GRIDs” [8]. The purpose was to

reduce the time used to systematically evaluate in one run all reaction channels of all stable isotopes corresponding to a given element,

and possibly neighboring elements using local and/or regional parameters [6].

Also a modified version of SCAT-2 [9] nuclear code that searches "best fit" nuclear optical model parameters was ported to the parallel environment using OpenMPI [10] library, the code being prepared to run on the in the EGEE FUSION VO Grid[9]. The performance tests show that the code is very scalable with the number of processors and thus perfect for the parallel and distributed computational environments.

Future evolution :

Porting the nuclear-reaction computer code EMPIRE-II to GRID infrastructure and the graphical integration into Genius web portal proved successful. However, there are still remaining issues regarding the MySQL EXFOR experimental database as well as an

eventual implementation of a facility for plotting online the calculated results in comparison with the experimental data.
The code SCAT-2 proved to be a highly scalable one, suitable for distributed architectures. It is thus supporting the next step of a concurrent study by taking into account all known experimental elastic-scattering angular distributions for a given isotope/element. The further use of dynamic chunks of data may improve the load balancing on heterogeneous environments, at the cost of the overhead of communication.
The reliability and cost effectiveness of future facilities relies on improved nuclear reaction models as well as faster computer codes and finally the covariance evaluation of nuclear models. In spite of it being a long-standing subject, only recent investigations in this respect have been motivated by the computational power of present computers. This work has showed that it is rather straightforward to integrate actual nuclear computer codes into GRID computational environment for assessment of large data sets. Following the achievement of this goal concurrently and also performing large scale computations with MPI enabled nuclear codes, the related outcome will be firsthand the reduced computational time, and also one of the first steps towards nuclear models uncertainties generation.

Primary authors : Mr. ROMAN, Faustin Laurentiu ("Horia Hulubei" National Institute for Physics and Nuclear Engineering (IFIN-HH))

Co-authors : Prof. AVRIGEANU, Marilena ("Horia Hulubei" National Institute for Physics and Nuclear Engineering (IFIN-HH))

Presenter : Mr. ROMAN, Faustin Laurentiu ("Horia Hulubei" National Institute for Physics and Nuclear Engineering (IFIN-HH))

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Mr. ROMAN, Faustin Laurentiu

Submitted on Tuesday 30 January 2007

Last modified on : Saturday 10 February 2007

Comments :

The authors acknowledge the hospitality and support of the ICTP/INFM-Democritos workshop on “Porting Scientific Applications on Computational GRIDs"l. Work supported in part by the EURATOM-MEdC Fusion Association and MEdC Contract No. CEEX-05-D10-48.

References:

1. V. Avrigeanu, M. Avrigeanu, F.L. Roman, R.A. Forrest, R.Eichin, H. Freiesleben, and K. Seidel, Sensitivity of activation cross sections of the hafnium, tantalum and tungsten stable isotopes to nuclear reaction mechanisms, in Proc. Int. Workshop on Nuclear Data Needs for Generation IV Nuclear Energy Systems, Antwerpen, Belgium, 5-7 April 2005, World Scientific, Singapore, p. 145.
2. WPEC, Subgroup 24: Covariance Data in the Fast Neutron Region; H. Leeb, M.T. Pigni, and I. Raskinyte, Covariances for Evaluations Based on Extensive
Modelling, in Proc. Int. Conf. on Nuclear Data for Science and Technology (ND2004), Santa Fe, New Mexico, 26 Sept.-1 Oct. 2004, AIP Conf. Proc. No. 769, 2005, p. 161; A.J. Koning, Generating Covariance Data with Nuclear Moldels, in Proc. Int. Workshop on Nuclear Data Needs for Generation IV Nuclear Energy Systems, Antwerpen, Belgium, 5-7 April 2005, World Scientific, Singapore, p. 152; M. Herman, P. Oblozinsky, D. Rochman, T. Kawano, and L. Leal, Fast Neutron Covariances for Evaluated Data Files, in Proc. PHYSOR-2006 Topical Meeting on Advances in Nuclear Analysis and Simulation, Vancouver, Sept. 10-14, 2006 (in

press).
3. Nuclear Physics and Related Computational Science R&D for Advanced Fuel Cycles
Ws., Aug. 2006, Bethesda, Maryland;
http://www-fp.mcs.anl.gov/nprcsafc/Report_FINAL.pdf .
4. F.L. Roman, Porting Nuclear Reaction Codes for Nuclear Data Evaluations on GRIDs,
in Proc. 2nd Int. Conf. on Distributed computing and Grid technologies in
science and
education", Dubna, Russia, 26-30 June, 2006 (in press), and the references therein.
5. M. Herman, P. Oblozinski, M. Sin, A. Trkov, A. Ventura, and V. Zerkin, Recent
developments of the nuclear reaction code EMPIRE, in R.C. Haight, M.B. Chadwick, T.
Kawano, and P. Talou, editors, Proc. of the Int. Conf. on Nuclear Data for Science
and Technology, page 1184, New York, USA, 2005, Santa Fe, Sept 26 – Oct 1, 2004,
AIP.
6. Gilda Virtual Organisation, http://gilda.ct.infn.it/.
7. https://glite-tutor.ct.infn.it/
8. F.L. Roman, "Porting Nuclear Reaction Codes for Nuclear Data Evaluations on
GRIDs", in ICTP/INFN-Democritos Ws. on “Porting Scientific Applications on Comp.
GRIDs”.
9. O.Bersillon, SCAT2:Un programme de modele optique spherique, Report CEA-N-2227
NEANDC(FR), INDC(E) 49 (1981); www.nea.fr/abs/html/nea-0829.html
10. OpenMPI library, www.open-mpi.org.
11. EGEE FUSION Virtual Organisation,
http://grid.bifi.unizar.es/egee/fusion-vo/index.html.

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 42

g-Eclipse - a framework for Grid users, Grid operators and Grid developers based on Eclipse

Description of the activity :

The g-Eclipse project aims to build an integrated workbench framework to access the power of existing Grid infrastructures. The framework will be built on top of the reliable eco-system of the Eclipse community to enable a sustainable development. The framework will provide tools to customize Grid users' applications, to manage Grid resources and to support the development cycle of new Grid applications.
In the first year of the project, support for the gLite middleware will be implemented

Grid added value :

As the g-Eclipse framework aims for middleware independence, the emerging eco system has the potential to be the reliable graphical user interface to access the power of grids in the future. The framework will be open to integrate existing and new tools.
The g-Eclipse project will address three major groups in the Grid domain: Grid users will benefit from the Desktop-like access to Grid resources; Grid operators and resource providers will be able to reduce the time-to-service by the Grid management and Grid site configuration tools; and Grid application developers will reduce the time-to-market for new Grid applications by accelerating the development and deployment cycle.
The g-Eclipse project addressed sustainability from the very beginning. The g-Eclipse project is now accepted as an offical technology project of the Eclipse Foundation (www.eclipse.org/geclipse).

Experience or proposed activity :

The g-Eclipse project started its collaboration with the EGEE project in July 2006. The project uses computing and storage resources of the EGEE infrastructure as well as different central services like Information Systems, Resource Brokers, Workload Management Systems. As the project focuses on the implementation of reusable user interfaces, it uses the existing services of the EGEE infrastructure.
The project is therefore strongly dependent on the exisiting Java API of all EGEE grid
services. Currently, such APIs do not exist for all services which might influence the project's progress.

Future evolution :

The g-Eclipse framework will build the integrated Grid environment (IGE) of the future for Grid end users, Grid resource operators and Grid developers by providing an open and extensible eco system based on the industry compliant Eclipse Platform.

Primary authors : Dr. KORNMAYER, Harald (FORSCHUNGSZENTRUM KARLSRUHE (FZK))

Co-authors :

Presenter : Dr. KORNMAYER, Harald (FORSCHUNGSZENTRUM KARLSRUHE (FZK))

Track classification : Related Projects

Contribution type : oral presentation

Submitted by : Dr. KORNMAYER, Harald

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Hi Massimo, hi all □
□
THIS is an application for oral presentation!!□
□
It would be fine to give an overview of the g-Eclipse project as an oral□
presentation.□
An additional demo submission will be done too!□
□
Best regards □
Harald □

Status : SUBMITTED

Track judgments :

Track : Related Projects

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 43

An Intuitive Web Interface for Massive Molecular Docking on the Grid

Description of the activity :

The avian flu data challenge in last spring has successfully demonstrated a high-througput docking service on the EGEE infrastructure. The next step is to make this service more accessible to real biology end-users rather than grid specialists. Collaborating with biomedical researchers, we have developed an intuitive web interface for biologists to access the massive molecular docking service on the EGEE infrastructure.

Grid added value :

Aiming at making the grid-enabled docking service more useful for biologists, we developed a web-based user interface on top of the DIANE framework (http://cern.ch/diane), trying to reduce researchers' effort in accessing the service and in analyzing the docking results. By hiding the technical details, the interface exposes only the configurations meaningful for biologists. Through this interface, biologists could set a filter on a compound library and select interesting targets and ligands by visually examining their structures. Leveraging on the interactive feedback feature of the DIANE framework, results of finished dockings are not only ranked in table; they are also transformed into 3-D structures so that biologists can immediately look into the docking poses and download corresponding results to their local disk for further analysis.

Experience or proposed activity :

This work is built on top of the DIANE framework. DIANE was originally developed for handling the distributed applications within a Master-Worker model. It provides an overlay system on top of the Grid system, in which the pull-mode scheduling and failure recovery mechanisms are implemented based on the CORBA protocol. On the other hand, the DIANE framework hides the details of the job operations on the Grid so that application developers can concentrate on the implementation of application logic and interface design. DIANE interfaces the EGEE infrastructure through a high-level grid tool, GANGA (http://cern.ch/ganga).

Future evolution :

As the efficiency of the grid-enabled docking service has been demonstrated by the data challenge, we expect a well customized interface can encourage biology researchers to use the service for their daily research. In developing the web interface, we experienced that adopting high-level Grid application frameworks (e.g. DIANE and GANGA) can dramatically reduce the effort in handling Grid jobs, concerning the work of failure recovery, and of application specific monitoring and bookkeeping.

Primary authors : Mr. LEE, Hurng-chun (ASGC) ; Mr. HO, Li-yung (ASGC) ; Dr. WU, Ying-ta (Genomic Research Center, Academia Sinica)

Co-authors : Mr. CHEN, Hsin-yen (ASGC) ; Dr. LIN, Simon (ASGC)

Presenter : Mr. LEE, Hurng-chun (ASGC)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : LEE, Hurng-chun

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 44

g-Eclipse - Access the power of the Grid

Description of the activity :

The g-Eclipse project aims to build an integrated workbench framework to access the power of existing Grid infrastructures. The framework will be built on top of the reliable eco-system of the Eclipse community to enable a sustainable development. The framework will provide tools to customize Grid users' applications, to manage Grid resources and to support the development cycle of new Grid applications.

In the first your of the project, the gLite middleware support will be implemented.

Grid added value :

As the g-Eclipse framework aims for middleware independence, the emerging eco system has the potential to be the reliable graphical user interface to access the power of grids in the future. The framework will be open to integrate existing and new tools. The g-Eclipse project will address three major groups in the Grid domain: Grid users will benefit from the Desktop-like access to Grid resources; Grid operators and resource providers will be able to reduce the time-to-service by the Grid management and Grid site configuration tools; and Grid application developers will reduce the time-to-market for new Grid applications by accelerating the development and deployment cycle. The g-Eclipse project addressed sustainability from the very beginning. The g-Eclipse project is now accepted as an offical technology project of the Eclipse Foundation (www.eclipse.org/geclipse).

Experience or proposed activity :

The features of the g-Eclipse framework will be demonstrated as a live demo by accessing EGEE Grid infrastructure. The demo will show the setup and configuration of the g-Eclipse framework as well as the "Frist Grid application cycle". This includes the generation of a VO-specific Grid project, the inclusion of remote Grid file systems, the generation of Grid job specifications with wizards and the submission of job. The demo will prove that the migration of a local application to
a Grid infrastructure will be much easier with the toolbox of the g-Eclipse framework.
Additionally the integrated features to the GGUS system and VOMS management system will be demonstrated.

Future evolution :

At the end of the demo, we will high-light problems in terms of currently missing Java
API or webservice specifications for the gLite central services.

Primary authors : Dr. KORNMAYER, Harald (FORSCHUNGSZENTRUM KARLSRUHE (FZK))

Co-authors :

Presenter : Dr. KORNMAYER, Harald (FORSCHUNGSZENTRUM KARLSRUHE (FZK))

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : Dr. KORNMAYER, Harald

Submitted on Tuesday 30 January 2007

Last modified on : Friday 16 February 2007

Comments :

Hi Massimo, hi all

THIS is an application for DEMO presentation!!

I hope find a nice booth for us in the demo area!

Best regards
Harald

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 45

Simulations and Offline Data Processing for the Auger Experiment

Description of the activity :

The Pierre Auger Cosmic Ray Observatory is studying the universe's highest energy particles, which shower down on Earth in the form of cosmic rays. Data recorded by the experiment are analysed and compared with results from simulations. The Pierre Auger Collaboration includes more than 250 scientists from 17 countries. All members of the collaboration should have easy access to data.

Grid added value :

There are two main features of the Grid important for the Auger collaboration. As a global collaboration, it benefits from a uniform access to data stored on Storage Elements and registered in a catalogue. CPU requirements on simulations of the ultra high energy cosmic rays are huge, possibility to use many CPUs connected via Grid enables to simulate events with higher precision. We expect tens of active grid users accessing data and/or running simulation jobs.

Experience or proposed activity :

We developed a framework for submission of many simulation jobs with varying input parameters. We will report on our activity of porting offline data processing to the Grid. Experience from a data management in the VO auger will be given.

Future evolution :

We had to establish our own rules for file naming schemes and registration to be able to manage data on various SEs. Generic rules imposed by middleware and tools for data consistency checks are needed.

Primary authors : Dr. CHUDOBA, Jiri (Institute of Physics)

Co-authors : SCHOVANCOVA, Jaroslava (CESNET)

Presenter : Dr. CHUDOBA, Jiri (Institute of Physics)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : CHUDOBA, Jiri

Submitted on Tuesday 30 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 46

Results on LHCb Data Challenge 06

Description of the activity :

The Large Hadron Collider (LHC) at CERN is the front end machine for the high energy physics (HEP) and will start operating in 2007. The expected amount of data that will be produced and that has to be analyzed is unprecedented. LHCb, one of the large experiment at the LHC, moved toward grid technologies to cope with their extraordinary data and CPU intensive needs. The integration of the experiment specific computing framework into the underlying production grid has not been always effortless.

Grid added value :

Grid technologies represent the only way to deal with HEP today's computing needs. The complexity of these new techniques brought the need of designing, for each experiment, a model for processing and analyzing the data. The 2006 data challenge – LHCb DC06 – is the latest of a series of big activities on the Grid and represents the final benchmark before the real data taking. Its goal is validating the computing model and the computing framework of LHCb but it is also the last opportunity for exercising the whole simulation chain on WLCG resources and testing the readiness of all resources involved. Over the past few years, LHCb has always been one of the top users of LCG-EGEE resources gathering considerable experience in distributed computing at a large scale. The central part of the system is DIRAC (Distributed Infrastructure with Remote Agent Control). It is the experiment gateway to the grid and its key-words are resilience, reliability and redundancy.

Experience or proposed activity :

The achieved maturity of the LHCb computing framework from one side and the knowledge acquired on Grid technologies by the community placed the DC06 experience in a privileged position. It represents indeed a lucid and objective outlook on the health status of the Grid few months before the first beam collisions. The aim of this work is to present this experience, its original objectives and how these have been adjusted in the time reflecting problematic encountered. A description of the DIRAC system and how it is evolving to cope with the limits of the back-end system, a discussion of the performances achieved, and the analysis of the problems observed are also given. The DC06 has started in August 2006. Six months after its start, DC06 is perceived as the set of all Monte Carlo production, reprocessing and analysis activities on WLCG. It highlighted that the WLCG service is enhancing but there is still big room for improvements.

Future evolution :

Data Management System and Workload System on WLCG are still too instable and inefficient. The reliability of all these services can be improved by instituting an operational infrastructure that monitors and guarantees problems are correctly addressed and fixed. Alternatively LHCb think that resources providers have to be motivated in chasing up issues. We triggered several times debugging sessions in tight collaboration with site managers that proved a prompt and efficient reaction.

Primary authors : Dr. SANTINELLI, Roberto (CERN/IT/GD)

Co-authors :

Presenter : Dr. SANTINELLI, Roberto (CERN/IT/GD)

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : Dr. SANTINELLI, Roberto

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 47

Grid Training Worldwide with GIILDA

Description of the activity :

GILDA is the the well know t-Infrastructure adopted by EGEE but also by several other EU Grid Projects worldwide. In this contribution the experience of running gLite tutorials in the context of EELA, EUCHINAGRID and EUMEDGRID infrastructure projects as well as in the context of some thematic projects like BIOINFOGRID and EMBRACE will be reported together with the new services available on GILDA to cope with this unprecedented scale of training activities.

Grid added value :

.

Experience or proposed activity :

.

Future evolution :

.

Primary authors : Prof. BARBERA, Roberto (UNIV. CATANIA AND INFN)

Co-authors :

Presenter : Prof. BARBERA, Roberto (UNIV. CATANIA AND INFN)

Track classification : Training and Support for Grid Users

Contribution type : oral presentation

Submitted by : BARBERA, Roberto

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Training and Support for Grid Users

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 48

The Current Status of the Regional Grid Projects going on in Sicily

Description of the activity :

Two big Projects, TriGrid VL (http://www.trigid.it) and PI2S2 (http://www.pi2s2.it), financed by Sicilian Regional Government and Italian Ministry of University and Research, respectively, with a total budget of about 15 M€, are changing completely the way of doing e-Science and industrial pre-competitive research in the South of Italy. All academic and research Institutes operating in Sicily as well as several SMEs are involved in the two Projects.

Grid added value :

The entire infrastructure of the two Projects mentioned above is built on top of the gLite Middleware and will eventually join the EGEE production service through the Italian Grid Infrastructure.

Experience or proposed activity :

In the present contribution the status of the various activities of the two Projects as well as the applications running on the e-Infrastructures will be presented providing the audience with a real example of a regional grid for communities belonging both to Research and Industry.

Future evolution :

.

Primary authors : Prof. BARBERA, Roberto (UNIV. CATANIA AND INFN)

Co-authors :

Presenter : Prof. BARBERA, Roberto (UNIV. CATANIA AND INFN)

Track classification : Related Projects

Contribution type : oral presentation

Submitted by : BARBERA, Roberto

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Related Projects

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 49

Experiences of using gLite infrastructure with g-Eclipse environment

Description of the activity :

The g-Eclipse project aims to build an integrated workbench framework to access the power of existing Grid infrastructures. The framework will be built on top of the reliable eco-system of the Eclipse community to enable a sustainable development. The framework will provide tools to customize Grid users' applications, to manage Grid resources and to support the development cycle of new Grid applications. In the first year of the project, the gLite middleware support will be implemented.

Grid added value :

As the g-Eclipse framework aims for middleware independence, the emerging eco system will provide reliable graphical user interface to access different grid middleware and infrastructures. The first supported infrastructure is gLite. The g-Eclipse project will address three major groups in the Grid domain: Grid users, Grid operators and Grid developers. All three groups of user will benefit from graphical user interface to gLite resources. Using the same intuitive environment user will be able to prepare, submit and manage their jobs, access remote data, access information about available resources, test specific services or sites, deploy and debug their application on the gLite grid. As the g-Eclipse project is now accepted as an official technology project of the Eclipse foundation (www.eclipse.org/geclipse) we expect to growing number of people interested in grid infrastructure and user friendly tools for working with Grid.

Experience or proposed activity :

The g-Eclipse project started in July 2006 its collaboration with the EGEE project. The aim of the project is to prepare general grid environment for users, but the first reference infrastructure is gLite. As the project focus on the implementation of reusable user interfaces, it is using the existing services of the EGEE infrastructure. The project is therefore strongly dependent from exisiting Java API of all EGEE grid services. We are using VOMS for VO management, RGMA for querying grid resources, WMS and CREAM to submit job and LB for retrieving job status. As g-Eclipse is the environment running on the client machine we cannot rely on gLite UI machine and we need to use only pure java or WS implementations. As such implementation is not available for LFS we cannot support it yet.
We encounter a lot of technical troubles, especially for services which are currently in beta state.

Future evolution :

The g-Eclipse framework will built the future the complete, user-friendly, integrated Grid environement (IGE) for Grid end users, Grid resource operators and Grid developers. We will add more EGEE tools and support for more gLite services as they will be available implementations with pure java API or with web services.
The documentation for gLite is very often not complete, but contact with gLite developers (both direct and via GGUS) are very helpful.

Primary authors : WOLNIEWICZ, Pawel (PSNC)

Co-authors :

Presenter : WOLNIEWICZ, Pawel (PSNC)

Track classification : Related Projects

Contribution type : oral presentation

Submitted by : WOLNIEWICZ, Pawel

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Another abstract about g-Eclipse was submitted by Harald Kornmayer with id=42.
But those one was more general and focus on g-Eclipse goals and status.
The current submission is about technical experiences and problems ancountered
when using gLite middleware from g-Eclipse environment.

Status : SUBMITTED

Track judgments :

Track : Related Projects

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 50

prolog-mpi: a System for Explicit Predicate Distribution in Prolog Systems

Description of the activity :

prolog-mpi provides an abstract distributed-computing interface to operators of the Prolog computer language. Prolog is a popular language in a variety of fields, especially in the mathematical fields of provability, artificial intelligence, and optimisation. The prolog-mpi primary executable, pl-mpi, is an extension of the familiar SWI-Prolog interpreter that operates upon custom predicates in order to distribute instructions and data to a network of compute nodes.

Grid added value :

The complexity of most academic research systems is considerable. Execution on conventional computer hardware is often unreasonable; thus, in order to execute such systems within a practical amount of time, grid architectures offer an environment of massive, scalable sequence parallelisation. Unfortunately, multicomputer environments are not available to all computer languages. The University of Latvia IMCS developed prolog-mpi in order to allow preexisting prolog systems to take advantage of explicit parallelism. The system provides a number of custom predicates that transparently handle both blocking and non-blocking distribution strategies. prolog-mpi's precursor, dnlp, was featured at the BalticGrid 2AHM, October 2006. It demonstrated considerable polynomial speed-up of Prolog systems previously constrained to uniprocessor environments or overly-fine-grained implicit distribution. prolog-mpi surpasses this system with a general interface for explicit parallelisation.

Experience or proposed activity :

The University of Latvia began developing prolog-mpi as a means to provide preexisting Prolog scripts with the ability of explicit distribution. To date, prolog-mpi has been successfully integrated into a significant NLP code-base: parallel regions are explicitly labelled and, when executed by pl-mpi, these regions are properly distributed and recombined. The prolog-mpi system is also bundled with a utility for automated testing, producing a range of run-time profiles for the purposes of correctness testing and benchmarks. In order to operate prolog-mpi, computing environments must provideMPI. Further, each system must have a functional SWI-Prolog dynamic library available for linking. With these requirements met, prolog-mpi may run both interactively or non-interactively. The system has unbounded scale in terms in terms of input size and number of participating nodes.

Future evolution :

When approaches were considered for parallelisation, choice of parallel interface was a matter of significant concern. MPI was chosen because of its support on local grids available to the IMCS, not due to an informed endorsement by an authoritative text. Unfortunately, current deployments of MPI on the BalticGrid lack the ability to span multiple grid sites, which artificially constrains processor availability.

Primary authors : Mr. DZONSONS, Kristaps (Researcher)

Co-authors : Mr. BARZDINS, Guntis (Professor)

Presenter : Mr. DZONSONS, Kristaps (Researcher)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Ms. KASKINA, Baiba

Submitted on Tuesday 30 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Submitted by Baiba Kaskina (baiba@latnet.lv)

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 51

Predicting Bounds on Queuing Delay in the EGEE grid

Description of the activity :

Predicting the performance of schedulers is a notoriously difficult task [1]. As a consequence, grid users might be tempted to work around the standard grid middleware by designing specific strategies, which would be counterproductive if generally adopted. On the other hand, Machine Learning has been successfully applied to performance prediction in distributed and shared environments [2,3]. This paper reports on experiments on predicting the basic parameters of scheduling in the EGEE framework.

Grid added value :

The expected running time (RT) of jobs and expected queuing delay (QD) are important inputs for grid global schedulers. Within gLite, QD is dynamically published by the Computing Elements into the grid information system, which is in turn queried by the scheduling agents called the brokers. At this time, little is known about the accuracy of the prediction of QD. In ordinary production, gLite uses the published QD for minimizing the expected job turnaround time, and errors in this prediction impact grid utilization. gLite also considers all jobs being equivalent, so it is difficult (without reconfiguring the site schedulers) to raise the priority of certain classes of jobs in situations such as social emergency, important events for a scientific community, or software prototyping. To overcome these problems, reinforcement learning has been proposed as a solution for time-constrained scheduling by coupling efficient prediction of QD and scheduling decisions.

Experience or proposed activity :

We carried out preliminary statistical analysis (including summary statistics, density estimation, and time series analysis) on scheduler logs of a site of the EGEE grid (the LAL node). We show that the experimental arrival process and service times are extremely far from simple standard models (the classical M/M/N Kendall queue model with Poissonian arrival times and exponential service time), and might in fact exhibit long-range correlation and periodic behaviour. The failure of linear autoregression suggests that non-linear methods are more appropriate in the time series analysis of the expected queuing delay. We are currently investigating such methods (neural networks, gaussian processes and hidden Markov models), which can be able to take into account both inter-arrival time and load.

Future evolution :

The major pitfall in analyses similar to our approach is the possible lack of representativity of the data. Further research in this direction could greatly profit from an easier access to the existing monitoring data (beyond isolated experiments). Furthermore, easier access would also reduce the associated cost of developing analysis software.

Primary authors : Mr. PEREZ, Julien (LRI) ; Dr. KéGL, Balázs (CNRS) ; Prof. GERMAIN-RENAUD, Cécile (CNRS LRI)

Co-authors :

Presenter : Mr. PEREZ, Julien (LRI) ; Dr. KéGL, Balázs (CNRS)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Mr. PEREZ, Julien

Submitted on Tuesday 30 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 52

An Intuitive Web Interface for Massive Molecular Docking on the Grid

Description of the activity :

The avian flu data challenge in last spring has successfully demonstrated a high-througput docking service on the EGEE infrastructure. The next step is to make this service more accessible to real biology end-users rather than grid specialists. Collaborating with biomedical researchers, we have developed an intuitive web interface for biologists to access the massive molecular docking service on the EGEE infrastructure.

Grid added value :

Aiming at making the grid-enabled docking service more useful for biologists, we developed a web-based user interface on top of the DIANE framework (http://cern.ch/diane), trying to reduce researchers' effort in accessing the service and in analyzing the docking results. By hiding the technical details, the interface exposes only the configurations meaningful for biologists. Through this interface, biologists could set a filter on a compound library and select interesting targets and ligands by visually examining their structures. Leveraging on the interactive feedback feature of the DIANE framework, results of finished dockings are not only ranked in table; they are also transformed into 3-D structures so that biologists can immediately look into the docking poses and download corresponding results to their local disk for further analysis.

Experience or proposed activity :

This work is built on top of the DIANE framework. DIANE was originally developed for handling the distributed applications within a Master-Worker model. It provides an overlay system on top of the Grid system, in which the pull-mode scheduling and failure recovery mechanisms are implemented based on the CORBA protocol. On the other hand, the DIANE framework hides the details of the job operations on the Grid so that application developers can concentrate on the implementation of application logic and interface design. DIANE interfaces the EGEE infrastructure through a high-level grid tool, GANGA (http://cern.ch/ganga).

Future evolution :

As the efficiency of the grid-enabled docking service has been demonstrated by the data challenge, we expect a well customized interface can encourage biology researchers to use the service for their daily research. In developing the web interface, we experienced that adopting high-level Grid application frameworks (e.g. DIANE and GANGA) can dramatically reduce the effort in handling Grid jobs, concerning the work of failure recovery, and of application specific monitoring and bookkeeping.

Primary authors : Mr. LEE, Hurng-chun (ASGC) ; Mr. HO, Li-yung (ASGC) ; Dr. WU, Ying-ta (Genomic Research Center, Academia Sinica)

Co-authors : Mr. CHEN, Hsin-yen (ASGC) ; Dr. LIN, Simon (ASGC)

Presenter : Mr. LEE, Hurng-chun (ASGC)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : LEE, Hurng-chun

Submitted on Tuesday 30 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 53

Grid reliability

Description of the activity :

We are offering a system to track the efficiency of different components of the GRID. We can study the performance of both the WMS and the data transfers

At the moment, we have set different parts of the system for ALICE, ATLAS, CMS and LHCb. None of the components that we have developed are VO specific, therefore it would be very easy to deploy them for any other VO.

Grid added value :

Our main goal is basically to improve the reliability of the GRID. The main idea is to discover as soon as possible the different problems that have happened, and inform the responsible. Since we study the jobs and transfers issued by real users, we see the same problems that users see. As a matter of fact, we see even more problems than the end user does, since we are also interested in following up the errors that GRID components can overcome by themselves (like for instance, in case of a job failure, resubmitting the job to a different site).

This kind of information is very useful to site and VO administrators. They can find out the efficiency of their sites, and, in case of failures, the problems that they have to solve.

The reports that we provide are also interesting for the COD, since the errors might not be VO specific

Experience or proposed activity :

All this system is based on studying the different actions that users do. Therefore, the first and most important dependency is on monitoring systems. The way we do it is to interface it with the DASHBOARD, which will hide the differences between the heterogeneous sources of data (like RGMA, ICXML or MonALISA).

Another service very important for the effectiveness of the Grid reliability is the submission and tracking of tickets, GGUS. This has already been tested with a manual procedure. Since the result was very encouraging, we are working on ways of automatizing this interaction.

Future evolution :

Themain problem that we have found so far is the lacking of communication between the new gLite RB and RGMA. Jobs that went through these resource brokers do not publish their status, thus making our taks imposible.

Another possible problem that we might encounter is the confidentiality of the data. To solve this, we are anonymising the jobs and transfers, since we are only interested in the different status that the job or transfer goes through.

Primary authors : SAIZ, Pablo (CERN) ; GAIDIOZ, Benjamin (CERN) ; ROCHA, Ricardo (CERN) ; ANDREEVA, Julia (CERN)

Co-authors :

Presenter : SAIZ, Pablo (CERN)

Track classification : Poster session

Contribution type : poster

Submitted by : SAIZ, Pablo

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 54

Providing computational capacity for demanding medical imaging applications using EGEE infrastructure

Description of the activity :

Advances in medical image analysis (MIA) have increased the need to integrate and deploy MIA software in daily clinical routine. We have developed a Distributed Workflow Management System (DWMS) that supports a wide portfolio of MIA applications in different CT and MRI application domains. The DWMS improves the interoperability between image acquisition devices, clinicians and researchers by making MIA applications available in a transparent way.

Grid added value :

Because of recent developments in multi-slice CT and dynamic MR imaging the MIA results are often not delivered within the required time frame of specific medical specialists due to limitations in network capacity to transfer the data and computational power to perform the MIA calculations. The trauma radiologist is in urgent need for quick MIA results. The processed CT angiography (CTA) data enable him to clearly visualize the bloodvessels in patients suspected from stroke or brain aneurism. The neurosurgeon uses processed functional MRI (fMRI) data as a pre-operative planning tool for patients with brain tumors. Ideally, the results of the fMRI analysis become available soon after acquisition. Access to EGEE infrastructure would provide adequate computational resources to perform these high priority and demanding MIA applications upon urgent requests from clinicians.

Experience or proposed activity :

Our requirements call for a limited number of computing nodes for both MIA applications (CTA and fMRI analysis) and relatively short computational times. These resources must be pre-allocated in order to start the computation immediately after submission of the MIA job. Exclusive claiming of Grid resources, however, is not supported in the current EGEE job submission model. In order to circumvent possibly queued urgent MIA jobs we want to have access to a very large number of HPC sites within the EGEE infrastructure to have virtually direct access to the requested computing nodes. It is our goal to use optical lightpaths to provide secure and private network connections with high bandwidth to the EGEE resources.

Future evolution :

In order to provide reliable MIA services to the clinic a number of requirements on quality of service and data privacy / integrity should be met. The availability of adequate Grid resources should be guaranteed at a 24/7-hour basis such that a maximum time of MIA completion can be guaranteed. The medical imaging data should be transfered securely to and from the resources within the EGEE infrastructure without excessive delay due to limitations on network bandwidth.

Primary authors : Dr. SNEL, Jeroen (Academic Medical Center Amsterdam)

Co-authors : Dr. OLABARRIAGA, Silvia (Academic Medical Center Amsterdam) ; Dr. BOUWHUIS, Maurice (SARA Amsterdam) ; Dr. GROSSO, Paola (UVA, Institute for Informatics, Amsterdam) ; Mr. ALKEMADE, Johan (Academic Medical Center, Amsterdam) ; Mr. GRATAMA VAN

ANDEL, Hugo (Academic Medical Center, Amsterdam) ; Dr. NEDERVEEN, Aart (Academic Medical Center, Amsterdam)

Presenter : Dr. SNEL, Jeroen (Academic Medical Center Amsterdam)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Dr. SNEL, Jeroen

Submitted on Tuesday 30 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 55

Running interactive and parallel applications on the Grid – live demo - Visualization of Plasma Particles in Fusion Devices application example.

Description of the activity :

We would like to present activity that aims supporting demanding interactive and parallel application with advanced visualization capabilities on the example of Visualization of Plasma Particles in Fusion Devices application.
This activity is done within int.eu.grid (http://www.interactive-grid.eu) project.
The infrastructure is complementary with the EGEE infrastructure and inter-operable with EGEE resources.
We would like to interest new communities in the presented valuable grid features.

Grid added value :

The main added features of this scientific initiative are:
-Distributed Parallel(MPI) Interactive Computing–interactive support for parallel applications, which may run remotely across several sites
-User friendly interactive access to grid infrastructures through a Grid Interactive Desktop with powerful visualization
Interactivity is an important feature that gives user possibility to interact with application in natural way giving the possibility to change parameters while application is running. This interoperable production-level e-Infrastructure for demanding interactive app. impact the daily work of researchers.
There is identified number of app. and communities that could benefit of this e-infrastructure. This includes: in medicine (Ultrasound Computer Tomography),in Environment (Environmental Assessment of Climate-Change), in Fusion (Visualization of Plasma Particles in Fusion Devices). Project provides support for interactive and parallel app. not only at technical level.

Experience or proposed activity :

The Computation and Visualization of Plasma Particles in Fusion Devices app. has been chosen as a good example of data intensive, interactive app. that requires powerful computing resources, advanced visualization and high network throughput. The purpose of this app. is exploiting the combined power of a high number of CPUs in order to visualize the behavior of plasmas inside fusion reactors. It is the most representative use case that makes usage of most interactive services developed in the int.eu.grid project. This concerns a user-friendly interactive desktop–the Migrating Desktop(MD), CrossBroker, advanced visualization services-GVid, usage of PACX-MPI on top of Open MPI and glogin. It exploits also such EGEE services, tools and components like LFC, CE, L&B, UI, shared resources like SE’s, WN’s. CrossBroker starts immediately jobs, and in combination with the MD video streaming service GVid provides true steering of the simulation and visualization of the fusion parallel app.

Future evolution :

There are several issues that could be improved. We would like to mention some of them (important for our community);It would be very useful to have web service access to lcg_utils and lfc functionality;the job submission time

could be reduced; the time interval between the actual job end and the availability of the end of the job information could be reduced
There are many appl. in different fields which relay on the possibility of having a fast answer that could benefit for this features.

Primary authors : Mr. PLOCIENNIK, Marcin (On behalf of the INTERACTIVE EUROPEAN GRID)

Co-authors :

Presenter : Mr. PLOCIENNIK, Marcin (On behalf of the INTERACTIVE EUROPEAN GRID)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : PLOCIENNIK, Marcin

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

More detailed use case,descriptions on applications, usage of EGEE □ services and components could be provided on request.

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 56

Implementation of Web Coverage Service functionalities over the EGEE Platform

Description of the activity :

CYCLOPS Project is a FP6 SSA which aims to bring together two important Communities: GMES and GRID, focusing on the operative sector of European Civil Protection (CP). In this context, an evaluation of the adoption of Grid technology by the CP community is required. In particular, in this work we tested the porting on the EGEE platform of Open Geospatial Consortium WCS, one of the most demanding services for geospatial data sharing which is a critical issue in most CP applications.

Grid added value :

An Open Geospatial Consortium Web Coverage Service (OCG-WCS) provides the user with a particular view of a selected geospatial resource. This operation often requires executing an intense computing work on a big binary file, which is usually stored locally to the WCS.
The European Civil Protection data infrastructure is intrinsically distributed in various locations.
In our case we need to execute the same job on a large amount of big resource files, geographically spread among different sites.
The EGEE platform gives our WCS the chance to subset the initial request in an arbitrary number of sub-requests, which can be submitted to different Computing Elements located close (in the LAN sense) to the physical replicas of our initial resource.

Experience or proposed activity :

We are developing a Java WCS that, from a user request, computes a particular subset, resampling or interpolating a geospatial data that resides in the Grid.
Our software splits this request in an arbitrary number of sub-requests, generates a JDL file describing a Direct Acyclic Graph (DAG), and sends it to the WMProxy component of the gLite-WMS using its Java API.
The WMProxy submits the sub-jobs to the various Computing Elements close to the replicas of the requested initial data, and makes sure that all sub jobs are executed successfully.
After all requests have been computed, a node of the DAG acting as “collector” job
takes care that all the results are correctly merged in a single file and sends it back to the WCS.

Future evolution :

The WCS is presently implemented as a synchronous service,while EGEE runs batch jobs.This limits our implementation to be exploited within use-cases involving complex job workflows where other parts of the computation are batch-like.In some cases the WCS computations require to access to relational databases,while the EGEE Data Management system is file-based.A full java access to Grid,without the need for the application to run on top of a fully equipped gLite User Interface would be preferable

Primary authors : Dr. ANGELINI, Valerio (CNR-IMAA) ; Dr. MAZZETTI, Paolo (CNR-IMAA) ; Dr. MAZZUCATO, Mirco (INFN-CNAF) ; Dr. NATIVI, Stefano (CNR-IMAA) ; Dr. VERLATO, Marco (INFN-Padova)

Co-authors :

Presenter : Dr. VERLATO, Marco (INFN-Padova)

Track classification : Related Projects

Contribution type : oral presentation

Submitted by : VERLATO, Marco

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Related Projects

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 57

The Medical Data Manager

Description of the activity :

Medical images acquired in radiology centers represent very large amounts of data. The medical image analysis community (physicians, researchers, etc) needs to access such large, distributed and sensitive data sets for various purposes (patient healthcare, epidemiology, etc). The aim of the Medical Data Manager is to provide an interface between medical data repositories inside hospitals (DICOM format) and the grid distributed data management system to respond these needs.

Grid added value :

The long term objective is to convince hospitals to use grid for their data management and exchange needs. The Medical Data Manager enables access to patient images (DICOM files) and associated metadata (patient records), which are stored in remote hospitals. The access to large image repositories of data collected over several countries will be useful for statistical and epidemiological studies. The grid access policy can be defined at site level and data encryption is used to protect sensitive data. The scale is potentially very large and the grid infrastructure could be a solution to provide reliable and secured connections. A wide scale deployment can only be envisaged once data security is reliable enough and trusted by end users.

Experience or proposed activity :

The Medical Data Manager gives secure and transparent access to patient images stored in DICOM format. It is interfaced to the gLite Data Management System and it includes some necessary additional components such as the SRM-DICOM interface and the AMGA metadata database. Because the search of medical records often involve more than a file search in a catalog, AMGA is used to search for files on their associated metadata. We developed SRM-DICOM interface, which provides read-only access to DICOM servers of hospitals. This interface uses the standard DICOM protocol in hospital and the standard SRM protocol of the grid data storage. It is integrated within gLiteIO and FiReMan, which provides access control. The last component, Hydra, ensures the encryption of data. The low availability of the gLite WMS on EGEE is a limitation to the extension of the MDM.

Future evolution :

The metadata and the data security are the two key components to answer user needs. In the future, medical metadata needs to be distributed over the acquisition sites. An AMGA server will be installed at each site to ensure a local control. The hierarchical organization of AMGA must be adapted to this purpose. The GFAL-based data management system currently deployed on EGEE does not provide today the expected level of functionality to replace the gLite WMS.

Primary authors : ROMAIN, Texier (EGEE - I3S - LRI)

Co-authors :

Presenter : ROMAIN, Texier (EGEE - I3S - LRI)

Track classification : Poster session

Contribution type : poster

Submitted by : ROMAIN, Texier

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 58

Implementation of a Grid-based telemedicine platform

Description of the activity :

One of the greatest opportunity offered by EGEE Grid infrastructure is the possibility to exchange information between different users and organizations easily and securely. This cooperation is very similar to the one used in medical practice to exchange patient information between specialists. In this context the cooperation between physicians is a key factor to bring expert second opinion in isolated places such as Burchina Faso.

Grid added value :

Our application has been developed to make the medical information exchange through the EGEE Grid possible using gLite services. All the medical information is characterized by the need to manage them keeping in mind the patient privacy. The EGEE Grid Services and the Web Service technology, provide all needed tools to develope a medical platform (MIP) to store and manage medical data guaranteeing the needed security levels both in terms of user authentication (and authorization) and data encryption (both during information transfer and storage). The spotted user target is composed by physicians and medical staff. The MIP infrastructure needs to be able to give them all the services described above without modifing their "modus operandi" inside their medical environment. The development and deployement of MIP services in more medical infrastructures creates a network for knowledge and information sharing across the Grid.

Experience or proposed activity :

The MIP application has been developed using Web Service technology to create an interface between end users and the EGEE Grid Infrastructure. This project organization has been adopted to keep the medical platform as easy to use as possible and to give users the possibility to access the medical information stored across the Grid in a secure way from any machine with internet access and password based authentication. This organization reduces the impact of the "Grid World" complexity on physicians. To develop MIP several grid services will be used. In fact, on one side we will use AMGA service to store all the medical metadata (i.e. patient info, exam details, etc.), to manage users authentication and to define different access permissions based on user role and information category. On the other side the application will be able to store the anonymized (and also encrypted) medical images in different SE using gLite data management capabilities.

Future evolution :

During design and first steps of development of MIP infrastructure we haven't encontered particoulr problems on metadata management. In fact, the AMGA implementation of metadata catalog is stable and provides all needed features and flexibility to be adapted to our purpose. On the other hand the data management of gLite middleware still lacks in support and flexibility for what concern data management of encrypted data.

Primary authors : DIARENA, Matteo (CNRS IN2P3 LPC Clermont-Ferrand) ; FESSY, Arnaud (CNRS IN2P3 LPC Clermont-Ferrand)

Co-authors :

Presenter : DIARENA, Matteo (CNRS IN2P3 LPC Clermont-Ferrand)

Track classification : Data Management

Contribution type : oral presentation

Submitted by : DIARENA, Matteo

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 59

SymGrid: Symbolic Computations on Grids

Description of the activity :

The EU SCIEnce infrastructure project targets users of symbolic computing systems such as Maple. We aim to provide Grid-enabled symbolic computations through providing: (1) computational steering tools integrating seamlessly with existing tools; (2) the ability for users to identify symbolic components to form part of a Grid-enabled application; (3) adaptive resource brokers supporting the irregular workloads found in symbolic computations; (b4) large-scale heterogeneous demonstrators.

Grid added value :

Significant uses of symbolic computation systems, such as Maple, are found in areas of physics, mathematics, biology, chemistry and engineering disciplines. Worldwide, it is estimated that there are at least 1 million users. While symbolic computations may have extremely high computational and data demands, at present, however, only a few symbolic computation systems provide any form of Grid accessibility. Most of these provide little more than wrappers to standard Globus library calls, and none allows simultaneous deployment of Grid services, access to external Grid services, and the coupling of symbolic components into a coherent Grid application. In contrast, SymGrid provides sophisticated interactive computational steering interfaces, with simple and high-level access to Grid services. Common data and task interfaces allow complex computations to be constructed by orchestrating heterogeneous distributed components from various systems into a single system

Experience or proposed activity :

The two current components of SymGrid framework are SymGrid-Services and SymGrid-Par. SymGrid-Services complies with the WSRF standard for Grid services based on Globus Toolkit 4. A middleware package was designed in order to allow the easy-to-use access in a uniform way from any computer algebra system to Grid and Web services. It consists of a set of Java classes and a set of computer algebra system libraries. SymGrid-Par allows symbolic computations to be executed as parallel computations on a computational Grid. The SymGrid-Par middleware is built on the GRID-GUM, Grid-enabled implementation of Glasgow Parallel Haskell that provides various high-level parallelism services providing a flexible, adaptive environment for managing parallelism at various degrees of granularity. GRID-GUM uses MPICH-G2. Moreover, GRID-GUM has been designed with novel dynamic load scheduling mechanisms for shared hierarchical heterogeneous computational Grids based on Globus Toolkit 2 or higher.

Future evolution :

The delays in porting the EGEE tools towards new WSRF-compliant Grid architectures have created several discussions and problems in connecting EGEE sites from SCIEnce with new non-EGEE sites build upon Globus Toolkit 4. The SCIEnce team hopes that the new version gLite will solve this issue.

Primary authors : Prof. PETCU, Dana (Institute e-Austria Timisoara, and Western University of Timisoara, Romania) ; Dr. HAMMOND, Kevin (University of St. Andrews, UK) ; Dr. TRINDER, Phil (Heriot-Watt University, UK)

Co-authors :

Presenter : Dr. HAMMOND, Kevin (University of St. Andrews, UK)

Track classification : Related Projects

Contribution type : oral presentation

Submitted by : Prof. PETCU, Dana

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Related Projects

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 60

gLibrary: Digital Asset Management System for the Grid

Description of the activity :

In every Grid Infrastructure a huge amount of distributed storage resources is available to save user's data.
However, finding a specific file is not always easy. gLibrary challenge is to offer an extensible, secure and
easy-to-use system to handle digital assets stored as files on a Grid, providing an intuitive cataloging system that
allows to find assets in seconds. Target communities are both end-users and organizations that need a secure way
to save, share and organize their assets.

Grid added value :

gLibrary is entirely based on the Grid services provided by the gLite middleware. This implies the exploitation of the
benefits that comes by grid technologies usage, such as its built-in security and the availability of huge storage
resource for saving data. Users do not have to care about the complexity of the underlying systems and the
geographical location of their data and they can consider the available Grid storage as a huge virtual disk.
Permissions can be set on the users' assets in order to grant or deny access to given users, groups or even whole
organizations.All entries in gLibrary are organized according to their type, a list of specific attributes to describe each
kind of asset to be managed by the system. These are the same attributes that can be queried by users. Types can
be defined according to the assets users want to manage. The flexibility and extensibility offered by this type system
allow different communities to adopt gLibrary for any cataloging purpose

Experience or proposed activity :

A first prototype of gLibrary has been implemented as a web application available at https://glibrary.ct.infn.it The
front-end was developed as a Web 2.0 application to offer a pleasant user experience. This implies the usage of
technologies like AJAX, Javascript and Java Applets to offer a desktop-like interface. The business logic of the
application is built as PHP 5 objects that interact with several gLite services, such as AMGA Metadata Catalogs, VOMS
servers, Information Systems, File Catalogs and Storage Elements. Since AMGA is the most used grid service by this
web application, to guarantee an immediate user interface response, it has been decided to develop a low level PHP
5 API whose role is to interact directly with the AMGA server daemons. All communication between clients and the
application server are sent over HTTPS connections. Authentication on grid services makes use of VOMS-enabled
proxy certificates and all file transfers are carried over GridFTP channels.

Future evolution :

An issue that we faced during gLibrary developing is the inconsistency of the gLite API and the non-availability of
multiplatform API: almost all of the gLite Data Management APIs are provided for Scientific Linux Cern as C
and Python/Perl libraries. Moreover. VOMS APIs are not fully completed, lacking in methods to handle attribute
chains and again available only to one platform. gLibrary goal is to be a multiplatform tool, allowing users to handle
their assets from any operating system.

Primary authors : Dr. CALANDUCCI, Antonio (INFN Catania) ; Dr. CHERUBINO, Claudio (INFN Catania) ; Dr. NEUMANN CIUFFO, Leandro (INFN Catania) ; Dr. SCARDACI, Diego (INFN Catania)

Co-authors :

Presenter : Dr. CALANDUCCI, Antonio (INFN Catania)

Track classification : Data Management

Contribution type : oral presentation

Submitted by : Dr. CALANDUCCI, Tony

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :
On behalf of GILDA team

Status : SUBMITTED

Track judgments :

Track : Data Management
Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 61

ASAP Distributed Analysis

Description of the activity :

ASAP is a job creation and management framework used by CMS physicists to perform data analysis using the Grid. ASAP hides many of the low-level details of the Grid so that physicists can distribute their work across the Grid to take advantage of the available CPU power and access the stored data. There is also a server side component to which users can delegate responsibility for their jobs. The server monitors the status of jobs and resubmits in the case of Grid or application failure.

Grid added value :

The Grid provides the computing and storage resources necessary for the analysis and storage of large volumes of High Energy Physics data.

Experience or proposed activity :

ASAP provides a layer over the standard grid job management tools which means that users do not have to be concerned with the details of job creation, submission and monitoring. The presence of the server side component increases the chances of the users jobs being successfully completed.

Future evolution :

The main concerns are the performance and reliability of jobs. In addition documentation is often missing or inadequate.

Primary authors : MUNRO, Craig (Brunel University)

Co-authors : Dr. KHAN, Akram (Brunel University) ; ANDREEVA, Julia (CERN)

Presenter : MUNRO, Craig (Brunel University)

Track classification : Poster session

Contribution type : poster

Submitted by : MUNRO, Craig

Submitted on Tuesday 30 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 62

A WSRF-based Computational Gateway to the EGEE Infrastructure for the Simulation of Cardiac Electrical Activity

Description of the activity :

The understanding of cardiac electrical behaviour is essential to comprehend the genesis of cardiac arrhythmias and to develop new therapeutic methods based on the use of antiarrhythmic drugs. Computer simulations are an essential tool, as the mathematical models that predict ionic currents and propagating wavefronts have reached a high level of electrophysiological detail. However, these simulations are both computationally and memory intensive, producing a large amount of simulation data.

Grid added value :

The electrical activity enables to analyse the influence of certain drugs to heart diseases as well as the influence of different cellular parameters to the electrical activity. Moreover, cardiac case studies are inherently composed of lots of independent multi-parametric jobs, and the Grid is the appropriate infrastructure to accelerate their execution.

On average, we have estimated a requirement of 96 CPUs per day producing a total amount of data in the order of 120 GBytes. This would include research lines that investigate the genesis of reentrant activity under acutely ischemic (phase Ia) conditions, and the effects of antiarrhythmic drugs such as lidocaine or pinacidil. Potential user community is composed by cardiac electrophysiologists interested in using detailed cellular models to study effects on electrical activity. In addition, this activity fits within the goals of the Virtual Physiological Human by providing an effective simulation tool for action potential propagation.

Experience or proposed activity :

We have deployed a WSRF-based Grid metascheduler (GMarte) that enables biomedical experts to access a computational infrastructure, composed of different clusters of PCs within our research group. This provides transparent access to computing power for users. However, the multi-user support of the metascheduler combined with large cardiac studies can often overload these computing resources.

Therefore, we plan to provide a computational gateway to LCG-2, so that those simulations that can not be executed in our Grid are automatically delegated to a Resource Broker for its execution on a production Grid. This would require coordinated data staging to Storage Elements, using the functionality provided by a File Catalog to manage different replicas of simulation data. The usage of MyProxy is ideal to automatically renew the proxy for long simulations. Also, the Logging & Bookkepping service is fundamental to integrate the lifecycle of tasks within our metascheduler.

Future evolution :

The application is MPI-based, thus requiring parallel executions on LCG-2 resources. In addition, it uses MPI-2 I/O techniques for parallel I/O, what demands a shared file system among all the worker nodes. On the other hand, the application is complex enough to try an in-place compilation before execution. Thus, we are producing statically-linked self-contained executable files which we expect to be able to run. We would be interested in using Java APIs to interact with LCG-2 services.

Primary authors : Mr. ALONSO, Jose Miguel (Universidad Politécnica de Valencia) ; Mr. MOLTó, Germán (Universidad Politécnica de Valencia) ; Prof. FERRERO (JR.), Jose María (Universidad Politécnica de Valencia) ; Prof. HERNáNDEZ, Vicente (Universidad Politécnica de Valencia) ; Prof. SAIZ, Javier (Universidad Politécnica de Valencia)

Co-authors :

Presenter : Mr. MOLTó, Germán (Universidad Politécnica de Valencia)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Mr. MOLTó, Germán

Submitted on Wednesday 31 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 63

ELFI - Linux Filesystem Interface to EGEE Grid Storage

Description of the activity :

ELFI was initially targeted at the EGRID community. EGRID is an Italian National pilot grid infrastructure for economic and financial computation. EGRID software runs on top of gLite middleware in the EGEE/INFN-GRID infrastructure.

Grid added value :

The finance research community uses medium-size programmes, frequently modified and developed by non-professional programmers. The complexity of the GRID data management is a significant barrier to the use of the EGEE infrastructure.

ELFI hides the complexity of grid data management behind the familiar filesystem paradigm.
End-users can therefore access data on the EGEE Grid just like they access local files: they can even run graphical file managers on remote Grid data, with no need to learn technical details about Grid data management protocols.

Simplicity and ease of access to the Grid is a key feature to widespread usage of the Grid, it should appeal to all VOs - thus we think that the entire EGEE user community can benefit from ELFI.

Experience or proposed activity :

ELFI is a Linux filesystem, that acts as an interface to the LFC catalog and EGEE SE (both "classic" and SRM v2 flavor).

With ELFI, you can see the entries in the LFC catalog as files in a locally-mounted filesystem, and directly operate on the replica contents: read/write operations on the local filesystem are acted as read/write operations on a remote SE via the GSI-RFIO protocol. All operations on the catalog or the SE have a local filesystem equivalent.

ELFI features:

* Transparent file access via LFN only
* Posix ACL support
* Posix IO operations (through gsi-rfio transport protocol)
* Classic SE, SRM v2.1.1 and SRM v2.2 support

The ELFI filesystem process runs entirely in user-space: it uses the standard FUSE (http://fuse.sf.net) userland filesystem framework.

Future evolution :

Two main issues have affected the development of ELFI:

1) there are two incompatible flavors of the RFIO protocol, but they are both advertised as "rfio" on the Information System - so there is no way to filter out incompatible servers.

2) the information system contains data that is often not easily parsed by a program - it looks like the information is put there mostly for *humans* to read.

Primary authors : Dr. MESSINA, Antonio (ICTP/EGRID) ; Dr. MURRI, Riccardo (ICTP/EGRID)

Co-authors :

Presenter : Dr. MESSINA, Antonio (ICTP/EGRID)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : MESSINA, Antonio

Submitted on Wednesday 31 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 64

MARATRA: A Production Fusion Physics System

Description of the activity :

The target of "MAssive RAy TRAcing in Fusion Plasmas" is the Fusion Physics scientific community. In particular, MARATRA aids those community members who are working in the optimization of plasma heating by electron Bernstein waves (EBW).

Grid added value :

A microwave beam can be modeled as a set of rays, each one with an independent trajectory. The more rays that can be traced, the more accurate the results are. A regular case study in a cluster environment involves tracing a factor of 10^2 rays. The proposed optimization procedure in the case of EBW heating needs to increase two orders of magnitude the number of rays used in conventional ray tracing. The possibilities offered by a large-scale Grid infrastructure, and in particular EGEE, is to increase this factor to 10^3 and 10^4.

Experience or proposed activity :

For porting this application onto the Grid, we relied on the execution services offered by FUSION VO resources and used the GridWay Metascheduler (http://www.gridway.org/) as a Resource Broker. In particular, its workload management capabilities have been proved to be very valuable in order to scale the input data. These valuable capabilities comprehend Opportunistic Migration (automatic resubmission of running jobs to better resources), Performance Slowdown Detection (migration when a intolerable performance loss for a job, which happens to be case in the present application) and Fault Detection & Recovery (job cancellation, remote system crash or outage, network disconnection, client fault tolerance). Also, GridWay has been chosen because it implements the Open Grid Forum (http://www.ogf.org/) DRMAA Standard (http://drmaa.org/), both C and JAVA bindings, providing compatibility of applications with DRM systems that implement the Standard, such as SGE, Condor, ...

Future evolution :

Basically, we found the queuing times high and, in many occasions, many computing resources (downtime and even DN not mapped) were not available. Thanks to its fault tolerance mechanisms, GridWay could handle these situations, making its combination with the EGEE infrastructure a very reliable one.

Primary authors : Mr. VAZQUEZ-POLETTI, Jose Luis (Universidad Complutense de Madrid (Spain)) ; Dr. HUEDO CUESTA, Eduardo (Universidad Complutense de Madrid (Spain)) ; Dr. SANTIAGO MONTERO, Ruben (Universidad Complutense de Madrid (Spain)) ; Dr. MARTIN LLORENTE, Ignacio (Universidad Complutense de Madrid (Spain)) ; Dr. CAPPA ASCASIBAR, Alvaro (CIEMAT (Spain)) ; Dr. TERESHCHENKO, Maxim A. (GPI (Russia)) ; Dr. CASTEJON MAGAñA, Francisco (CIEMAT (Spain))

Co-authors :

Presenter : Mr. VAZQUEZ-POLETTI, Jose Luis (Universidad Complutense de Madrid (Spain))

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Mr. VAZQUEZ, Jose Luis

Submitted on Wednesday 31 January 2007

Last modified on : Monday 12 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 65

Protein Clustering on the EGEE

Description of the activity :

The target of this activity is the Bioinformatic scientific community, and in particular, those members who use a tool called "CD-HIT" which performs protein clustering on a protein sequence database. This consists in removing redundant sequences at a given sequence similarity level and generating a new database with the representatives only. This activity was proposed by CNIO (Spanish National Cancer Research Centre) and started in the context of the BioGridNet Program (www.biogridnet.org).

Grid added value :

As protein databases are growing up day after day, the clustering process on interesting datasets in a single machine is not feasible due to memory constrains. A Grid environment allows an adaptive database distribution in order to optimize its overall analysis. The complexity of the workflow inherent to "CD-HIT" needs a robust
framework able to handle it. In addition, this framework may be successfully used in other applications which result in a same type of workflow.

Experience or proposed activity :

For porting this application onto the Grid, we used the GridWay Metascheduler (http://www.gridway.org/) and relied on the execution services offered by BIOMED VO. GridWay is being used as a Resource Broker because its workflow management capabilities and interoperability have been proved to be very valuable. On the other hand, we are working with its Open Grid Forum (http://www.ogf.org/) DRMAA Standard (http://drmaa.org/) implementation (both C and JAVA bindings). Finally, GridWay has also been chosen because it allows interoperability with EGEE and GRIDIMadrid (http://www.gridimadrid.org/), which is a Globus-based regional testbed.

Future evolution :

We plan to start with production input data proposed by CNIO. In particular, with the analysis of various meta-genomes, starting with the first published one from Sargasso Sea.

Primary authors : Mr. VAZQUEZ-POLETTI, Jose Luis (Universidad Complutense de Madrid (Spain)) ; Dr. HUEDO CUESTA, Eduardo (Universidad Complutense de Madrid (Spain)) ; Dr. SANTIAGO MONTERO, Ruben (Universidad Complutense de Madrid (Spain)) ; Dr. MARTIN LLORENTE, Ignacio (Universidad Complutense de Madrid (Spain)) ; Dr. FERNANDEZ GONZALEZ, Jose Maria (CNIO (Spain)) ; Dr. VALENCIA HERRERA, Alfonso (CNIO (Spain))

Co-authors :

Presenter : Mr. VAZQUEZ-POLETTI, Jose Luis (Universidad Complutense de Madrid (Spain))

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Mr. VAZQUEZ, Jose Luis

Submitted on Wednesday 31 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 66

A GSI-secured jobmanager for connecting PBS servers in independent administrative domains

Description of the activity :

The "RemotePBS" jobmanager is aimed at security-conscious PBS server managers of major non-Grid computing facilities. It enables them to securely connect existing PBS servers to the Grid, even if the PBS server is in an separate administrative domain to the Grid servers. The PBS server admins have full control over authorisation of grid-authenticated users accessing their resources. The software has been designed to support execution of grid applications at large-scale computing centres.

Grid added value :

The EGEE middleware assumes the gatekeeper and PBS system enjoy an implicit trust relationship, an anathema to security-conscious admins of supercomputing centres. Also the EGEE model assumes the PBS is configured to suit EGEE and is typically installed on a gatekeeper. Major sites already have PBS servers, often tuned over years. To have access to these sites, these issues must be redressed. Our jobmanager does this.

The jobmanager, with modified information publishers, allows multiple PBS servers to be attached to the Grid via a single gatekeeper. Interactions between the CE and PBS are GSI-secured. The design makes it easy for PBS servers to manage existing nodes that only need to have standard WN software installed. It is extensible to allow some batch related information to be passed into the job. In addition, the environment can be customized whenever any enqueued Grid jobs are executed. The burden and cost of running a full Grid site is reduced for the PBS administrator.

Experience or proposed activity :

RemotePBS is installed at four Grid-Ireland sites in EGEE. We have tested over 6000 jobs, including MPI codes, with good results. A stress-test run of 3000 small jobs completed with a 100% success rate. We have identified a number of key weaknesses in current jobmanagers. The "RemotePBS" manager is still in an experimental stage, and a number of bug fixes and enhancements are likely. Given the lightweight requirements required from the remote resources, it is hoped that jobmanager would connect to the Grid.

Future evolution :

Current Grid middleware does not support the Grid model we had envisaged deploying at a national level. Lack of MPI support, portability of middleware, and the facility to connect remote queue managers are still issues. The gLite CE assumes an implicit unsecured trust between the CE and the queue manager and does not accommodate independent administrative domains. In addition, only one remote queue manager per CE is allowed. We expect to to implement a comparable gLite CE version presently.

Primary authors : Mr. WALSH, John (EGEE SA1 & SA3/Grid-Ireland/Trinity College Dublin)

Co-authors : Dr. COGHLAN, Brian (EGEE/Grid-Ireland/Trinity College Dublin) ; Dr. CHILDS, Stephen (EGEE SA1/Trinity College Dublin) ; Dr. KENNY, Eamonn (EGEE SA3/Trinity College Dublin/WebCom-G)

Presenter : Mr. WALSH, John (EGEE SA1 & SA3/Grid-Ireland/Trinity College Dublin)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : WALSH, John

Submitted on Wednesday 31 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 67

Managing Mass Storage with Grid Jobs

Description of the activity :

GridPP is the UK Grid for particle physics, and covers the UK participation in the Large Hadron Collider Computing Grid (LCG) and also supports other scientific communities, including the EGEE biomedical community. Users are located not just in the UK, but across the world: each community has a need to transfer data to or from collaborator sites in the UK, and to analyse the data with grid jobs and to store the output. The work presented here is also relevant to the UK National Grid Service, NGS

Grid added value :

One thing all these communities have in common is the need to store and manage scientific data. Standardised interfaces are essential to ensure that the data can be accessed regardless of the type of storage system: from a single disk running on a PC, over large dedicated disk servers, to the largest tape robots. From the Grid's view, a storage system is a Storage Element, or SE. Each SE provides control interfaces, data transfer interfaces, and information systems, which in turn are used by client applications or higher level Grid services to manage data consistently. The storage group in GridPP supports different storage solutions to enable UK Grid communities to participate in global Grid collaborations with SEs that interoperate with other SEs across the world. Variations in sites' infrastructure are essentially hidden from the grid clients, and SEs fit together to form a global data grid. The NGS is evaluating these SEs as well.

Experience or proposed activity :

We aim to demonstrate interoperability between different storage systems by showing data being accessed or transferred by Grid jobs on behalf of the user, or by higher level transfer services that schedule and manage data replication between the sites. Common use cases include regular transfers of data from one site to another, replication of "popular" files to improve access time, and fetching data to local disk before the job runs for efficient access. This demo thus does not just depend on the SEs, but also on the job submission system and high level transfer services. We also depend on the networks for the transfer, and aim to show different transfer rates achieved, depending on a site's network bandwidth and how busy the SE is. We aim to demonstrate this with a live demo, showing the details of the file transfers as they happen on the Grid.

Future evolution :

Providing a standard interface to storage is essential but it must also have sufficient functionality to support the different use cases. Traditionally, getting SEs to interoperate has taken lots of effort, particularly with specifications being loose or leaving things optional. A dependable interoperating core is essential, and this requires lots of testing. Clients needing advanced functionality must be prepared to cope with systems that don't provide it.

Primary authors : Dr. JENSEN, Jens (CLRC-RAL)

Co-authors : Dr. COWAN, Greig (Edinburgh) ; Dr. HODGES, Matthew (CLRC-RAL) ; Dr. KANT, David (CLRC-RAL) ; Mr. GARRETT, Barney (Edinburgh) ; Dr. MOONT, Gidon (Imperial College)

Presenter : Dr. JENSEN, Jens (CLRC-RAL) ; Dr. COWAN, Greig (Edinburgh) ; Dr. HODGES, Matthew (CLRC-RAL) ; Dr. KANT, David (CLRC-RAL) ; Mr. GARRETT, Barney (Edinburgh) ; Dr. MOONT, Gidon (Imperial College)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : JENSEN, Jens

Submitted on Wednesday 31 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 68

Grid Computing for Kinetic Transport in Magnetic Fusion Devices

Description of the activity :

The study of ion trajectories in low collisionality plasmas is very important for understanding transport in tokamaks and stellarators. Several issues make this study useful to understand the confinement in those devices. One of them is the kinetic transport and the behaviour of particles in a given magnetic configuration. We make a transport estimate for the TJ-II device, which is a medium size flexible heliac characterized by a complex magnetic configuration.

Grid added value :

The customary neoclassical transport estimates assume that the transport coefficients depend only on the local plasma characteristics. In TJ-II, the ion particle orbits include large radial excursions in a single collision time. The very same particle thus visit plasma regions of widely differing conditions, which invalidates the local approximation. Therefore, global plasma characteristics must be taken into account in more accurate transport estimates for TJ-II. With this in mind, one can find an equivalence between a Fokker-Planck equation (which in our case describes the evolution of the particle distribution a plasma in the presence of electric field and collisions) and a Langevin equation. This approach it is ideally suited for massive parallel computing in Grids. We expect the Langevin approach to be useful to study transport in devices where deviations from the local hypothesis are sizeable.

Experience or proposed activity :

We have developed a computer code, ISDEP, that solves the guiding-centre equations in the presence of collisions. One million particles are followed in a realistic TJ-II magnetic configuration for a time (for the first time in TJ-II) comparable to the discharge duration. The large number of trajectories as well as a careful data analysis allow us to obtain accurate estimates of the time evolution of several quantities of interest. The method is specially appropriated for a complex magnetic configuration like that of TJ-II. Global features of transport, not present in the customary neoclassical models, appear: a monotonic increasing of heat and particle fluxes with minor radius, the non-diffusive character of transport, the appearance of asymmetries on the magnetic surfaces and the non-Maxwellian character of the distribution function.

Future evolution :

The next step will be to let the background plasma vary due to the evolution of the test particles followed by the code. The approach consisting on keeping constant the background does not allow the study of evolving plasmas and dissipates any perturbation. The application must, hence, be modified to be able to be run iteratively and self-consistently in the Grid.

Primary authors : Mr. VELASCO, José Luis (Instituto de Biocomputación y Física de Sistemas Complejos / Departamento de Física Teórica, Universidad de Zaragoza)

Co-authors : Dr. CASTEJON, Francisco (Instituto de Biocomputación y Física de Sistemas Complejos / Laboratorio Nacional de Fusión (Asociación Euratom/Ciemat)) ; Dr. FERNáNDEZ, Luis Antonio (Instituto de Biocomputación y Física de Sistemas Complejos / Departamento de Física Teórica I, Universidad Complutense de Madrid) ; Dr. GUASP, José (Laboratorio Nacional de Fusión (Asociación Euratom/Ciemat)) ; Dr. MARTIN-MAYOR, Víctor (Instituto de Biocomputación y Física de Sistemas Complejos / Departamento de Física Teórica I, Universidad Complutense de Madrid) ; Dr. TARANCóN, Alfonso (Instituto de Biocomputación y Física de Sistemas Complejos / Departamento de Física Teórica, Universidad de Zaragoza)

Presenter : Mr. VELASCO, José Luis (Instituto de Biocomputación y Física de Sistemas Complejos / Departamento de Física Teórica, Universidad de Zaragoza)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Mr. VELASCO, José Luis

Submitted on Wednesday 31 January 2007

Last modified on : Friday 09 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 69

Application Architecture for High Performance Microarray Experiments over the Hellas-Grid Infrastructure

Description of the activity :

Microarray experiments permit a genome-scale evaluation of gene functions and are therefore among the most topical and prominent developments of biomedical research. Our target user group includes: Doctor, researchers, biologists, bioinformatics specialists carrying out DNA microarray experiments.

Grid added value :

Economic parameters are a major obstacle to microarray experiments, since some processes are extremely computationally demanding. This is particularly true for the normalization process, which deals with the fact that every repeat experiment will give rise to a certain amount of variations. Variations can be minimized based on statistical methods, which in turn allows one to compare the expression levels between multiple microarray experiments. Normalisation procedures rely on the fact that gene expression data can follow a normal distribution. Thus the entire distribution can be transformed about the population mean and median without affecting the standard deviation. cDNA arrays can contain up to 25,000 gene-complementary seqsequences while high-density oligonucleotide arrays can hold more than 100,000 proceses.
Grid computing can accelerate the computations associated with microarray normalization, while at the same time providing access to vast amounts of federated storage.

Experience or proposed activity :

Bioinformatics applications in general and microarray experiments in particular are perfectly tailored to Grid infrastructures. Motivated by this fact, University of the Aegean proposed to 'Gridify' a selected number of microarray analysis and normalization applications. These applications focus on cDNA arrays. The target Grid infrastructure was the Hellas Grid portion of the pan-European Grid infrastructure developed for e-science in the scope of the EGEE (Enabling Grids for E-Science in Europe) project and its successors. The idea was proposed in the scope of the Grid-App tender issued by the Greek General Secretariat of Research and Technology (GSRT) and was warmly embraced by the evaluators.

The application will leverage a host of Hellas Grid (i.e. gLite) middleware middleware services inlcuding security, information services, data management, job submission, and resource management . The paper elaborates on the use of these services based on our experiences.

Future evolution :

Our recently started developments involve the following steps:
I)Transforming our microarray applications (from MATLAB, which is not currently supported on HellasGrid.) to supported executables such as the open-source GNU Octave (forge) .
II)Exploiting appropriate gLite middleware components in the areas of security (GIS), information services (R-GMA), data management (LFC), job submission (WMS), and resource management.
III) Implementing an access portal based on UI images available.

Primary authors : Prof. SOLDATOS, John (Athens Information Technology)

Co-authors : Prof. MAGLOGIANNIS, Ilias (University of the Aegean) ; Dr. CHATZIOANNOU, Aris (University of the Aegean) ; Mr. KANARIS, Ioannis (University of the Aegean) ; Mr. MYLONAKIS, Vasileios (National Technical University of Athens)

Presenter : Prof. SOLDATOS, John (Athens Information Technology)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Prof. SOLDATOS, John

Submitted on Wednesday 31 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 70

Application of a GRID Technology for a Monte Carlo Simulation of Elekta Gamma Knife

Description of the activity :

Number of patients cared with radiotherapy methods has recently increased due to the development of linac and multi leaf technologies. The high precision in the dose distribution possible with the new machines needs the development of more accurate codes to determine the treatment plan. MonteCarlo methods allow big precision in the dose distribution calculation but are very CPU time consuming. The application makes use of the power of the GRID to validate treatment plan systems (TPS).

Grid added value :

The GRID can represent the solution for the large computing requirements of Monte Carlo applications in radiotherapy which are embarrassingly parallel. The GRID can, in fact, be of great utility to provide physicians with a fast computing resource able to calculate detailed and precise Monte Carlo based treatment plans without requiring the hospital to set-up and manage a dedicated computing centre.
In this contribution we refer to a Monte Carlo application created to simulate the Elekta Gamma Knife, a stereotactic radiotherapy machine to treat the brain lesions. The application makes use of the Geant4 toolkit and reproduces all Gamma Knife features (like collimators' geometry) so to simulate a realistic energy deposition.

Experience or proposed activity :

In order to run a Geant4 application on the GRID, the first step was to install the used libraries at the various sites. The second step was to prepare a shell script to automatically submit jobs and download the outputs at their end. In this respect, it has been very useful to set a threshold in the number of jobs "done" before starting
the downloading & merging phase. The last step was to choose the best quality factor between the number of events and the number of jobs to be submitted.
The key services, that are essential to run a TPS Monte Carlo application on GRID, are a low-latency Workload Management System and a more reliable monitoring system of available computing resources.

Future evolution :

Sometimes jobs stay "Scheduled" for a very long time even if there are some free CPUs
available. WMS, on matchmaking phase, does not take in account the real status of resources. In fact, during the submission of a large production, the matchmaking of the n-th job does not take into account that the (n-1) jobs have just been submitted, so jobs are not uniformly spread across the CEs.
Another problem is the low efficiency (about 60%) due to many jobs aborted for unknown reasons.

Primary authors : Dr. GIORGIO, Emidio (INFN Catania)

Co-authors :

Presenter : Dr. GIORGIO, Emidio (INFN Catania)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : GIORGIO, Emidlo

Submitted on Wednesday 31 January 2007

Last modified on : Saturday 10 February 2007

Comments :

The primary author is intended to be just as the presenter. Authors are □
□
R. Barbera – Univ. of Catania and INFN□
G.A.P. Cirrone – LNS-INFN □
G. Cuttone – LNS-INFN□
F. Di Rosa – LNS-INFN□
E. Giorgio – INFN Catania□
F. Romano – LNS-INFN□
G. Russo – LNS-INFN□
M. G. Sabini – LNS-INFN, A.O. Cannizzaro

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 71

Porting a complex workflow on EGEE infrastructure: The case study of WIEN2k

Description of the activity :

The target community is material science. The application called WIEN2k performs electronic structure calculations of solids using density functional theory. It is based on the full-potential (linearized) augmented plane-wave ((L)APW), one among the most accurate schemes for band structure calculations. WIEN2k is used by more than 1000 institutes worldwide in academia and industry. This work aims at enabling exiting tools (w2web web portal) to be able to use EGEE infrastructure.

Grid added value :

WIEN2k is composed of several packages which form a complex grid workflow, including complex data-dependencies (which are separated from the control-flow), parallel sections and loops (not supported by basic job submission capabilities of glite, currently limited to a DAG for example). Our experience will give a useful input to help porting other scientific applications with similar characteristics. In addition, WIEN2k is a licensed application and existing solutions, e.g. creating a full VO per application, are not satisfactory. Institutions having already a license would like to use EGEE infrastructure, but not necessarely be bound to a VO, and at the same time, the deployed code on the Grid should not be accessible to non-licensed users. Our proposed solution is based on a combination of existing capabilities of the middleware, using certificate roles and access control.

Experience or proposed activity :

This is work-in-progress and part of NA4 effort. We are using our own glite testbed based on virtual machine technology. Using our own infrastructure was necessary, because of licenses issues. Once validated, we plan to deploy our solution in Gilda, and finally on VOCE. One important step will be to show the WIEN2k community the complete integration of the glite-enabled tools into the tool (e.g. w2web web portal) they are familiar with. The key services for implementing our solution will be VOMS and the deployment capabilities based on certificate roles. Another key element for the implementation of the workflow engine is the Ganga toolkit.

Future evolution :

One of the missing functionalities which would simplify the development of the WIEN2K workflow is the support of loops and the ability to separate control-flow from data-flow. Based on our experience with the ASKALON grid environment (which supports both features, but currently is not yet glite-enabled), we are trying to provide a similar but simplified solution for the WIEN2k glite porting.

Primary authors : Dr. VILLAZON, Alex (University of Innsbruck)

Co-authors : Mr. SIDDIQUI, Mumtaz (University of Innsbruck) ; Mr. JUNAID, Malik (University of Innsbruck) ; Prof. FAHRINGER, Thomas (University of Innsbruck)

Presenter : Dr. VILLAZON, Alex (University of Innsbruck)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Dr. VILLAZON TORRICO, Alex Stephan

Submitted on Wednesday 31 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 72

Data-intensive automated construction of phenomenological plasma models for the advanced tokamaks control

Description of the activity :

The activity aims to offer a method for extraction of the most significant dynamical input-output dependencies from the large data sets using the Grid. This work has mainly been done to generalize cumulated observations in the nuclear fusion science for the purpose of model-based controllers improvement, though an approach can be employed for any domain where a short-range prediction is needed.

Grid added value :

The main problem is that a mass of data acquired from large experiments could hardly be processed. System identification and generalization of observations problems are well-known in data mining. When the studied processes are complex and involve a great number of variables, solving this task by the ordinary means is too time-consuming. Grid infrastructure enables an implementation of methods for the conversion of multidimensional raw data into useful information. It seems encouraging for the technical domains where reliable real-time prediction and control is critical.

Experience or proposed activity :

To solve these multipurpose optimization problems the authors have exploited some kind of genetic algorithm (GA). GA implies a large number of independent calculations, so it can be easily parallelized by means of grid computing.
A testing model of plasma transport inside ITER has been found. Plasma transport has been simulated by ASTRA code under various physical conditions (initial plasma state, controlled plasma heating) in 100,000 runs. Then, the optimal structure of ordinary differential equations (ODE) describing plasma transport depending on external influence and plasma state has been derived. The ASTRA simulations and the optimization were performed by GA method using Russian Data Intensive Grid infrastructure.
To implement the Grid computing technique heterogeneous software has been glued by Python and Shell scripts. Though such an approach is flexible enough, it becomes rather tangled with the growth of number of software components.

Future evolution :

1. It would be nice to have some automatic regular middleware means for dynamical
scheduling of new jobs based on the results of already completed ones. A good example of dynamical finding of good-enough solution is the computer chess.
2. Native storages representation. There are many experimental non-grid databases
using a variety of data formats. While solving some task only a small part of the data is usually needed, so it is unpractical to transfer all the data into storage and computing e

Primary authors : Mr. MARUSOV, Nikolay (RRC Kurchatov Institute) ; Mr. VOZNESENSKY, Vladimir (RRC Kurchatov Institute)

Co-authors : Mr. SEMENOV, Igor (RRC Kurchatov Institute) ; Prof. SOLDATOV, Aleksey (RRC Kurchatov Institute)

Presenter : Mr. MARUSOV, Nikolay (RRC Kurchatov Institute)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : NIKOLAY, Marusov

Submitted on Wednesday 31 January 2007

Last modified on : Saturday 10 February 2007

Comments :

By now in the Nuclear Fusion science, only a small part of the data acquired from the experiments can be manually processed due to the great amount of data and the difficulty of interpretation. Data flow to be acquired from the International Thermonuclear Reactor (ITER) by estimation will be extremely high, of the order of 1GB per second.

This presentation is devoted to an approach for the generalization of hot plasma behavior from the all available data by means of the grid.

In the nowaday fusion community there are no conventional approaches for getting robust plasma behavior models suitable for immediate analytical synthesis of feedback controllers. The complexity of mathematical formulation, lack of analytical methods and engineering limits impel using of simplified or phenomenological models of processes in plasma. On the other hand, optimal control methods for nonlinear multilinked objects described by ODE are well-developed. That is why the authors have developed a set of tools for construction of such ODE description based on the modeling or experiments results.

Such construction implies at least two problems to solve: 1) discovery of the statistical input-output dependencies and 2) structural system identification. Both problems are typical to the statistics domain and can be solved with the established methods, but high dimensionality and great amount of data lead to considerable computational cost. In either case an optimal ODEs structure of the studied system should be found.

To solve these multipurpose optimization problems the authors have exploited some kind of genetic algorithm (GA). GA implies a large number of independent calculations, so it can be easily parallelized by means of grid computing.

A testing model of plasma transport inside ITER has been found. Plasma transport has been simulated by ASTRA code [1] under various physical conditions (initial plasma state, controlled plasma heating) in 100,000 runs. Then, the optimal structure of ODE describing plasma transport depending on external influence and plasma state has been derived. The ASTRA simulations and the optimization were performed by GA method using Russian Data Intensive Grid infrastructure.

The proposed technique makes possible organizing the successive improvement of model-based feedback controllers both for existing and prospective facilities as the experimental data base grows.

[1] http://edoc.mpg.de/282186

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 73

AMGA WI: the AMGA Web Interface to Remotely Access Metadata

Description of the activity :

The AMGA Web Interface Project is carried out by INFN Catania and IR&T Engineering s.r.l.(a SME located in Catania, http://www.irt-engineering,.com). INFN leads research activities to port several Industrial Use Cases over the Grid and the IR&T aims to design a Digital Archive for Cultural Heritage that adopts Grid as a Content Management System. Since metadata is at the base of cataloguing activity we considered AMGA as support for archiving functionalities in our feasibility study.

Grid added value :

Usually, the users use a suite of command line tools installed into specified machines called User Interfaces. Typically, this approach is for expert users only and it brings several limits: user account on a UI, troubles with net access and firewalls, VPN client installation on the user machine. Moreover, commands must be typed exactly requiring a pretty good knowledge of syntax and a steep learning curve is necessary to get new users started. AMGA WI is designed for the end user to provide a visual mode to work on the AMGA Server. The web interface is very user friendly and there are not dependencies from the UIs. The interaction is immediate by means of simple and schematic service presentation pages. A simple internet connection is enough to use the service and wizards represents a fast and simple way to train and teach users quickly. However, the most important point is that no syntax knowledge is required and users do not loose the high level conception of their activities.

Experience or proposed activity :

AMGA WI allows users to access to the AMGA Server from any platform and from anywhere (via internet HTTPS connection). The user just needs a web browser, an internet connection and a valid VOMS proxy to be authenticated to the AMGA Server. After a successful login he will able to interact with the AMGA Server and use the same functionalities of AMGA as he/she would have been using the command line client. The user is able to browse the hierarchy of AMGA collections, to inspect their schema and permissions and to list their entries. He/she also can create a new collection, to define a metadata schema for it and modify this one adding/deleting attributes. Finally, he/she can add/edit/delete entries. The application provides the user with a tool to define and execute queries in order to find entries that match against specified conditions. All previous operations are performed according to the role that the user has in the VO he/she belongs to (this is done using the VOMS features).

Future evolution :

AMGA WI is built on top of the AMGA API. We did not have encounter many issues in using the gLite middleware during the development of the AMGA WI application. We had just some troubles (they revealed already known bugs) with the API but their developers gave us fast support. The only issue that deserves to be underlined regards the impossibility to design a “Connection Pool” for the AMGA Server since each AMGA session is tightly dependent on the security credentials of the current logged user

Primary authors : Dr. SCIFO, Salvatore (INFN)

Co-authors : Dr. MILAZZO, Vincenzo (IR&T engineering s.r.l.)

Presenter : Dr. SCIFO, Salvatore (INFN)

Track classification : Data Management

Contribution type : --not specified--

Submitted by : Dr. SCIFO, Salvatore

Submitted on Wednesday 31 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Application (prototype) is deployed within the GILDA test bed than it is □
available at the following URL : https://amga.ct.infn.it:8443/amgawi/.□
User Manual is published here : □
https://grid.ct.infn.it/twiki/bin/view/TRIGRID/AMGAWI.□
□
Best Regards,□
Salvatore Scifo

Status : SUBMITTED

Track judgments :

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 74

DILIGENT - from document into computers to content across infrastructures

Description of the activity :

DILIGENT aims at develop a service layer on top of grid infrastructure enabling several and different communities to share and manage content. Two case history will be provided one from environmental science, coordinated by the European Space Agency) and one from cultural heritage, coordinated by Scuola Normale Superiore di Pisa.
The service layers has been developed in compliance with WSRF references and is based on gLite 3.0

Grid added value :

DILIGENT results will be of paramount relevance for the grid community as well as the Digital Library community. Concerning the first one, the problem of managing and share contents (thus including any type of data in this broad concept) is common to any scientific community. Other than the two communities included in the project proposal, a third one will be targeted in the next months. In particular the Exploitation team will analyze the potential impact of the usage on DILIGENT services in specific experiments of the BioMed community.
Moreover DILIGENT is one of the first example of usage of the grid technology in conformity of WSRF references, the full exploitation of grid-services supplied by gLite and the extension of AuthN&AuthZ mechanism to provide a more-fine-grained access to single items.

Experience or proposed activity :

Currently (Jan '07) DILIGENT is about to publish the alpha version of the system whose code that for obvious reasons of confidentiality will keep private to the consortium. Nevertheless the two scenarios have implemented two different public portals to show the initial features of the system, try it and provide initial feedbacks to the development teams.
Further services will be added in the next months and will be accessible to any other communities interested in see tangible results of the project.

Future evolution :

All technical problems met have been already reported trough the different mailing list and working groups and almost all have been promptly solved. Some issues remain open as the performance of some core services, the policies for short term job submission, the adoption of de-facto standards for developing grid-services.

Primary authors : Mr. MANIERI, Andrea (Engineering Ingegneria Informatica s.p.a.)

Co-authors :

Presenter : Mr. MANIERI, Andrea (Engineering Ingegneria Informatica s.p.a.)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : MANIERI, Andrea

Submitted on Wednesday 14 February 2007

Last modified on : Wednesday 14 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 75

GSAF: GRID Storage Access Framework

Description of the activity :

The GSAF Project is carried out by INFN Catania and IR&T Engineering s.r.l.(a SME located in Catania, http://www.irt-engineering,.com). INFN leads research activities to port Industrial Use Cases over the Grid and the IR&T aims to design a web applications that adopts Grid as a Content Management System. We analyzed the high level requirements of a classic web application in order to identify common points and shared problems and isolate peculiarities for porting it over the Grid infrastructure.

Grid added value :

Every service of DMS is independent from each other and works in a "stand alone" mode. Then, applications must refer to those services in a decoupled way. This fragmentariness obliges software engineers and web designers to consider a vertical architecture in order to access any single service and this always induces the same problem: application must take care of coherence and synchronization of data manipulation. From the end user side it is possible (using CLI tools), to manage objects across file catalogue rather than the metadata catalogue or storage element but all logical relationships must be kept clearly in the mind and can not be demanded to any framework.. We focused our effort on both of these two different aspects. We created an Object Oriented framework to help developers to write Java applications that use Grid as Data Access Layer. Subsequently, we tried to satisfy end user needs developing an easy-to-use web interface to access Grid DMS remotely.

Experience or proposed activity :

GSAF hides the complexity and the fragmentation of the underlying APIs (one for each data service) and ensures both coherence and flexibility to the underlying application. GSAF, both as framework and application, interacts to the Storage Element, File Catalogue, and Metadata Catalogue allowing both applications (through the APIs) and users (by means of the web interface) to perform the following functionalities: 1) collection management (collection tree browsing, collection creation, collection deletion); 2) entry management (entry list, entry search, entry creation, entry deletion); 3) metadata schema management (attribute list, attribute creation, attribute deletion); 4) File management (upload to SE, download from SE, register file on File Catalogue, deregister file from File Catalogue); 5) Access Control (login based on VOMS proxy file); 6) ACLs for collections (owner permission list, change access mode, change owner, group list, group add, group deletion).

Future evolution :

1)AMGA does not allow separation between user X.509 credentials and the software connection object. 2)There were not Java GFAL APIs and we wrote them by ourselves wrapping the C APIs. 3)There is some unexpected incompatibility between GFAL and lcg-* tools, uploading a file to a SE and registering its entry into LFC at the same time looks impossible. 4)Existing APIs seem to support only a single user logged while we need to bound the proxy file with the working session of the user.

Primary authors : Dr. SCIFO, Salvatore (INFN)

Co-authors :

Presenter : Dr. SCIFO, Salvatore (INFN)

Track classification : Data Management

Contribution type : --not specified--

Submitted by : Dr. SCIFO, Salvatore

Submitted on Wednesday 31 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Application (prototype) will be deployed within the GILDA test bed soon □
and its URL will be notified.□
□
User Manual is published here : □
https://grid.ct.infn.it/twiki/bin/view/TRIGRID/GSAF.□
□
Best Regards,□
Salvatore Scifo□

Status : SUBMITTED

Track judgments :

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 76

AWare - An easy Way to Access GRID REsources

Description of the activity :

AWare is a European Grid technology project aimed at the development of a opensource,userfriendly solution,to create,store and manage complex job workflows.There are many communities and related Grid projects within EGEE(eg ALICE,LHCb,BioMed,INFN,etc)already using the glite middleware and many more potential users from diverse industrial & business community(eg manufacturing,automotive,financial services,etc)that could benefit from a framework able to manage complex job flows and their lifecycle

Grid added value :

A-WARE (An easy Way to Access GRID REsources) will develop a stable, supported, commercially exploitable and high quality technology, framework to manage workflows, from design to storage,including submission, monitoring,flexible retrieval,data management and incarnation on virtualized grid resources. All these functions will be performed in a very simple, interoperable and application independent way. The guiding principle of the project as stated in the AWare technical annex, will be to exploit and leverage the maturity of the grid middleware in particular gLite middleware support as one of the project outcomes. Another guiding principle of the project is the development of a technology help diffuse and widen the adoption of Grid technology via a web thinclient approach, based on userfriendly interfaces, that hides the complexity of the underlying Grid middleware to occasional as well as frequent users. Finally industrial exploitation and sustainability are key aspects of the project

Experience or proposed activity :

An evident synergy point is that both the A-WARE framework and the Grid portal of Gilda community (GENIUS) are both developed and running on the EnginFrame system, with common goals: to improve user friendliness, to focus on innovative and flexible workflow management systems and hides the complexity of the Grid from users. Genius is a powerful Grid Portal that allows scientists to exploit Grid resources and allows to expose GLite-enabled applications via Web-browser as well as Web Services. To facilitate the exploitation, particular care will be put on simplifying the deployment of Grid Applications under typical research scenarios, such as security, single-sign-on systems, ACL management, etc. Finally the project will aim to adopt as well as influence common used standards in all key project areas. A-WARE will primarily target the Grid Foundations topic of the above-mentioned strategic objective..

Future evolution :

To date, the EGEE community does not have an established solution to use and fully support complex and generic workflows. A-WARE is able to integrate different workflow engines (such as MOTEUR, etc.) for different standards (like BPEL, SCUFL, etc.). With all these features A-Ware satisfies most of requirements and needs of different user communities that aim at a powerful workflow management.

Primary authors : Dr. VENUTI, Nicola (NICE srl)

Co-authors : Dr. MELATO, Maurizio (NICE srl) ; Dr. CURTONI, Andrea (NICE srl) ; Dr. TORTEROLO, Livia (NICE srl)

Presenter : Dr. VENUTI, Nicola (NICE srl)

Track classification : Workflow ; Data Management

Contribution type : oral presentation

Submitted by : Dr. VENUTI, Nicola

Submitted on Wednesday 31 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 77

AWare - An easy Way to Access GRID REsources

Description of the activity :

AWare is a European Grid technology project aimed at the development of a opensource,userfriendly solution,to create,store and manage complex job workflows.There are many communities and related Grid projects within EGEE(eg ALICE,LHCb,BioMed,INFN,etc)already using the glite middleware and many more potential users from diverse industrial & business community(eg manufacturing,automotive,financial services,etc)that could benefit from a framework able to manage complex job flows and their lifecycle

Grid added value :

A-WARE (An easy Way to Access GRID REsources) will develop a stable, supported, commercially exploitable and high quality technology, framework to manage workflows, from design to storage,including submission, monitoring,flexible retrieval,data management and incarnation on virtualized grid resources. All these functions will be performed in a very simple, interoperable and application independent way. The guiding principle of the project as stated in the AWare technical annex, will be to exploit and leverage the maturity of the grid middleware in particular gLite middleware support as one of the project outcomes. Another guiding principle of the project is the development of a technology help diffuse and widen the adoption of Grid technology via a web thinclient approach, based on userfriendly interfaces, that hides the complexity of the underlying Grid middleware to occasional as well as frequent users. Finally industrial exploitation and sustainability are key aspects of the project

Experience or proposed activity :

An evident synergy point is that both the A-WARE framework and the Grid portal of Gilda community (GENIUS) are both developed and running on the EnginFrame system, with common goals: to improve user friendliness, to focus on innovative and flexible workflow management systems and hides the complexity of the Grid from users. Genius is a powerful Grid Portal that allows scientists to exploit Grid resources and allows to expose GLite-enabled applications via Web-browser as well as Web Services. To facilitate the exploitation, particular care will be put on simplifying the deployment of Grid Applications under typical research scenarios, such as security, single-sign-on systems, ACL management, etc. Finally the project will aim to adopt as well as influence common used standards in all key project areas. A-WARE will primarily target the Grid Foundations topic of the above-mentioned strategic objective.

Future evolution :

To date, the EGEE community does not have an established solution to use and fully support complex and generic workflows. A-WARE is able to integrate different workflow engines (such as MOTEUR, etc.) for different standards (like BPEL, SCUFL, etc.). With all these features A-Ware satisfies most of requirements and needs of different user communities that aim at a powerful workflow management.

Primary authors : Dr. VENUTI, Nicola (NICE srl)

Co-authors : Dr. MELATO, Maurizio (NICE srl) ; Dr. CURTONI, Andrea (NICE srl) ; Dr. TORTEROLO, Livia (NICE srl)

Presenter : Dr. VENUTI, Nicola (NICE srl)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : Dr. VENUTI, Nicola

Submitted on Wednesday 31 January 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 78

BMPortal – A Bio Medical Informatics Framework

Description of the activity :

Bioinformatics and Medical Informatics are converging disciplines. Today, traditional clinical health records are going to evolve in a new branch often identified as Bio Medical Informatics in order to include genomic data for each patient. Both genomic and modern medical data acquisition and processing demand large storage facilities and harness lot of computing power. Distribution of data is a key issue. Both researcher and medical doctors require an ubiquitous access to patient's data.

Grid added value :

Today EGEE provides enough computing power, storage resources and security features to guarantee adequate adoption of it in modern medical distributed context. Certificate based security and advanced feature on data and metadata federation, coupled with strong and fine grained security policies on users and groups available through VOMS, allow to replicate on a EGEE VO real world organizations with their own hierarchies. Distributed data storage and replication could trigger the adoption of all –online patients archives and health records-avoiding low latencies in accessing
old patient data stored on magnetic or optic backup devices. The cost per Gbyte and the availability of low cost commodity hardware could help health care providers in defining new kind of storage management systems, without concentrating patient records on centralised data centre.EGEE data access utilities are used to access data remotely, reducing copies of data in conformance of most common privacy medical issues

Experience or proposed activity :

Distributed medical imaging process of rare data has been proven to be successful and of clinician real interest. The unprecedented Grid possibilities to access distributed data in respect to key security and privacy features has been proven to be the “killer feature” of our previous EGEE experiences.The Grid has been proven to
be useful also when low amount of CPU power is used,due to its data and metadata management functionalities.Activities to deploy test application in production contexts (hospitals) are already planned and will be carried out in next months.There is however still the need for low latency (realtime or interactive) job execution tools, where the CPU power is not relevant but the response time is perceived as the key factor. Medical doctors are demanding high responsive tools, provided through transparent, user friendly infrastructures. There is still the need of a dedicated framework for the deployment of bioinformatics and medical informatics integrated platforms

Future evolution :

The key target of our ongoing activities is the development of a Grid-based platform for the management of bioinformatics and medical informatics data. The platform will serve users through a Grid portal based on Enginframe product from NICE,that has been proved to be a reliable framework for interfacing Grid applications. New tools dedicated to medical data and metadata management will be developed.GILDA testbed has been used for testing.GILDA support has been revealed essential for it.

Primary authors : Ms. TORTEROLO, Livia (Bio-Lab, DIST, University of Genoa) ; Mr. PORRO, Ivan (Bio-Lab, DIST, University of Genoa)

Co-authors : Mr. VENUTI, Nicola (NICE srl) ; Mr. GATTI, Stefano (Aitek S.p.A.) ; Mr. CALANDUCCI, Antonio (INFN Catania) ; Mr. SCIFO, Salvatore (INFN Catania) ; Prof. FATO, Marco (Bio-Lab, DIST, University of Genoa)

Presenter : Ms. TORTEROLO, Livia (Bio-Lab, DIST, University of Genoa)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Mrs. TORTEROLO, Livia

Submitted on Wednesday 31 January 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 79

The Simple API for Grid Applications (SAGA)

Description of the activity :

Grid applications need programmatic access to compute and data resources. Grid middleware like gLite (or Globus) provide custom API's to their services. These interfaces are driven by the provided middleware and services, rather than application needs.

Grid added value :

The Simple API for Grid Applications (SAGA) defines such a programmatic interface. SAGA is being standardized within the Open Grid Forum (OGF). SAGA is focused on and has been derived from application use cases. SAGA emphasizes the simplicity of the API, and a consistent presentation of its functionality areas, namely job submission, file and replica management, data access and streaming, application monitoring, and remote procedure calls.

SAGA implementations present the API while binding to a variety of Grid middleware systems. This enables portability of applications across different grids, middlewares, and their versions.

Experience or proposed activity :

We will demonstrate a number of small, SAGA-based utilities for data and replica management and for job submission and control for GT4-based middleware grids. (Bindings to gLite are currently under development and will likely not be ready for demonstration at the time of the meeting.)

Future evolution :

No issues as of now.

Primary authors : Dr. KIELMANN, Thilo (Vrije Universiteit) ; Mr. MERZKY, Andre (Vrije Universiteit)

Co-authors :

Presenter : Dr. KIELMANN, Thilo (Vrije Universiteit)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : Dr. KIELMANN, Thilo

Submitted on Wednesday 31 January 2007

Last modified on : Sunday 18 February 2007

Comments :

This proposal had been discussed before with Erwin Laure.

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :

Judged by :

Date :
Comments : ""

Abstract ID : 81

SALUTE – GRID Application for problems in quantum transport: recent developments

Description of the activity :

SALUTE (Stochastic ALgorithms for Ultra-fast Transport in sEmiconductors) is a Grid application developed for solving computationally intensive problems which describe ultra-fast carrier transport in semiconductors. These problems are important for understanding the phenomena occurring in semiconductors at phemtosecond and nanometer scale.
Physicists, mathematicians and engineers are interested in either development of the simulation methods or their application for simulation of nano-devices.

Grid added value :

SALUTE is a computationally intensive application, which needs vast amounts of CPU power for achieving the desired accuracy and resolution. Its user community currently consists of scientists and engineers from 2 countries, and we expect more users and functionalities to be added. The application can be used to study the properties and the phenomena relevant to several semiconductor materials, used currently by the industry, from GaAs to GaN. The use of the Grid will provide not only CPU power, but also a platform for sharing the achieved results and avoiding of duplication of effort when using the application. The application can use up to thousands of CPUs, and necessitates adequate storage and networking capacities. We expect to be able to add more algorithms and functionality to SALUTE and gradually grow it into a main building block for a powerful Grid-based device simulator.

Experience or proposed activity :

SALUTE application currently runs on the SEE-GRID pilot grid infrastructure, which use the EGEE gLite middleware. The new implementation scheme of SALUTE employs: CE (lcg or glite flavour), WNs for computations, gliteWMS for workload distribution SE, FTS, AMGA for data and metadata management.
We built a custom server offering secure WebServices. Communication with the portal and the Grid is done from a graphical user interface. We had to develop a failover scheme, in order to cope with unreliable sites and services, which in principle is cumbersome and should be avoided. In order to achieve accurate results, we submitted thousands of jobs, and our observation was that the core services were mostly reliable, with the exception of gliteWMS and gliteCE. Our application stretched the SEE-GRID infrastructure in several areas – computation, communication, storage, exposing problems in reliability that would otherwise go undetected by functionality tests.

Future evolution :

We believe that the area of nanoscience should be better represented on the EGEE infrastructure, because of its importance for the future technological developments. We expect to increase our collaboration with users from other European countries and achieve greater accuracy in our simulations by using the resources of the EGEE production infrastructure. The incomplete support of MPI is a hindrance to many applications. We would also welcome better data management and better documentation.

Primary authors : Prof. ATANASSOV, Emanouil (IPP-BAS)

Co-authors : Prof. GUROV, Todor (IPP-BAS) ; Prof. KARAIVANOVA, Aneta (IPP-BAS)

Presenter : Prof. KARAIVANOVA, Aneta (IPP-BAS)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : KARAIVANOVA, Aneta

Submitted on Friday 02 February 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 82

Distributed Optimal Genetic algorithm for grid applications Scheduling

Description of the activity :

The work is developed as a Grid application/service in the See-Grid project, by a team in the National Center for Information Technology (NCIT) of the University "Politehnica" of Bucharest. NCIT is member of the RoGrid consortium, which participates in See-Grid and in EGEE-2. NCIT is involved in projects that are developed in cooperation with CERN and Caltech (RoDiCA), develops national Grid research projects GridMOSI, MedioGRID, and contributes to See-Grid and EGEE projects.

Grid added value :

DIOGENES (DIstributed Optimal GENEtic algorithm for grid application Scheduling) provides a solution for Grid meta-scheduling, aiming to achieve a distributed, fault-tolerant, scalable and efficient method for optimizing task assignment in Grid. The DIOGENES scheduler uses a combination of genetic algorithms and lookup services for obtaining a scalable and highly reliable optimization tool. Gridification topics that are covered are: application monitoring, running job, data management and access, workflows. The social impact is more efficient use of SEE-GRID VO's resources and the scientific impact represent new optimization method for Grid scheduling using Genetic Algorithms The current research is focused on multi-criteria optimization of Grid scheduling, complex task dependencies, backup and recovery from service failures (re-scheduling). We also provide a method for efficient integration with the Globus Toolkit, based on Open Grid Service Architecture.

Experience or proposed activity :

DIOGENES doesn't control the clusters / resources directly. It is more natural to consider it closer to Grid applications. It is responsible for the management of jobs, such as allocating resources needed for any specific job, managing the tasks for parallel execution, managing of data, and correlation of events. To fulfill its functions, DIOGENES needs information coming from information and monitoring services available in the platform, in this case in EGEE. The results of testing DIOGENES strategies in a Romanian cluster of SEE-GRID demonstrate very good behaviour in comparison with other scheduling approaches.

Future evolution :

EGEE includes the services requested by DIOGENES. For the moment we don't foresee difficulties in using it on this infrastructure. We'll use DIOGENES in projects like GridMOSI – VO in Grid technology for high performance MOdeling, SImulation, and optimization, and MedioGRID – Parallel and distributed processing of geographic and environmental data. We will pre-test new strategies developed in DIOGENES for applications built in these and other projects and eventually run on EGEE.

Primary authors : Prof. CRISTEA, Valentin (University "Politehnica" of Bucharest) ; Mr. POP, Florin (University "Politehnica" of Bucharest) ; Ms. STRATAN, Corina (University "Politehnica" of Bucharest)

Co-authors :

Presenter : Mr. POP, Florin (University "Politehnica" of Bucharest) ; Ms. STRATAN, Corina (University "Politehnica" of Bucharest)

Track classification : Poster session

Contribution type : poster

Submitted by : POP, Florin

Submitted on Sunday 04 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 83

UK e-Science Update

Description of the activity :

The community includes a wide range of researchers in all disciplines. The UK is integrating its provision according to a 10-year plan to provide e-Infrastructure, integrating the investments in multiple institutions via the National Grid Service (NGS). The software for production use that is developed in the UK is coordinated and supported by OMII-UK. The UK e-Infrastrucure will be represented by an NGI to participate in EGI. The focus on the broad e-Science community that are potential users.

Grid added value :

The grid provides a low-level technological infrastructure underpinning a wide range of higher-level services. A survey of users' requirements undertaken recently identifies those services that users require to be more pervasive, accessible and usable. They need a variety of forms of support as well as an integrated e-Infrastructure to enable seamless use. The current community of users can be reviewed, to understand their requirements. This is best done by catagorising the projects and presenting representative projects across a wide range of disciplines. Some of these requirements are already met by deployed UK e-Infrastructure. Many are not. The EGEE services could meet some of the needs. The talk will treat this as an open question.

Experience or proposed activity :

This talk is about the UK e-Science experience with particular reference to those among the UK e-Science community who do not use EGEE services. Therefore, it is not appropriate to identify EGEE key services. A higher-level discussion is more appropriate. Can we gain from a European-wide collaboration towards satisfying a broad spectrum of research requirements?

Future evolution :

The primary issue is a generic one - a requirement on research leaders to produce immediate results makes them consider local bespoke solutions more of han generic shared solutions. In most disciplines there are very few researchers who can afford to take a long-term "jam-tomorrow" approach. It is therefore essential to offer "jam today" as well (see Carole Goble's talk at the last EUF) as well. The challenge is to find a sensible incremental approach.

Primary authors : ATKINSON, Malcolm (UK e-Science Envoy & e-Science Institute)

Co-authors :

Presenter : ATKINSON, Malcolm (UK e-Science Envoy & e-Science Institute)

Track classification : Related Projects

Contribution type : --not specified--

Submitted by : ATKINSON, Malcolm

Submitted on Monday 05 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Related Projects

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 84

WISDOM production environment

Description of the activity :

During 2005 and 2006, three biomedical data challenges were run on the EGEE grid: two on malaria and one on avian flu. These deployments, based on relevant biological needs, were successfully achieved using most of the available resources on the Biomed virtual organisation. As a total, almost 600 years of computations were achieved during these 3 deployments using the WISDOM production environment.

Grid added value :

With WISDOM, we wanted to produce a straightforward application, "easy" to use for
non grid experts, and being able to integrate any type of docking software with it. The system was designed to deploy high-throughput experiments on the grid, and is being reengineered to offer a fully interoperable web services interface, with connections to databases to store and query, in almost real-time, the statistics and results.
One of the major added values of this new architecture is that the whole system can be easily integrated in workflow engines that just call the ad-hoc operations. The developments were focused on fault-tolerance, flexibility and scalability but several issues arose during the experiments.

Experience or proposed activity :

The environment is made of a set of scripts that generate the jobs, submit the files and check regularly their status while they are on the Grid through the workload management system. Given the status of the jobs, several actions are taken. The main one is the job resubmission if a job has been aborted by the workload management system, cancelled by the user or has failed because problems occurred during job run. The environment can also cancel a job automatically if the job stayed for too long in a queue; in this case the job is also resubmitted after it has been cancelled.
The submission process is performed by a java multithreaded submission engine, that can submit multiple jobs on several resource brokers in a round robin. The job results are stored directly on the grid storage elements, and the useful scoring information are registered directly in a relational database.

Future evolution :

The grid is still a very unpredictable system and there are many single points of failure that makes the efficiency decrease as the grid become more and more overloaded. For instance, resource brokers are still the source of a lot of troubles because they can become easily overloaded by job submissions, and are sometimes inefficient in the scheduling of the jobs.

Primary authors : Mr. SALZEMANN, Jean (IN2P3/CNRS)

Co-authors : Mr. KASAM, Vinod (IN2P3/CNRS) ; Mr. BLOCH, Vincent (IN2P3/CNRS)

Presenter : Mr. SALZEMANN, Jean (IN2P3/CNRS) ; Mr. KASAM, Vinod (IN2P3/CNRS) ; Mr. BLOCH, Vincent (IN2P3/CNRS)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : Mr. SALZEMANN, Jean

Submitted on Monday 05 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 85

Dissemination and Exploitation of Grids in Earth Science

Description of the activity :

Earth Science (ES) is an all-embracing term for sciences related to the planet Earth, covering a large and diverse user community. The major disciplines use physics, geology, mathematics, chemistry, and biology to build a quantitative understanding of the principal areas or spheres of the Earth system. Examples of thematic areas are: atmospheric sciences, hydrology, and geophysics. The DEGREE project is a consortium of ES and computer science partners aiming at promoting the uptake of Grid in ES

Grid added value :

Like other sciences ES needs to deal with enormous amounts of data and large computational needs. But what makes ES different? ES deals with geospatial data in combination with time components (4D) in various scales and resolutions, consists of many different domains, scattered among all countries in numerous institutes, using complex applications. For ES e-science can be an essential improvement in research, operation and business, especially when Grid services can be coupled to existing ES services. But is e-science mature and does it cover all ES requirements? DEGREE aims at answering these questions by capturing ES requirements and analyzing Grid solutions. To capture evolving ES requirements for Grid services, over 20 application scenarios were collected and grouped into families of applications. This will ease communication, maintenance and tracing of requirements More ES applications scenarios will be added; updating and analyzing requirement progress will be a continuous effort

Experience or proposed activity :

Three families have been identified, following a scheme with three different levels of complexity. In this case complexity is referring to the complexity of requirements placed on Grid technology by the application, not of the application itself. The first family is called the simple applications, the second the complex applications and the third family is the complex workflow applications. In short specific ES Grid requirements are: Reliability (QoS); real-time and instantaneous access; The need to access licensed software; Data policies on input and output data i.e. complicated security requirements; data is scattered around various institutes in various formats, in various databases and has metadata attached to it in various forms, i.e. data management (accessibility, harmonization) is essential in ES. Workflow orchestration is vital to cover requirements from the third family of applications. Key services for ES are workflow management and data and metadata management.

Future evolution :

The current maturity of GRID systems implies that (potential) users are troubled with in-depth knowledge to perform basic actions as a hindrance for wide utilization in the ES community. In order to make e-science take the next step in ES, there is a need for standardization on Grid service level so existing and for coupling easily new web services to the Grid. Portals improving the accessibility to data, computing and results will greatly improve ES research and Grid infrastructure usage.

Primary authors : Mr. SOM DE CERFF, Wim (KNMI)

Co-authors : Dr. PETITDIDIER, Monique (IPSL) ; Prof. HLUCHY, Ladislav (UISAF) ; Mr. LONJARET, Mathieu (IPSL) ; Mr. LINFORD, Julian (ESA-ESRIN) ; Mr. SCHWICHTENBERG, Horst (SCAI) ; Dr. TRAN, Viet (UISAF) ; Dr. FUSCO, Luigi (ESA-ESRIN) ; Dr. ZHIZHIN, Mikhail (Geophysical Centre, Russian Academy of Science)

Presenter : Mr. SOM DE CERFF, Wim (KNMI)

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : Mr. SOM DE CERFF, Wim

Submitted on Wednesday 07 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 86

The EGEE user support infrastructure

Description of the activity :

User support in a grid environment is a challenging task due to the distributed nature of the grid. The variety of users and VOs adds further to the challenge. One can find support requests by grid beginners, users with specific applications, site administrators, or grid monitoring operators. With the GGUS infrastructure, EGEE provides a portal where users can find support in their daily use of the grid. The current use of the system has shown that the goal has been achieved with success.

Grid added value :

The grid user support model in EGEE can be captioned "regional support with central coordination". Users can submit a support request to the central GGUS service, or to their Regional Operations' Centre (ROC) or to their Virtual Organisation helpdesks. Within GGUS there are appropriate support groups for all support requests. The ROCs and VOs and the other project wide groups such as middleware groups (JRA), network groups (NA), service groups (SA) and other grid infrastructures (OSG, NorduGrid, etc.) are connected via a central integration platform provided by GGUS.
GGUS central helpdesk also acts as a portal for all users who do not know where to send their requests. They can enter them directly into the GGUS system via a web form or e-mail.
This central helpdesk keeps track of all service requests and assigns them to the appropriate support groups. In this way, formal communication between all support groups is possible.

Experience or proposed activity :

GGUS provides a single entry point for reporting problems. In collaboration with the EGEE EIS team, the EGEE UIG, and NA3, GGUS offers a portal where users can find documentation, and powerful search engines to find answers to resolved problems and examples. Wiki pages are compiled for frequent or undocumented problems/features arising from tickets. GGUS is interfaced with other grids' support infrastructures such as in the case of OSG. Also, GGUS is used for daily operations to monitor the grid and keep it healthy. Finally, GGUS is used also to follow and track down problems during stress testing activities such as the HEP experiments production data challenges and the service challenges.
The daily operations are being dealt with in the framework of the Executive Support Committee (ESC). The ESC is a body that has representatives from all of the ROCs of EGEE. This organisation meets monthly to discuss the development of the support system and to decide on actions and priorities.

Future evolution :

The current model foresees only basic interfaces to the existing VO user support infrastructures. A move towards a further integration of the existing structures is needed for a better overall user support. The plan is hence to understand in detail how VOs provide support to their users so that GGUS can better suit their needs. The scalability of GGUS is constrained by the availability of supporters. This will rapidly become a constraint to growth unless more dedicated supporters are found.

Primary authors : ANTONI, Torsten (GGUS, INSTITUT FüR WISSENSCHAFTLICHES RECHNEN, FORSCHUNGSZENTRUM KARLSRUHE)

Co-authors : BOSIO, Diana (CERN) ; MILLS, Alistair (CERN)

Presenter : ANTONI, Torsten (GGUS, INSTITUT FüR WISSENSCHAFTLICHES RECHNEN, FORSCHUNGSZENTRUM KARLSRUHE)

Track classification : Training and Support for Grid Users

Contribution type : oral presentation

Submitted by : ANTONI, Torsten

Submitted on Wednesday 07 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Training and Support for Grid Users

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 87

Docking and Moleculardynamik - Complete 15 stage workflow with gLite WMS

Description of the activity :

Docking is the method of first choice for rapid in silico screening of large ligand databases for drug research, since it is based on a rational physical model and very fast. Thus, in 2005 during the first WIDSOM data challenge this method has been employed to screen the ZINC database for compounds potentially inhibiting the Malaria parasite plasmodium falciparum by blocking one of its particular proteases. As a result, guanidinium compounds, so far not known to inhibit the tested plasmepsin hav

Grid added value :

Drug design for neglected deseases is adressed in the WISDOMI and II.
The docking process is now combined with Moleculardynamics, which is heavy computing intensive and may be provide more accurate results.

Experience or proposed activity :

Therefore, a workflow combining the docking program FlexX2,3 and parts of the molecular dynamics (MD) package AMBER 84 has been set up. Since MD calculations tend to become quite consumptive, the workflow has been implemented using the EGEE-middleware gLite5 for performing world wide large scale studies in order to rerank the hits of the previous simple docking experiment. With the invention of the Workload Management Service gLite is able to handle so-called DAG (Directed Acyclic Graphs) jobs. Our directly with gLite features deployed workflow consist of 15 stages were each stage consist of subjobs and the number of subjobs depend on the size of the input. No portal and no workflow engine is used.
Each stage executes the different programs like FlexX, Amber etc. Data is held on the storage elements and Metadata is stored in database.

Future evolution :

We will present and first results generated during WISDOM2 for demonstrating both, the benefits as well as the limits of the chosen approach. We will discuss also the problems encountered with experiments, where a DAG Job, controlled by WMS starts hundreds of follow up jobs automatically over many days and the dependencies on the stability of the resources. It is of interest to run the Moleculardynamik part of the workflow on DEISA. We are in discussion with FzJ to use it as Demo.

Primary authors : ASTRID, Maass (SCAI Fraunhofer)

Co-authors : SCHWICHTENBERG, Horst (SCAI Fraunhofer) ; KALLUMADIKAL, Jisamma (SCAI Fraunhofer)

Presenter : ASTRID, Maass (SCAI Fraunhofer)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : HORST, Schwichtenberg

Submitted on Wednesday 07 February 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 88

G-ICS: Grid – Instrument Control System

Description of the activity :

The G-ICS is relevant for astronomers who use (or plan to use) the Grid to run remote observations from their home Institutes making use of remote observing facilities (Telescopes and related instrumentation). The G-ICS allows to configure the astronomical instrumentation as a new embedded resource in Grid. Astronomers may submit jobs to accomplish tasks like retrieving scientific and telemetric data under acquisition in quasi-real-time and sending requests to change the instrumentation status.

Grid added value :

People maintaining observing facilities (telescopes and related instrumentation) may resort to G-ICS enabled Grids to support the concepts of remote monitoring/control without the need of developing dedicated systems from their own. The Grid already supplies the underlying technology, once a Grid node is installed at the observing facility and the astronomer has access to a Grid-UI. There are no limitation on the number of telescopes and/or instruments a single G-ICS enabled Grid can support. Any observing facility, moreover, may be gridified in this way, provided that a Grid driver is written for the specific telescope and/or instrument, so the G-ICS may be beneficial for a huge community of users distributed worldwide. G-ICS allows the porting in Grid any kind of scientific instruments (i.e. networks of sensors) and this makes it suitable not only for astronomers but for any other scientific community and to carry out tasks like the environment quality monitoring.

Experience or proposed activity :

To support the remote monitoring/control of astronomical (but more generally scientific) instrumentation, a Grid infrastructure must be G-ICS enabled, that is the underlying middleware was modified in some of their components, mainly the GRAM and the GIS. The GRAM is extended in a way that the Job Manager is joined by a ICSC (Instrument Control System Connector) Manager that takes charge of jobs whose target is the remote monitoring/control of instrumentation. The ICSC Manager activates when needed a specific ICS instance for any instrument to monitor/control. Being instruments a new resource of the Grid like a CPU or a storage space, a set of meta-data characterizing this kind of resource have to be defined and fed to the GIS so that users are able to discover instrument resources via the RB. The solution proposed for instruments is similar to what has been done for databases in Grid. Some work was already done on LCG-based Grids; we plan to port the same solution in gLite too.

Future evolution :

The current version of the EGEE Grid middleware does not support instruments as embedded resources in Grid, but the porting of the G-ICS in the current LCG-based production Grid and in gLite is in our planning. We don’t expect particular problems from this point of view. At present however it is difficult for us to say whether and when the G-ICS technology might be ported on the EGEE Grid infrastructure.

Primary authors : Dr. VUERLI, Claudio (INAF-OA Trieste)

Co-authors : Dr. TAFFONI, Giuliano (INAF-OA Trieste) ; Mr. BARISANI, Andrea (INAF-OA Trieste) ; Dr. MANNA, Valeria (INAF-OA Trieste) ; Dr. PASIAN, Fabio (INAF-OA Trieste)

Presenter : Dr. VUERLI, Claudio (INAF-OA Trieste)

Track classification : Interactivity and Portals

Contribution type : oral presentation

Submitted by : Dr. VUERLI, Claudio

Submitted on Wednesday 07 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Interactivity and Portals

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 89

Collaborative environment for applications in BalticGrid

Description of the activity :

The BalticGrid, having related infrastructure to EGEE, is implementing operational gLite-based architecture, also for applications. Sveral Special Interest Groups (SIG) are being developed as a public service based on grid technology.
The main task of SIGs is to enable group-to-group communication of users, having similar research interests. The functionality of SIG introduces new possibilities to users including sharing among user group the desktop, user resources, data and files.

Grid added value :

The relatively new idea, expressed by SIG, emphasizes service for group of users, working in a tight collaborative way. It is designed to give communication/sharing tools inside user group, in ways of functionality, listed above and complemented by video-audio teleconferencing. The SIG software is designed and implemented as high level e-services; it provides sharing of resources (and working together) to Grid users: variety of data sources of interest; description of computing processes; explicit knowledge-processing services; description of networks existing between scientists; knowledge descriptions that can be asserted or generated in their own right. The special value to SIG is added by the possibility to have joint knowledge transformations for user groups. Such capabilities of SIGs are achieved by envisaging procedures of on-line forum, mailing lists, on-line documents and data repositories, surveys, news, events.

Experience or proposed activity :

The design and implementation of SIG in BalticGrid project is applied to user groups, having similar or related interests in the following research areas: Baltic Sea eco-system modelling, Text annotation service, Text-to-Speech service, Stellar spectra computing, Atomic and nuclear computing, Computer modelling of heterogeneous processes.
SIG will be also used to establish a special form of application support, acting as an interactive tool for application developers (especially in the process of user interface design and gridification process). The software for SIG is designed in an object-oriented way and includes cooperating components: a content management system (CMS), selective open source software components from Access Grid, a specially designed software component (gridcom), implemented by BalticGrid software developers. The SIG software is now applied to the areas of application mentioned above.

Future evolution :

The high level e-services offered by SIG software to grid users will be made as wide applicable as possible. It will
be independent from computer platform and is expected to serve architecture of many applications. Also
technically it will use the certification procedure common to EGEE and other grids. The special impact may be made
for social/humanitarian grid applications, where groups of scientists are usually collaboratively intensive (not
computationally ones).

Primary authors : Dr. JUOZAPAVICIUS, Algimantas (associate professor)

Co-authors : Mr. ANBINDERIS, Tomas (research fellow) ; Mr. PIATOV, Danila (research fellow)

Presenter : Dr. JUOZAPAVICIUS, Algimantas (associate professor)

Track classification : Related Projects

Contribution type : oral presentation

Submitted by : Dr. JUOZAPAVICIUS, Algimantas

Submitted on Wednesday 07 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Related Projects

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 90

MOTEUR grid-enabled data-intensive workflow manager

Description of the activity :

MOTEUR (http://egee1.unice.fr/MOTEUR) is a workflow manager designed to support data-intensive applications taking advantage of grid resources to transparently distribute the computations over a large set of resources. Currently, MOTEUR is used within the biomed VO to deploy applications to medical image analysis. However, MOTEUR scope is broader and it may be of interest for enacting many workflow-based scientific applications on the EGEE infrastructure.

Grid added value :

Many scientific applications involve multiple computations over large data sets. In the medical imaging community for instance, the processing of large image databases is needed for large scale studies such as epidemiological or statistical studies. MOTEUR is a service-based application enactment engine. It interfaces to different application services (Web Services, GridRPC servers) and provide a generic Web Service interface to embed non specific code. It benefits from a SOA design to provide a flexible and compact application description framework nicely decoupling processings from data sets. It is exploiting the Scufl (Taverna's description language) iteration strategies to enable the description of complex data processing patterns. It transparently exploits different levels of parallelism and distributes computations over grid resources to optimize performances. In particular, it exploits the large scale data parallelism available in many scientific data analysis procedures.

Experience or proposed activity :

MOTEUR is interfaced to different middlewares including gLite. From the application side, MOTEUR enacts a workflow of application web services. Unlike traditional web services, MOTEUR execution web service interfaces to the gLite workload Management System in order to trigger application computations on the grid infrastructure. At the higher level, the MOTEUR workflow engines interprets the workflow description graph. At run time, this graph combined with input data sets is used to instantiate individual computing tasks described through JDLs. The execution service submits, monitors and retrieve each task, concurrently whenever possible. It improves reliability by supporting jobs resubmissions on failure. MOTEUR uses a data provenance history tree to keep track of data transformations and to ensure coherent parallel execution. The architecture of MOTEUR is very modular and the grid interface can easily be adapted or exchanged without modifying the core workflow engine.

Future evolution :

MOTEUR is providing a flexible and high level framework for enacting scientific workflows. It enables a service-oriented description of applications on the batch-oriented EGEE infrastructure. The computing tasks triggered by MOTEUR may be of variable length. the current gLite WMS only provides limited support for efficient execution of short tasks that can easily penalize the total execution time. We have been successfully using SDJs to drastically improve performances in some cases though.

Primary authors : Dr. MONTAGNAT, Johan (CNRS)

Co-authors : Mr. GLATARD, Tristan (CNRS / INRIA) ; Dr. LINGRAND, Diane (CNRS / I3S)

Presenter : Dr. MONTAGNAT, Johan (CNRS)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Dr. MONTAGNAT, Johan

Submitted on Thursday 08 February 2007

Last modified on : Thursday 08 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 91

A workflow engine for grid-oriented applications

Description of the activity :

Two industrial grid projects are planning to use the EGEE infrastructure in addition to their own regional infrastructure based on Globus Toolkit 2 and 4. These projects provide hardware resources and software services for small and medium enterprises (SME) in the scope of bio-informatics (RUGBI) and of simulation of polymer injection (OpenPlast).

Grid added value :

These activities require a huge amount of computing power and expensive software. The associated investment is not always made beneficial for such enterprises, because these activities often occur in the context of intermittent short-term projects.
The grid enables SME to improve the accuracy of their computations and simulations with a reasonable cost, by sharing hardware and software resources.

Experience or proposed activity :

Job submission, file transfer and storage. No need for special services, but the core services have to be secure and efficient.

Future evolution :

The main issue is security, especially privacy. Indeed, for SME, confidentiality is a concern not only for data, but also for any information that would give some hints to guess what the enterprise is working on and what its strategy is. Even log messages may contain sensitive information for example.
Other important issues are the cost of deployment and operation of grid components on SME sites, and their capability to integrate with site local policies and software.

Primary authors : Mr. REYNAUD, Sylvain (CNRS)

Co-authors :

Presenter : Mr. REYNAUD, Sylvain (CNRS)

Track classification : Poster session

Contribution type : poster

Submitted by : Mr. REYNAUD, Sylvain

Submitted on Thursday 08 February 2007

Last modified on : Sunday 18 February 2007

Comments :

The poster will focus on the workflow engine developed in the context of □ these projects rather than on these projects.

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 92

Tycoon market-based system: integration with EGEE

Description of the activity :

Tycoon is a market-based system for trade of resources. The idea is to integrate Tycoon with EGEE, obtaining a dynamic Grid ready to share resources with others (Universities, research centers, etc.). Tycoon will run as a new service, deploying and destroying virtual machines (Worker Nodes and Computing Elements) on demand.

Grid added value :

Tycoon will give to EGEE a dynamic, flexible, secure and transparent platform/service to trade resources more efficiently. It could attract small and medium VOs to the EGEE project, they could take advantage of the Grid resources and obtain extra computing power while EGEE could obtain credits/money for its service.

Experience or proposed activity :

Our goal is to integrate Tycoon as a new service of EGEE, we don't have special requirements, we don't need to modify the EGEE middleware, Grid special services, etc. Standard and "tycoonized" nodes (Computing Elements and Worker Nodes) must coexist in the same Grid, working without problems.

Future evolution :

Our main issue would be to ensure the security of the whole system (EGEE platform). Tycoon will deploy Worker Nodes and Computing Elements on-the-fly, reconfiguring the Grid adding nodes on demand using virtual machines. This dynamic part of the Grid must not be a security problem.

Primary authors : Mr. DANA PEREZ, Jose Miguel (CERN)

Co-authors : Mr. GREHANT, Xavier (CERN)

Presenter : Mr. DANA PEREZ, Jose Miguel (CERN)

Track classification : Workflow ; Data Management

Contribution type : oral presentation

Submitted by : JOSE, Dana Perez

Submitted on Thursday 08 February 2007

Last modified on : Saturday 10 February 2007

Comments :

Bob Jones (CERN) asked for this presentation.

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 93

Modeling star evolution on the GRID

Description of the activity :

In Astronomy, the possibility to use a large set of stellar evolution models spanning a wide range of stellar masses
and initial chemical composition, is a necessary pre-requisite for any investigation aimed at interpreting observations
of Galactic and extra-Galactic stellar populations. FRANEC is a numerical code designed to address this problem. We
present the activity done to use the EGEE computational Grid environment to produce a large set of evolutionary
tracks to compare with observat

Grid added value :

In these last years, "FRANEC" has been extensively used for computing models of Stars (for a wide range of mass
and chemical composition, and in all their evolutionary stages) that have been compared with the most recent
empirical data from both ground-based telescope and space ones such as the Hubble Space Telescope. It is used to
produce a library of stellar evolutionary predictions: the BaSTI archive. BaSTI is the most updated database of stellar
models all around the world and it is a formidable tool for population synthesis investigations. It is particularly
important to complete as much as possible the archive. This effort would require a huge amount of computing
power. Moreover, it is crucial to fulfill the requests of astronomers that need a model not already present in the
archive with "on-the-fly" runs. Grid computing is an answer to FRANEC data generation and processing needs.
Grid may be used by Astronomers share simulated data and analyze the results obtain

Experience or proposed activity :

FRANEC is a serial code, each instance of this code produce a stellar model for a star of a given mass. It is written in
Fortran. To create evolutionary tracks it is necessary to use a computational environment that is able to manage
workflows operations. The EGEE resource Broker is able to manage DAGs and parametric Jobs, and we use this
capability to run swarm of FRANEC runs. A Grid File catalog is mandatory to share data between different
researcher. FRANEC-G (FRANEC on GRID) will be available as a command line interface that, starting from a
configuration file, automatically distributes jobs on the GRID. Moreover, we are planning to redesign the BaSTI
archive to allow astronomers to access any stellar model though a portal (www.oa-teramo.inaf.it/BASTI) . If the
model has been already simulated it is uploaded otherwise a Grid job is submitted. This job is a swarm of FRANEC
runs. The asynchronous webservices technology is used to give the result back to the Astronomer.

Future evolution :

Porting FRANEC code on EGEE did not implies deep modification in the code structure,

data access is easely
implemented using a specific application layer that interacts with the Grid services. However, the evolution of our
work focused on integrating FRANEC-G on a Web Portal which is not Grid specific. The portal shuld have both
database access and Grid "access". This involvese the possibility to access through web services the EGEE resources.

Primary authors : Dr. TAFFONI, Giuliano (INAF - SI) ; Prof. CASSISI, Santi (INAF-OATERAMO)

Co-authors : Dr. VUERLI, Claudio (INAF-OATS) ; Prof. PASIAN, Fabio (INAF-OATS) ; Dr. PIETRINFERNI, Adriano (INAF-OATERAMO) ; Dr. MANNA, Valeria (INAF-OATS) ; Mr. BARISANI, Andrea (INAF-OATS)

Presenter : Prof. CASSISI, Santi (INAF-OATERAMO)

Track classification : Experience with application domains – setting up and production

Contribution type : --not specified--

Submitted by : Dr. TAFFONI, Giuliano

Submitted on Thursday 08 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 94

GOSSIP: a galaxy SED fitting tool on the GRID

Description of the activity :

To understand galaxy evolution, it is of fundamental importance to derive physical parameters from the observable quantities, like the electro-magnetic spectrum of emitted light (SED).This is usually accomplished by fitting the SED against a set of theoretical models,to derive from the best-fitting model the global physical properties of the galaxy.When such process is applied to a large data set comprising some thousands of galaxies,it gives strong constraints on universe evolution models

Grid added value :

For each galaxy, about 40000 synthetic models have to be compared to find the one best fitting the observed SED, and the current version of GOSSIP requires about 180 seconds (even if much effort has been put in optimizing the program to increase speed). For cosmological studies, the same procedure has to be applied to samples of the order of 10000 galaxies, thus making up a total of about 500 hours on a single CPU (about 20 days!) for each run.
Subdividing the task on multiple CPU using the EGEE infrastructure can considerably speed up the procedure and would allow astronomers to make multiple fitting runs in a timely fashion to really find the best fitting model and parameters combination, and to improve the galaxy properties determination.

Experience or proposed activity :

The EGEE implementation of GOSSIP use the standard services to subdivide the fitting procedure on multiple CPUs assigning to each of them a subset of galaxies to be fitted. The number of objects that each CPU has to process is adaptively controlled during the task execution in GRID, using AMGA to store information about the global procedure.
In this way task balancing is automatically achieved, each CPU fits only what is needed and the global procedure ends contemporary on each node.

Future evolution :

We haven't experienced severe technical issues actually using the Grid for Planck. The most challenging issue comes from the fact that the python version installed in the EGEE resources was not compatible with the one used to make GOSSIP. Some modules was re-written and some modified in C code implemented as external fuctions.
A benefit may come from the use of a new version of the C-fitsio libraries that supports direct access on the files saved on the storage elements.

Primary authors : Dr. TAFFONI, Giuliano (INAF - SI) ; Dr. FRANZETTI, Paolo (IAFS-Milan)

Co-authors : Dr. VUERLI, Claudio (INAF-OATS) ; Dr. MANNA, Valeria (INAF-OATS) ; Mr. BARISANI, Andrea (INAF-OATS) ; Prof. GARILLI, Bianca (IASF-Milan) ; Prof. PASIAN, Fabio (INAF-OATS)

Presenter : Dr. TAFFONI, Giuliano (INAF - SI) ; Dr. FRANZETTI, Paolo (IAFS-Milan)

Track classification : Experience with application domains – setting up and production

Contribution type : --not specified--

Submitted by : Dr. TAFFONI, Giuliano

Submitted on Thursday 08 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 95

GRelC Data Access Service: Extreme Performance Managing Grid Databases

Description of the activity :

The activity aims at integrating within the EGEE infrastructure the GRelC DAS, a GSI enabled web service allowing high performance access in grid environments both to relational and non-relational DBs.This service currently runs on the GILDA t-Infrastructure and fully interacts with several gLite services.The target community includes scientists belonging to different domains within the EGEE collaboration that need for their applications to transparently and securely access huge distributed DBs.

Grid added value :

The GRelC DAS: (i) could attract new communities that usually work intensively with databases (e.g. financial applications exploiting datawarehouse and data mining); (ii) help existing ones (bioinformatics, astrophysics, etc.), in performing new tests and experiments; and (iii) improve e-Science research results. Due to the transversal role of this service, many experiments and Virtual Organizations within the EGEE project would benefit from it. Within such a multifaceted environment new kinds of queries could also be invented to perform distributed computation based on a work sharing approach.
Moreover, replication (cross-DBMS) facilities that we developed within our service could help grid users and site admins to ease (i) deployment of data sources, (ii) data migration and (iii) database backup.

Experience or proposed activity :

We present the GRelC DAS, a data grid access service developed by the GRelC Team (SPACI Consortium and Univ. of Salento).Deployment and experimental results related to the GILDA test-bed as well as porting issues on gLite middleware will also be presented.The GRelC DAS is a GSI enabled web service addressing extreme performance, interoperability and security.It provides an uniform access interface to relational and non-relational data sources. The GRelC DAS supports both basic and advanced functionalities. The aim of this service is to efficiently, securely and transparently manage databases on the grid across VOs, with regard to emerging and consolidating grid standards and specifications and interoperation with gLite. We will detail the key components of the GRelC DAS architecture, presenting queries (with advanced functionalities such as chunking and compression),user management, technological choices, heterogeneous DB access and security issues (GSI support, VOMS integration, etc).

Future evolution :

Several EGEE-VOs need access to relational DBs for their experiments. Even though in the last few years many efforts have been concentrated in this direction, performances have not been completely addressed. Within the EGEE middleware there is a lack of services concerning this topic and current solutions do not fully address user's/VO's requirements. The GRelC DAS tries to bridge this gap providing a well suited solution strongly integrated with the existing gLite components.

Primary authors : Dr. FIORE, Sandro (SPACI Consortium and University of Salento, Italy) ; Prof. ALOISIO, Giovanni (SPACI Consortium and University of Salento, Italy) ; Prof. BARBERA, Roberto (INFN Sez. di Catania, Italy)

Co-authors : Dr. CAFARO, Massimo (SPACI Consortium and University of Salento, Italy) ; Mr. NEGRO, Alessandro (SPACI Consortium and University of Salento, Italy) ; Mr. VADACCA, Salvatore (SPACI Consortium and University of Salento, Italy) ; Dr. MIRTO, Maria (SPACI Consortium and University of Salento, Italy) ; Dr. GIORGIO, Emidio (INFN Sez. di Catania, Italy)

Presenter : Dr. FIORE, Sandro (SPACI Consortium and University of Salento, Italy)

Track classification : Data Management

Contribution type : oral presentation

Submitted by : Dr. FIORE, Sandro

Submitted on Thursday 08 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 96

Simulations of the ESA Planck mission in EGEE

Description of the activity :

Planck is a cosmological mission of ESA composed of the LFI instrument covering frequencies between 30 and 70 GHz and the HFI instrument that works between 100 and 857 GHz. The two created consortia gather about 700 people whereof approximately 250 deal with the pipelines data reduction software. In 2004 the Planck simulations was approved by EGAAP and officially supported by EGEE. The applications aims at porting the whole Planck simulations software on the EGEE Grid infrastructure.

Grid added value :

Planck simulations are highly computing demanding and produce a huge amount of data. Such resources cannot be usually afforded by a single research institute, both in terms of computing power and data storage space. In our case the federation of resources coming from different providers plays a crucial role to tackle the shortage of resources available at single institutions. Other added values that it is worth to mention are: a) the Grid authentication/authorization mechanism; b) the opportunity to easily monitor/control accesses to software and data; c) The federation of all Planck software developers in a single VO fostering in this way the scientific collaboration. Up to now a restrict number of Planck developers made use of the Grid to run their software. The number of potential users however is high and, once trained, it is expected that many of them will port their software in Grid.

Experience or proposed activity :

During the first simulation tests, a complete simulation of the Planck/LFI instrument was run on a single, dual-CPU, workstation and in Grid involving 22 nodes, one for each detector of the LFI instrument. The gain obtained by using the Grid was of ~15 times. Moreover we ran simulations in Grid many times and every time with a different number of worker nodes; we achieved the best results when we assigned each LFI detector to a different WN; the execution time is in this case ~30 times faster with respect to the worst case where all LFI detectors are assigned to the same WN. Results are promising so both partial and full simulations of the mission for both instruments will be ported in EGEE. Because one of the most challenging aspects of Planck simulations is the amount of intermediate data to store and their movement in Grid we expect full support in terms of data storage and pieces of Planck software installed on WNs, since we will probably move code instead of intermediate data.

Future evolution :

We haven't experienced severe technical issues actually using the Grid for Planck. The most challenging issue comes from the fact that the movement of intermediate data produced by simulations in Grid is practically hindered. Waiting for more capable network infrastructures we circumvent the problem through an accurate setup of the application. In particular the code (a piece of pipeline) has to be moved where data reside so that data movements doesn't take place over the WAN.

Primary authors : Dr. TAFFONI, Giuliano (INAF-OA Trieste) ; Dr. VUERLI, Claudio (INAF-OA Trieste)

Co-authors : Mr. BARISANI, Andrea (INAF-OA Trieste) ; Dr. MANNA, Valeria (INAF-OA Trieste) ; Dr. ZACCHEI, Andrea (INAF-OA Trieste) ; Dr. PASIAN, Fabio (INAF-OA Trieste)

Presenter : Dr. TAFFONI, Giuliano (INAF-OA Trieste) ; Dr. VUERLI, Claudio (INAF-OA Trieste)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : Dr. VUERLI, Claudio

Submitted on Thursday 08 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations
Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 97

CNES GRID EXPERIENCES AND PLANS FOR SPACE APPLICATIONS

Description of the activity :

Today, space data processing centers are usually organized in a centralized workflow of operations as acquisition, archiving, processing and distribution functions. Scientific laboratories achieve processing for level 1, 2, 3 products while CNES and other space agencies are responsible for the system architecture.

Grid added value :

In case of future ground segments or new scientific programs which require huge storage and processing capabilities, the Grid concept can be an opportunity to deploy operational distributed applications through Europe (and beyond) for which CNES can gather and share its computational and storage resources. This is why CNES has started a new study with CS in order to examine the gLite grid technology and its potential use in data processing centers where actors and processing resources are geographically distributed.

Experience or proposed activity :

This study consists also in gridifying a typical scientific application and in providing a set of recommendations in order to facilitate the deployment gridification phase.

Future evolution :

The presentation will be focused on the first technical results in terms of architecture definition, security and CNES experience feedback

Primary authors : Mr. COURQUET, Joel (cs si) ; Mr. WALLUT, Jean-marie (CNES)

Co-authors :

Presenter : Mr. COURQUET, Joel (cs si)

Track classification : Workflow ; Data Management

Contribution type : oral presentation

Submitted by : Mr. COURQUET, Joel

Submitted on Friday 09 February 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :

Comments : ""

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 98

The UNOSAT-GRID Project: Access to Satellite Imagery through the Grid Environment

Description of the activity :

UNOSAT is a United Nations activity to provide access to satellite images and geographic system services for humanitarian operations for rescue or aid activities. UNOSAT is implemented by the UN Institute for Training and Research (UNITAR) and managed by the UN Office for Project Services (UNOPS). In addition, partners from different organizations constitute the UNOSAT consortium. Among these partners, CERN participates actively providing the required computational and storage resources.

Grid added value :

The bottleneck of the UNOSAT activity is the storage and processing of large quantities of images that their members need to manage. The fast and secure access to these images from any part of the world is mandatory during these activities. Based on two successful CERN-GRID/UNOSAT pilot projects (data storage/compression/download and image access through mobile phone), the GRID-UNOSAT project has consolidated the considerable work undertaken so far in the present activity. The use case we have demonstrated is the delivery of satellite images from the GRID to a portal (web and portable devices). This use case, applied for the moment to UNOSAT, can also be used by many communities requiring a fast and reliable access to geographical images from any portable device.

Experience or proposed activity :

During the project development we have enabled the selection and download of satellite images starting on a portable device (using the GPS coordinates provided by the devide itself). The system provides seamless access to valuable satellite images while preserving the security requirements of the data provider and of the EGEE infrastructure (using X509 certificates). The system uses EGEE services already used by other applications and in this presentation we show how we have orchestrated them. The satellite images are catalogued by the AMGA (metadata) and LFC (location) services. The handling of images (compression/decompression, cropping, etc) is provided by the computational GRID resources via the EGEE workload management system. This work is being performed in close collaboration with the NICE Company, providing their EnginFrame technology (used also by Genius EGEE portal, for a development and deployment environment for portal applications)

Future evolution :

The fast access and processing of the images is the key point for the UNOSAT project. It requires to build the support for a reliable storage and workload management system in the EGEE production is system to be ready in case of peak activity.

Primary authors : Dr. MENDEZ LORENZO, Patricia (CERN IT/PSS-ED) ; Dr. KOBLITZ, Birger (CERN IT/PSS-ED) ; Dr. LAMANNA, Massimo (CERN IT/PSS-ED) ; Mr. MEYER, Xavier (CERN/HESGE) ; Mr. LAZEYRAS, Michel (HESGE-Geneva) ; Ms. BJORGO, Einar (UNOSAT/CERN) ; Mr. RETIERE, Alain (UNOSAT/CERN) ; Mr. FALZONE, Alberto (NICE-Italy) ; Mr. VENUTI, Nicola (NICE-Italy) ; Mr. MACCARONE, Salvo

(NICE-Italy) ; Mr. UGOLOTTI, Beppe (NICE-Italy)

Co-authors :

Presenter : Dr. MENDEZ LORENZO, Patricia (CERN IT/PSS-ED)

Track classification : Poster session

Contribution type : poster

Submitted by : Dr. MENDEZ LORENZO, Patricia

Submitted on Friday 09 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 99

The usage of the gLite Workload Management System by the LHC experiments

Description of the activity :

The LHC experiments will study the physics of p-p interactions at a centre-of-mass energy of 14 TeV using the LHC accelerator at CERN. The primary purpose of their research is the discovery of the Higgs boson and of new physics at the TeV scale.

Grid added value :

The LHC experiments are undoubtely the most demanding communities for the EGEE infrastructure. They have thousands of collaborators who expect to run their physics analyses on the data that will be collected starting from late 2007. It is expected that each experiment will need to run hundreds of thousands of jobs per day, including: event reconstruction from raw data, analysis on reconstructed data, and Monte Carlo simulation. The management of such amounts of jobs is an extremely complicated task and the gLite WMS is being considered as a candidate to perform it.

Experience or proposed activity :

This report describes the experience of the ATLAS and the CMS experiments using the gLite Workload Management System as a tool to submit both simulation and analysis jobs. In fact, this experience has led to a significant improvement of the WMS performance and reliability due to a close interaction with the gLite developers. Many limitations have been hit, and solved or alleviated. An evaluation of the current status of the WMS and how it matches with the experiment's expectations will be given.

Future evolution :

The most severe limitations found during this work relate to the ability of the WMS of working reliably and unattended for long periods of time. Though a lof of progress has been made, some issues remain. There are proposals to solve them, and they will be discussed.

Primary authors : Dr. SCIABA', Andrea (CERN) ; Dr. CAMPANA, Simone (CERN)

Co-authors :

Presenter : Dr. SCIABA', Andrea (CERN)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Dr. SCIABA', Andrea

Submitted on Friday 09 February 2007

Last modified on : Saturday 10 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 100

A Grid Software for Virtual Eye Surgery Based on Globus 4 and gLite

Description of the activity :

"Grid-Enabled SEE++" is a grid-based simulation software that supports the diagnosis and treatment of certain eye motility disorders (strabismus). The overall goal is to to develop an efficient grid-based tool for "Evidence Based Medicine", which supports the surgeons in choosing optimal surgery techniques for the treatments of different syndromes of strabismus.

Grid added value :

In Globus Toolkit 4, we developed a parallel version of the simulation of the Hess-Lancaster test (typical medical examination). By this, we speeded up this simulation by a factor of 14-17.

Furthermore, we reported the prototype implementation of a medical database component for "Grid-Enabled SEE++". Our next steps concentrate on developing a distributed grid-enabled database system.

Finally, we designed a so called grid-based Pathology Fitting algorithm, which would be able to determinate (or at least estimate) automatically the pathological reason of a patient's strabismus. Since the outcome of this algorithm strongly depends on the initial estimation for the pathological case, we propose to exploit the grid in the following way:

-by searching in the grid-based SEE++ medical databases for similar pathological cases and

-by starting concurrent pathology fitting processes with these cases as the starting points of the optimizations (parameter study).

Experience or proposed activity :

Since we met with some limitations of Globus 4, we also designed a version of "Grid-Enabled SEE++" compatible with gLite.

We use some kind of server jobs (as executers for parallel Hess calculations) started via the WMS service. To return the allocated port numbers, we investigate and exploit the interactive job submission feature of gLite.

We may exchange the access layer developed earlier for the SEE++ medical databases to an AMGA-based solution.

Pathology Fitting is proposed to execute on gLite as parametric jobs. Each job will be started with different initial parameters founded in the SEE++ medical databases.

We plan to apply the R-GMA information system as well, such that our system will be able to discover automatically the available databases and the executer jobs on the grid.
An important security concept is the managing of the Virtual

Organizations, because we have to be sure that the published medical data will be hosted only by certain trusted grid nodes.

Future evolution :

The fact that "Grid-Enabled SEE++" is an interactive application with many fine-grained jobs (the users change the eye parameters by a manual trial and error method and they wait for the result of a simulation before starting a subsequent one) may make the software an interesting testcase for gLite. We expect to report on the differencies between gLite and Globus Toolkit based on the observations with the two versions of our software system.

Primary authors : Dr. BOSA, Karoly (RISC, JKU)

Co-authors : Dr. SCHREINER, Wolfgang (RISC, JKU) ; Dr. BUCHBERGER, Michael (Upper Austrian Research) ; Mr. THOMAS, Kaltofen (Upper Austrian Research)

Presenter : Dr. BOSA, Karoly (RISC, JKU)

Track classification : Poster session

Contribution type : poster

Submitted by : KAROLY JOZSEF, Bosa

Submitted on Friday 09 February 2007

Last modified on : Friday 16 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 101

Multiple Ligand Trajectory Docking - A Case Study of Complex Grid Jobs Management

Description of the activity :

Interaction between large biomolecules and smaller bio-active ligands lies on the foundation of many biological properties and is of huge interest in the bio-molecular and pharmaceutical research. The role of the ligand is to influence the reaction that occurs in an active site of a biomolecule. The interaction is usually studied using multiple ligand trajectory docking, a computationally intensive process aimed to find energetically favorable orientation of the ligand within an active site.

Grid added value :

The search is done on snapshots taken from the molecular dynamics trajectory describing the dynamic behavior of the biomolecule. For each snapshot and ligand the best position of the ligand is calculated, yielding a matrix containing energies of snapshot/ligand interactions. The minima correspond to the most favorable ligands and provide insight into the specific shape of the biomolecule.

The concrete problem shown in the demo deals with one 2 ns acetylcholinesterase trajectory and 3 ligands, requiring some 6000 CPU hours on an average compute server. A realistic studies use more and longer trajectories (tens ns) and higher number of potential ligands (tens to hundreds).

Performing such computation is infeasible without the Grid infrastructure. Also, managing its results is a non-trivial problem, undoable without semi-automatic support tools. A sophisticated job submission system coupled with a community-wide provenance of already run jobs is a necessary prerequisite.

Experience or proposed activity :

We present a specific thin graphical application built on top of the gLite Job Provenance service (JP) and Charon Extension Layer (CEL). JP keeps long-term track of execution of Grid jobs, including annotations, and provides data mining capabilities on these data. CEL toolkit is a universal framework creating a layer upon the basic Grid middleware making the process of job preparation and manipulation a relatively easy task. CEL provides a command-line interface and offers full control over submitted jobs.

The application supports selection of subsets of both trajectory snapshots and specific ligands, queries JP, and displays a 2D array of finished Grid jobs matching the criteria. Their results can be examined, including 3D visualization of the emerging structures. Bunches of jobs can be prepared in order to fill empty cells of the matrix to complete the overall docking analysis. Also, the already finished jobs can be taken from JP, cloned and re-run with modified input parameters.

Future evolution :

Job Provenance can handle arbitrary annotations assigned to Grid jobs. Such annotations can be understood to belong to the job outputs too. The demonstration is based on this approach. However, native data item annotations are not supported by JP by design. Presence of general data annotation/provenance service would streamline the solution of the presented problem.

Primary authors : PETREK, Martin (CESNET) ; WIESNER, Jiri (CESNET) ; KMUNICEK, Jan (CESNET) ; KRENEK, Ales (CESNET)

Co-authors : FILIPOVIC, Jiri (CESNET) ; SUSTR, Zdenek (CESNET) ; KULHANEK, Petr (CESNET) ; SITERA, Jiri (CESNET) ; MATYSKA, Ludek (CESNET)

Presenter : PETREK, Martin (CESNET) ; KMUNICEK, Jan (CESNET) ; KRENEK, Ales (CESNET)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : KMUNICEK, Jan

Submitted on Saturday 10 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 102

SIG - collaboration tool for grid applications

Description of the activity :

The BalticGrid project is implementing applications in a framework of gLite-based architecture. The Special Interest Group (SIG) is being developed as a high level e-service for such applications. The main task of SIG is to enable close communication inside a group of users, having similar interests. The functionality of SIG for groups of users includes such possibilities: sharing applications, desktop, user resources, data and files, teleconferencing.

Grid added value :

The SIG is primarily oriented and designed to give communication/sharing tools for group of users, in order to implement functionality, listed above. The SIG software presents high level e-services, enabling sharing (and working together) to Grid users: repository of data sources and data sets; interactive forums for research processes; knowledge descriptions that can be asserted or generated in their own right. The special value to SIG is added by the possibility to have joint computing and similar data transformation procedures for user groups. Such capabilities of SIGs are achieved by envisaging suitable software components. The applications are: Baltic Sea eco-system modelling, Text annotation service, Text-to-Speech service, Stellar spectra computing, Atomic and nuclear computing, Computer modelling of heterogeneous processes. SIG will be also used to establish a special form of application support, acting as an interactive tool for application developers.

Experience or proposed activity :

The software for SIG is designed by using object-orientation modelling for software process and includes operational components (as cooperating software objects): an adopted open source content management system (CMS), selective open source software components from the Access Grid, a specially designed software components (gridcom component). It is operational under Internet browser (in conventional way). The software is implemented by BalticGrid software developers. The key component for the worflow in SIG is an interactive repository, which provides data, input and output streams for computing processes, input and output media streams, specification of computing applications, some additional information.

Future evolution :

The high level e-services implemented by SIG software will be distributed (or coupled with suitable applications to be served) as wide as possible. It will be independent from computer platform and is expected to serve architecture of many applications. Also technically it will use the certification procedure common to

EGEE and other grids.

Primary authors : Mr. ANBINDERIS, Tomas (research fellow)

Co-authors : Dr. JUOZAPAVICIUS, Algimantas (prof.) ; Mr. PIATOV, Danila (research fellow)

Presenter : Mr. ANBINDERIS, Tomas (research fellow)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : Dr. JUOZAPAVICIUS, Algimantas

Submitted on Sunday 11 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 103

Bogazici University Kandilli Earthquake Seismic Grid Data Server

Description of the activity :

Researchers that use earthquake seismic data.

Grid added value :

Massive time-series earthquake seismic data is generated from several stations all over Turkey. A researcher that wants to use this data, need to download data from a web server and use his own computer to process it. Grid provides both the data as well as the computational power to process this data. Grid also makes it possible perform performance optimizations such as file caching.

Experience or proposed activity :

Hourly earthquake seismic data in SAC file format is mirrored on the grid automatically from about one hundred stations. Data is stored on storage elements in files and made available through the File Catalog Service. Therefore, tools that will enable easy acccess to these files becomes important. Since several applications may be using these files, file caching also becomes quite important.

Future evolution :

MPI currently works on a single cluster on a grid. It will be nice if we could combine and use several grid nodes/cluster in an MPI program.

Primary authors : Prof. OZTURAN, Can (Computer Eng. Bogazici University, Istanbul, Turkey)

Co-authors : Ms. UNAT, Didem (Computer Eng. Bogazici University, Istanbul) ; Mr. YıLMAZER, Mehmet (Kandilli Observatory and Earthquake Research Center, Bogazici University, Istanbul) ; Mr. BALKıR, Atilla Soner (Computer Eng. Bogazici University, Istanbul, Turkey) ; Mr. ŞENAY, Emin (Computer Eng. Bogazici University, Istanbul, Turkey)

Presenter : Prof. OZTURAN, Can (Computer Eng. Bogazici University, Istanbul, Turkey)

Track classification : Poster session

Contribution type : poster

Submitted by : Prof. OZTURAN, Can

Submitted on Monday 12 February 2007

Last modified on : Friday 16 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 104

ELIS@: A gateway platform for Grids (developed by CS in collaboration CNRS)

Description of the activity :

ELIS@ offers to users with no specific computing skills, such as end-users from fusion or plastic communities, an abstraction layer to simplify the access and to use the best grid according to their needs.
The ELIS@ platform is a set of grid services used through a web portal to easy access to grid resources. The services use an interoperable workflow engine to submit jobs to the Globus middleware, and soon to EGEE with gLite and DEISA with Unicore.

Grid added value :

The Fusion community has shown its interest in accessing to either super computers through DEISA for large simulation or to EGEE for pre and post treatment.
From another scale, SMEs for instance from plastic industry need to share external computing resources.
Both communities need to access in the easiest way to the best grid according to their needs.
CS in collaboration with CNRS with the co-operation from end users such as Pole Européen de Plasturgie and ITER people from Cadarache, have developed a generic platform composed with a set of grid services used through a web portal to easy access to grid resources.

Grid added value
The services use an interoperable workflow engine to submit jobs to the Globus middleware, and soon to EGEE with gLite and DEISA with Unicore.

Experience or proposed activity :

The presentation will be focused on the architecture definition, and first prototype feedback.

Future evolution :

The services use an interoperable workflow engine to submit jobs to the Globus middleware, and soon to EGEE with gLite and DEISA with Unicore.

Primary authors : Mr. DEMESY, Nicolas (CS)

Co-authors : Mr. ZAQUINE, Gabriel (CS)

Presenter : Mr. DEMESY, Nicolas (CS)

Track classification : Interactivity and Portals

Contribution type : oral presentation

Submitted by : Mr. DEMESY, Nicolas

Submitted on Monday 12 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Interactivity and Portals

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 105

PFGR - Protein Folding on the EGEE Grid

Description of the activity :

PFGR is a grid-enabled system that aims to harness the power of the EGEE infrastructure to run simulations based on the "biased exchange metadynamic (BEM) algorithm" which is routenely used in the biosimulation community to investigate protein folding mechanisms.

Grid added value :

At the moment PFGR is used by skilled people of the biophysics community and it under testing and improvement in order to make it a final tool for a larger (and less skilled) biophysics community.

It has to be noted that the computational characteristics of the BEM algorithm fits almost perfectly on the EGEE infrastructures and the computing power available would allow, along with the EGEE grid computational resources, to fold the structure of complex systems (like small DNA chains, by example).

Experience or proposed activity :

PFGR is a combination of a pre-existent folding program with an added layer that allows it to exchange data on the grid transparently and without MPI (thus enabling multi CE execution and dynamic recruitment of resources).

The system is completed with an easy to use (though still in development) user interface which guides the user through most of the workflow.

The key functionality used by PFGR is outbound connectivity on the WNs (to enable comunication) and the data access (that actually uses the very low level GSI protocol, but can be easily reverted to use the SRM protocol).

PFGR features:

* Put Biased Meta dynamic folding on the grid.
* Dynamic recruitment of resources.
* Simulations can use multiple CE at a time.
* Easy to use user interface.
* Real time logging of the simulation.

The PFGR protein folding algorithm is based on the Gromacs software (with substantial modifications made by Alessandro Laio et al.) and the communication layer is built in python 2.5

Future evolution :

The main issue affecting the developement of PFGR has been the the old version of the software installed on the various WNs.
The Scientific Linux distribution presently installed on WNs is quite old missing important software (compilers and basic developement tools) and with out-of-date packages (like for istance python). This combination makes the development and the optimization for heterogeneuos resources complex and time consuming.

Primary authors : COZZINI, Stefano (CNR-INFM Democritos and ICTP)

Co-authors : DI MEO, Riccardo (EGRID project ICTP) ; LAIO, Alessandro (SISSA and CNR-INFM Democritos) ; PIETRUCCI, Fabio (SISSA)

Presenter : COZZINI, Stefano (CNR-INFM Democritos and ICTP)

Track classification : Workflow ; Data Management

Contribution type : oral presentation

Submitted by : Dr. COZZINI, Stefano

Submitted on Monday 12 February 2007

Last modified on : Wednesday 14 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 106

The EGRID Web Portal: an access point to the EGRID facility.

Description of the activity :

EGRID is a project funded by the Italian Ministry of Education (MIUR) with the aim of building a pilot grid facility for the Italian research community of Economics and Finance.
Our web portal was developed to address some of their specific usability needs.

Grid added value :

The portal is designed to give our users a straightforward access to the grid.
Its easy to use interface allows to focus on the problem without spending too much time on learning the tool itself. The graphical interface is closer to the way users work and it makes available all grid tools through a Java enabled web browser. The EGRID portal data management is centred
around the Logical Filename Catalogue (LFC) and it hides the physical replica mechanism. More over, it allows the graphical set up of ACLs on files in grid storage in the EGEE infrastructure, for members of the same VO.
This portal can be used by all other communities with similar data management needs.

Experience or proposed activity :

The portal is based on the Gridsphere framework. In particular it uses the MTA SZTAKI P-Grade portlets for the workflow job submission and for downloading certificates from a MyProxy server. Data management and certificate upload were instead developed by EGRID to meet Economics and Finance research requirements.
The data management is based on ELFI, a Linux filesystem interface centred around logical names. The key feature of the infrastructure was the existence of an SE that is able to enforce the logical ACLs. EGRID co-developed the StoRM SRM server, the only SE that meets this requirement.
The portal's main features include:

-data management through a GUI file browser;

-GUI access to file permissions (rwx) and extended ACL;

-the personal certificate and private key never leave the user's computer;

-simplified input and output data files for job submission, since only logical file names are needed.

Future evolution :

The information system contains data that are not always accurate: because the update rate is too low, and also because information may be misleading. For ex. the status of a submitted job is not updated in real time.
There are misconfigured sites that act as "Black Holes" where all jobs end-up and get aborted. It would be useful if the RB isolated these sites.
Our portal is affected because it displays outdated information, and because it occasionally gives an anomalously hight rate of failures

Primary authors : Mr. SPONZA, Massimo (ICTP - EGRID Project)

Co-authors :

Presenter : Mr. SPONZA, Massimo (ICTP - EGRID Project)

Track classification : Interactivity and Portals

Contribution type : oral presentation

Submitted by : Mr. SPONZA, Massimo

Submitted on Monday 12 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Interactivity and Portals

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 107

Using Grid services for Sciamachy data sharing and processing

Description of the activity :

The Netherlands Sciamachy Data Center (NL-SCIA-DC) provides access to satellite data from the GOME and Sciamachy instruments. GOME and Sciamachy both measure trace gases like Ozone, NO2 and aerosols, which are important for climate and air quality monitoring. The NL-SCIA-DC serves the atmospheric research community not only with the data but also with services like processing and data mining.

Grid added value :

The data services will be upgraded to offer additional functionality to end-users of Sciamachy data. One of the functionalities provided will be the possibility to select and process Sciamachy products using different data processors, which can lead to intensive use of the limited computing and data storage resources available to NL-SCIA-DC. Through our partners SARA and NIKHEF the NL-SCIA-DC can use facilities provided by the Dutch Grid infrastructure. The NL-SCIA-DC will also run tests to be coupled to the ESA G-POD Grid infrastructure. This future cooperation will give the NL-SCIA-DC users access to a wider range of instrument data available at G-POD, and G-POD can offer our data services to the G-POD users.

Experience or proposed activity :

Our aim is to add user-requested functionality to NL-SCIA-DC for which currently the computer and/or storage resources are not available to the partners KNMI and SRON. Grid technology will helps us to extent the functionality of NL-SCIA-DC without losing the current high-standard of security and to make efficient use of the resources available at KNMI, SRON and elsewhere.

Future evolution :

The end result will be an integrated 'griddified' NL-SCIA-DC, integration of two organization domains (SRON and KNMI) using standard and open general purpose protocols and interfaces, delivering non-trivial quality of service to the Dutch and international atmospheric research community.

Primary authors : Mr. VAN DE VEGTE, John (KNMI) ; Mr. SOM DE CERFF, Wim (KNMI) ; Dr. VAN HEES, Richard (SRON)

Co-authors : Mr. VAN DER NEUT, Ian (KNMI) ; Dr. WOLFRAT, Jules (SARA) ; Dr. GROEP, David (NIKHEF) ; Mr. COLIN, Olivier (ESA-ESRIN) ; Dr. PITERS, Ankie (KNMI)

Presenter : Mr. VAN DE VEGTE, John (KNMI)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Mr. SOM DE CERFF, Wim

Submitted on Monday 12 February 2007

Last modified on : Monday 12 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 108

Todays Status of Data Management in Earth Science

Description of the activity :

Earth Science Applications in various domains like Remote sensing in Earth Observation, Ocean and Atmosphere Environments, Geology, Seismology highly depend on Data. Data flow, access and provisioning is not consistent. Only a few applications are deployed in the last years to grid environments and this was mainly done on the EGEE infrastructure. There is an ongoing process of deploying complex scenarios on Grid infrastructure resulting in an extensive list of requirements to the middleware deve

Grid added value :

The Requirements of Earth Science addresses different developers. Important for the data management are middleware and database developers, but also developers of Ontologies and Metadata definitions. These and the implementation of ES scenarios (Use Cases) on the Infrastructure is interesting for other groups particular for business applications – not only from ES.

Experience or proposed activity :

Metadata management and access to databases are evaluated and in production via AMGA and OGSA DAI since the beginning of EGEE-II. The result of these implementations have shown that these abstraction layers are either too slow or not based on any standard.

Future evolution :

By the Use Cases of the System Support Action DEGREE, we will show what ES needs as Abstraction Layer for data management

Primary authors : HORST, Schwichtenberg (SCAI Fraunhofer) ; Dr. PETITDIDIER, Monique (IPSL) ; Dr. CIGLAN, Marek (IISAS)

Co-authors :

Presenter : Dr. PETITDIDIER, Monique (IPSL)

Track classification : Data Management

Contribution type : oral presentation

Submitted by : HORST, Schwichtenberg

Submitted on Monday 12 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 109

ALICE Data Challenges experience on LCG infrastructure

Description of the activity :

The ALICE experiment at CERN, a general-purpose heavy-ion experiment designed to study the physics of strongly interacting matter and the quark-gluon plasma in nucleus-nucleus collisions at the Large Hadron Collider. The collaboration currently involves more than 900 physicists and senior engineers, from both the High Energy and Nuclear Physics communities. ALICE plans to use Grid Computing for all of its needs of data reconstruction, analysis and Montecarlo simulation productions.

Grid added value :

The experiment is expected to record data at the rate of 100MB/s for proton-proton collisions, and up to 1.25 GB/s in heavy-ion mode, thus producing several PB of data per year that need to be safely stored, reconstructed and analysed. Furthermore, physicists need to produce large amounts of simulated Montecarlo data. To cope with such a huge need of processing power while efficiently exploiting the available resources, the computing model is intermediate between the MONARC multi-tiered architecture, assigning different roles to T1s and T2s, and the so-called 'cloud model', in which T1s and T2s differ only for the Quality of Service and the archival
storage capability. Its unique feature is the single point of access to resources through the AliEn system, which hides some of the complexity of the Grid environment from the end users.

Experience or proposed activity :

ALICE gained a lot of experience in running Montecarlo productions and subsequent reconstructions and analyses in a Grid environment, all its Physics Data Challenges since 2003 having been performed on some Grid infrastructure. The last one, PDC06, produced about 0.5 PB of data (40 millions of events) in 685k jobs across 55 sites. AliEn, the ALICE-developed distributed computing environment currently provides users transparent access to several Grid flavours, first of all the EGEE/LCG infrastructure but also OSG and NDGF. Having a general high-level interface for both production and general users, the key features for a Grid to be successfully exploited by ALICE is the stability of a few simple low-level services (WMS, Data Movement services) and reliable tools for data management and access, that need to support a large and distributed community, as well as support from sites and infrastructure experts to cope with problems and bugs in a timely fashion.

Future evolution :

The major challenge will be unscheduled distributed analysis by physicists, which will stress both data access and management services (chaotic access to distributed data). ALICE is working on a solution based on the xrootd protocol, coupled with a Mass Storage or Disk Pool manager (CASTOR, dCache, DPM) and with SRM capabilities for file transfers. Furthermore, there will be a transition from a small group of expert users to a wider community, that will increase the need for efficient support.

Primary authors : Dr. BAGNASCO, Stefano (Istituto Nazionale di Fisica Nucleare (INFN))

Co-authors :

Presenter : Dr. BAGNASCO, Stefano (Istituto Nazionale di Fisica Nucleare (INFN))

Track classification : Experience with application domains – setting up and production

Contribution type : --not specified--

Submitted by : BAGNASCO, Stefano

Submitted on Monday 12 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Authorship is on behalf of the ALICE collaboration.

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 110

Status of Fusion Activities in the Grid

Description of the activity :

ITER project is going on and the ITER Team is being arranged and installed in the site in Cadarache. The Fusion research continues to improve the understanding of basic Plasma Physics in order to improve the confinement in future reactors. Several pilot applications suitable for running in the grid has been chosen to produce a demonstration effect in Fusion research community.

Grid added value :

The Fusion VO is working with more that 1,500 CPUs and can be used to run codes. Several applications that are running in EGEE infrastructure are already in production phase and will give us scientifically relevant results:

The Kinetic Transport application consists of a code that follows the trajectories of a large amount of ions that suffer collisions with other ions in complex 3D magnetic structures. An ion-electron collision term has been introduced and it is being checking now. The grid allows us follow a large number (10^7) particles to give us relevant results.

Maratra, which produces massive ray tracing, is useful to study the microwave heating in plasma confinement devices. The basis is to simulate a microwave beam by a bunch with a large number of rays (typically about 104 rays), which implies the necessity of a large capability distributed computation.

Stellarator Optimization. The genetic algorithm that is running in the Russian grid and is able to produce an opt

Experience or proposed activity :

The three above mentioned applications have been developed in close collaboration among the people who are porting the applications to the grid and those who work on Physics, which implies the appearance of synergies. Two methods have been used up to now: Gridway and gLite.

An issue for porting new applications to the grid could be that many of them need to deal with MPI.

Future evolution :

The main intention is exploit the former applications to produce relevant results from the scientific point of view, in order that a demonstration effect happens and the fusion community realises about the advantages of using the grid. In order to extend the use of grid technologies and infrastructure in the fusion community, DKES (Drift Kinetic Equation solver), a very common code to study collisional Neo Classical Transport in tokamaks and stellarators

Primary authors : Dr. CASATEJóN, Francisco (CIEMAT)

Co-authors : Dr. CAMPOS, Isabel (IFCA) ; Dr. CAPPA, Álvaro (CIEMAT) ; Dr. FERNáNDEZ, Luis Antonio (Facultad de Físicas. Universidad Complutense) ; Dr. HUEDO, Eduardo (Facultad de Informática. Universidad Complutense) ; Dr. LLORENTE, Ignacio Martín (FAcultad de Informática. Universidad Complutense) ; Dr. MARTíN, Víctor (Facultad de Físicas. Universidad Complutense) ; Dr. MONTERO, R. S. (FAcultad de Informática. Universidad Complutense) ; Dr. MIKHAIL, Mikhailov (Kurchatov Institute) ; Dr. TARANCóN, Alfonso (BIFI) ; Dr. TERESHCHENKO, Max (Institute of General Physics) ; Dr. VAZQUEZ-POLETTI, José Luis (FAcultad de Informática. Universidad Complutense) ; Dr. VELASCO, José Luis (BIFI) ; Dr. VOZNESENSKY, Vladimir (Kurchatov Insitute)

Presenter : Dr. CASATEJóN, Francisco (CIEMAT)

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : Dr. CASATEJóN, Francisco

Submitted on Monday 12 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 111

G-DSE: Grid – Data Source Engine

Description of the activity :

The G-DSE was developed to allow astrophysical applications running in Grid to access data in astronomical databases, so it was mainly thought to meet a specific problem of the astrophysical community. On G-DSE enabled Grids astronomers are able to discover astronomical databases as any other Grid resource, access them and also use abstract database machines to process data. Thanks to the G-DSE two basic tools for astronomers (the Grid and the Virtual Observatory) can be naturally integrated.

Grid added value :

Astronomical applications (often related to large scale projects like the simulations of the ESA Planck mission) are both computing intensive and data intensive. The usage of the Grid is the right answer to overcome the shortage of computing power resources paining each institution/consortium participating to such projects. The G-DSE makes possible to use the same technology (the Grid) to access computing power and astronomical databases. The IVO (International Virtual Observatory) is a world-wide effort of the astronomical community to define standards for the federation of astronomical databases. The IVO is a Grid of databases devoid of computational resources. The G-DSE is able to bridge the Grid and the IVO, so a new, large community, can benefit of the Grid technology. Although the G-DSE was thought to solve an issue related to astronomical applications, it can be used to access generic databases via Grid, so other scientific communities can benefit of it.

Experience or proposed activity :

The G-DSE was thought and developed in the framework of GRID.IT, a multidisciplinary project for an Italian Grid infrastructure. The first experiments with the G-DSE took place on the INFN Grid infrastructure. Because the development of the G-DSE was resumed only recently, an effort was firstly made to guarantee the compatibility with the current production Grid infrastructure. The next step is to produce a gLite-compatible version of the G-DSE. Tests carried out with the first versions were extremely encouraging. A test campaign is now in progress between INAF, INFN and SPACI for a comparison of different technologies, namely OGSA-DAI, GRelC and G-DSE, aiming at integrating the Grid and Databases.

Future evolution :

The G-DSE is the low-level middleware component of a Grid Query Element. In practice, it involves modifications on the basic components of a Grid resource (GRAM and GIS) so that the new "Database" Grid resource can be discovered and used as any other resource. On the basis of the G-DSE design no modification on the high level services is required. We don't expect technical problems in replicating this also in gLite.

Primary authors : Dr. TAFFONI, Giuliano (INAF-OA Trieste) ; Dr. VUERLI, Claudio (INAF-OA Trieste)

Co-authors : Mr. BARISANI, Andrea (INAF-OA Trieste) ; Dr. MANNA, Valeria (INAF-OA Trieste) ; Dr. PASIAN, Fabio (INAF-OA Trieste)

Presenter : Dr. TAFFONI, Giuliano (INAF-OA Trieste)

Track classification : Data Management

Contribution type : oral presentation

Submitted by : Dr. VUERLI, Claudio

Submitted on Tuesday 13 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 112

EGEE/XtremWeb : Bridging Instititional and Desktop Grids

Description of the activity :

LAL provides resources to its physicists both through the EGEE infrastructure, an institutional grid, and through Xtremweb, a desktop grid. Xtremweb developed by LRI-INRIA and LAL-IN2P3 aims to construct large-scale distributed systems by aggregating individual resources in virtual fault-tolerant clusters. It is similar concept to BOINC developed at Berkeley. Currently, these two sets of resources serve different groups within the laboratory and present markedly different interfaces to their

Grid added value :

We propose to bridge the EGEE and Xtremweb infrastructures by making the Xtremweb system appear as a standard Computing Element on the EGEE infrastructure, bringing more resources into the EGEE infrastructure from a currently unutilized pool of resources. The bridging could potentially increase overall efficiency by off-loading smaller tasks to the desktop systems and freeing larger, "institutional" resources for larger tasks. Primarily, however, the merged system provides an unique interface to the users, freeing them from costly ports of their software to different systems.

Experience or proposed activity :

We have built a prototype that successfully links the EGEE grid infrastructure to resources managed by Xtremweb. Currently the prototype uses Xtremweb resources that have the gLite client software installed to ensure users have the environment that they expect. A standard gatekeeper (LCG CE) acts as the interface and rudimentary state information is published to permit grid-level scheduling.

Future evolution :

Three areas requires further work. The most trivial is binding for the more recent gLite CE. Then, to fully take advantage of the two systems, the gLite developers must make the software portable enough to run on the volatile, heterogeneous resources typical of desktop grids. Lastly, an efficient grid meta-schedulling needs predictive mechanisms that forecast the size and type of volatile resources. This is a general issue that must be solved regardless of the desktop grid implementation.

Primary authors : Dr. LODYGENSKY, Oelg (LAL - IN2P3 - CNRS) ; Dr. LOOMIS, Charles (LAL - IN2P3 - CNRS)

Co-authors :

Presenter : Dr. LODYGENSKY, Oelg (LAL - IN2P3 - CNRS)

Track classification : Poster session

Contribution type : poster

Submitted by : Dr. LODYGENSKY, Oelg

Submitted on Tuesday 13 February 2007

Last modified on : Friday 16 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 113

The EELA Project Applications

Description of the activity :

EELA runs HEP, Biomed, e-Learning and Climate applications. It uses its own gLite-based e-Infrastructure to deploy EGEE applications (ALICE,LHCb,GATE,WISDOM) and new EELA ones: BiG (alignment of molecules), Phylogenesys (history of an organism), VoD (interactive multimedia server), LEMDist (access to distributed instrumentation), CuGfL (Learning Management System), SATyrus (neurosymbolic application), PILP (Inductive Logic Programming), CAM (Global Climate model) or WRF (Regional Climate model)

Grid added value :

The added value for the EGEE applications has been demonstrated. BiG and Phylogenetic processes are very computing-intensive for finding homologous of sequences in huge data bases (with a size higher than a gigabyte) and for calculating large-scale phylogeny studies (a medium size of 50 sequences x 300 nucleotides takes 50 hours on a state-of-the-art PC). Concerning e-Learning tools (VoD, LEMDist, CuGfL), clients can use any UI available and registered on the EELA VO to retrieve any videos or execute any experiments that may be located elsewhere; even more, EELA plans to use these applications in their tutorials in a future. For HPC tools (PILP, SATyrus) they are more efficient with a distributed use, so Grid is very convenient. Climate applications (CAM, WRF) deal with a large number of datasets stored locally, so grid technology can offer a solution to access them in a transparent way. The whole Latin American scientific community is very interested in these applications

Experience or proposed activity :

All the applications are gLite compatible. BiG and Phylogenetics are going to be run through a web portal with access to the EELA Grid performed through the Gate-to-Grid (an EELA Grid node which provides a WSRF-Based Web interface) using a MyProxy repository for the security. VoD uses a multimedia server installed on entry points of the EELA infra-structure with a VO-box so it retrieves the data from the storage servers. CuGfL uses a LMS-Moodle architecture able to work on the Grid. LEMDist uses several layers; the Grid one uses gLite services. PILP extracts relevant knowledge from structured data starting one random search per machine. SATyrus is multistart metaheuristics: multiple networks are produced, each mapped onto the available grid machine and each starting from a different point in the state space of solutions. The Climate applications need LFC, MPI, AMGA and R-GMA

Future evolution :

No limitations are expected

Primary authors : Dr. MAYO, Rafael (CIEMAT) ; Mr. CASADO, Jesus (CIEMAT)

Co-authors :

Presenter : Dr. MAYO, Rafael (CIEMAT) ; Mr. CASADO, Jesus (CIEMAT)

Track classification : Related Projects

Contribution type : oral presentation

Submitted by : Dr. MAYO, Rafael

Submitted on Tuesday 13 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Related Projects

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 114

Implementation and performance analysis of the LHCb LFC replica using Oracle streams technology

Description of the activity :

The presentation will describe the architecture and the deployment of the LHCb read only File Catalog for the LHC Computing Grid (LFC) replica implemented at the Italian INFN National Center for Telematics and Informatics (CNAF) and evaluates a series of tests on the LFC with replica. The LHCb computing model foresees the replication of the central LFC database in every Tier-1, in order to assure more scalability and fault tolerance to LHCb applications.

Grid added value :

Scientific data intensive applications use a large collections of files for storing data. In particular, as regards the HEP community, data generated by large detectors will be managed and stored using databases. The intensive access to information stored in databases by the Grid computing applications requires a distributed database replication in order to guarantee the scalability and, in case of failure, redundancy. Besides the results of the tests will be an important reference for all the Grid users.

Experience or proposed activity :

This talk will describe the replica implementation of LHCb LFC database that is foreseen to be done to each LHCb Tier-1 site. The LFC database replication is realized by means of Oracle Streams technology, a high-speed tool that allows synchronization and full consistency of replicated database. The streams technology takes care of propagating information from one database to another. In order to verify the requirements asked by the LHCb Collaboration a series of tests have been performed. The tests are designed to evaluate performances, delay on the propagation of the streams and the scalability of LFC. The tests show the robustness of the replica implementation with performance going beyond the experiment's requirements. In our tests we also took into account these issues verifying the ability of the LFC to retry a connection if the Oracle server becomes unavailable, as well as the ability of the LHCb applications to choose an LFC replica and if this replica fails.

Future evolution :

The LHCb LFC replica implementation will allow to the LHCb users to perform data analysis with high performance, scalability and fault tolerance. This results of this experience will be useful for many other EGEE applications based on Oracle databases. Applications need to be aware of the underlying replicated environment in order to take advantage of the added fault tolerance and scalability.

Primary authors : CARBONE, Angelo (INFN-Bologna & CNAF-Tier1) ; MARTELLI, Barbara (CNAF-Tier1) ; PECO, Gianluca (INFN-Bologna)

Co-authors : BONIFAZZI, Federico (CNAF-Tier1) ; DUELLMANN, Dirk (CERN) ; DA FONTE PEREZ, Eva (CERN) ; BARANOWSKI, Zbigniew (Warsaw University of Technology) ; VAGNONI, Vincenzo (INFN-Bologna)

Presenter : CARBONE, Angelo (INFN-Bologna & CNAF-Tier1)

Track classification : Data Management

Contribution type : oral presentation

Submitted by : CARBONE, Angelo

Submitted on Tuesday 13 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 115

Access to user data via HTTP and HTTPS

Description of the activity :

.

Grid added value :

We describe the advantages of HTTP and HTTPS as data transfer protocols for users of grid environments, with particular reference to GridFTP, concentrating on the widespread support for the HTTP protocol in all languages, the robustness and maturity of their implementations, the straightforward way in which X.509, GSI and VOMS security credentials may be transmitted via HTTPS, and the existing support for read/write access, transfer continuation, third-party transfers, and multistream data flows
which are implicit in the HTTP(S) IETF standards and common implementations. In addition, we discuss how the GridSite/Apache framework which is used by EGEE for some web services can also support data transfers, and the ease with which users can share data in ad-hoc or dynamic groups on the Grid using these tools. Finally, we present new comparisons of the performance of HTTP and other data transfer protocols across the production network.

Experience or proposed activity :

.

Future evolution :

.

Primary authors : Dr. MCNAB, Andrew (UNIVERSITY OF MANCHESTER) ; Dr. LI, Yibiao (UNIVERSITY OF MANCHESTER)

Co-authors :

Presenter : Dr. MCNAB, Andrew (UNIVERSITY OF MANCHESTER)

Track classification : Workflow ; Data Management

Contribution type : oral presentation

Submitted by : Dr. MCNAB, Andrew

Submitted on Tuesday 13 February 2007

Last modified on : Wednesday 14 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Track : Data Management

Judgment :

Judged by :
Date :
Comments : ""

Abstract ID : 116

Data Management in LHCb: consistency, integrity and coherence of data

Description of the activity :

The Large Hadron Collider (LHC) at CERN will start operating in 2007.
The LHCb experiment is preparing for the real data handling and analysis via a series of data challenges and production exercises. The aim of these activities is to demonstrate the readiness of the computing infrastructure based on WLCG (Worldwide LHC Computing Grid) technologies, to validate the computing model and to provide useful samples of data for detector and physics studies.

Grid added value :

DIRAC (Distributed Infrastructure with Remote Agent Control) is the gateway to WLCG. The Dirac Data Management System (DMS) relies on both WLCG Data Management services (LCG File Catalogues, Storage Resource Managers and File Transfer Service) and LHCb specific components (Bookkeeping Metadata File Catalogue). Although the Dirac DMS has been extensively used over the past years and has proved to achieve a high grade of maturity and reliability, the complexity of both the DMS and its interactions with numerous WLCG components as well as the instability of facilities concerned, turned frequently into unexpected problems in data moving and/or data registration. Such problems make it impossible at all times to have a coherent picture of experimental data-grid across various services involved.

Experience or proposed activity :

The LHCb policy on these issues has been addressed towards an investment in resources targeting
the minimization of the number of occurrences involving data corruptions, data missing, data incoherence and inconsistencies among Catalogues and physical storages, both through safety measures at data management level (failover mechanisms, check sums, roll back mechanisms) and through expensive background checks. The data integrity and the consistency checks activity are presented here. The goal of this activity is to be able to maintain a consistent picture of the main catalogues (Bookkeeping and LFC) and the Storage Elements, primarily among them, and at a second order with the computing model.
While reducing actively the number of these interventions still represents the main goal of the DMS in LHCb, the outcome of these checks represents also a lucid evaluation of the quality of service offered by the underlying Grid infrastructure.

Future evolution :

The planned activity on data integrity, consistency and coherence in the Grid is addressed for the
development, in a near future, of a generic tool suite able to categorize, analyze and systematically cure the disparate problems affecting data management. The advantages are: the efforts made to solve immediate problems can be embedded in more generic and higher level tools; and fixes to some problems can be applied to DIRAC as well to avoid repetitions of problems.

Primary authors : BARGIOTTI, Marianne (European Organization for Nuclear Research (CERN))

Co-authors :

Presenter : BARGIOTTI, Marianne (European Organization for Nuclear Research (CERN))

Track classification : Workflow

Contribution type : oral presentation

Submitted by : BARGIOTTI, Marianne

Submitted on Tuesday 13 February 2007

Last modified on : Wednesday 14 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 117

The Gap between Grid Users and Infrastructure providers - a GridPP perspective

Description of the activity :

GridPP is the major provider of EGEE resources in the UK and Ireland region. The community represented spans individual site administrators to regional coordination bodies.

Grid added value :

The Grid is what the authors of this abstract are providing so this question is not possible to answer directly. Our project existence is based on the defined need for distributed computing that can cope with the data output and processing associated with experiments that will come on line with the Large Hadron Collider at CERN later this year. There are thousands of scientists associated with this project, hundreds of whom are in the UK. In addition we work with scientists from other areas of High Energy Physics and Science to explore the potential value of Grid computing.

Experience or proposed activity :

GridPP was setup primarily to provide, and help enable High Energy Physics applications to make use of, Grid resources. In practice the deployment and operations team of GridPP (which also form a large part of the EGEE UK&I ROC) spend time ensuring that a much broader community can use the Grid. Thus we can relay experience working with HEP groups outside of the LHC world, non-HEP groups such as the biomedical community and business applications (such as Cambridge Ontology and TOTAL E&P).

Future evolution :

- Requirements of large vs small VOs and how they interact with the grid resource providers
- Common problems in specifying job parameters
- VO local disk requirements
- Centralised monitoring of jobs
- Feedback on job failure reasons
- Keeping the grid secure
- Testing sites vs VO environment needs
- Working with users to resolve problems (via GGUS)
- VOMS groups and roles

Primary authors : Dr. COLES, Jeremy (Physics Department)

Co-authors :

Presenter : Dr. COLES, Jeremy (Physics Department)

Track classification : Workflow ; Data Management

Contribution type : oral presentation

Submitted by : COLES, Jeremy

Submitted on Tuesday 13 February 2007

Last modified on : Wednesday 14 February 2007

Comments :

As the above fields do not lend themselves to our purpose in wishing to submit
an abstract here is what we would like to propose:

The UK GridPP deployment and operations team together with members
of the UK & Ireland ROC have been very active in providing resources
for WLCG and EGEE. This talk will examine some of the findings and
issues identified by these teams that relate to users. Topics such as
the need for clear communication between sites and VOs and the timely
follow up of problem tickets are examined. Once a VO is setup problems
can occur due to the misconfiguration of site middleware components
and in this context the current gap between automated infrastructure
testing and experiment environment testing is investigated. The talk
concludes with a clear list of things that need to be addressed to
make grid computing sustainable for both users and the infrastructure providers."

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 118

Parallel Programming and Multi-Cluster Execution Using “Grid Library/Service Applications”

Description of the activity :

The availability of the Grid infrastructure enables a wide new area of scientific, technical and other problems to be solved using Grid Applications. It is obvious that this infrastructure can be used by scientists, researchers, artists and others in any field of human creative effort.

Grid added value :

However, all of the mentioned communities are, in this moment of time, limited to the use of already gridified applications. Development of a new Grid application is a very tedious and time/effort/money consuming process. A common approach is to coordinate the development, from an initial idea, through the development of algorithms up to the gridification self, between the scientists/researchers/artists who necessitate the use of the emerging application and the computer experts, which have to implement it. This process usually, after a lot of strain, results in an Application, i.e. a gridified programmed
system of algorithms, usable for a more or less, but nevertheless very limited, specific set of data and results. A novel approach is being proposed, which we, at the RBI, started implementing, where not fully fledged Applications are being developed, but gridified "Library Applications" and "Service Applications".

Experience or proposed activity :

A "Library Application" mentioned above is actually a specific, non-user community targeted (or multi-user community targeted) algorithm or algorithmic system gridified in such a way that it's primary purpose is not to be used by any human user (although it is open to such uses, too), but to be used as a selfstanding grid "Application" (note the quotes) which is actually a sub-programme (sub-application) of the user Application, i.e. the end-user Application uses the "Library Application" in a very similar way a computer programme uses programme libraries. However, the scope of the "Library Application" is much wider, as it does not provide specific functions, but a very complex algorithm execution, including, where necessary, high level programibility (e.g. a pattern recognition system, or a mathematical system). "Service Applications" are specific services (e.g. a quantum random number generator) wrapped up into a "Library Application".

Future evolution :

This approach enables much easier user PROGRAMMING of the Grid, as the LAPIs (Library/Service Application Programming Interface) enable a coherent and consistent use of a huge variety of algorithms and complex programmable executioners executing in parallel on different parts of the Grid (enabling also use of very specific equipment, e.g. vector processors, quantum processors...). The main Application would actually be just a "driver" for the necessary algorithmic "applications"

Primary authors : Prof. ŠOJAT, Zorislav (Ruđer Bošković Institute)

Co-authors : Mr. ĆOSIć, Tomislav (Ruđer Bošković Institute) ; Dr. SKALA, Karolj (Ruđer Bošković Institute)

Presenter : Prof. ŠOJAT, Zorislav (Ruđer Bošković Institute)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : SOJAT, Zorislav

Submitted on Tuesday 13 February 2007

Last modified on : Tuesday 13 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 119

Modeling network for File Transfer Service : a network view from the application side.

Description of the activity :

Our work aim is to provide a finer network model of channels in File Transfer Service (FTS). Moreover, our work could be helpful to any other data intensive application which might use simultaneous bulk data transfers from different sources and destinations. So, scientific communities who could be interested by our work are both File Transfer Service developers, as well as developers intending to optimize their bulk data transfers.

Grid added value :

We provide a finer model of channels by identifying and inferring performances of the network. In the past, tools have been developed and has been widely used in order to predict performances of client/server communications data transfers. However, nowadays grids connect up to thousands communicating resources that may interact in a partially or totally coordinated way. Consequently, applications running upon this kind of platform often involve massively concurrent bulk data transfers. This implies that the client/server model is no longer valid. One must discover and evaluate impact of concurrent data transfers incoming from separate sources and outgoing to separate destinations. From an FTS point of view, this mean discover possible physical bottleneck common to logically separated channels. This is a significant shift between client/servers approaches and those dedicated to the Grid.

Experience or proposed activity :

Our ongoing research is focused on two main axis : first, design an accurate model that matches the need of File Transfer Service or any other data intensive application. So far, we have a proposal of such a model. In order to validate and/or refine such a model it is essential to have feedback from potential end users of it. The other axis is to develop both efficient measurement methods and reconstruction algorithms. We have now prototypes of such methods and algorithms that have been validated via simulation under both NS and SimGrid simulator. As those methods relies on some assumptions on both network state and acceptable performances of measurement procedure, we need to evaluate what are the expectations of end users in terms of delay induced by such methods and exchange knowledge about real experiment upon the network.

Future evolution :

By using EGEE infrastructure and meet its developers, we expect to have means to evaluate the validity of our approach and validate methods and tools we develop in a real world scenario.

Primary authors : Mr. BOBELIN, Laurent (CSSI)

Co-authors :

Presenter : Mr. BOBELIN, Laurent (CSSI)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Mr. BOBELIN, Laurent

Submitted on Tuesday 13 February 2007

Last modified on : Tuesday 13 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 120

Applying Grid Technologies to In Silico Oncology

Description of the activity :

In silico oncology is an emerging interdisciplinary field aiming at mathematically describing and computationally simulating the multiscale biological mechanisms that constitute the phenomenon of cancer and its response to therapeutic techniques. Within this framework, the In Silico Oncology Group, National Technical University of Athens, has already developed a four-dimensional simulation model of glioblastoma multiform response to radiotherapy.

Grid added value :

Due to the hypercomplexity of the problem, high-performance computing infrastructures are necessary. In order to simulate numerous candidate therapeutic scenarios as fast as possible, grid technologies seem to be particularly effective. In addition, as tumor response to radiotherapy is a highly non linear phenomenon, parallel executions of the simulation code for a large number of sets of parameters are highly desirable, in order to gain better insight into the dynamics of the system.

Experience or proposed activity :

In order for In Silico Oncology to be efficiently transferred to the EGEE infrastructure, certain aspects need to be addressed regarding its adaptation to
the grid programming model. First and foremost, suitable workflows need to be developed for the coordination of the grid-enabled application components responsible for job-simulation submission and monitoring, resource monitoring, data management and result retrieval, which will provide some basic quality of service. A first approach towards grid-enabling In Silico Oncology is to execute several simulations in parallel, thus reducing the overall time that a researcher or a doctor has to wait for different simulation results. Jobs-simulations may be efficiently scheduled by the gLite workload management system, according to system loading criteria and data locality.

Future evolution :

Beyond the basic functionality described above, it is also important that further mechanisms for fault tolerance and quality of service are developed and incorporated to the application. Fault tolerance is a feature that is not supported inherently by the grid middleware at present and that is highly desirable for a grid-based application. QoS may be achieved by taking into consideration workload and resource capacity estimation, resulting in more complex scheduling patterns.

Primary authors : Mr. ATHANAILEAS, Theodoros (National Technical University of Athens) ; Dr. DIONYSIOU, Dimitra (National Technical University of Athens) ; Dr. STAMATAKOS, Georgios (National Technical University of Athens) ; Dr. MOURAVLIANSKY, Nikolaos (National Technical University of Athens) ; Prof. KAKLAMANI, Dimitra (National Technical University of Athens) ; Prof. UZUNOGLU, Nikolaos (National Technical University of Athens)

Co-authors :

Presenter : Mr. ATHANAILEAS, Theodoros (National Technical University of Athens)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : ATHANAILEAS, Theodoros

Submitted on Tuesday 13 February 2007

Last modified on : Wednesday 14 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 121

Accounting in EGEE/WLCG

Description of the activity :

The accounting activity collects the accounting data of all sites participating in the EGEE and WLCG infrastructures as well as from sites belonging to other grids that are collaborating with EGEE (OSG, Nordugrid, ..).
The data are then available for use, in different views through the Accounting Portal, by: all scientific communities using the named Grid Infrastructures; the owners and administrators of the constituent resources, and the management of the Grid Infrastructures.

Grid added value :

The accounting statistics available through the Accounting Portal help the community to understand how the grid is being used by the different VO and users. How the resources are being distributed among the different VO and sites. How many resources (memory, disk, cpu, etc.) are needed to perform the different scientific tasks. The efficiency of the different jobs and the reliability of the sites. To establish SLA and to check them. The added value obtained is an increase in efficiency of use of the resources on the various Grids and an optimisation of resources delivered to the various scientific communities.

Experience or proposed activity :

The accounting activity comprises the following services:
- R-GMA: The collection of accounting usage records is done through R-GMA, an implementation of the Grid Monitoring Architecture (GMA) proposed by the Global Grid Forum (GGF).
- Apel: Apel is a log processing application which is used to interpret gatekeeper and batch system logs to produce accounting records.
- Accounting Enforcement Task: To ensure the "data rightness" it is required to supervise and analyse the accounting data. Sites with problems publishing data are contacted using the EGEE GGUS support and ticketing system.
- Accounting Portal: Accounting statistics are available through the accounting portal for the analysis of the different grid users, VO admins and site administrators. The APG advisory board ensures the development of the accounting portal meets the requirements of the grid community.

Future evolution :

Future work:
- Storage Accounting: availability and usage of disk and tape storage.
- Adoption of more OGF Standards. The OGF Usage Record is already used. Resource Usage Service web services interfaces for data transfer and interrogation are being developed.
- Include information about the reliability of the infrastructure (cancelled/failed/successful jobs) and the efficiency of the jobs/sites.
- Include information about the differences between committed (MoU), installed and used resources.

Primary authors : Mr. REY MAYO, Pablo (CESGA) ; Dr. LOPEZ, Javier (CESGA) ; Dr. KANT, Dave (RAL-LCG2) ; Dr. GORDON, John (RAL-LCG2) ; Dr. FERNANDEZ, Carlos (CESGA)

Co-authors :

Presenter : Dr. LOPEZ, Javier (CESGA) ; Dr. FERNANDEZ, Carlos (CESGA)

Track classification : Grid Monitoring and Accounting

Contribution type : oral presentation

Submitted by : PABLO, Rey Mayo

Submitted on Tuesday 13 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Grid Monitoring and Accounting

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 122

ARDA Dashboard Data Management Monitoring

Description of the activity :

The Atlas DDM (Distributed Data Management) system is responsible for the management and distribution of data across the different grid sites. The data is generated at CERN and has to be made available as fast as possible in a large number of centers for production purposes, and later in many other sites for end user analysis. Monitoring their data transfer activity and availability is an essential task for both site administrators and end users doing analysis in their local centers.

Grid added value :

Data management using the grid depends on a complex set of services. File catalogs for file and file location bookkeeping, transfer services for file movement, storage managers and others. In addition there are several flavors of each of these components, tens of sites each managing a distinct installation - over 100 at the present time - and in some organizations data is seen and moved in larger granularity than files - usually called datasets, which makes the successful usage of the standard grid monitoring tools a non straightforward task.
The dashboard provides a unified view of the whole data management infrastructure, relying mostly on the Atlas data management (DDM) system to collect the relevant information regarding dataset and file movement between the different sites, but also retrieving information from the grid fabric services where appropriate. This last point makes it an interesting tool also for other communities that rely on the same lower level grid services.

Experience or proposed activity :

Focusing mostly on data management on the grid, the most relevant services for this area of the dashboard are the transfer services and storage managers. It is essential that all information can be easily and quickly propagated to the dashboard service, either directly or via the DDM services, so that end users can have an almost real-time view over their activities and production systems can rely on the system views provided by the monitoring.
File transfer information is transient in most cases, and taken from the main transfer tool being used - the File Transfer Service (FTS). Storage and storage space information lies in the Storage Resource Managers (SRM), which should be able to provide a unique and implementation independent over the physical data and available space. Information regarding file and system meta data is expected to be kept consistent everywhere, and any changes to be propagated to the interested services - like the dashboard.

Future evolution :

The main problem dealing with grid data management tools today is tracking the source of errors. It is very complicated to understand the cause of a file transfer failure, even to identify the service or service class that is responsible for the error, or to distinguish a service problem from a user mistake. We expect the FTS and the new SRM interface to be able to provide and expose to end users better and more consistent error categories, essential to reduce the effort needed today.

Primary authors : ROCHA, Ricardo (CERN) ; GAIDIOZ, Benjamin (CERN) ; ANDREEVA, Julia (CERN) ; SAIZ, Pablo (CERN)

Co-authors :

Presenter : ROCHA, Ricardo (CERN)

Track classification : Data Management

Contribution type : oral presentation

Submitted by : BRITO DA ROCHA, Ricardo

Submitted on Tuesday 13 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 123

BioinfoGRIDBlast: a new approach at automatic database management, update and replication for Grid-enabled BLAST

Description of the activity :

The EGEE platform is most useful for research groups producing uneven computational workloads. In these cases the cost of joining the EGEE Grid is substantially lower of that of an owned cluster. This can provide benefits to many research groups such as bioinformatics research groups. BioinfoGRID EU project is already using the Grid for computationally intensive tasks. A dramatic performance gain is relatively easy to obtain for applications which can be trivially parallelized.

Grid added value :

Blast if a typical use case for a computationally intensive application which can leverage the Grid obtaining dramatic speedups compared to local execution. Blast is also very widely used in Bioinformatics research groups. A working, centrally maintained and optimized grid Blast installation can provide significant benefits for the bioinformatics community. Problems to be faced when porting Blast to the Grid platform are: Grid job management, databases updates management, availability of all versions of several databases, amount of replication for each of the databases (storage costs are significant). We addressed such problems in our BioinfoGRIDBlast. BioinfoGRIDBlast sports an automatic management system for keeping the databases constantly updated and to dynamically adjusts the amount of replication for each database. The amount of replication is based on the recent usage amount for each database; this constantly optimizes the balance of storage costs vs grid availability.

Experience or proposed activity :

BioinfoGRIDBlast is an effort for providing the user a Blast system over the Grid with automatic database management.
The core for BioinfoGRIDBlast provides workload slicing into small jobs, and jobs tracking and management over the Grid. On top of this the following functionalities have been implemented:
1) An updater engine maintains the BLAST biological databases on the Grid constantly updated. The updates are polled and fetched from the FTP sites of origin for each database.
2) Older versions of the biological databases are also kept available on the Grid,but instead of uploading each version in full BioinfoGRIDBlast implements a patch system based on xdelta to dramatically reduce storage costs.
3) A dynamic Replication Engine (RE) keeps track of the usage for each database. The RE then constantly adapts the number of replicas for each database, keeping it proportional to the database's recent usage. This keeps Blast responsiveness high while keeping storage costs bearable.

Future evolution :

We noticed a few tricky commands such as lcg-cp not exiting after download is completed. These can be worked around relatively easily once they are known. The efficiency of the queueing system might be improved: sometimes not the optimal queues are chosen by the Broker.

Primary authors : Mr. TROMBETTI, Gabriele (CNR-ITB)

Co-authors : Mr. MERELLI, Ivan (CNR-ITB) ; Dr. ORRO, Alessandro (CNR-ITB) ; Dr. MILANESI, Luciano (CNR-ITB)

Presenter : Mr. TROMBETTI, Gabriele (CNR-ITB)

Track classification : Data Management

Contribution type : oral presentation

Submitted by : Mr. TROMBETTI, Gabriele

Submitted on Tuesday 13 February 2007

Last modified on : Sunday 18 February 2007

Comments :

We seek approval as either oral (preferred) or poster. We do not seek approval as "Demo Presentation" because of the wait times involved in this kind of features, which would make a demo unpractical.

Status : SUBMITTED

Track judgments :

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 124

CDF Monte Carlo Production using LcgCAF, the portal to the LCG Grid resources.

Description of the activity :

LcgCAF is a portal used by CDF experiment to access Grid resources by using the LCG/EGEE Middleware components. This is a reimplementation of the old CDF portal used to access dedicated pools.
LcgCAF is constituted by a set of services each of them responsible for accepting, submitting and monitoring CDF user jobs during theirs lifetimes in the Grid environment.
These services are running on a CDF User Interface, which is the CDF submission point, and then delegate the jobs to the Grid.

Grid added value :

To cope with increased demand on computing resources for Monte Carlo production the CDF collaboration, constitued by more than 800 physicists, had to move beyond the usage of dedicated farms and start exploiting Grid resources.
In Europe this can be done via LcgCAF, a unique portal for CDF for the European Grid.
The full functionalities of the WMS are exploited either for job submission, job resubmission in case of failures and the job monitoring.
The CDF code is not available on Grid sites but CDF uses Parrot, a virtual file system, to make it available on each worker node.
With LcgCAF CDF has to maintain only the portal and it exploits the Grid resources and the services availables respect to the old configuration when it had to maintain dedicated farms.
This portal is used by all the CDF community as of today and can be used with small modifications also by other experiments which are not fully supported by the Grid.

Experience or proposed activity :

LcgCAF is in production for Monte Carlo production for CDF since November.
Simulated data are produced on Grid sites and the shipped to Fnal.
One of the major requirements for a running experiment that has data every day and need Monte Carlo data every day is the stability.
In LcgCAF the key Grid service is the WMS. The WMS has to be fast enough to submit jobs.
The WMS has to up and running always. Time by time the WMS gets down and all the jobs submitted can not be monitored and if the down is not recoverd in time the job can be lost and the bookkeeping becomes really difficult.
In these months CDF experienced several problems with the WMS whose performances were not optimal.
The VOMS is the other service used for the users authentication on which LcgCAF is based and had excellent performances in the period of running.

Future evolution :

A WMS stable and matching the declared performances is the basic requirement for substitute the dedicated farms with the Grid resources in Europe for CDF experiment.
CDF would use the storage elements to keep the simulated data before transfering files to Fermilab if a SRM that manages disk space is available.
At the moment a home-made interactive monitor has been developed since nothing like that is on the market but it has several failures due to the difficulty of

interacting with worker node

Primary authors : Dr. LUCCHESI, Donatella (INFN and University Padova) ; Dr. PAGAN GRISO, Simone (INFN and University Padova) ; Dr. COMPOSTELLA, Gabriele (INFN and University Trento)

Co-authors : Dr. SFILIGOI, Igor (Laboratori Nazionali di Frascati) ; Dr. SARKAR, Subir (INFN Pisa) ; Dr. JEANS, Daniel (INFN National Center for Telematics and Informatics) ; Dr. DELLI PAOLI, Francesco (INFN Padova)

Presenter : Dr. PAGAN GRISO, Simone (INFN and University Padova)

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : Dr. LUCCHESI, Donatella

Submitted on Tuesday 13 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 125

Grid Portal for seismic data processing and geophysical simulations

Description of the activity :

BEInGRID, Business Experiments in GRID, is the European Union's largest integrated project funded by the Information Society Technologies (IST) research, part of the EU's sixth research Framework Programme (FP6). The aim of our experiment is to validate the technical and economic viability of geo-science processing on the Grid using the gLite middleware.

Grid added value :

Modern seismic data processing and geophysical simulations require greater amounts of computing power, data storage and sophisticated software. The research community hardly keeps pace with this evolution, resulting in difficulties for small or medium research centres to exploit their innovative algorithms.
Grid Computing is an opportunity to foster sharing of computer resources and give access to large computing power for a limited period of time at an affordable cost, as well as sharing data and sophisticated software.
The capability to solve new complex problems and validate innovative algorithms on real scale problems is also a way to attract and keep the brightest researchers for the benefit of both the academic and industrial R&D geosciences communities.

Experience or proposed activity :

The main business objective of this BE is to validate the economics viability by creating a model for economics suitable for EGEODE Virtual Organization and to explore new business generated by deploying grid technology in Oil & Gas, in term of cost saving and support of IT.
The technical objective of this BE is to validate the technical viability of geosciences processing on the grid by consolidating EGEODE actual applications, in using and validating gLite outside the general EGEE infrastructure.

The experiment is based on 3 main phases
* Consolidation of applications and services, additional porting, install tools to gather information for setting up the economic model, feedback to BEINGRID core tools.
* Run the applications, gathers figures, simulate and design the economic model.
* Integrate last versions of software (both middleware and applications), experiment the economic model, final report.

Future evolution :

One barrier to deploy such simulation techniques is also the complexity of access to grid environment by end-users who are not computer scientists. This is why we will implement the EnginFrame Portal to give an ubiquitous Grid access.

Primary authors : Mr. RUFFINO, Francesco (NICE srl)

Co-authors :

Presenter : Mr. RUFFINO, Francesco (NICE srl)

Track classification : Related Projects

Contribution type : oral presentation

Submitted by : Mr. RUFFINO, Francesco

Submitted on Tuesday 13 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Related Projects

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 126

Virtual Organisations management : the project view

Description of the activity :

The VO Managers group aim is to help virtual organisations work with EGEE. Virtual organisations not only can use the infrastructure, but it is also very important that there is communication between groups of users and the infrastructure : users experience is improved, and the infrastructure has a better knowledge of how to best satisfy its users.

Grid added value :

The VO managers group intends to develop relationships with all VOs, and intends to help their managers whenever it is possible. We want VO management to become as easy and simple as possible, so that people can concentrate on real problems. Managing a VO is done at a user level, but also at a project level : both the VO and the project may have requirements on each other, and collaboration of both is a primary concern for the whole community.

Experience or proposed activity :

The first step in managing a VO is registering it at the EGEE level, on the CIC portal. Therefore, efforts are made on the registration procedure, in coordination with the OAG, to have new VOs get quickly started. There is no need in registering if the information gathered at that step is not used : we therefore are providing efforts so that this useful registration data can be made available and used by various tools, which in turn will render the VO lives easier. Efforts are also made to collect users and groups requirements : either feedback is collected on specific subjects, or VO managers can directly contact the VO Managers group. The key here is the communication : we can only address problems that are reported.

Future evolution :

We hope that in the future, sites basic configuration concerning VOs will be pulled from the CIC portal, so that trivial configuration problems will be eradicated. Issues such as data integrity checks will have to be addressed for the CIC data to be trusted. Finally, the challenge for the group is to become more dynamic and visible to the VO responsibles...

Primary authors : Mr. SCHAER, Frederic (CEA)

Co-authors : Mr. LOOMIS, Charles (CNRS)

Presenter : Mr. SCHAER, Frederic (CEA)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Mr. SCHAER, Frederic

Submitted on Tuesday 13 February 2007

Last modified on : Wednesday 14 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 127

EuroVO Data center alliance

Description of the activity :

The Virtual Observatory(VObs)is a world-wide effort of the astronomical community to enable the international utilization of astronomical archives structured in Databases as an integrated and interoperating resource:it is a knowledge Grid.The Euro-VO Data Center Alliance aims at coordinating and assisting European Data Centres to take up VO standards.It collects all the DCs of Europe.One of the activities of this project is the coordination between the VObs and the Grid systems.

Grid added value :

Astronomical applications are both computational and data intensive. Astronomical data is mainly structured in database systems. The VObs is a "grid" of databases devoid of computational resources. The Euro-VO Data Center Alliance organizes the activities necessary to allow VObs users to exploit (through the data centres involved in the project) the processing capabilities offered by the computational Grid and in particular by EGEE. This activity will involve a part of the Astronomical community in the Grid computing. Moreover, DCs have a number of computational needs related to data reduction and analysis activity. The theoretical data, produced by numerical simulations, is also available in the VObs. Theoretical data post-pressing is in charge of the theoretical DCs and the EGEE grid may be a useful computational resource. Finally, the Grid may be used as a collaboration tool to share data and exchange experiences through DCs.

Experience or proposed activity :

We plan to explore the use of the Grid for DCs through two use cases well representative of the typical needs of a DC. A data reduction use case: a DC gets data from archive using standard VObs protocols (DAL), puts the data on a file space, processes data on the EGEE, saves the output to file space and finally gets output for the user, eventually redirecting them to a visualization tool. We identify two possible data processing activities, one involves catalogues and cross matching tools and a second one involves imaging using Sextractor. The second use case is based on BaSTI.It is a database of stellar evolution models for a large range of masses and chemical compositions. It is created using a stellar evolution code FRANEC that requires long computational time to produce a model.To produce on demand models and to make them available to the VObs users it is necessary to make interoperable VObs and Grid.The EGEE services used will be the workload system and the file catalogue

Future evolution :

This activity is focused on the possibility to make interoperable two communities, the Grid and VObs. One of the key

issues is to identify a common authentication and authorization mechanism that allows DCs and Astronomers
through the DCs to access grid resources. The usage of Grid services via the DCs does not represent in our opinion a
challenging task.

Primary authors : Dr. TAFFONI, Giuliano (INAF - SI)

Co-authors : Dr. VUERLI, Claudio (INAF-OATS) ; Dr. MANNA, Valeria (INAF-OATS) ; Mr. BARISANI, Andrea (INAF-SI) ; Prof. PASIAN, Fabio (INAF-OATS) ; Dr. SMAREGLIA, Riccardo (INAF-OATS)

Presenter : Dr. TAFFONI, Giuliano (INAF - SI)

Track classification : Workflow

Contribution type : --not specified--

Submitted by : Dr. TAFFONI, Giuliano

Submitted on Tuesday 13 February 2007

Last modified on : Wednesday 14 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 128

How Business and Industry can profit on using the EGEE Grid

Description of the activity :

Business and industry can use vast resources (machines, CPUs, storage, applications) linked by the EGEE Grid in the on-demand / ASP model. Apart from solving peek compute capacity issues,
businesses can be granted access to unique, specialized software and architectures normally not available in-
house. The solution from GridwiseTech includes the portal, the security framework, application-specific
user interface framework, and programmer API's integrated directly with LCG and gLite framework

Grid added value :

EGEE presents potential to the business and industry through its infrastructure,resources and applications. Allowing
commercial entities use those resources is only the first step in the long and fascinating journey. Eventually, EGEE
could become a virtual hub for the industry and spark the world's largest business market for resources and
applications accessible through virtualized infrastructure (just like CERN initiated the WWW). The current EGEE grid
could provide a critical mass to start this process.
However, some businesses complain that the infrastructure is only theoretically accessible, while others are
unaware of the chance.
Through our work we have resolved a number of technical, architectural and formal obstacles on the way to allow
external commercial entities to use resources from EGEE and associated providers, using LCG, gLite and our client
interface framework. The shining example of our customer, Philips Research, confirms the business case.

Experience or proposed activity :

We are in contact with Erwin Laure through the EGEE business associates programme. He is helpful in resolving
issues like that; we will route any new issues through him.

Future evolution :

Possibly, the main challenge is the business-unfriendly structure of some EGEE procedures and policies, which may
disappoint potential customers/users. As mentioned above, we are in ongoing discussion with Erwin on those issues.

Primary authors : Mr. PLASZCZAK, Pawel (GridwiseTech)

Co-authors :

Presenter : Mr. PLASZCZAK, Pawel (GridwiseTech)

Track classification : Workflow ; Data Management

Contribution type : --not specified--

Submitted by : Mr. PLASZCZAK, Pawel

Submitted on Wednesday 14 February 2007

Last modified on : Wednesday 14 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 129

Beliefs About the Future of IT and How They Relate to the Enterprise Grid Vision

Description of the activity :
N/A

Grid added value :
N/A

Experience or proposed activity :

- What the baby's brain tells the CIO (Chief Information Officer)
- Beliefs about the future of Information Technology – challenges and opportunities
- Grids for the global enterprise – an ideal IT world? Thesis, anti-thesis and perspectives for action

The presentation looks at high-level IT challenges and trends. The beliefs are about growth of IT capabilities, flattening world, explosive edge, power of the user and more.
Grid for the masses beyond the scientific community is still equally promising and challenging. The timeline is hard to predict. However, in my view, looking at the current eScience achievements, the journey from "System-Level Science" to "System-Level Business" is about to start.

Future evolution :
see above

Primary authors : Dr. KUBLI, Rolf (EDS Information Business GmbH)

Co-authors :

Presenter : Dr. KUBLI, Rolf (EDS Information Business GmbH)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Dr. KUBLI, Rolf

Submitted on Wednesday 14 February 2007

Last modified on : Wednesday 14 February 2007

Comments :
Enterprise IT/business technology view
Author is Chief Technology Officer EDS Switzerland and EDS Fellow

Status : SUBMITTED

Track judgments :

Track : Workflow
Judgment :
Judged by :

Date :
Comments : ""

Abstract ID : 130

Ganga - an Optimiser and Front-End for Grid Job Submission (Demo)

Description of the activity :

The presentation will introduce the Ganga job-management system (http://cern.ch/ganga), developed as an ATLAS-LHCb common project. The main goal of Ganga is to provide a simple and consistent way of preparing, organising and executing analysis tasks, allowing physicists to concentrate on the algorithmic part without having to worry about techical details.

Grid added value :

Ganga provides a clean Python API that reduces and simplifies the work involved in preparing an application, organizing the submission, and gathering results. Technical details of submitting a job to the Grid, for example the preparation of a job-description file, are factored out and taken care of transparently by the systems. By changing the parameter that identifies the execution back-end, a user can trivially switch between running an application on a portable PC, running higher-statistics tests on a local batch system, and analysing all available statistics on the Grid.

Experience or proposed activity :

Although Ganga is being developed for LHCb and ATLAS, it is not limited to use with HEP applications, and already has several non-HEP users. These include users on projects in bio-medicine, engineering, and (Grid) software testing.

Future evolution :

Ganga is a higher-level Grid tool and therefore tries to circumvent typical problems when submitting jobs to the Grid, easing the user experience. Ganga has a plug-in mechanism, so that it can be highly customised to suit the needs of a given user community.

Primary authors : Dr. MAIER, Andrew (CERN)

Co-authors :

Presenter : Dr. MAIER, Andrew (CERN)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : MAIER, Andrew

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 131

Report on the implementation of research applications on the Grid; Experiences and results so far

Description of the activity :

Since the establishment of the GridAUTH technical support team we have come across a number of scenarios regarding the gridification of several applications. We have collaborated with a number of research teams coming from various scientific disciplines, such as the atmospheric physics, earth sciences, gravitational physics, and computational chemistry groups. These teams participate actively in the Earth Science Research VO, the Gaussian VO and the South Eastern Europe VO.

Grid added value :

The distributed and ubiquitous computing model, that the Grid has to offer, matches the modern needs of complex research project activities. Thus the available computational and storage resources make the Grid an exceptional environment for large scale computing. In addition, the global character of the Grid makes it easy for distant scientific groups to collaborate by sharing data and resources at the same time.
One such collaboration our team is a part of is the CECILIA project, a multinational collaboration which primarily aims to model climate changes in the Central and Eastern European region. Our involvement in this project is, together with the atmospheric physics group, to design and implement a functional workflow on the existing EGEE infrastructure using the available computing tools and APIs.
Other collaborations, with the earth sciences and computational chemistry groups, have been established and the results of our workflow implementations are already on the way.

Experience or proposed activity :

Until now our line of work and approach towards the gridification of a given algorithm has been following the "top-down" paradigm. The first steps are to understand the researchers' needs in computer hardware and software components and then implement a simple test case into a basic workflow. Our aim is to build upon that simple test case and produce a practical and functional workflow covering both the set-out criteria and the needs of the project at the same time.
In the meantime we try to encourage the groups to participate in the local Grid community though our web based forum and the weekly user group meetings we organize. We have found that open discussion of technical issues can be beneficial to all the research groups and can lead to the extraction of interesting workflow patterns that could be applied successfully in multiple use cases.

Future evolution :

```
Through our experiences with the aforementioned research groups we have come to
understand a great deal
regarding the researchers' needs both in terms of {hard/middle/soft}ware components
and in terms of Grid
services. Unfortunately, the Grid has not yet reached the state of functionality it
deserves.
Thus, there are still many issues to be overcome regarding the existence of
commercial software and regarding the
availability of services towards the users.
```

Primary authors : Mr. KOROSOGLOU, Paschalis (Aristotle University of Thessaloniki)

Co-authors : Mr. KANELLOPOULOS, Christos (Aristotle University of Thessaloniki) ; Mr. TRIANTAFYLLIDIS, Christos (Aristotle University of Thessaloniki) ; Mr. ZILASKOS, Dimitrios (Aristotle University of Thessaloniki)

Presenter : Mr. KOROSOGLOU, Paschalis (Aristotle University of Thessaloniki)

Track classification : Workflow ; Data Management

Contribution type : oral presentation

Submitted by : PASCHALIS, Korosoglou

Submitted on Wednesday 14 February 2007

Last modified on : Wednesday 14 February 2007

Comments :

In the context of this presentation we wish to share our experiences so far as the technical support team; the obstacles we have faced, the way we did (or did not) overcome them and the so far "success stories".

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 132

The Grid Security Vulnerability and Risk Assessment activity EGEE-II.

Description of the activity :

The Grid Security Vulnerability Group is composed of security experts in EGEE drawn from many regions. The purpose of the activity is to find existing security vulnerabilities in the deployed infrastructure, assess their risk and prioritize their resolution.

Grid added value :

The aim is to "incrementally make the Grid more secure and thus provide better sustainability of the deployed infrastructure". This is to make sure the infrastructure is available for legitimate users, and prevent its use or damage by those who should not use it. The Grid has large resources and as such is an attractive target for attack. This work is relevant to the user community because users need to know what to do if they find, or suspect they have found, a vulnerability within the Grid Middleware or deployment.

Experience or proposed activity :

We have setup a process for handling vulnerabilities, and a strategy for Risk Assessments, along with the appropriate infrastructure. Issues of varying risk have been effectively processed.

Future evolution :

This activity needs to be publicised within both the user community and the sites, to ensure that issues found are handled by the appropriate process and resolved in a timely manner. Testing and code walkthroughs are another aspect of vulnerability detection, including attacks using automated tools. We also plan to provide guidelines for developers to help them avoid developing vulnerable software.

Primary authors : Dr. CORNWALL, Linda Ann (RAL)

Co-authors :

Presenter : Dr. CORNWALL, Linda Ann (RAL)

Track classification : Poster session

Contribution type : poster

Submitted by : CORNWALL, Linda Ann

Submitted on Wednesday 14 February 2007

Last modified on : Friday 16 February 2007

Comments :

I'd like to present a poster at the User Forum in order to advertise to □ users what they should do if they think they might have found a □ vulnerability, and inform them that we do have this activity and how it □ works. □

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 133

Supporting Earth Science community through DILIGENT

Description of the activity :

Earth Science scientists need to access data and tools within a multi-institutional, heterogeneous and large-scale context. The analysis and the generation of objective facts on the Earth status (i.e. Earth Observation, EO) require integration of specific data product, handling of information in multiple forms and use of storage and computing resources in a seamless, dynamic and cost effective way.DILIGENT supports the Earth Science community with the facilities needed to implement such scenario

Grid added value :

Earth Science users are concerned with the acquisition, integration and analysis of information available at different sites. Such processes run according to specific time constraints to meet operational needs, e.g. integration of EO products in forecast models. Large environmental initiatives bring together operational EO-based services, value added service providers and users to establish objective, timely, accurate global information. The integration of technical knowledge from global to local scale aims to support the definition and implementation of environmental policies. While end-users are currently accessing the needed resources through different interfaces and applications, Grid-enabled versions of current EO services have a direct impact on such users allowing them to have a single access point to all resources.
Diligent through EGEE resources and infrastructure enables activities like the dynamic integration of processing results and the creation of environmental reports

Experience or proposed activity :

DILIGENT puts its bases on the integration of digital libraries and Grid technologies. It builds on the EGEE services and part of its test-bed aims at providing Earth Science users with a framework addressing key factors to the EO activities, like management of very large and distributed virtual organisations; seamless access to and handling of distributed and heterogeneous data and services; creation of virtual thematic digital libraries; on-demand and efficient processing of huge amounts of information; definition of ad-hoc user defined workflows of services together with scalable and reliable executions; storage of data as well as of the dependencies between them, and traceability of the operations performed. DILIGENT offers end-users with a uniform information space where services to discover and access information sources are easily accessible; Grid resources used to process data and maintain intermediate results are managed transparently to the end-users.

Future evolution :

EGEE provides the largest production grid infrastructure aiming to serve scientists from different disciplines. Due to the heterogeneity of its clientele it implements generic models and solutions to virtualise access to resources. Computing and storage capacities are the two most important resources of any e-Science application but data constitute an additional essential dimension to meet community specific operational ambitions to promote cross-fertilization and wide collaboration processes

Primary authors : Dr. PAGANO, Pasquale (CNR-ISTI) ; Dr. GUIDETTI, Veronica (ESA-ESRIN)

Co-authors :

Presenter : Dr. PAGANO, Pasquale (CNR-ISTI)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : Dr. PAGANO, Pasquale

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

We intend to deliver an oral presentation and a demo presentation

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 134

Comparative evaluation of tools providing access to different types of data resources exposed on the Grid

Description of the activity :

The problem of managing and accessing huge datasets distributed across multiple sites and stored into heterogeneous databases is common to several research areas.
We report on the comparative evaluation of three tools to access different types of data resources exposed onto Grids: G-DSE (http://wwwas.oats.inaf.it/grid/index.php?option=com_content&task=view&id=70&Itemid=8),
GRelC (http://grelc.unile.it)
and OGSA-DAI (http://www.ogsadai.org.uk/).

Grid added value :

The Grid infrastructure has proven to be capable of providing storage and computing resources for the computational needs of the modern research. However eScience projects need also a way to access widespread databases within a computational grid environment, through a set of secure, interoperable and efficient data grid services.
The evaluation test, reported here, addresses the needs of the bioinformatics community engaged, through the BioinfoGRID (http://www.bioinfogrid.eu/) and the LIBI (http://www.libi.it/) projects, in the adoption of a grid infrastructure layer at the base of their research activities and of the Astrophysical community of the INAF (Istituto Nazionale di Astrofisica) (http://www.inaf.it/) interested to access data in astronomical databases from the GRID,
The access to data from the Grid is also a crucial problem for the adoption of the grid technology to provide services in public administration (EGG project).

Experience or proposed activity :

A test bed, which includes several sites in the Italian production Grid infrastructure, Grid.it, has been set up. In each site the instances of the G-DSE, GRelC and OGSA-DAI servers have been installed together with several instance of their respective clients.
A set of databases provided by the bioinformatics and astrophysical communities were exposed over the grid, some of them located on the same LAN as the access tool, others are accessed remotely.
The test plan spans from very simple queries, use cases provided by the bioinformatics and astrophysical communities, up to stress tests with multiple queries submitted simultaneously from different sites.

Future evolution :

All the three tools relay on the basic authentication mechanism provided by GSI to grant the authorization for accessing the Grid exposed data.
However the Virtual Organization Membership Service, presently employed on the EGEE infrastructure, provides detailed information on the user's relationship with the Virtual Organization and could allow conditional access depending from the user group, role and capability.
An evolution of the tools to become VOMS compliant is highly desirable.

Primary authors : MAGGI, Giorgio (INFN)

Co-authors : TAFFONI, Giuliano (INAF) ; VUERLI, Claudio (INAF) ; BARISANI, Andrea (INAF) ; PASIAN, Fabio (INAF) ; MANNA, F. (INAF) ; GISEL, Andreas (CNR-ITB) ; GIORGIO, Emidio (INFN) ; AIFTIMIEI, Cristina (INFN) ; ATUL, Jain (INFN+Politecnico Bari) ; BARBERA, Roberto (INFN+Università Catania) ; CAROTA, Luciana (INFN) ; DONVITO, Giacinto (INFN) ; GHISELLI, Antonia (INFN) ; LA ROCCA, Giuseppe (INFN) ; MAZZUCATO, Mirco (INFN) ; PIERRO, Antonio (INFN) ; VERLATO, Marco (INFN) ; FIORE, Sandro (Università Lecce) ; ALOISIO, Giovanni (Università Lecce) ; CAFARO, Massimo (Università Lecce) ; VADACCA, Salvatore (Università Lecce) ; NEGRO, Alessandro (Università Lecce) ; DEL FREO, Federico (EGG project)

Presenter : MAGGI, Giorgio (INFN) ; DONVITO, Giacinto (INFN)

Track classification : Data Management

Contribution type : oral presentation

Submitted by : MAGGI, Giorgio

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 135

The Arts and Humanities e-Science Support Centre

Description of the activity :

This poster will explain the role, activities, and context of the Arts and Humanities e-Science Support Centre (AHeSSC), which is based at King's College London and funded
by JISC to support, co-ordinate and promote e-Science in all arts and humanities disciplines. The poster will present to the wider EGEE community the organisational and institutional setup in this very unique area of e-Science in the UK and the wider European context.

Grid added value :

AHeSSC works with so-called early adopters: arts and humanities scholars who have recognised the benefits of e-Science and begun to apply the methods and technologies in their work. In the poster, the relationship between AHeSSC and this community will be outlined with a series of illustrated examples. We also showcase our case-study based outreach work and local training event, with which we try to extend the group of early adopters.
The added value of humanities and arts research on the Grid is the new perspective derived from the specific requirements of this research community with an emphasis on novel usage of the Grid for event-based data as in e.g. performances and interaction.

Experience or proposed activity :

AHeSSC exists within a broader national and international landscape of e-Science/Grid. In the UK this consists of centres of activity in HEIs, and organisations which provide grid and related services. The poster will outline ways in which AHeSSC engages with these and with other European services (e.g. TextGrid and DARIAH), and in particular identify lessons that can be learned for supporting the arts and humanities academic communities, such as the importance of usability and accessibility of resources, as well as basic training needs beyond the direct work on the Grid. The poster will showcase a map of national activities with relevance to the arts and humanities which AHeSSC is producing, and what we understand to be the developing culture of collaboration on e-Infrastructures, within a research area that is only beginning to see the benefits of collaborative knowledge production and that is traditionally characterised by a high degree of silo activities.

Future evolution :

The main purpose of AHeSSC is to provide practical assistance to arts and humanities researchers wishing to use UK and European Grid infrastructure, tools and technology. The issues we encountered stem therefore from our advisory, training, and outreach activities. It is often not easy to showcase with concrete case studies the benefit of the infrastructure to arts and humanities. A technology lookup service would be highly beneficial. Also crucial is easier collaboration with computing experts.

Primary authors : Dr. BLANKE, Tobias (AHeSSC, King's College London)

Co-authors : Dr. DUNN, Stuart (AHeSSC, King's College London)

Presenter : Dr. BLANKE, Tobias (AHeSSC, King's College London) ; Dr. DUNN, Stuart (AHeSSC, King's College London)

Track classification : Poster session

Contribution type : poster

Submitted by : BLANKE, Tobias

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 136

Trials and Tribulations of Grid Training

Description of the activity :

The EGEE project creates a need for provision of training covering a wide range of topics and user communities. The presentation will summarise experiences of trainers and present and future requirements for training by the wide variety of users of the EGEE infrastructure.

Grid added value :

The presentation will examine current models used for giving courses in a grid environment. The demand for different types of training courses - Induction, Application Development, site administrator, Training Trainers - will be explored.

Experience or proposed activity :

The presentation will also include an overview of practical experiences with grid users and their needs. User feedback is an important aspect of training provision but it is often difficult to obtain comprehensive and reliable data. Approaches to improving this feedback will be explored.

Future evolution :

The presentation will conclude with a look to the future of grid training in the light of the ever-expanding subject areas deploying grid applications. The provision of online resources such as e-Learning modules, digital libraries and online user documentation will be explored and mechanisms for support of these resources discussed.

Primary authors : Dr. HAMILL, Brendan (University of Edinburgh) ; Dr. MINETER, Mike (University of Edinburgh)

Co-authors :

Presenter : Dr. HAMILL, Brendan (University of Edinburgh) ; Dr. MINETER, Mike (University of Edinburgh)

Track classification : Training and Support for Grid Users

Contribution type : oral presentation

Submitted by : HAMILL, Brendan James

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Training and Support for Grid Users

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 137

Workflow level parametric study support by the P-GRADE portal

Description of the activity :

The P-GRADE portal is already used by several EGEE and EGEE-related communities. It is used by specific, scientific communities like the EGRID (Economics Grid) and BioInfoGrid as well as by regional communities like VOCE, SEEGRID and BalticGrid. Besides, many national Grid installed it as their Grid portal: Bulgarian Grid, CroGrid, HunGrid, SwissGrid, Turkish Grid, UK NGS, etc. Recently, the GIN VO of OGF uses the portal as their resource testing portal.

Grid added value :

Large workflow applications often require to run the same workflow with many different input parameter sets. This was a clear message from our user communities. Therefore we have decided to extend the portal with the capability of supporting this kind of applications.

We introduced a new workflow concept in P-GRADE where the workflow is partitioned into three parts that are executed in three separate phases. The first partition consists of Generator jobs (either user written or special portal provided ones) that run in the 1st phase to generate the necessary input files for the core workflow part. The 2nd partition consists of the core workflow that is executed with all the input parameter sets generated in the 1st phase. The third partition contains Collector jobs that collects and processes the results of phase 2.

The generated input and output files are stored in SEs and the portal uses the broker services of glite in order to orchestrate the parallel execution of workflows.

Experience or proposed activity :

A parametric study application, particularly if it has to execute a workflow as many times as many input parameter sets are available, requires the orchestration of large number of grid resources. Therefore the grid broker plays a key role in such a service. The robustness of the broker is particularly important when thousands of jobs should be handled in a reliable way. The current middleware is not robust enough to handle thousands of jobs without errors and hence we extended the portal with the following features:

1. The user can suspend (and resume) the execution of any workflow. An often required feature when - for example - the selected Grid resource seems too slow to perform the needed job, and rescue would be advisable.

2. From now on the user does not have to wait for getting partial results until the last job of his workflow terminates. The results of terminated jobs can be downloaded immediately. So, workflows can be aborted if partial results make them useless.

Future evolution :

P-GRADE portal enables grid interoperability at the job, workflow and parameter study level. It means that jobs of a parameter study workflow can be simultaneously executed in different grids (LCG-2, glite, GT2, GT4) connected by the portal. However, to achieve load-balancing between grid resources and

grids we need a meta-broker that can connect brokers of different grids. We are working on the development of such a meta-broker. It will be first connected to glite WMS and GTBroker.

Primary authors : Prof. KACSUK, Peter (MTA SZTAKI)

Co-authors : Mr. FARKAS, Zoltan (MTA SZTAKI)

Presenter : Prof. KACSUK, Peter (MTA SZTAKI)

Track classification : Workflow ; Data Management

Contribution type : oral presentation

Submitted by : KACSUK, Peter

Submitted on Wednesday 14 February 2007

Last modified on : Wednesday 14 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 138

DIRAC Workload Management System

Description of the activity :

DIRAC (Distributed Infrastructure with Remote Agent Control) is the Workload and Data Management system (WMS) for the LHCb experiment. The DIRAC WMS offers a transparent way for LHCb users to submit jobs to the EGEE Grid as well as local clusters and individual PCs. This paper will describe workload management optimizations which ensure high job efficiency and minimized job start times.

Grid added value :

The computing requirements of the LHCb experiment can only be fulfilled through the use of many distributed compute resources. DIRAC provides a robust platform to run data productions on all the resources available to LHCb including the EGEE Grid. More recently, user support was added to DIRAC that greatly simplifies the procedure of submitting, monitoring and retrieving output of Grid jobs for the LHCb user community.

Experience or proposed activity :

DIRAC submits Pilot Agents to the EGEE Grid via the gLite WMS as normal jobs. Pilot Agents then request jobs from the DIRAC Workload Management System after the local environment has been checked. Therefore DIRAC realizes the so-called PULL paradigm which ensures a high efficiency for LHCb Grid jobs.

Future evolution :

The possibility of using generic VO Pilot Agents is very exciting and DIRAC is ready to exploit tools such as glexec in order to optimize workloads. This would allow DIRAC to work in a 'filling' mode by which multiple jobs may be requested for execution by Agents deployed to Grid Worker Nodes in a secure way.

Primary authors : Dr. PATERSON, Stuart (CERN)

Co-authors :

Presenter : Dr. PATERSON, Stuart (CERN)

Track classification : Workflow ; Data Management

Contribution type : oral presentation

Submitted by : Dr. PATERSON, Stuart

Submitted on Wednesday 14 February 2007

Last modified on : Wednesday 14 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 139

Users, Usability and Grids: Introducing Pegasus, a social study of the development and use of GridPP.

Description of the activity :

Pegasus (www.pegasus.lse.ac.uk) is a project looking at the development and use of GridPP by the particle physics community as it prepares for the LHC experiment. The research studies the techniques, practices, and infrastructure involved in the development of GridPP with the aim of providing qualifying guidence to others contemplating the development of complex Grid infrastructure. Our study is qualitative, with the aim of discussing and understanding how Grids come into being and

Grid added value :

Pegasus draws on the Information Systems field, which itself consists of a range of disciplines. We draw on the social sciences to provide qualitative reserach methods which enable us to understand the working practices, collaboration and communication involved in developing GridPP. We also draw on Software Engineering as it provides us with comparisons with existing methods and practices for developing complex IT infrastructure (for example Open-Source, Agile Methods and Globally distributed development practices).

At the User Forum we propose to provide a presentation based on the argument that in order to consider usability we need to reconsider what we mean by Users in the context of the Grid. (**CONTINUED IN NEXT BOX**)

Experience or proposed activity :

The Oxford English Dictionary defines 'user' as "one you uses or employs a thing", or "a person who takes narcotic drugs". Both these definitions suggest a passive, accepting role for the 'user' as a person who employs and is controlled by a thing; only able to employ (or reject) the features it affords. Indeed in the case of narcotics (and many computer systems) the user is a passive, subdued, and manipulated victim.

For Grids however we argue that this relationship between the 'user' and technology is far too simplistic. Through a discussion of a social science study of the development and use of GridPP for users of the LHC particle physics experiments, we aim to unpick and question the role of 'users', and subsequently the aspiration to make Grids 'usable', in terms of the working practices of scientists shaping Grids to reflect their needs. (**CONTINUED IN NEXT BOX**)

Future evolution :

We discuss how the experiments conducted as the LHC impose themselves on the requirements for Grids, and act as an organising vision (Swanson & Ramiller, 1997) for Grid innovation and development.
We discuss the negotiation process inherent in the development of any standard, and we explore the importance of 'power-users' as spanning the gap between developers and so called users.

Swanson, E. B. and Ramiller, N. C. (1997) The organizing vision in information

systems innovation. Org Sci. 8-5

Primary authors : Dr. VENTERS, William (London School of Economics) ; Dr. YINGQIN, Zheng (London School of Economics)

Co-authors :

Presenter : Dr. VENTERS, William (London School of Economics)

Track classification : Workflow ; Data Management

Contribution type : oral presentation

Submitted by : Dr. VENTERS, William

Submitted on Wednesday 14 February 2007

Last modified on : Wednesday 14 February 2007

Comments :

NOTE: We have discussed our somewhat different type of presentation □ and work with Massimo Lamanna and with Owen Appleton. Massimo in □ particular seems keen that we submit even though our work is somewhat □ unauthodox.

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 140

Users on FermiGrid, Open Science Grid -- Experiences of Campus to National Infrastructures

Description of the activity :

Fermilab supports a scientific program which in-cludes experiments and scientists located across the globe. In order to better serve this community, Fermilab supports data analysis of all its users through a grid infrastructure called FermiGrid [1] and through FermiGrid to make leading contributions to the general of the Open Science Grid.
We will report on the user experience of using a campus infrastructure interfaced to a national cyberinfrastructure - the successes and the problems

Grid added value :

OSG is a generally used national cyberinfrastructure in the US. As part of its mission it is working with Campus and other local groups to enable them to fully participate and benefit from the technologies and resources.

Experience or proposed activity :

OSG and EGEE share common communities who submit work to resources across the 2 infrastructures.
Local and global interoperability is a shared goal.

Future evolution :

Will do this later

Primary authors : Mrs. PORDES, Ruth (FERMILAB)

Co-authors : Ms. BERMAN, Eileen (Fermilab)

Presenter : Mrs. PORDES, Ruth (FERMILAB)

Track classification : Related Projects

Contribution type : oral presentation

Submitted by : Mrs. PORDES, Ruth

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Related Projects
Judgment :
Judged by :

Date :
Comments : ""

Abstract ID : 141

Elmer - finite element package for the solution of partial differential equations

Description of the activity :

Elmer is a versatile finite element package for the solution of partial differential equations. It is particularly well suited for the study of generic 2D and 3D problems involving coupling between different physical phenomena. It has already been used to solve several multi-physics problems.

The planned ELMER use under the NA4 for the Earth Sciences Applications includes environmental models, focusing in glasiological and ground water simulations.

Grid added value :

ELMER combines the basic fields of classical physics: computational fluid dynamics, computational solid mechanics, computational electromagnetics, heat transfer, structural mechanics. As a multiphysics environment, the main added value is in coupling these phenomena. There are existing applications in: crystal growth, mems, acoustics, microfluidics. Earth sciences is a rising application area: glasiology, ground water modelling. The scale of activity is to model local, transitory events as case studies. The potential user community are earth scientist, either in academic or governemental institutions. The relevance comes from the growing need of environmental models, and the need to couple several phenomena.

Experience or proposed activity :

Elmer has been applied to some mediumsize problems in glaciology e.g. in nonlinear flow problems.Models such as thermo-mechanical coupling including a numerical correct treatment of pressure-melting point limit, anisotropy as well as prognostic runs have been implemented. A Poisson type of equation solved with Elmer has proved to excellently scale up to 200 processors on an AMD Opteron-Infiniband cluster.

Elmer is dynamically developing system, so there is possibility for updates. The EGEE should enable these updates easily. New physical models are linked as dll:s (dynamically linked libraries). Some problems might arise from running ELMER in heterogenious MPI environments.

Future evolution :

Coupling flow, transport, temperature etc. with ELMER multiphysics and multiscale capabilities enable versatile problem staments and models for coupled systems. The future is in larger spacial and temporal simulations.

Primary authors : Dr. RåBACK, Peter (CSC - Scientific Computing Ltd.)

Co-authors : Dr. FORSSTRöM, Pirjo-leena (CSC - Scientific Computing Ltd.) ; Dr. LYLY, Mikko (CSC - Scientific Computing Ltd.) ; Dr. GRöHN, Matti (CSC - Scientific Computing Ltd.)

Presenter : Dr. RåBACK, Peter (CSC - Scientific Computing Ltd.)

Track classification : Poster session

Contribution type : poster

Submitted by : MATTI, Grohn

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 142

Real Time Monitor

Description of the activity :

The Real Time Monior (RTM) is both an eye catching demonstartion and a historical job analysis tool. By gathering information on active jobs from Resource Brokers, the RTM Java application provides a powerful visual way to show public audiences the current reality of a geographically distributed and actively functioning EGEE Grid. The RTM also summarises jobs as they finish, and publishes daily summary reports. We are also involved in the WLCG Monitoring Working Group

Grid added value :

This is a Grid monitoring tool, and so I do not know how to answer this question. There is no RTM without a Grid...?

Experience or proposed activity :

The RTM has been in existence for several years, having undergone various changes.

The current version relies on querying Resource Brokers directly (as have all previous versions), and relies on the BDII to provide information about the geographical location and resources of EGEE sites. The use of the BDII has meant requesting corrections in GlueSchema information for many sites, and this is ongoing as new sites continually join the EGEE framework.

The RTM has been in high demand for public demonstrations, and EGEE associated projects have also requested to be represented (SEEGRID,EELA,EUMEDGRID,EUCHINAGRID,TRIGRID,GILDA).

Future evolution :

The use of the L&B API for gathering information from the Resource Brokers would be preferable, but so far the API is not suited to the types of query necessary for the RTM.

Involvement in the WLCG Monitoring Working Group means we hope to be able to provide more information back to Site administrators in a meaningful way, to improve the EGEE Grid fabric reliability.

Primary authors : Dr. MOONT, Gidon (GridPP/Imperial)

Co-authors : Dr. COLLING, David (GridPP/Imperial)

Presenter : Dr. MOONT, Gidon (GridPP/Imperial)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : Dr. MOONT, Gidon

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 143

Web3D tools and Grid Services for Compchem VO users

Description of the activity :

The Compchem VO users perform the study of the properties of molecular systems on the EGEE Grid production environment performing massive computational campaigns. We are developing some tools and Web Services based on Web3D and visualization technologies to interact with the simulators running on the Grid. This enables the researchers to interact with the Grid during the simulation and to represent the outcomes of the simulation in an effective and more productive way.

Grid added value :

Our activity is based on EGEE infrastructure and services for making efficient simulations in the shortest time. To this end the tuning made in EGEE to support efficiently user applications requiring various types of MPI APIs is really important. In fact, most of the applications we use for making simulations of molecular systems, Carbon Nano Tubes and biological systems are structured in parallel and require an efficient parallel environment to reduce the wall clock time required for a simulation. The EGEE Grid environment represents for our community the most suitable environment to share competencies and resources in a secure way and the most powerful environment to perform massive computational campaigns to determine the properties of molecular systems in a fastest way. The availability of large computational facility scattered on the most powerful and reliable research centers is of invaluable importance for our community of users.

Experience or proposed activity :

In collaboration with the COST Action D37, called GridChem, we are developing some visualization tools and Web3D based Web Services to strengthen the user interaction with the Grid, where the Simulators are running. In this way the user is enabled to interact with the Simulator as far as the simulation progresses instead of waiting for its end. This facility increases the interest of the researcher in the Grid approach and in some circumstances avoid her/him exhausting waiting times due to wrong operating conditions or closed paths. The Grid environment is becoming a rich environment where sophisticated and complex simulation environments could be implemented combining various simulation environments and open services. Furthermore the graphic, multimedia or virtual reality based interaction with the Grid environment and will reinforce the user's experience of the computational power of the Grid.

Future evolution :

The COST Action GridChem will develop the visualization and Web3D Services in the EGEE Grid environment. Most of the partners of the six Working Groups of the Action are joining Compchem VO and a significant increase of the number of users and of the available simulation codes is expected in the next couple of years. We are confident that the collaboration of Compchem users in the dissemination of computer codes and visualization web Services will be of great importance for the EGEE community.

Primary authors : Dr. GERVASI, Osvaldo (Dept. of Mathematics and Computer Science, University of Perugia)

Co-authors : Dr. ARTECONI, Leonardo (Dept. of Chemistry, University of Perugia) ; Dr. GENTILI, Francesca (University of Perugia)

Presenter : Dr. GERVASI, Osvaldo (Dept. of Mathematics and Computer Science, University of Perugia)

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : Dr. GERVASI, Osvaldo

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 144

Experiences in managing web application software resources for the astronomical community in an EGEE compliant grid infrastructure

Description of the activity :

The INAF VO collects individuals and institute that compose the Italian National Institute for Astrophysics. Projects and research in this area involve the use of a grid computation system in order to exploit the data and computational facilities of such an infrastructure. Astronomers deal with a massive amount of data both textual data and binary data (i.e. images and spectra) coming from celestial bodies observations made with ground or spatial telescopes. Such information is stored worldwide

Grid added value :

By starting from the inherent distributed nature of such information, the potential user community is the INAF scientists which work in different ways on the data in the databases. They have to extract useful information with several queries and retrieve requests, since they search for the “best” data. This operation may include a comparison between different sources (i.e. different observations stored in time and period of the same sky’s portion). Otherwise they elaborate images to refine and extract information following a pre-determined schema (pipeline). Finally data are processed by using different software programs (all running on Linux-compliant platform). Moreover data may not be always public available since each institution that owns them may put restrictions. The implementation of a security infrastructure like the grid ones may take account of these requirements and provide a mechanism to set rights and privileges for different users. The added value of the grid is having a

Experience or proposed activity :

Practical experiences regard the deploying and managing of web applications and web services application in an EGEE-compliant grid infrastructure. The work deal with the development of a set of web services performing the different tasks. By starting from services interacting with database, some web applications have been developed. Essentially based on the Java platforms (java web services), they may be easily deployed on the application container (Tomcat) present on grid nodes (i.e. the Computing Element), but problems were outlined in their use. The main problems encountered from the “application layer” point of view according to grid infrastructure, and in the optics of a decentralized management of resources are discovery of such resources and inherent security issues. By considering such web applications like resources to be shared by the VO, difficult is in having an appropriate and exhaustive resource schema and encoding mechanism for such “software” grid resources. According t

Future evolution :

Future evolution are related to the possibility of integration between different discovery and composition method into the infrastructure. An example is the use of gLite service discovery library that in its scope has the development of

different plug-ins able to interact with different front-end. By maintaining the UDDI registry as a mechanism for registering web services application, it may be used by other grid components. Moreover an enhanced semantic registry may be established for automa

Primary authors : PASTORE, Serena (INAF)

Co-authors :

Presenter : PASTORE, Serena (INAF)

Track classification : Poster session

Contribution type : poster

Submitted by : PASTORE, Serena

Submitted on Wednesday 14 February 2007

Last modified on : Friday 16 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 145

E-Marketplace Model Integrated with Logistics (EMMIL)

Description of the activity :

The EMMIL model represents a new type of B2B electronic marketplace with a new kind of bid structure that allows buyers, sellers and third party logistics service providers (3PL) to take part in the same transaction aiming at buying-selling and delivering tangible goods of homogeneous kinds. The new structure involves an optimization process that is much more complex than the ones currently in use.This project is carried out on the SEEGRID infrastructure using the gLite middleware.

Grid added value :

The mathematical foundation of the EMMIL model has been laid down for the buyer-oriented type of EMMIL, where a powerful buyer issues a request for proposal followed by bidding of sellers and 3PLs. The new three sided auction model involves a new cost model for transportation separating fix and variable costs.

The suggested model leads to a non-linear optimization problem where the objective function is discontinuous so the non-linear solvers do not work. Linearisation is possible at the price of a massive increase in the number of variables. The mixed integer programming problem we get can be so big that it results in a computation time infeasible for a web auction. To shorten the processing time a parallel computational model has been developed and a grid based solution has been tested.

The potential user community consists of companies who want to buy, sell and deliver products via an e-marketplace.

Experience or proposed activity :

The grid version of the EMMIL model is developed under the P-GRADE portal release 2.5 that can support parameter sweep applications. The organization of EMMIL on the Grid is basically a parameter sweep application consisting of three main parts. The first part contains the so-called generator jobs that are used to generate and place the required input parameter sets into a given storage element of the grid.

The 2nd part is a workflow to be executed as many times as many input sets are generated by the generator jobs. We use a realistic business constraint, namely that despite the high number of bidding sellers the buyer wants to establish partnership with only a limited number of partners. The 3rd part finds the best solution and generates output.

The key services are the broker and the built-in workflow manager of the portal. The portal hides the orchestration of the jobs and workflows in the parameter sweep application.

Future evolution :

The first grid solution implements the simplified auction model where only line-haul delivery is allowed and full market is assumed. It is also assumed that a best 3PL can be found from each seller. In the following versions this limitations will be resolved one by one. Meanwhile running the application we found

several problems in the infrastructure: 1. File catalogue entries may become incoherent with accessible files on SEs. 2. Unspecified stochastic refusal of correct jobs from the WMS.

Primary authors : Mr. HERMANN, Gabor (Computer and Automation Institute of the Hungarian Academy of Sciences (MTA SZTAKI)) ; Dr. KACSUKNE-BRUCKNER, Livia (IBS Hungary)

Co-authors : Mr. SIPOS, Gergely (Computer and Automation Institute of the Hungarian Academy of Sciences (MTA SZTAKI)) ; Mr. FARKAS, Zoltan (Computer and Automation Institute of the Hungarian Academy of Sciences (MTA SZTAKI)) ; Mr. KOZLOVSZKY, Miklos (Computer and Automation Institute of the Hungarian Academy of Sciences (MTA SZTAKI))

Presenter : Mr. HERMANN, Gabor (Computer and Automation Institute of the Hungarian Academy of Sciences (MTA SZTAKI))

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : Mr. HERMANN, Gabor

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 146

DIRAC Data Management System

Description of the activity :

The LHCb experiment being built to utilize CERN's flagship Large Hadron Collider will
generate data to be analysed by a community of over 600 physicists worldwide. DIRAC, LHCb's Workload and Data Management System, facilitates the use of underlying EGEE Grid resources to generate, process and analyse this data in the distributed environment. The Data Management System, presented here, provides real time, data driven distribution in accordance with LHCb's Computing Model.

Grid added value :

The data volumes produced by the LHC experiments are unprecedented, rendering individual institutes and even countries, unable to provide the computing and storage resources required to make full use of the produced data. EGEE Grid resources allow the processing of LHCb data possible in a distributed fashion and LHCb's Computing Model is based on this approach. Data Management in this environment requires reliable and high-throughput transfer of data, homogeneous access to storage resources and the cataloguing of data replicas, all of which are provided by EGEE infrastructure and utilized by LHCb.

Experience or proposed activity :

During WLCG's Service Challenge 3 (2005) the LHC experiments tested their bulk data transfer frameworks for the distribution of files from CERN to their associated Tier1 sites. LHCb's Data Management System was interfaced to gLite's File Transfer Service
(FTS) and showed, during optimal performance, it could meet the requirements of the Computing Model for data distribution from Tier0. During LHCb's Data Challenge 2006 the Data Management System components for data driven distribution was used to transfer over half a million files, using FTS, and their registration in grid catalogue (LFC). Future developments of the Data Management System will embrace the new functionality provided by the Storage Resource Manager (SRM) v2.2 implementations allowing LHCb to fully manage space reservation, pinning and storage classes of its data.

Future evolution :

The DIRAC Data Management System has recently been reviewed to allow transfer requests present on site VOBOXes to be processed using FTS to increase the efficiency and redundancy of data distribution. This development as well as the systems mentioned above will be presented here.

Primary authors : SMITH, Andrew Cameron (CERN - University of Edinburgh)

Co-authors :

Presenter : SMITH, Andrew Cameron (CERN - University of Edinburgh)

Track classification : Data Management

Contribution type : oral presentation

Submitted by : SMITH, Andrew Cameron

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 147

An advanced Storage Monitoring tool

Description of the activity :

All the Storage Elements on the grid are lacking in monitoring and accounting features. This implies that could become difficult to take trace, in an easy way, of what files are stored into the system, which user is storing files, how many bytes were written by each user and so on. This problem is much more evident nowadays as a storage system is composed by a large amount of pools, such as new SRM software.

Grid added value :

We have developed a monitoring system that is capable to provide much more information on the behaviour of the Grid Storage Elements. Thanks to this monitoring tool it is possible to see which user is using the system, what he is doing, if there are some errors etc.
With this tool it is also possible to measure how much a dataset is accessed by the user; the storage occupancy for each VO or user supported.
For the user point of view this kind of tool is really useful in order to see how his application performs (how many bytes are red or written by the application).
The tool supports all the most important SRM flavours installed on EGEE infrastructure: such as CASTOR, dCache, DPM.
Also the ClassicSE can be supported as well.

Experience or proposed activity :

This monitoring tool is now installed on the INFN-BARI farm that uses dCache as Storage Element, at INFN-T1 (CNAF) where CASTOR is installed, and was tested on INFN-LNL on DPM SRM.
The result of the last several month of run show that the tool meets the need of both site admin and user that wants to know the much more information from their storage system.
The reliability of the system, especially with dCache, seems really good and it is also simple to build graph and detailed query in order to have all information needed.
The tool is used successfully from some farms in every day work to cross check information coming from other monitoring system in order to better understand the storage system behaviour.
It is also very useful in order to have a quick look on resource usage both per User and VO base.

Future evolution :

In the existing infrastructure it is missing some advanced monitoring system that provides detailed information about transfers and other operation on Storage Elements. With the tool presented all farms can provide some advanced functionality both in monitoring and in accounting their storage system.

Primary authors : Dr. DONVITO, Giacinto (INFN-BARI)

Co-authors : Dr. CUSCELA, Guido (INFN-BARI) ; Dr. DUDHALKAR, Vihang (INFN-BARI + Politecnico di Bari) ; Prof. MAGGI, Giorgio Pietro (INFN-BARI + Politecnico di Bari) ; Dr. PIERRO, Antonio (INFN-BARI)

Presenter : Dr. DONVITO, Giacinto (INFN-BARI)

Track classification : Grid Monitoring and Accounting

Contribution type : oral presentation

Submitted by : Dr. DONVITO, Giacinto

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Grid Monitoring and Accounting

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 148

Job management and control in Earth Science

Description of the activity :

This work is a part of DEGREE project and is focused on requirements of Earth Science applications on job management. Earth Science community covers a wide range of applications from space industry to meteorology, hydrology, seismology, geology and environmental applications. The community consists of research institutes, European organizations and industries, and actively participate on EGEE I and EGEE II projects.

Grid added value :

The work in WP3 in DEGREE project is focused on analyzing requirements of ES applications on job management and control, identifying the missing functionalities of available middleware and tools and preparing test cases for assessing usability of the middleware and tools from the view of ES applications. It will deliver the requirements on job management of the wide application range in Earth Science community to middleware developers, interact with them using concrete use cases from applications, provides test suites for testing functionality of the middleware according to the requirements and motivate the research and developments in Grid middleware

Experience or proposed activity :

The requirements of job management and control in Earth Science communities can be classified into following areas: workflow management, fault tolerance, near-realtime job execution and job monitoring.
For workflow management, although a built-in workflow manager in gLite and several independent workflow managers exist, ES still needs an upper layer defined as an 'intelligent assistant' interface, including knowledge representation and reasoning tools, to facilitate scientific workflows for complex applications. Fault tolerance and near-realtime job execution are very important for several applications in Earth Science. ES application also require more detail and up-to-date information about job running on the Grid.

Future evolution :

The discussion would be useful for detailing requirements of ES applications:
- Fault tolerance: guaranty that the job will be sucessfully executed. If a fault occurs, more information about the reasons and status of job at the moment are required for debugging
- Real-time and near-realtime job executions for some groups of interactive and critical applications

Primary authors : TRAN, Viet (Institute of Informatics, Slovakia)

Co-authors :

Presenter : TRAN, Viet (Institute of Informatics, Slovakia)

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : TRAN, Viet

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 149

The UK National Grid Service and EGEE

Description of the activity :

The mission of the UK NGS is to provide coherent electronic access for UK researchers to all computational and data based resources and facilities required to carry out their research, independent of resource or researcher location. This access will be based upon integrated open standards and the NGS will operate the core services required to exploit local, National and International Partner facilities such as EGEE.

Grid added value :

To date the NGS currently has approximately 500 users and has been successful in attracting users from a wide range of disciplines. Examples of current work include projects in chemistry, engineering, Census data analysis, archaeology, medical imaging, molecular dynamics, integrated biology and biological processing. Users come from over 25 different institutions, mostly, but not exclusively from the UK. By utilising the grid these communities are enabled to co-operate in international collaborations across wide area infrastructure such as EGEE with a reducing amount of effort and the option to access resources that were previously unavailable to their community.

Experience or proposed activity :

Part of the success depends on providing end user support, documentation and training. The NGS Support Centre is a virtual, distributed centre comprising CCLRC, the White Rose Grid at the University of Leeds, the University of Manchester, the University of Oxford and the National e-Science Centre at the University of Edinburgh. The support centre operates a central helpdesk for the NGS that is closely linked to the UK and Ireland Regional Operations Centre. Through this link the NGS helpdesk is integrated into the European wide support structure co-ordinated by EGEE. Through the National e-Science Centre the NGS leverages training and related materials developed for the EGEE project.

Future evolution :

Looking ahead the NGS aims to improve interoperability with EGEE. This has proved difficult due to incompatibilities between the NGS and EGEE infrastructure. Combined with tight EGEE timescales this has led to technical choices and software products which are difficult for the NGS to deploy. However, the EGEE and NGS experience is teaching us which aspects are absolutely vital for large scale production grids. From the user perspective differences between NGS and EGEE are gradually reducing.

Primary authors : Dr. RICHARDS, Andrew (CCLRC)

Co-authors :

Presenter : Dr. RICHARDS, Andrew (CCLRC)

Track classification : Related Projects

Contribution type : oral presentation

Submitted by : Dr. RICHARDS, Andrew

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Related Projects

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 150

Grid Computing for Phylogeny and Molecular Databases

Description of the activity :

At the Institute of Zoology and the Institute of Ecology at the University of Innsbruck molecular studies play a crucial role in scientific work. Common aim of our cooperation is the usage of grid infrastructure for molecular biology applications: We want to 1) apply BLAST software (Altschul et al, Basic Local Alignment Search Tool, J. Mol. Biol. 215:403-410 (1990)) and 2) introduce software for the reconstruction of phylogenetic trees.

Grid added value :

The usage of the EGEE grid infrastructure for biological applications will ease the usage of molecular data. 1) BLAST software allows a rapid sequence comparison and directly approximates local similarity. Regularly large scale comparisons will facilitate finding genes of interest in the vast number of – for instance - EST sequences (expressed sequence tag). 2) Molecular sequence data can be reconstructed by use of statistical analyses like maximum likelihood or pair wise maximum likelihood distances. Depending on the number and the length of amino acid sequences this reconstruction demands enormous computing power.

Experience or proposed activity :

We tested WU-BLAST 2.0 (Washington University BLAST) with larger numbers of sequences on AustrianGrid infrastructure and plan to run WU-BLAST 2.0 with large databases on the EGEE grid.
Concerning reconstruction of phylogenetic trees two applications are now tested and prepared for use on grid infrastructure: TREE-PUZZLE 5.2 and software of the ARB project (Technical University Munich, Germany). Both routines can handle large data sets of molecular sequences.

Future evolution :

As molecular biology is very fast evolving the demand of high throughput computing will increase in the next years. The more molecular sequences are known the more detailed analyses can be taken but the more time and computing power will be demanded. Therefore it is necessary for molecular biology groups to have access to a grid infrastructure to be able to perform statistical analyses and phylogenetic reconstructions.

Primary authors : Dr. NIMETH, Katharina (University Innsbruck)

Co-authors :

Presenter : Dr. NIMETH, Katharina (University Innsbruck)

Track classification : Poster session

Contribution type : poster

Submitted by : Dr. NIMETH, Katharina

Submitted on Wednesday 14 February 2007

Last modified on : Friday 16 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 151

Application of Gaussian package on the EGEE Grid to chemical reaction study

Description of the activity :

A vast number of methods available for computational chemistry caused that community is accustomed to software packages rather than single programs. To attract the community a port of many freely available chemical packages is now available on the Grid. In order to make the Grid even more attractive for the community we present solution which enables use of commercial chemical software on the Grid which makes it a perfect platform for chemistry application.

Grid added value :

Among existing chemical applications commercial ones are the most needed. This is due to availability of many computational methods, fast development of new ones and better user support. However, the main reasons blocking wide use of commercial packages on the Grid are their strict license requirements. These include not only a code protection against unauthorized use but what is equally important check of usage patterns compliance. Confirmation of fulfilling the license requirements is the key step in porting software to the Grid and then using it. In case of Gaussian program our solution was based on the negotiations with Gaussian Inc,. As a result a new virtual organization Gaussian was created. Due to availability of Gaussian for all EGEE users other disciplines such as medicine, biochemistry or physics may benefit from it. This is also a sign for business companies to attempt porting of their software to the Grid.

Experience or proposed activity :

The Grid advantages in chemical applications were demonstrated by us several times. This mainly includes the numerical frequency calculations, molecular dynamics simulations and vibrational averaging study with help of freely available chemical software. In this report we apply commercial software installed on the Grid to chemical reaction study. Chemical reactions are the most interesting and the most demanding computational chemistry challenges. In general such a study requires detailed insight not only in to the structure and properties of products and substrates but also into so called reaction path which describe how changes in substrates structure lead to the product. We shortly characterize a whole procedure and demonstrate how researchers can benefit from the Grid at each computation step.

Future evolution :

Presented solution is currently being applied to other commercial chemical packages and we recommend similar procedure to be employed to other commercial. With a little of negotiation with commercial companies all the license issues may be solved making the desired product available on the Grid. There are however, still unsolved issues which will need to be addressed in the future like `floating' type licenses for which
the Grid solution is highly needed.

Primary authors : Dr. STERZEL, Mariusz (ACC Cyfronet AGH)

Co-authors : Mr. SZEPIENIEC, Tomasz (ACC Cyfronet AGH) ; Dr. PISKORZ, Witold (Dep. of Chemistry, Jagiellonian University)

Presenter : Dr. STERZEL, Mariusz (ACC Cyfronet AGH)

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : MARIUSZ, Sterzel

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 152

Performance Analysis of Atmosphere and Ocean Model on EU-IndiaGrid Infrastructure

Description of the activity :

C-DAC, India and ICTP, Italy are collaborating in EUIndia Grid project -an interoperable Grid infrastructure for the European and Indian Scientific Community. In this project, earth and atmospheric Science has been identified as a pilot grid application. In this presentation we give a brief overview of the EU-India Grid Infrastructure and the performance of an atmospheric model RegCM3 (Regional Climate Model Version 3) and an Ocean Model ROMS (Regional Ocean Modeling System) on it.

Grid added value :

RegCM3 is a regional climate model developed at International Center for Theoretical Physics (ICTP) and is used widely for downscaling the climate scenarios generated by Global climate models. The code is parallelized using MPI and is scalable on a cluster of computers. The ocean model ROMS has been developed by Rutgers University and UCLA and is widely used for simulating the ocean conditions. Under the EU-IndiaGrid initiatives it is planned to develop a coupled ocean atmosphere regional model for generating realistic climate scenarios which will help the International community have a better understanding of the effects of various green house gases on the climate. We discuss the design of computational workflow and the performance of the RegCM3 and ROMS model on the EUIndia grid resources.

Experience or proposed activity :

At the moment the ReGCM Code is running in parallel on the GARUDA Indian GRID and we are currently porting and evaluating and comparing performance of the code on both the infrastructures. The code is a parallel code and it therefore fundamental to have an efficient MPI implementation above the grid easily usable. Another key issue about Earth and Atmospheric simulation is the handling of I/O requirements: beside the size of the output generated there could be also privacy and security issues associated with them.

Future evolution :

The first attempt to use the grid revealed as main issue the Information System. As a matter of fact the IS provided by EGEE infrastructure seems a little bit weak in this respect not indicating with enough details the computational resource MPI-enabled on the GRID. It is for instance not easy to spot out which are the cluster and/or the MPI resource that are based on high speed network (like for instance Myrinet and/or Infiband cards).

Primary authors : Dr. COZZINI, Stefano (CNR-INFM Democritos and ICTP)

Co-authors : DIMEO, Riccardo (ICTP) ; KAGINALKAR, Akshara (CDAC-Pune) ; VENKATA, Ratnam (CDAC-Pune)

Presenter : Dr. COZZINI, Stefano (CNR-INFM Democritos and ICTP)

Track classification : Workflow

Contribution type : --not specified--

Submitted by : Dr. COZZINI, Stefano

Submitted on Wednesday 14 February 2007

Last modified on : Wednesday 14 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 153

Supporting High Performance Computing Applications over gLite

Description of the activity :

Being WLCG Tier-1 site in Asia, ASGC not only fulfill the commitment to WLCG but also provides reliable and stable computing resources supporting variant e-Science applications in this region. In order to attract the more scientists to take the advantage of grid technology, one of our focuses is to engage high performance computing users in grid. This is very important for development and deployment of e-Science applications since HPC is essential to many research disciplinary.

Grid added value :

User interface is the first hurdle in comparison to traditional batch or interactive cluster computing mechanism. Other key factors such as the requirements of heavy I/O, user home directory management, application deployment, real-time job monitoring, compiler and library, and the user support are all vital to the successful migration to Grid. The objective of our works is to improve the Grid computing environment and even the middleware and fabrics to have better support for generic MPI HPC applications.

Experience or proposed activity :

We build and evaluate HPC environment from the building blocks, especially in parallel computing. For this purpose, high bandwidth and low latency of data communication is very important. We install high performance network device – Infiniband (IB). There is another important issue is the high performance IO, especially global file system. We survey nfsRDMA, GPFS and GPFS over IB, and evaluate a better solution for HPC users in Taiwan.
Most of the user should prepare job description language (JDL) for job submission, transfer large inputs to resource broker (RB) by gridftp and retrieve results by edg commands. In order to reduce end-user's effort, we set a convenient wrapper to help HPC users. Therefore, the users can choice the IB or GbE communication channel to execute parallel job in one argument. In our setting, user submit jobs from UI which can map the same user id in CE, the UI and CE share a global file system. This will help system become more suitable for HPC users.

Future evolution :

In the future, the more HPC patterns of different applications and requirements will be elaborated. Grid portal would be devised to be the major user interface in our plan. Workflow management for process optimization, such as the best inter-connection selection, I/O management and compiler coordination etc., would be developed and integrated with gLite.

Primary authors : Mr. CHEN, Hsin-yen (ASGC)

Co-authors : Mr. WANG, Chi-wei (ASGC) ; Dr. SHIEH, Danny (ASGC) ; Mr. YEN, Eric (ASGC) ; Dr. LIN, Simon (ASGC)

Presenter : Mr. WANG, Chi-wei (ASGC)

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : Mr. CHEN, Hsin-yen

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 154

Large scale pharmacological experiment on the Grid - protein folding in EUChinaGrid.

Description of the activity :

The contemporary pharmacology facing the problem of the individual therapy design, requires the understanding of the activity of molecules (such as proteins). in the case of
proteins the activity is implied by the 3-dimensional structure, therefore the simulation of the folding process is crucial in the research. In the scope of EUChinaGRID project the large number of "never born proteins" of 70 amino acids in polypeptide chain is planned to be predicted using 2 methods.

Grid added value :

As the protein folding simulation is a compute intensive task and the number of interesting molecules may be in the order of millions, therefore a usage of the Grid infrastructure comes as a natural solution for providing computing power. The storage requirements of the application are not negligible as well, and may reach an order of terabytes. In addition to raw computing power offered by the Grid, the collaboration aspect becomes important as well. Since there are many research groups developing various computational models and producing their simulation data, making them available on the Grid enables collaborative work on model verification and testing. In our project results obtained using two methods (one using a standard Rostetta tool and other using a newly developed model by JUMC team from Krakow) and leading to similar structure of protein are going to be synthesized and examined experimentally by another team.

Experience or proposed activity :

Our application requires simulating of miilions of proteins using our software, therefore the first challenge was to grid-enable the application. This task was accomplished by preparing a self-contained package of the software which could be deployed and run as a grid job with minimal requirements. For storing the results the grid storage is used and all resulting files are registered using grid catalogues (namely LFC). Another challenge was the management of such a large experiment and for that purpose the database system integrated with a portal and with automating job submission system has been developed. The portal is based on Gridsphere framework and uses GridwiseTech LCG_API to interface with the EGEE middleware. The portal may be used for the user-friendly definition of experiments and is integrated with the scientific visualization and analysis tools.

Future evolution :

When developing scientific experiments, which are applications using the Grid from the client-side, there is a problem of lacking or insufficient standard client API to the underlying middleware. It should be possible to have a lightweight and easy to install client library for EGEE in all important programming languages. Currently available solutions are discouraging the developers of scientific applications from integrating with and from using the EGEE software in their experiments.

Primary authors : Mr. WISNIOWSKI, Zdzislaw (Department of Bioinformatics and Telemedicine – Collegium Medium – Jagiellonian University, Krakow, Poland) ; Mr. MALAWSKI, Maciej (AGH University of Science and Technology, Krakow, Poland)

Co-authors : Mr. SZEPIENIEC, Tomasz (ACC CYFRONET-AGH, Krakow, Poland) ; Mr. KOCHANCZYK, Marek (Department of Bioinformatics and Telemedicine – Collegium Medium – and Faculty of Physics, Astronomy and Applied Computer Science - Jagiellonian University,

Krakow, Poland) ; Mr. FLIS, Lukasz (Faculty of Physics, Astronomy and Applied Computer Science) ; Ms. PRYMULA, Katarzyna (Department of Bioinformatics and Telemedicine – Collegium Medium – and Faculty of Chemistry - Jagiellonian University, Krakow, Poland) ; Dr. JURKOWSKI, Wiktor (Department of Bioinformatics and Telemedicine – Collegium Medium – Jagiellonian University, Krakow, Poland) ; Dr. BRONIATOWSKA, Elzbieta (Department of Bioinformatics and Telemedicine – Collegium Medium – Jagiellonian University, Krakow, Poland) ; Ms. PIWOWAR, Monika (Department of Bioinformatics and Telemedicine – Collegium Medium – Jagiellonian University, Krakow, Poland) ; Prof. ROTERMAN-KONIECZNA, Irena (Department of Bioinformatics and Telemedicine – Collegium Medium – Jagiellonian University, Krakow, Poland)

Presenter : Mr. WISNIOWSKI, Zdzislaw (Department of Bioinformatics and Telemedicine – Collegium Medium – Jagiellonian University, Krakow, Poland) ; Mr. FLIS, Lukasz (Faculty of Physics, Astronomy and Applied Computer Science)

Track classification : Poster session

Contribution type : poster

Submitted by : Mr. MALAWSKI, Maciej

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 155

Grid application to model spacecraft reentry: calculation of N+N2 rate coefficients

Description of the activity :

We are carrying out the present work in the Compchem VO to perform the realistic ab initio simulations of a system playing an important role in the optimization of the spacecraft reentry technology. The accurate evaluation of the N+N2 system requires a massive computational campaign of hundred of thousand long running sequential jobs.
The EGEE production Grid environment is giving us the necessary computational resources via some scripts to manage the job submission and monitoring.

Grid added value :

The added value of the Grid for our activity is twofold. On one side the Grid is needed to sustain the distributed execution of the massive number of jobs on target nodes equipped with high performance computing capabilities and large RAM memory.
On the other side the Grid allows to build applications out of the synergy between various competences and know how.
This means that not only most often applications need to be run with the contribution of several laboratories of the Grid virtual community but also that the workflow of the application must be designed so as to synchronise the contributions and evaluate the real support given. This last aspect is very important: As a matter of fact not only the community has organized itself in the Compchem Virtual Organization but it has also proposed an algorithm that is being tested to establish a credit system for rewarding individual contributions.

Experience or proposed activity :

The accurate evaluation of state-to-state rate coefficients of atom diatom collisions at high temperature requires full quantum calculations for several vibrational (v) and rotational (j) states and a converged sum on the fixed total angular momentum value J reactive probabilities.
The method adopted to compute the quantum reactive probabilities of N + N2 is based on the time dependent formalism.
To start the calculations a coordinate grid large enough to include both reactant and product asymptotes is constructed. On this grid the potential and the complex system wavepacket are calculated for a given value of the total angular momentum quantum number J and its projection K.
The system wavepacket is assembled and is assigned a relative kinetic energy amount towards the strong interaction region. The wavepacket is then propagated in time an the overlap of the wavepacket with each of the open product channels is calculated and stored in order to calculate the outcomes at the end of the r

Future evolution :

The stategy adopted to calculate the state-to-state rate coefficients is to build several PHP programs in order to manage the data partitions ot the input parameters, the submission of the related series of jobs and their monitoring in a user-friendly, very efficient and productive way.

During the computational campaign we met some troubles with some sites unable to schedule Compchem jobs, and, only sporadically, some strange and unpredictable behavior of the EGEE servers.

Primary authors : Prof. LAGANA', Antonio (Dept. of Chemistry, University of Perugia, Perugia (Italy))

Co-authors : Prof. GARCIA PARA, Ernesto (Dept. Physical Chemistry, University of the Basque Country, Vitoria (Spain)) ; Dr. ARTECONI, Leonardo (Dept. of Chemistry, University of Perugia, Perugia (Italy)) ; Dr. GERVASI, Osvaldo (Dept. of Mathematics and Computer Science, University of Perugia, Perugia (Italy))

Presenter : Prof. LAGANA', Antonio (Dept. of Chemistry, University of Perugia, Perugia (Italy))

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : Dr. GERVASI, Osvaldo

Submitted on Wednesday 14 February 2007

Last modified on : Monday 19 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 156

Data and the Grid: the role of a Database in a Distributed Environment

Description of the activity :

Overview

The creation, analysis and manipulation of data and meta-data lie at the heart of Grid Computing. This grid data takes many forms and is widely distributed - characteristics that are not normally associated with database management systems (DBMS). However databases are today playing a major part in the development of grid computing.

Grid added value :

This presentation from Oracle Corporation, will explore the use of DBMS within grid computing. A range of data distribution models will be covered together with data structure and access methods. Examples of these models from both the Scientific and Enterprise Grid worlds will be presented.

Topics that will be covered are:

§ How to architect data
§ How to distribute data
§ Current experiences and common approaches

Attendees will be invited to participate and contribute to the open discussion during and following the speakers' contributions with the aim of:

§ Exchanging ideas and knowledge
§ Outline possible solutions or steps forward
§ Discussing open or specific issues in Data Distribution environments

Experience or proposed activity :

Target Audience

§ DBA, Architects, Service Managers, and Developers

Future evolution :

Speaker

Graeme Kerr, Technical Architect, Oracle in R&D Programme.

Primary authors : Mr. KERR, Graeme (Oracle Corporation)

Co-authors :

Presenter : Mr. KERR, Graeme (Oracle Corporation)

Track classification : Data Management

Contribution type : oral presentation

Submitted by : Mr. KERR, Graeme

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

I found it difficult to fit my Presentation summary to the above questions. I hope the above is clear enough.

Status : SUBMITTED

Track judgments :

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 157

The S.Co.P.E. Project

Description of the activity :

The University of Naples "Federico II" is one of the great Universities of Italy and is employed in the main fields of the fundamental research, in strong collaboration with the INFN Naples Unit. More than 1000 people that include Professors, researchers, phd students and technicians work in the computational sciences for many different applications that we divided in three different areas: - Microcosmic and Macrocosmic Sciences, Life Sciences, Matter and Environmental Sciences.

Grid added value :

We plan to create a great example of Metropolitan Area Grid between all the University department's distributed in 4 big site in the City of Naples, in order to have a singole interface to the heterogeneous resources. The first step are the implementation of a new computational centre in the site of M.S.Angelo and the integration with the already available and distributed resources. The goal is obtain an improvement of the scientific results by using in a more efficient way the existing hardware and promoting a grid dissemination. At first, the system will be used from all the researchers of the University "Federico II", INFN and of other local research institutes and from local industrial partners, respectively for fundamental research applications in the 3 fields of Microcosmic and Macrocosmic Sciences, Life Sciences, Matter and Environmental Sciences and for business applications.

Experience or proposed activity :

We plan to implement in Metropolitan scale all the main Grid services available in gLite. In this context the Workload Management Services will be very important to facilitate the access to the Grid and the Grid monitoring to manage the system.
Very important will be the choice of the best interface to simplify the access to the Grid at the new user (shell and graphical interface available).
Finally the S.Co.P.E. Project promote the implementation in local scale of new services or new implementation of already existing services in a limited and simplex testbed environment in order to contribute at the Grid reserach.

Future evolution :

I think that will be useful a metadata system based on a relational model. Currently we can add only a label for each file register in a catalog.
However from the experience in data analysis in the within of the VIRGO Experiment, appare the necessity of having more fields appears in order to execute file searches based on the metadata.

Primary authors : Dr. PARDI, Silvio (University of Naples "Federico II" - C.S.I.)

Co-authors :

Presenter : Dr. PARDI, Silvio (University of Naples "Federico II" - C.S.I.)

Track classification : Poster session

Contribution type : poster

Submitted by : Dr. PARDI, Silvio

Submitted on Wednesday 14 February 2007

Last modified on : Friday 16 February 2007

Comments :

More authors must be addedd.. I think that towmorrow I can comunicate all the
authors. Thank you
Silvio

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 158

The Arts and Humanities e-Science Initiative

Description of the activity :

This poster will explain the activities and context of the Arts and Humanities e-Science Initiative in the UK, funded by the AHRC, EPSRC and JISC. It will firstly present to the wider EGEE community last year's activities of workshops discussing opportunities of e-Science in A+H, of the A+H e-Science Scoping Study, and of small-scale demonstrators. Secondly, the award holding projects for the major funding scheme for A+H e-Science will be presented, as they start their work in summer 2007.

Grid added value :

The work of the so-called early adopters of A+H scholars suggests that the benefits of integrating this particular group of researchers into the wider e-Science community are mutual. A+H researchers face a mushrooming of the size of digital information for their research, as a result of the general availability of digital information and specialised digitisation programmes that preserve artistic and cultural assets in a digital format. This makes it necessary to find tools and methodologies that allow the exploitation of the digital information. The e-Science community itself can benefit from the unique challenges of A+H data. Relative fuzziness of the data and the interpretive character of research in the A&H will enable a further advance on other e-Science projects.

Experience or proposed activity :

Current discussions within the A+H initiative tackle questions about the shape the new e-infrastructure for humanities research could take and the technologies that could help address the specific issues of arts and humanities data. On the data consolidation level, technologies like the Storage Resource Broker (SRB) for easy distributed file access or OGSA-DAI for a middleware to assist on distributed database access, are considered to be highly useful. SRB, Shibboleth as well as Semantic Web technologies are tested in early adaptor projects for their use in providing access and organizing arts and humanities data. As humanities data is so dispersed, metadata technologies could help cross-institutional boundaries. Moreover, it has been shown that the development of ontologies can assist understand better research questions in the domain. Probabilistic indexing and record linking services have also been successfully used in A+H projects.

Future evolution :

The base e-Infrastructures still lack user-friendliness, and efforts have to be made to make them accessible to less technical users, e.g. by portals, rich clients, or Shibboleth. More user-centric approaches will leverage the results of these efforts. Another reoccurring topic in the early arts and humanities e-Science projects is how to foster and enhance interdisciplinarity. To make the Grid work in A+H, it is essential to reorganize the interoperability and reusability of their production.

Primary authors : Dr. BLANKE, Tobias (King's College London)

Co-authors : Dr. DUNN, Stuart (AHeSSC, King's College London)

Presenter : Dr. BLANKE, Tobias (King's College London) ; Dr. DUNN, Stuart (AHeSSC, King's College London)

Track classification : Poster session

Contribution type : poster

Submitted by : BLANKE, Tobias

Submitted on Wednesday 14 February 2007

Last modified on : Friday 16 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 159

A Molecular Dynamics study of propane on the EGEE platform

Description of the activity :

Molecular based simulations of complex systems can be performed thanks to the increasing availability of computer power on distributed platforms.
A molecular dynamics study of the Propane bulk system in liquid phase is carried out by pergorming parallel calculations. The calculations were aimed at estimating in an a priori fashion some thermodynamics properties of the system.
To this end the DL POLY[1] software package was used.
[1] W. Smith, T.R. Forester, J. Mol. Graph. 14 (3), (1996), 136

Grid added value :

Computational tasks were distributed using the EGEE platform.
Preliminary indications on the performance of the used platform were obtained running a test case on six di□erent EGEE clusters. All the calculations were ran for each cluster sequentially on one node and in parallel on 2 and 4 nodes in order to evaluate the elapsed time of each simulation and
the related speed-up.
Elapsed times and related speed-ups will be illustrated at the meeting.

Experience or proposed activity :

The parallel performance of some clusters is very close to the ideal value and deviations from it occurring in the other clusters are mainly due to the time sharing regime adopted by them.
From these preliminary indications is clear that the parallel performance of each EGEE cluster strictly depend on the adopted regime.
A statistical analysis carried out on 50 parallel jobs shows that more than 70% of the jobs ran properly and only 26% was aborted. Abortion is mainly due to comunications errors between the nodes of the same cluster (62%) but also to internal errors of DL POLY occurred during
the running (23%) and to the RB scheduler (15%).
The calculated value of the density of the system is similar to the data given in literature (582 Kg m(E-3) at T=230 K at P=1.013 bar[2]) when going from 200 K to 230
K. At the same time we computed also high values of pressure and correlated statistical errors.
[2] Air liquid group website: http://www.airliquide.com

Future evolution :

The problem of high values of pressure and correlated statistical errors computed can be resolved increasing the simulation time in order to obtain a better average value of the pressure and/or bringing small modification of the force field of the system in
order to normalize the e□ect of the pressure.
We are planning a more massive use of the EGEE infrastructure in order to build the related phase diagram of the propane bulk system.

Primary authors : Dr. COSTANTINI, Alessandro (University of Perugia)

Co-authors : Prof. LAGANà, Antonio (University of Perugia)

Presenter : Dr. COSTANTINI, Alessandro (University of Perugia)

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : Dr. COSTANTINI, Alessandro

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 160

The use of GRID tools for automatic file transfer of ARGO-Yangbajing experiment data.

Description of the activity :

The aim of ARGO-YBJ experiment is the study of cosmic rays registered at the Yangbajing Laboratory altitude, at 4300m a.s.l. in Tibet. ARGO-YBJ is an Italian - Chinese collaboration that involves researchers of INFN in Lecce, Napoli, Roma, Torino, Pavia, of IASF in Palermo, of IHEP in Beijing and of the Chinese Universities in ShanDong, Chengdu, Lhasa, KunMing, ZengZhou and HongKong.

Grid added value :

Some aspects of the gamma ray astronomy, like Gamma Ray Burst phenomena, require the access and the processing of the data in the shortest possible time. The use of simple procedures, based on FTS services, to transfer data from the remote experimental site to the SEs at the processing sites and the automatic introduction of these data into the LFC catalogues allows us a fast processing and access to the interesting data. Moreover we implemented a procedure, also based on GRID tools, to mirror the data available at each moment at Chinese and Italian processing sites, to allow the both communities to cooperate eficiently in analysing the data and at the same time to realize a backup of the data. We are going to process around 200 TBytes of raw experimental data, producing 30 TBytes of reconstructed events.

Experience or proposed activity :

We designed and realized a set of scripts, that use the FTS PERL API and a dedicated DataBase, to perform the experiment data file transfer from a SE, gathering the files from ARGO-YBJ data acquisition system, to one of the prosessing site (IHEP in Beijing or CNAF in Bologna). The file transfer uses the FTS channels, open previously between the SE at Yangbajing and the SEs at IHEP and CNAF. We tested the whole system on a local testbed at INFN Roma 3, that was extended afterwards to the IHEP and CNAF sites. For monitoring purposes a graphical interface was also developed.
An efficient data mover system needs an stable and efficient FTS and SRM services. In our particular case an efficient link and routing between italian and chinese sites are mandatory.

Future evolution :

The future evolution of our procedures foresees the confrontation of SRM file transfer comands in respect of FTS comands. In the past we experienced some problems in using the FTS JAVA API due mainly to the insufficient "how to use" documentation, and this explain why we decided to use FTS PERL API instead. The achievement of good performance in our automatic file transfer system strongly depends from an efficient and reliable network link to the CSTNET chinese research network.

Primary authors : Dr. CELIO, Paola (Istituto Nazionale Fisica Nucleare - Sezione Roma III)

Co-authors : ANDRONICO, Giuseppe (INFN Catania) ; ARDIZZONE, Valeria (INFN Catania) ; BARBERA, Roberto (Physics and Astronomy Department Catania) ; BUDANO, Antonio (INFN Roma 3) ; CELLINI, Simone (INFN Roma 3) ; GALEAZZI, Fulvio (INFN Roma 3) ; GARGANA, Riccardo (INFN Roma 3) ; GIORGIO, Emidio (INFN Catania) ; LAROCCA, Giuseppe (INFN Catania) ; RUGGIERI, Federico (INFN Roma 3) ; STANESCU, Cristian (INFN Roma 3)

Presenter : Dr. CELIO, Paola (Istituto Nazionale Fisica Nucleare - Sezione Roma III)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : Dr. STANESCU, Cristian

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 161

Atlas event production on the EGEE infrastructure

Description of the activity :

ATLAS is one of the four LHC (Large Hadron Collider) experiments at CERN, is devoted to study proton-proton and ion-ion collisions at 14TeV. ATLAS collaboration is composed of about 2000 scientists spread around the world. The activity of the experiment requirements for next year is of about 300TB of storage and a CPU power of about 13 Mski2sk, and is relying on GRID philosophy and EGEE infrastructure. Simulated events are distributed over EGEE by the Atlas production system.

Grid added value :

Data has to be processed and must be accessible by a huge number of scientists for analysis. The throughput of data for Atlas experiment is expected to be of 320 MB/s with an integrated amount of data per year of ~10Pb.
The processing and storage need a distributed share of resources, spread worldwide and interconnected with GRID technologies as the requirements are so demanding for the LHC. In that sense event production is the way to produce, process and store data for analysis before the experiment startup, and is performed in a distributed way.
Tasks are defined by physics coordinators and then are assigned to Computing Elements spread worldwide.
Some of the jobs that build up the tasks need input data as well to produce new output, this means the jobs may need input from external sites and store remotely. For that reason sites are connected by File Transfer Service (FTS) channels that links the Storage Elements (SE) interface for each site.

Experience or proposed activity :

ATLAS is using the services provided by the EGEE middleware. Event simulation jobs are sent to the LCG (LHC Computing Grid) GRID by glite-WMS (Workload Management System) and Condor-G and using the dispatching tools of the CE's.
Event simulation jobs perform the Data Management as well, request the inputs and stores the outputs on the desired SE's, file location and information is managed with distributed LCG File Catalogues (LFC).
On the other hand, asymmetric file movement is performed by the ATLAS specific software on Distributed Data Management (DDM), which takes care of the file movement on top of the FTS services.
Services which are causing problems are basically the Storage Elements, the system is strongly dependent on the inputs for the event simulation jobs and failing to retrieve it produces job failures, while failures in storing the outputs due to SE's instabilities leads to the loss of the CPU consumed by the job and the consequent failure.

Future evolution :

From the event simulation is expected that glite-WMS handles the jobs in a more reliable way, and concerning the
CE's perhaps introduce different implementations that would have no scalability limitations.
Certainly we hope new implementation of the SRM (Storage Resource Manager) interface which would solve stability
problems mainly in the stage-in and stage-out of the files needed and produced by the jobs respectively.

Primary authors : Dr. ESPINAL, Xavier (PIC/IFAE)

Co-authors : Dr. CAMPANA, Simone (CERN) ; Dr. PERINI, Laura (CERN) ; Dr. ROD, Walker (TRIUMF)

Presenter : Dr. ESPINAL, Xavier (PIC/IFAE)

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : Dr. ESPINAL, Xavier

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 162

The gLite Workload Management System

Description of the activity :

The EGEE-II Joint Research Activity 1 (JRA1) is responsible for the re-engineering of the gLite middleware. One of the fundamental functionalities provided is a system for the submission of job and the management of the workload. The institutes involved in this activity are INFN, Datamat and CESNET. The challenge and originality of the activity lies in selecting, potentially re-engineering and integrating a set of reliable production-quality services

Grid added value :

The gLite Workload Management System (WMS) is a collection of components providing a service responsible for the distribution and management of tasks across resources available on a Grid. The main purpose is to accept a request of execution of a job from a client, find appropriate resources to satisfy it and follow it until completion.
Several kinds of jobs (or aggregate of jobs) are allowed:
Normal - simple batch job
MPICH - a parallel application to be run on the nodes of a cluster using the MPICH implementation
interactive - job whose standard streams are forwarded to the submitting client in order to let him interact
Directed Acyclic Graphs (DAGs) - a set of jobs where the input, output or execution of one of more jobs may depend on one or more other jobs
parametric - allows submission of a large number of jobs by specifying a parameterized description
collection - a possibly big number of independent jobs that can be specified within a single JDL description

Experience or proposed activity :

Intense testing and constant bug fixing activities have been performed over the last months in order to improve job submission rate and service stability. Several new functionalities were tested and adopted so far (20% faster submission achieved)
WMProxy user front-end implements an interoperable WEB service interface, allowing users either to implement its own language-independent client, or to adopt the multi language (C++, Java & Python) provided APIs, or to use the C++ based WMS command-line User Interface.
Integration of the Service Discovery functionality within the UI provides the user with a new set of possible service endpoints by performing queries to external databases without needing the manual reconfiguration
Automatic job’s sandbox files archiving and compression, along with the possibility to make jobs share the same sandbox, dramatically reduced network traffic
Job Perusal functionality allows users to monitor actuoal output files producing during job lifecycle

Future evolution :

The future developments of the gLite WMS will be focused on improving its portability and usability. Dependencies on external software will be reduced on the client part (User Interface) to improve portability. Improved logging and error reporting as well as an improved monitoring system will make easier to maintain and use the service. The reduction of the resources needed by the WMS processes will increase the stability and throughput.

Primary authors : Mr. MARASCHINI, Alessandro (DATAMAT) ; Mr. GIACOMINI, Francesco (IFNN) ;
Mr. PACINI, Fabrizio (DATAMAT)

Co-authors : Mr. SGARAVATTO, Massimo (INFN) ; Mr. MONFORTE, Salvo (INFN) ; Mr. PAPPALARDO, Marco (INFN)

Presenter : Mr. MARASCHINI, Alessandro (DATAMAT) ; Mr. GIACOMINI, Francesco (IFNN)

Track classification : Workflow

Contribution type : --not specified--

Submitted by : MARASCHINI, Alessandro

Submitted on Wednesday 14 February 2007

Last modified on : Thursday 15 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 163

The NGS JSDL Repository and Grid Job Submission Portlet

Description of the activity :

The portal serves the scientific research communities of the UK NGS (National Grid Service). The main activities facilitated by the portal include: a) Job submission/monitoring for the execution of jobs described in JSDL; b) JSDL GUI editing for constructing, viewing and validating JSDL. Documents may also be uploaded into the portlet repository as a means of initialising new jobs and sharing JSDL and expertise between users. c) Saving and browsing JSDL in the portlet repository.

Grid added value :

The portal provides a means of sharing and accessing JSDL application descriptions between Grid users in a platform agnostic and middleware independent manner (XML Job Submission Description Language). Consequently, users will benefit from the associated expertise and resources described in the JSDL (specialised applications can be described and published by domain-experts and resource administrators). Application descriptions have being configured by NGS Operations for the purpose of demonstrating Grid job submission concepts, and to build a repository of pre-configured applications on the NGS. The portlet is being extended to support additional middleware to facilitate generic Grid use and widespread Grid adoption. The portlet is available for re-use and customisation in new Grid projects requiring a simple-to-use web Portal framework. The portlet is JSR-168 compliant for re-use in portal containers.

Experience or proposed activity :

The work is aimed to encourage a wider audience to use Grid resources by simplifying job submission, file staging and data management. The scale of adoption by the NGS user community and other Grid communities will be an invaluable measure of success. User communities include new/ inexperienced users, domain-specific Grid user communities and experimental facility users.

Future evolution :

The aim is produce a portlet that can be extended and not tied to any particular set of middleware through the use of generic JSDL. The main issues experienced to date therefore, are associated with the need to add (specific) support for different middleware providers. The other main issue is related to accommodating different authentication mechanisms and single-sign procedures encountered across Grids (especially Grids of different middleware).

Primary authors : Dr. MEREDITH, David (CCLRC NGS)

Co-authors : Dr. RICHARDS, Andrew (CCLRC NGS) ; Dr. MANIOPOULOU, Mina (CCLRC NGS) ; Mr. CHURCHILL, Jonathan (CCLRC NGS)

Presenter : Dr. MEREDITH, David (CCLRC NGS)

Track classification : Related Projects

Contribution type : oral presentation

Submitted by : Dr. MEREDITH, David

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

If accepted, an oral presentation with demo would be preferable. If not, other□ formats are also great. □

Status : SUBMITTED

Track judgments :

Track : Related Projects

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 164

Astrophysical applications in an EGEE compliant grid: the TriGrid Project

Description of the activity :

The INAF community starts to have a great interest in the usage of grid facilities. We presents some scientific projects that are using the EGEE compliant TriGrid infrastructure. Model of the current Stellar Perturbations on the Oort Cloud. Extremly Faint Galaxies in Astronomical Images extraction from images. COROT Mission: to search extrasolar planets. Cosmological Simulations: simulations of the interaction of a Black Hole jet and evolution of clusters of galaxies.

Grid added value :

Model of the current Stellar Perturbations: The basic calculations that were made for this project do fit very well the grid distributed calculus: many different trajectories can be calculated at once.
Faint Galaxies in Astronomical Images: The added value of the grid is very large in that we have a problem which can be solved on a typical workstation for a single image, but in the long run we need to do that thousands of times. Therefore our application is quite suitable to distributed computing as for the grid case.
COROT Mission: Thousands of simulated light curves, will be used to test the performance of different modeling techniques using a Monte Carlo approach.
Cosmological Simulations: As AGN are the most brilliant radiation sources in the early Universe, their study reveals essential features of the formation and evolution of our Universe. Moreover the added value of the grid is that it allow us to find the computational resources for medium-size simulations.

Experience or proposed activity :

The availability of a platform with hundreds of CPUs make the usage of the infrastructure (for applications that not require a particular performance) attractive: the COROT experiment and the model for Solar System are examples of this. Another very useful aspect is the usage of IDL, one of the most popular software for image manipulation. IDL license was installed on a system of the TriGrid and it makes the system available for many astronomers that use IDL to run MACROS (programs) for data analysis. The test we performed were successful and all the TriGrid nodes will be provided with an IDL RUNTIME license. The situation of codes that require HPC resources is more complicated. Cosmological simulations are often requiring resources that are not immediately available with the middleware and a large disk space for large datasets. Moreover the usage of an extended platform to run highly demanding codes in terms of number of CPUs and storage make the usage of the GRID attractive.

Future evolution :

Cosmological codes require HPC resources in the GRID. Mainly this requires: the usage of compilers specially optimized for the CPU and the platform where the code will run. Simulations and analysis codes can require to download several tens of files each having huge dimension: for example a typical N-Body simulation could require up to thirty files each having 15 GBytes.
The last aspects concern big runs that can require from 52 to 128 CPUs for several days (60 or more days).

Primary authors : Dr. BECCIANI, Ugo (INAF-OA Catania)

Co-authors : Dr. ANTONUCCIO, Vincenzo (INAF-OA Catania) ; Dr. BORGANI, Stefano (University of Trieste) ; Dr. LANZA, Antonino Francesco (INAF-OA Catania) ; Dr. SCARAMELLA, Roberto (INAF-OA Roma)

Presenter : Dr. BECCIANI, Ugo (INAF-OA Catania)

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : Dr. VUERLI, Claudio

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 165

Monte Carlo production moves to grid @ COMPASS

Description of the activity :

COMPASS is a fixed target experiment running since 2002 at CERN SPS. Its main goal is to shed light on the nucleon spin structure, and to perform hadron spectroscopy. With data acquisition rates up to 70MB/s, during 100 days a year, the experiment collects 10^10 events a year (300 TB/year) and needs a massive Monte Carlo production, carried on by the 30 collaborating institutes. Among them is Trieste.

Grid added value :

The current local data simulation at each institute (for example in Trieste a 50 kSI2k farm has been set up) benefits from a move to grid technologies in a variety of ways. The new production infrastructure, hinged on a file catalogue (implemented using gLite LFC and AMGA), allows first of all to ease the handling of the huge amount of simulated datasets. Data aggregation based on metadata, seamless dataset relocation and transparent data access are the most relevant and needed.
Grid technologies help the world-wide distributed institutes of the collaboration to share the MC data they hold, whether the data are replicated or processed where they reside according to advisability or actual feasibility. Simulated data production itself can benefit from a grid integration of the available computing resources in terms of usage exploitation.

Experience or proposed activity :

The MC data production model assumes the local farm as an essential element of the MC data service, with the data produced and analysed at home. Outbound (grid) production is performed only in case the local resources are insufficient. We have started to test DIANE as a scheduler to provide an effective means to produce large simulation samples, integrating local dedicated resources with the grid in a seamlessly way.
In the framework we have thought of, data production and data aggregation are performed through a system which holds and manages the information about the current production status of each dataset that a user requests to the MC data service. Each dataset, exposed to the user as data characterised by a set of MC parameters relevant to a specific study, actually translates into a certain set of production tasks. Data aggregation can return the dataset that comply to a specific query or can be set to trigger a new data simulation.

Future evolution :

We plan to expose the system later in spring 2007 to the first users. The framework will be set up with all basic functionalities (production jobs managing, dataset cataloguing and data aggregation), which will grow in number and complexity of capabilities up to a scientist-friendly interface that will hide the technicalities of grid computing.

Primary authors : DUIC, Venicio (Università degli Studi di Trieste & Istituto Nazionale di Fisica Nucleare (INFN - Trieste)) ; MARTIN, Anna (Università degli Studi di Trieste & Istituto Nazionale di Fisica Nucleare (INFN - Trieste)) ; SBRIZZAI, Giulio (Istituto Nazionale di Fisica Nucleare (INFN - Trieste))

Co-authors :

Presenter : DUIC, Venicio (Università degli Studi di Trieste & Istituto Nazionale di Fisica Nucleare (INFN - Trieste))

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : DUIC, Venicio

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 166

A VO-oriented AuthN/AuthZ approach

Description of the activity :

The grid infrastructure is used by Virtual Organizations (VOs). VOs group together real organizations or businesses that are willing to share resources to gain computational power or provide services that would be unobtainable otherwise. Currently, most VOs belong to the field of High energy physics, however applications in other fields are emerging (e.g., bionformatics, astronomy, etc...)

Grid added value :

We are creating an integrated administration service for authentication and authorization targeted at VO and site administrators. An integrated solution allows for easier management of authorization decisions and policies on grid resources at the VO and site level. The solution we are proposing is implemented on top of a few well-defined interfaces that group the administration work and will be discussed in detail in the talk.

Experience or proposed activity :

Our experience with the management of the INFN T1 at CNAF shows that an integrated solution for policy management and attribute authority administration is needed to simplify the daily setup and operation of the grid infrastructure. Currently, no available Grid middleware provides a similar service to VO and site administrators. This results in administrators being required to go manually from site to site in order to adjust the local setups to new VO requirements. This is especially cumbersome when there is no conceptual need to change a configuration at the site level, e.g. when a VO just wants to change the way its groups map into usage classes.

Future evolution :

In this talk, we present a set of interfaces for policy management, targeted at VO and site administrators, that simplify the writing, validation and distributed deployment of complex authorization policies and does not have the shortcomings (e.g., policy propagation delays and inconsistency, manual configuration, etc..) of the currently available solutions.

Primary authors : Dr. CIASCHINI, Vincenzo (INFN CNAF) ; Dr. GHISELLI, Antonia (INFN CNAF) ; Dr. CECCANTI, Andrea (INFN CNAF) ; Dr. VENTURI, Valerio (INFN CNAF) ; Dr. FERRARO, Andrea (INFN CNAF) ; Dr. ANDREOZZI, Sergio (INFN CNAF) ; Dr. SALOMONI, Davide (INFN CNAF) ; Dr. ITALIANO, Alessandro (INFN CNAF)

Co-authors :

Presenter : Dr. CIASCHINI, Vincenzo (INFN CNAF)

Track classification : Workflow ; Data Management

Contribution type : oral presentation

Submitted by : Dr. CIASCHINI, Vincenzo

Submitted on Wednesday 14 February 2007

Last modified on : Wednesday 14 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 167

Grid Application Platform (GAP)

Description of the activity :

Grid Application Platform (GAP) is a light-weight framework for developing problem solving applications. GAP was designed by layered architecture to make the system easy to extend. Compared to traditional grid services, it provides a simpler way for both users and developer to use grid and create grid application. Avian Flu docking services were integrated to GAP with improved portal-based user interface, and have been promoted to wider auto-docking applications in Taiwan and Asia Pacific.

Grid added value :

The added value of GAP is that the GAP is developed with the following aspects:

- The GAP is easy to use for not only the end-users but also the grid application developers. GAP provides higher level of Java API which maps the problem domain model to programming domain model very easily.
- The GAP is easy to evolve for adapting new IT technologies, and the accommodation must be transparent to both developers and users. The GAP abstracts the difference of grid middleware with an unified interface and could be extended with new middleware.
- The GAP is light-weight in terms of the deployment effort and the system overhead.

Its goal is to provide problem domain models for grid application and prevent developers from reinventing the wheels.

Experience or proposed activity :

GAP consists of Core Framework(CF), Application Framework(AF), and Presentation Framework(PF), to simplify staged works in developing scientific applications. For reliability and scalability, Service Oriented Architecture is deployed to make the framework scalable from a single box to highly distributed network.

CF introduces an abstraction layer for bridging utility applications deployed on distributed environments. Handling of computing jobs and Grid users are simplified by a set of well-defined Java APIs.

AF provides an action-based approach for problem solving. Applications are modeled by a collection of commands corresponding to certain actions. The outcome is a collection of reusable commands which can be reused for new application development.

PF can adopt any java-based presentation frameworks to meet user requirement by flexible UI customization. For example, BeanShell is used for developing an interactive shell, while Java Server Face could be used for web interface design.

Future evolution :

GAP is adopted to the drug discovery, digital archive application, earthquake data center services and so on. The first version of Avian Flu drug discovery application is scheduled to be released in March, 2007. GAP is designed for multiple Grid middleware although the current focus is gLite. Once the interoperation specification is defined, the wrapper of GAP could be adapted to that standard very easily. Overhead would be an issue and should be evaluated to spot the bottleneck in the future.

Primary authors : Mr. CHIU, Shih-chun (ASGC)

Co-authors : Mr. LEE, Hurng-chun (ASGC) ; Mr. HO, Li-yung (ASGC) ; Mr. CHEN, Yu-hsuan (ASGC) ; Mr. UENG, Wei-lung (ASGC) ; Mr. CHEN, Hsin-yen (ASGC) ; Mr. YEN, Eric (ASGC)

Presenter : Mr. CHIU, Shih-chun (ASGC)

Track classification : Interactivity and Portals

Contribution type : --not specified--

Submitted by : Mr. CHEN, Hsin-yen

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Interactivity and Portals

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 168

End-user viewpoint of the EGEE infrastructure using GridICE

Description of the activity :

Large Grid infrastructures such as EGEE are characterized by a huge amount of geographically dispersed heterogeneous resources. In this context, end-users ave not a clear knowledge of the type, state and features of the resources constituting the Grid. GridICE is a monitoring tool for Grid systems based on requirements given by different types of users. During the second phase of EGEE, this tool has been expanded in its functionalities to simplify the capability of tracking the user activity.

Grid added value :

The monitored data is collected from each site by means of the Information service. The available measurements include the GLUE Schema and extensions have been defined to satisfy new requirements. In particular, the new measurements are related to Grid jobs (e.g., Grid and local ID, memory usage, CPU usage) and to summary information of batch systems (e.g., number of total slots, number of worker nodes that are down). End-users have therefore the possibility to obtain information related to Grid resources inventory and availability. Moreover, they are able to track their Grid activities in terms of job state, resource consumption and associated VOMS attributes.

Experience or proposed activity :

The different aggregations and partitions of monitoring data are based on the specific needs of different users categories. Being able to start from summary views and to drill down to details, it is possible to verify the composition of virtual pools or to sketch the sources of problems. A complete history of monitoring data is also maintained to deal with the need for retrospective analysis. Both VO managers and end-users could navigate GridICE Web pages to verify the available resources and their status before to start the submission of a huge number of jobs.
The recent activity was targeted at comparing the information presented by GridICE with that obtained from the batch systems and other monitoring tools in order to improve the data quality of GridICE data. Furthermore, the non-intrusiveness and stability of the tool have been substantially improved even in a very large computing centers such TIER1's.

Future evolution :

In our future work, we will concentrate on improving the user experience, both at content and presentation levels (e.g., by adding VOMS group-based aggregations). Furthermore, we will cooperate to improve access to GridICE data from external applications (e.g. LCG Dashboard).

Primary authors : Mrs. AIFTIMIEI, Cristina (INFN, Padova, Italy, on leave from NIPNE-HH, Romania) ; Dr. ANDREOZZI, Sergio (INFN-CNAF, Bologna, Italy) ; Mr. CUSCELA, Guido (INFN, Bari, Italy) ; Mr. DAL PRA, Stefano (INFN, Padova, Italy) ; Mr. DONVITO, Giacinto (INFN, Bari, Italy) ; Dr. DUDHALKAR, Vihang (INFN, Bari also affiliated to Politecnico di Bari, Italy) ; Mr. FANTINEL, Sergio (INFN, Padova also affiliated to INFN-LNL, Padova Italy) ; Mr. FATTIBENE, Enrico (INFN-CNAF, Bologna, Italy) ; Prof. MAGGI, Giorgio Pietro (INFN, Bari also affiliated to Politecnico di Bari, Italy) ; Mr. GIUSEPPE, Misurelli (INFN-CNAF, Bologna, Italy) ; Mr. PIERRO, Antonio (INFN, Bari, Italy)

Co-authors :

Presenter : Mr. GIUSEPPE, Misurelli (INFN-CNAF, Bologna, Italy)

Track classification : Grid Monitoring and Accounting

Contribution type : oral presentation

Submitted by : Mr. MISURELLI, Giuseppe

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Grid Monitoring and Accounting

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 169

CSTGrid: a whole genome comparison tool for the identification of coding and non-coding conserved sequences.

Description of the activity :

One of the major task of genomic analysis is the identification of genes and regulatory elements of gene expression. Among the different methods used to solve this task, comparative analysis have shown their strength and reliability.
CSTminer is a comparative analysis tool developed to compare two or more sequences to identify conserved sequences and to classify them as coding (likely genes) or non-coding (potential regulatory regions) upon the computation of a coding potential score (CPS).

Grid added value :

The availability of a collection of evolutionary conserved sequences among different organisms could be of great importance for scientific community for the functional annotation of genomes and for the identification of regulatory elements.
Given the reliability and sensitivity of CSTminer software it would be very useful to perform genome wide comparisons of several genomes. Unfortunately given the complexity of CSTminer alignment step and the size of the genomes (3 and 2.6 Gbp for Human and Mouse genomes respectively) it would be impossible to carry out such analysis even on very powerful servers.
This is the reason why we exploited the possibility to use the EGEE infrastructure and the power of a Grid approach.

Experience or proposed activity :

In order to have a more efficient jobs we grouped together a lot of comparison (about 1000), in this way we can reduce the overhead do to files transfer and environment setting-up.
We have developed a job submission tool (JST) that allows the submission of large number of jobs in an almost unattended way. It is based on the concept of “task” to
be executed. Only if all steps are correctly executed, the status of that particular task on the central DB is updated to “Done”. In this way the central DB provides a
monitoring of the task execution and no manual intervention is required to manage the resubmission of the failed tasks.
The most important services needed in order to run this application is WMS in order to submit jobs optimizing the utilization of each resource available.
We also exploit successfully the data transport services using gsiftp in order to copy files both for input and output.

Future evolution :

We can foresee to build a web interface in which the user can easily submit its own comparison with between a given genome and a new one (provided by the user itself).
In this way this application can be established as a service for the user even if not expert of the EGEE infrastructure.

Primary authors : Dr. DONVITO, Giacinto (UNIVERSITà DEGLI STUDI DI BARI) ; Dr. MIGNONE, Flavio (Università di Milano)

Co-authors : Prof. MAGGI, Giorgio Pietro (INFN-BARI + Politecnico di Bari) ; Prof. PESOLE, Graziano (Università di Bari) ; Dr. GRILLO, Giorgio (CNR-ITB) ; Dr. LIUNI, Sabino (CNR-ITB)

Presenter : Dr. MIGNONE, Flavio (Università di Milano)

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : Dr. DONVITO, Giacinto

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production
Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 170

First bioinformatics services of the LIBI project on EGEE infrastructure: MrBayes and PAML

Description of the activity :

The tools offered (PAML and MrBayes) are commonly used in study of Genomics, Evo-Devo (a.k.a. Evolutionary Development), Systematic Biology, Barcoding (a.k.a.Species identification). In short in all the fields of biology where researchers need the inference of molecular evolution parameters as rates of nucleotide substitution (synonymous and non synonymous), or the topology of the phylogenetic tree that connects the different sequences.

Grid added value :

The availability of these tools on public servers, that offer them in distributed and parallel versions, offer new opportunities to researchers. In fact, researchers have four kind of difficulties in trying to infer parameters of molecular evolution: first, often a single problem requires large effort of computing power usage; second, in Genomics, but also in other fields, the inference needs to be repeated across the different part of a genome for thousand of times; third, several evolution models often need to be tested and compared and this requires to repeat even long analyses several times, lastly often hundred of simulated data sets need be analyzed to confirm the result of the inference on the observed one.

Experience or proposed activity :

For both the MrBayes and PAML applications we have executed thousands of jobs both MPI and Standard ones. For the MPI jobs we have we have analyzed as the ratio between the queue time and CPU time needed varies depending on the number of CPU used by the job.
The most important services needed in order to run this application is WMS in order to submit jobs optimizing the utilization of each resource available. A specific requirement was also used, in the case of MrBayes application, in order to find the farm configured to accept MPI jobs.
Another important service is the Data Management. We exploit successfully both data transport using gsiftp and LFC (LCG File Catalog) in order to simplify the storing and reading of the files.

Future evolution :

It seems that the services provided are enough to allow the proficient run of many distributed instance of these applications. For the parallel runs, the status of the infrastructure reveals that some more work is needed on the farm's configuration to make parallel jobs execution more efficient. For example dedicated queue (specifically devoted to MPI jobs) or specific scheduling policies could be adopted.

Primary authors : Dr. DONVITO, Giacinto (UNIVERSITà DEGLI STUDI DI BARI) ; Dr. VICARIO, Saverio (CNR - ITB Bari + EEB - Yale University)

Co-authors : Dr. TURI, Antonio (University of Salento) ; Dr. JEAN, Atul (INFN-BARI + Politecnico di Bari) ; Dr. MIRTO, Maria (SPACI Consortium and University of Salento, Italy) ; Dr. MY, Salvatore (INFN-BARI + Politecnico di Bari) ; Prof. ALOISIO, Giovanni (SPACI Consortium and University of Salento, Italy) ; Prof. SACCONE, Cecilia (CNR - ITB Bari + Università di Bari)

Presenter : Dr. DONVITO, Giacinto (UNIVERSITà DEGLI STUDI DI BARI)

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : Dr. DONVITO, Giacinto

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 171

Monitoring, accounting and automated decision support for the ALICE experiment based on the MonALISA framework

Description of the activity :

We are developing a general purpose monitoring system for the ALICE experiment, based on the MonALISA framework. MonALISA (Monitoring Agents using a Large Integrated Services Architecture) is a fully distributed system with no single point of failure that is able to collect, store monitoring information and present it as significant perspectives and synthetic views on the status and the trends of the entire system. Furthermore, agents can use it for taking automated operational decisions.

Grid added value :

Monitoring information is gathered locally from all the components running in each site. The entire flow of information is aggregated on site level by a MonALISA service and then collected and presented in various forms by a central MonALISA Repository. Based on this information, other services take operational decisions such as alerts, triggers, service restarts and automatic production job or transfer submissions.

Experience or proposed activity :

The system monitors all the components: computer clusters (all major parameters of each computing node), jobs status and consumed resources (CPU, both in time and SpecInt2k units, memory, disk usage), jobs network traffic while reading/writing files with xrootd, services availability with details in case of failures (both AliEn and LCG services, proxies lifetime), storage monitoring with detailed information on number of files, available space, or staging and migrating operations, FTD/FTS transfers. The system is reliable and functional for more than two years, representing the main view towards the ALICE Grid.

Future evolution :

Our focus is now on using the monitoring information for the development of higher level services that can take more intelligent operational decisions.

Primary authors : Mr. CIRSTOIU, Catalin (CERN) ; Mr. GRIGORAS, Costin (CERN) ; Dr. BETEV, Latchezar (CERN)

Co-authors : Mr. SAIZ, Pablo (CERN) ; Dr. PETERS, Andreas Joachim (CERN) ; Mr. MURARU, Adrian (CERN) ; Mr. VOICU, Ramiro (Caltech) ; Dr. LEGRAND, Iosif (Caltech)

Presenter : Mr. CIRSTOIU, Catalin (CERN)

Track classification : Grid Monitoring and Accounting

Contribution type : oral presentation

Submitted by : CIRSTOIU, Catalin

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Grid Monitoring and Accounting

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 172

VOMS : cross middleware virtual organization management through standards compliance

Description of the activity :

The concept of Virtual organizations defined as dynamic collections of individuals, institutions, and resources emerged as central in the 'Grid Problem': flexible, secure, coordinated resource sharing across dynamic, multi institutional collaborations. VOMS was born with the aim of supporting the dynamic, fine grained, and multi-stakeholder access control needed to enable sharing over virtual organizations.

Grid added value :

We are going to make VOMS available to a larger community using diverse middleware stacks with different security infrastructures. This will be achieved by supporting the two major standards for expressing attributes (RFC3281 and OASIS SAML). The ongoing concurrent standardization of Grid services interfaces such as for job submission services (e.g. OGF OGSA-BES) will assure to virtual organization members a transparent Grid experience over sites using different middleware distributions (e.g. gLite, UNICORE, Globus Toolkit).

Experience or proposed activity :

Within the OMII-Europe project, we are enhancing VOMS to support authorization standards emerging from the Open Grid Forum as well as from other standardization bodies. Besides the current widely deployed and used RFC 3281 Attribute Certificate based interface, OMII-Europe is developing an interface for VOMS based on the OASIS Security Assertion Markup Language (SAML) set of specifications. As a result VOMS will be available accross different security infrastructures. For instance, we are also enhancing UNICORE to integrate VOMS for authorization, having execution of jobs on UNICORE sites denied/allowed based on the user's VO attributes. OMII-Europe is also working on supporting standardization of the interface for job execution components of two majors Grid distributions, gLite CREAM-BES, and UNICORE OGSA-BES.

Future evolution :

Being available over different security infrastructures, VOMS aim to provide users of virtual organizations' users and administrators a management system that is unique and consistent accross different middleware stacks. In order to provide a complete authorization framework to users and resource owners, the possibility of using structured policies must be assured. Components in the gLite middleware stack such as gJAF and gPBox aim to complement the authorization capability offered by VOMS.

Primary authors : VENTURI, Valerio (INFN)

Co-authors : RIEDEL, Morris (FZJ)

Presenter : VENTURI, Valerio (INFN)

Track classification : Workflow ; Data Management

Contribution type : oral presentation

Submitted by : VENTURI, Valerio

Submitted on Wednesday 14 February 2007

Last modified on : Wednesday 14 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Track : Data Management

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 173

RDIG MONITORING AND ACCOUNTING

Description of the activity :

The developed monitoring system allows to keep an eye on parameters of Grid sites' operation in real time. There is also a option to keep track of a history of sites usage. The system is based on MonALISA package (Monitoring Agents in Large Integrated Systems Architecture) and our own developments. It permits to get an information on resources of computational sites, virtual organizations activities and some parameters of network channels.

Grid added value :

Additional task was an accounting system. It stores the data on resource utilization on Grid sites by virtual organizations and single users. The derivable parameters are jobs count, consumed CPU time, average job waiting time, used physical memory. Information is taking from R-GMA (Relational Grid Monitoring Architecture) system and stored in local Oracle DB. Web interface allows to select and group parameters by different criteria such as period of time, virtual organization, Grid site.

Now RDIG Monitoring and Accounting Web site is functioning well and is in use.

Experience or proposed activity :

We have experience in monitoring and accounting of grid sites using LDAP, R-GMA, MonALISA.

Future evolution :

Following parameters were chosen for Russian Grid segment:
number of busy, free and down CPUs;
amount of running and waiting jobs for each virtual organization (VO);
used and available disc space for each VO;
main servers loading (e.g. for Computing Element);
Round Trip Time (RTT) in networks between Resource Centers.
The new monitoring system based on current one is developed now. It will include different subsystems such as job monitoring, network monitoring, storage monitoring and other impro

Primary authors : Mr. TKACHEV, Igor (Joint Institute for Nuclear Research)

Co-authors : Mr. BELOV, Sergey (Joint Institute for Nuclear Research)

Presenter : Mr. TKACHEV, Igor (Joint Institute for Nuclear Research)

Track classification : Grid Monitoring and Accounting

Contribution type : oral presentation

Submitted by : Mr. TKACHEV, Igor

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Grid Monitoring and Accounting

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 174

Large Earthquakes Source & Mechanism Determination

Description of the activity :

The Institut de Physique du Globe de Paris (IPGP) is the primary user of the application itself while its results are used by a large tectonic and seismological community of scientists. The application mainly uses data originating from the IPGP-administrated worldwide Geoscope sensor network. Source Mechanism of large Earthquakes (magnitude > 6.5) is systematically determined on the occurence of such an event.

Grid added value :

Determining the Source Mechanism of large Earthquakes requires an overall computation time ranging from 400 to 1600 hours depending on several factors such as the quantity and quality of collected data and the size of the Earthquake. The grid enables this computing time to be apparently reduced to about half a day, providing the results shortly after the shaking. No other computing facilities allow this celerity.

These results can be used by the seismological community as inputs for further studies, such as prediction of static ground deformation through synthetic SAR (Small Aperture Radar satellite) interferometry imaging, or refenrence focal mechanism in kinematic source tomography.

Experience or proposed activity :

Speed in determining the results is the key factor that the Grid brings to the application. Latencies are caused by waiting / aborted / lost jobs and were reduced by a close monitoring of the submitted jobs. Now a (master) software has been developped that sends calculations to effectively running jobs (workers) and ensures that all needed computations are completed, without other user action than jobs submission itself. The required feature is outbound connectivity from the worker nodes, however bulk job submission also significantly shrinked the duration of this first phase of the grid process.

Future evolution :

In the beginning of the deployment of the application to EGEE, naïve grid usage led to an unsatisfying situation were tasks required from the user to obtain the application results were much too heavy and frequently needed. This has yet been solved with an external (from the

middleware
point of view) sotfware based on a master/worker architecture.

Primary authors : CLéVéDé, Eric (CNRS / IPGP) ; WEISSENBACH, David (CNRS / EGEE)

Co-authors :

Presenter : CLéVéDé, Eric (CNRS / IPGP) ; WEISSENBACH, David (CNRS / EGEE)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : DAVID, Weissenbach

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 175

Alternative Splicing Prediction on EGEE infrastructure

Description of the activity :

Alternative splicing (AS) is increasingly emerging as a major mechanism in the expansion of transcript and protein complexity in eukaryotes. Computational resources for AS prediction available on the web are limited to the analysis of single genes and their related transcripts. Among these class of algorithms we are planning to use ASPic (Alternative Splicing Prediction) resource on the EGEE infrastructure.

Grid added value :

The use of bioinformatics applications on the grid allows the mining and alignment of large amount of sequences. In order to analyse transcriptome and proteome complexity of multicellular organisms, we are facing the problem of determining the alternative splicing pattern of a huge list of genes and their collection of related transcribed sequences. We are planning to do test experiments on the EGEE grid performing thousands of relatively small independent tasks, each of which costs at most minutes or hours, to analyze gene classes involved in human health and disease. This could provide increasingly important results in many areas of basic and applied biomedical research.

Experience or proposed activity :

We are planning to distribute the large amount of task that this application requires in order to reduces the total running time; we must also solve problems like dependencies of the each task and data transfer for both input and output files. We must address also checking of failures and retries in order to make possible the running of the entire analysis as a Grid jobs.
The most important services needed in order to run this application is the WMS in order to submit jobs optimizing the utilization of each resource available, taking care of using the ranking that allow the job to run faster.
Another important service is the Data Management. We will exploit both data transport using gsiftp and LFC (LCG File Catalog) in order to simplify the storing and reading of the files.

Future evolution :

It seems that the services provided are enough to allow the proficient run of many distributed instance of these applications.
In the future we will provide also a web interface in order to make this service available to all users that do not know the details of the EGEE infrastructure.

Primary authors : Dr. DONVITO, Giacinto (INFN-BARI)

Co-authors : Dr. CASTRIGNANO', Tiziana (Consorzio Interuniversitario per le Applicazioni di Supercalcolo per Universita` e Ricerca) ; Dr. D'ONORIO DE MEO, Paolo (Consorzio Interuniversitario per le Applicazioni di Supercalcolo per Universita` e Ricerca) ; Dr. D'ANTONIO, Mattia (Consorzio Interuniversitario per le Applicazioni di Supercalcolo per Universita` e Ricerca) ; Dr. MIGNONE, Flavio (Università di Milano) ; Dr. RIZZI, Raffaella (DISCO, University of Milan Bicocca) ; Dr. BONIZZONI, Paola (DISCo, University of Milan Bicocca) ; Prof. MAGGI, Giorgio Pietro (INFN-BARI + Politecnico di Bari) ; Prof. PESOLE, Graziano (Università di Bari)

Presenter : Dr. DONVITO, Giacinto (INFN-BARI)

Track classification : Experience with application domains – setting up and production

Contribution type : --not specified--

Submitted by : Dr. DONVITO, Giacinto

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 176

Automatic Detection of Gravitational Arcs

Description of the activity :

Gravitational arcs are highly distorted and often magnified images of very distant galaxies. These images are caused by the gravitational lens effect and are extremely useful tools for various astrophysical purposes, e.g. mass determinations of the lens, investigations on the very early universe and even cosmological studies. However, these objects are usually very thin and faint structures and are therefore very hard to detect, in particular on Wide Field images.

Grid added value :

We have developed a new algorithm for an automated detection of these useful objects. This algorithm consists of mainly four steps: (1) a small Gaussian presmoothing, (2) histogram modification for defining the expected dynamic range of the arcs, (3) anisotropic filtering, which is a direction depended smoothing for enhancing arc-like structures, and, finally, (4) the selection of the arc candidates, which is done by defining selection criteria. All four steps depend on various parameters having large influence on the final detection success. However, this set of parameters has to be optimised for each individual combination of the used telescope/detector setup, as images of the same object taken with different instruments differ, e.g. in depth or resolution. Therefore the use of EGEE resources offer the possibility to perform parameter studies on each single setup. Once performed we expect this parameter set to be very useful for all astrophysicists working working on this topic.

Experience or proposed activity :

We have access to astrophysical images taken with various instrument/ telescope combinations containing proven gravitational arcs. Therefore we are able optimize the parameter sets to each individual of these instrument/telescope combinations we have in use. This will be done by varying the parameter values within meaningful ranges, which are already roughly constraint. In particular, we expect the job type Parametric" offered in the GLite middleware to be very useful for our purposes. The future use of the EGEE infrastructure with its tremendous computing power offers also the possibility for a much larger number of parameter value combinations compared to Beowulf clusters. Also the step size of the value variation can be decreased dramatically.

Future evolution :

We will further constrain the parameter space in a first step. This is necessary as an inappropriate set of values may cause a dramatic increase in computing time. The second step will then be the main parameter study. Both steps have to be done for each individual instrument separately.

Primary authors : Dr. KAUSCH, Wolfgang (Institut für Astro- und Teilchenphysik, Universität Innsbruck)

Co-authors : Dr. LENZEN, Frank (Institut für Informatik, Universität innsbruck) ; Prof. SCHINDLER, Sabine (Institut für Astro- und Teilchenphysik, Universität Innsbruck)

Presenter : Dr. KAUSCH, Wolfgang (Institut für Astro- und Teilchenphysik, Universität Innsbruck)

Track classification : Poster session

Contribution type : poster

Submitted by : Dr. KAUSCH, Wolfgang

Submitted on Wednesday 14 February 2007

Last modified on : Friday 16 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 177

On demand Web-Based Configuration-Driven Bug-Patch Track and Build System

Description of the activity :

Software deployment are difficult processes when people are located in different sites and involved in developing several software components subject to frequent changes and improvements. Therefore, it is extremely important to setup a local daily build system in order to allow each site Release Menager to handle the whole project and to better collaborate with the central Release Manager. In addition, by setting up this system, Testers are able to verify the latest functionalities.

Grid added value :

The local build system we have planned, called Goldrake, is an automatic build system able to store in a database information related to bugs, patches and tags in order to trace the whole history of a candidate release in order to monitor the whole life of a project. The clear advantage here is given by the introducing of a database organized to collect information about a specific software project, improving the cross-activities within the co-workers taken part in the project, and providing useful information in case of build and test problems.
Goldrake is an automatic and on demand web-based configuration-driven remote build system, able to interact with different type of build system, such as gLite, and to support several build mechanisms such as ant and autotool. It allows the site Release Managers to be in harness with Developers and Testers, but in particular to be reactive respect to the central Release Menager, whose activity represents the success of the software deployment.

Experience or proposed activity :

We already had experienced having to setup the local build system in the INFN INFN National Centre for Telematics and Computer Science (CNAF) in order to support developers involved Grid projects, such as DataGrid, EGEE I and EGEE II. Goldrake officially started in 2005 when the build system used in DataGrid project was integrated with a Web Application. In 2006 it was extended with simple monitor checking. During these years our experience with Goldrake has been focused on evaluating and improving Goldrake's applicability to various system, and testbeds, as well as Goldrake's functionality in automating the software development process. The system is used by the Release Managers and Testers of the Italian (IT) team that is a group of developers collaborating in the middleware JRA1 of EGEE II project. The IT team has the responsibility of some components that belong to the subsystem controlled by different teams and fully manages other subsystems.

Future evolution :

As stressed by the previous points, the setting up of a local build system is extremely important for the success of the Grid software deployment. Goldrake is currently a backup of the gLite sytem. In order to continuously allow this goal we plan to extend the current Goldrake to support the ETICS system, being the new official build, configuration, integration and testing system used by EGEE II project.

Primary authors : RONCHIERI, Elisabetta (INFN CNAF)

Co-authors : LOPS, Roberto (INFN CNAF) ; GIACOMINI, Francesco (INFN CNAF)

Presenter : RONCHIERI, Elisabetta (INFN CNAF)

Track classification : Poster session

Contribution type : poster

Submitted by : Miss. RONCHIERI, Elisabetta

Submitted on Wednesday 14 February 2007

Last modified on : Friday 16 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 178

PlugAndPlay User Interface: a simple grid access designed on the basis of grid user requirements.

Description of the activity :

PlugAndPlay User Interface(PnP UI) is a set of scripts allowing to access grid resources by any scientific/technical community in a simple and comfortable way, and with a valid proxy certificate. It is provided as a simple unique directory with all the required files, scripts and library, it doesn't require root privileges and can be used in AFS context. It was tested on the major Linux distributions. PnP UI can be configured by the final user through a text file to set required grid services.

Grid added value :

The development activity of the PnP UI is intended to easily extend all the grid advantages to wide and fast increasing communities. As it can be enabled on different RedHat platform without root privileges, the user can skip some disadvantages related to the classic UI: contacting a site manager to ask for an account on existing User Interface; transferring his data on an external and shared machine with necessary limits of storage space; compromising the management and the privacy of his scientific, technical or business data; risking to slow down his work because of high user traffic on the shared UI.
PnP UI supports high potential diffusion among grid users also because the RPM absence makes it easily customized to be compliant with different systems. If libraries are missing on any RedHat platform or required for specific applications, they can be easily added without changing the previous system configuration.

Experience or proposed activity :

PnP UI is released in TAR and AFS versions: it can be used inside a shell just running an init script or with an active AFS client, or downloading the tar package containing the middleware. Although PnP UI is intended to be used by any not-root user of any RedHat platform, however the RPMs absence has made also possible to extend the scripting procedure for a configuration by root in order to enable PnP UI for all the user of a certain server node. This operation can be useful on an e-mail server. User configuration values can be easily set on a text file for all the types of PnP UI access: in the TAR version the configuration file is contained in the package, in the AFS and root version it appears automatically in the user home directory. PnP UI is tested on Scientific Linux, but exploiting the experience of the users, it has been made compliant with Fedora and Debian. It is totally integrated with gLite middleware.

Future evolution :

The diffusion of the PnP UI can help to extend the grid user community and integrate in the EGEE infrastructure new and different applications (also easily including in the UI PnP the required libraries). It fulfils security requirements both on the user side and on the grid managers checking the right proxy details. During each user access, PnP UI also updates usage data collected on a Storage Element for statistics. The user easily can set grid services required by a Virtual Organizations(VO).

Primary authors : Dr. CAROTA, Luciana (INFN-CNAF-Bologna, Italy) ; Dr. FERRARO, Andrea (INFN-CNAF-Bologna, Italy)

Co-authors :

Presenter : Dr. CAROTA, Luciana (INFN-CNAF-Bologna, Italy)

Track classification : Poster session

Contribution type : poster

Submitted by : Dr. CAROTA, Luciana

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 179

Monte Carlo mass production of the H1 and ZEUS experiments

Description of the activity :

The presented activity is performed in the context of high-energy physics(HEP) research.
The H1 and ZEUS experiments study electron-proton collisions at the presently running HERA accelerator located at DESY, Hamburg, Germany.
In order to analyze the vast amounts of collected data, large sets of simulated, Monte Carlo (MC), events are required. A large fraction of these MC events has been produced using the EGEE infrastructure.

Grid added value :

The increased event rates after the upgrade of the HERA accelerator lead to more demands in computing and storage resources. In order to meet these demands with new resources becoming available via the Grid, both experiments have established MC mass production systems based on EGEE/LCG components. Development and production have been performed in the HONE and ZEUS VOs, which are meanwhile supported by various sites: about 25 sites in Europe for HONE and for ZEUS about 40 sites including two OSG sites in the USA.
Both experiments have produced together over one billion MC events on the Grid, where the processing of one event takes roughly 10s corresponding to over 300 CPU years used. Thus the Grid has become a vital part of the computing strategy for both experiments. The production experiences are fruitful for future HEP experiments and also other communities with need for big data productions.

Experience or proposed activity :

The development of the production system and the necessary infrastructure has been a common effort among the experiments and the DESY IT department, who provides the central services for the HONE and ZEUS VOs, e.g. VOMS, LFC, RB/WMS and a reliable mass storage system with tape back-end for the simulated events.
The main difficulty for the mass production was found to be unreliable data handling within the Grid. Additional functionality in terms of retries and fall-backs had to be implemented in the production frameworks to guarantee a continuous, efficient production.
In general there was a fast and constitutive response by the administrators from the participating sites. some of the observed problems have already been addressed in the middleware development.

Future evolution :

Although the middleware has evolved, the reliability for data handling is not yet sufficient for mass productions. The main reasons are missing features in the client tools, e.g. automatic retries and fall-backs, and the general stability of storage elements.
Another area for improvement is the resource brokering, where treatment of free resources and data availability is too static and not flexible enough.

Primary authors : Dr. WISSING, Christoph (Deutsches Elektronen-Synchrotron (DESY), Germany) ; Dr. STADIE, Hartmut (Deutsches Elektronen-Synchrotron (DESY), Germany)

Co-authors : Mr. ERNST, Michael (Deutsches Elektronen-Synchrotron (DESY), Germany) ; Dr. MANKEL, Rainer (Deutsches Elektronen-Synchrotron (DESY), Germany) ; Dr. FERRANDO, James (University of Glasgow, United Kingdom) ; Mr. WRONA, Krzysztof (Deutsches Elektronen-Synchrotron (DESY), Germany) ; Mr. KARBACH, Moritz (University of Dortmund, Germany) ; Mr. LOBODZINSKI, Bogdan (Deutsches Elektronen-Synchrotron (DESY), Germany) ; Mr. VOROBIEV, Maxim (ITEP Moscow, Russia) ; Dr. PREDA, Titi (Nat. Inst. for Phys. & Nucl. Engin., Bucharest, Romania) ; Dr. GELLRICH, Andreas (Deutsches Elektronen-Synchrotron (DESY), Germany)

Presenter : Dr. WISSING, Christoph (Deutsches Elektronen-Synchrotron (DESY), Germany) ; Dr. STADIE, Hartmut (Deutsches Elektronen-Synchrotron (DESY), Germany)

Track classification : Experience with application domains – setting up and production

Contribution type : oral presentation

Submitted by : CHRISTOPH, Wissing

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Experience with application domains – setting up and production

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 180

Grid efficiency for high throughput and data-intensive analysis of bacterial genomic sequences

Description of the activity :

This work handles the grid performances of high throughput comparison of genomes, producing in parallel tens of Gigabytes of data. The community is composed of Bioinformatics and grid partners. Input/output data were stored on Storage Elements as flat files, according to a logical/physical directory structure. The grid efficiency was depending on the balance between wide parallel resource access and data access bottle necks. Methods and algorithms were designed for supporting high throughput.

Grid added value :

Technologies for genome sequencing disclosed large amount of knowledge about genes and proteins. The accumulation of genomic data in public databases allows a comprehensive comparative large scale investigation with the aim of solving many important tasks in molecular biology, such as functional and structural genomic annotation. Starting from these comparative analysis we can also infer phylogenetic relationships between species. In this context the grid added value consists in distributed parallel computing resources and data storage for making possible data-intensive and high throughput comparisons which cannot be supported by local farm. Once extracted the genomic data from the public database of the National Center for Biotechnology Information, our task included two massively parallel steps: one dedicated to protein sequence comparisons and the second checking the consistency of the output. They required about 2 millions and 8 thousand millions of independent runs respectively.

Experience or proposed activity :

For supporting both high parallelism for independent runs and manageable data format for a large and complex information, I/O data were stored on grid Storage Elements according to a specific name-space structure of Logical File Names, which are mapped on their physical paths by a grid File Catalog. Whenever possible, the high computing parallelism was reduced by properly grouping a set of elementary steps in single executable nodes, running on single grid Worker Nodes. Once defined parallel executable nodes, the grid middleware allows the users to group and manage many of them through a single structure. The grid Workload Management System (WMS) supports jobs of Collection and DAG type, where many nodes are grouped in a single job and are distributed by the WMS on different grid computing resources. In this task we successfully used both few Collection JDLs from hundreds to thousands of nodes, and hundreds of DAG JDLs including thousands of nodes.

Future evolution :

The grid File Catalog was able to store millions of paths: large file divisions of output support high parallelism for independent runs. However, as output files become input files next task steps, multiple file access on shared storages could cause reading bottle necks. The grid efficiency depends on the balance between wide resource parallel access and emergence of reading bottle necks. This balance can be optimized by specific algorithms or by future improvings of

storage access.

Primary authors : Dr. CAROTA, Luciana (INFN-CNAF-Bologna, Italy) ; Dr. BARTOLI, Lisa (Biocomputing Group, University of Bologna) ; Dr. FARISELLI, Piero (Biocomputing Group, University of Bologna) ; Dr. MARTELLI, Pier Luigi (Biocomputing Group, University of Bologna) ; Dr. MONTANUCCI, Ludovica (Biocomputing Group, University of Bologna) ; Prof. CASADIO, Rita (Biocomputing Group, University of Bologna)

Co-authors : Dr. DONVITO, Giacinto (INFN-Bari,Italy) ; Prof. MAGGI, Giorgio (INFN-Bari,Italy)

Presenter : Dr. CAROTA, Luciana (INFN-CNAF-Bologna, Italy)

Track classification : Poster session

Contribution type : poster

Submitted by : Dr. CAROTA, Luciana

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Poster session

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 181

A Climate Application in the GRID: on behalf of EELA Project

Description of the activity :

El Niño phenomenon is a key factor for Latin-American climate prediction. El Niño has
a special interested due to its direct effects in the Pacific coast of South America and in particular in Peru and Chile. Moreover, research institutes from Peru and Chile (EELA LA partners) run global and regional climate models and need to compare their results with other simulations performed by international centres in the El Niño area.

Grid added value :

The climate applications has been deployed on the pilot EELA infrastructure for both production and dissemination purposes. The applications deployed for the EELA Climate task have been selected from the different available tools in the climatic community (global and regional climate models, data analysis tools, etc.) which are necessary to predict, analyze and understand the El Niño phenomenon. This tools perform a cascade of applications starting with a global climate model (CAM), which feeds input data to a regional model (WRF) to focus in world regions of interest. A third step is performed to make a data-mining analysis from both models. Climate applications (CAM, WRF, SOM) deal with a large number of datasets stored locally, so grid technology can offer a solution to access them in a transparent way.

Experience or proposed activity :

Due to the special characteristics of the workflow of the application, experiments lasting beyond proxy certificates lifetime, control of jobs not supplied by the middleware, etc., the cascade of applications makes an efficient use of the middleware. This control is based on LFC and AMGA (R-GMA is planned) to store all input and output datasets, and to control the status of the different experiments submitted to the GRID.

Future evolution :

Some limitations has been found from the middleware, specially
with long lasting jobs, for example a typical experiment for a
100 years in a CAM run it requires 7 months of CPU. This issue
has been solved and implemented in the application. No middleware component allows to access, process and querymetadata thata are contained in the files selfdescribing itself, but a first approach has been implemented in the application to solved this problem.

Primary authors : Dr. COFINO, Antonio S. (University of Cantabria)

Co-authors : Mr. CARRILLO, Mauricio (SENAMHI) ; Mr. BAEZA, Claudio (Universidad de Concepción) ; Dr. FERNANDEZ, Jesus (University of Cantabria) ; Mr. MIGUEL, Richard (SENAMHI) ; Dr. ABARCA, Rodrigo (Universidad de Concepción) ; Dr. GUTIERREZ, Jose M. (University of Cantabria) ; Dr. MAYO, Rafael (CIEMAT)

Presenter : Dr. COFINO, Antonio S. (University of Cantabria)

Track classification : Related Projects

Contribution type : --not specified--

Submitted by : Dr. COFINO, Antonio S.

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Related Projects

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 182

Connecting the LOFAR telescope to the GRID

Description of the activity :

LOFAR consists of thousands of sensors (antennas) interconnected through fiber connections. The LOFAR telescope is centred in the Netherlands, but remote antenna stations are being built, or considered to be built, in a number of countries across Europe. The LOFAR user community consists of Radio astronomers across the world accessing the LOFAR system and scientific data through science centers. The opportunities of the seemingly close match to the Grid infrastructure will be explored.

Grid added value :

The aspects of LOFAR well suited to be supported by a GRID infrastructure are setting up scientific data centers and providing computational facilities for scientific analysis of the data provided by these data centers. Current prognosis for the LOFAR long term storage capacity is that it will require a growth of Petabytes per year. Even simple operations on such amounts of data will provide a challenge for computational resources. The data access patterns depend strongly on the scientific applications, and requirements for the storage and computational resources will be diverse. Furthermore, the astronomical scientist in general should not worry about the technical implications of these requirements and a transparent GRID based framework could provide the appropriate means for scientists to be able to concentrate on the astronomical aspect of their research. For international collaborations, another case to be considered is the realtime data reduction of (international) datastreams.

Experience or proposed activity :

Initiatives are taken to connect the LOFAR telescope to the GRID. The success will depend on the reliability, flexibility, and scalability of the resources. The central LOFAR system can only allow for a time/space limitted buffer for scientific data before it being distributed to science data centers. For the scientific user, GRID resources will have to provide a secure, and flexible environment for astronomical data analysis without exposure to technical details of the specific facility that is being used. The computational environment should allow for transparant deployment of, partly legacy, applications that expect a specific environment. The large quantities of data necessitate an inteligent management of data distribution, minimizing large data transfers.

Future evolution :

LOFAR challenges for GRID:
Being able to provide scalable and transparent storage taking into account storage of proprietary data.
Computational facilities providing an environment where diverse astronomical applications run efficiently.
Transparently and intelligently couple storage to computational facilities avoiding large data transfers.
Reliable and efficient data distribution mechanism preventing a "buffer overrun" of the LOFAR central systems.
Realtime processing of Tb/s datastreams.

Primary authors : Dr. HOLTIES, Hanno (ASTRON)

Co-authors :

Presenter : Dr. HOLTIES, Hanno (ASTRON)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Dr. HOLTIES, Hanno

Submitted on Wednesday 14 February 2007

Last modified on : Wednesday 14 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 183

EGEE grid infrastructure monitoring based on Nagios

Description of the activity :

We extended Nagios monitoring framework with grid specific features in order to implement efficient grid monitoring system. Main goal of this system is to achieve better availability of grid hosts and services, by precise problem detection and instant notification. The most important extensions we implemented are sensors for various EGEE services, advanced sensor hierarchy and certificate based authorization on web interface. This system is intended for various types of grid operators.

Grid added value :

This system has been deployed for monitoring core EGEE services (e.g. BDII, Resource Broker) in Central Europe (CE) Federation since May 2006. In September 2006, system was extended to monitor all grid sites in CE Federation. Currently the system is monitoring 67 nodes and over 550 services. System status can be seen on the following address: http://nagios.ce-egee.org. The system has been very well accepted in CE Federation. Since deployment, it has been used by core services managers and the first line of support personnel. Beside services monitoring, our system is used for certification (e.g. functionality testing) of new middleware installations on sites. Grid sites are provided with two options: receiving instant notifications and retrieving information through web interface. Beside EGEE monitoring we use this system for monitoring resources on Croatian national grid CRO-GRID, where further extension in form of automatic recovery mechanism was implemented and successfully utilized.

Experience or proposed activity :

Our monitoring system does not require any additional services or software to be deployed on EGEE resources. Single Nagios service which performs all the checks is deployed on the central server. So far, we have developed over 20 sensors for various EGEE services. Some sensors perform basic checks, while others perform complex functional checks (e.g. job submission). Sensors are organized hierarchically in a way that simple checks are performed more often and in the case when simple check fails, dependant complex check is not performed. In order to form as accurate picture of overall system as possible we utilize several available EGEE information services. The most important service we use is Grid Operational Database (GOCDB), from which we gather all the information about hosts, services and site administrators. GOCDB is also used for importing scheduled downtimes to our system. Beside GOCDB, we utilize central and site BDII services for getting additional information about sites.

Future evolution :

Currently our system performs only external monitoring of hosts and services. However, the system could be easily extended to monitor local fabric. Such deployment would enable combining external and internal view of individual hosts and services and thus getting more accurate status of monitored objects. This would also enable utilizing system's mechanisms for automatic recovery of services. Furthermore, our system is open and capable of supporting new EGEE grid services as they emerge.

Primary authors : Mr. IMAMAGIC, Emir (University Computing Centre (SRCE))

Co-authors : Mr. DOBRENIC, Dobrisa (University Computing Centre (SRCE))

Presenter : Mr. IMAMAGIC, Emir (University Computing Centre (SRCE))

Track classification : Grid Monitoring and Accounting

Contribution type : oral presentation

Submitted by : IMAMAGIC, Emir

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Since the subject of this presentation is grid monitoring system which is deployed as a grid service, and it is therefore not a user application, it was not possible to provide suitable answer to all above questions.

Status : SUBMITTED

Track judgments :

Track : Grid Monitoring and Accounting

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 184

Web Services running Protein Sequence Alignment on Grid: Application to Hepatitis C Viruses.

Description of the activity :

Bioinformatics analysis of data produced by high-throughput biology, for instance genome projects, is one of the major challenges for the next years. Some of the requirements of this analysis are to access up-to-date databanks and relevant algorithms. Since 1998, we are developing the Network Protein Sequence Analysis (NPS@) Web server , that provides the biologist with some most common resources for protein sequence analysis, integrated into several pre-defined and connected workflows.

Grid added value :

GPS@ Web portal and gBIO-WS make the remote access and bioinformatics job submission easier on the grid. We have used, as testcase, the ClustalW multiple alignment tool run on a remote Grid platform, to analyze the variability of a subset of sequences. Biologists can submit bioinformatics jobs on the Grid by using their usual Web client, but also integrate these grid services within complex workflow combining different databases and tools. They, then, benefit from the large-scale computing resources of the Grid, from their usual and local working environment. Grid computing and storage facilities will also permit GPS@ and gBIO services to scale to thousands of daily user as much as aligning complete genomes or proteomes. In this testcase, we consider a common task for bioinformaticians working on Hepatitis C Virus: doing a multiple alignment of sequences issued from different strains, where user is uploading its own sequence databank of HCV sequences.

Experience or proposed activity :

We have put online two multiple alignments Web Services on the CNRS IBCP servers. One is accessible through a classical Web interface, the other one can be used through a SOAP client such as Taverna or Triana, but also a user one built with gSOAP, perl SOAP::Lite or Java. These Web services can process the submitted alignment on two different computing environments: a local cluster, and the grid platform of the EU-EGEE project. The GPS@ sub-process that have submitted the job, is monitoring the job with trhe EGEE commands. When achieved, the GPS@ automat downloads the result file containing the multiple alignment computed by ClustalW, and processed it in a HTML page showing, in a colored and graphical way, the list of aligned protein sequences. This is directly inherited from the original NPS@ portal, providing biologists with a well-known interface and way of displaying results.

Future evolution :

Future works will be done about applying this Grid WebServices interface to other programs. We will, for example, work to put online, as Web Services, a selected panel of other protein alignment methods, but also similarity searching programs, like BLAST or SSEARCH, raising the issues of large and numerous databases management in Grid environment.

Primary authors : Dr. BLANCHET, Christophe (CNRS IBCP)

Co-authors : Dr. COMBET, Christophe (CNRS IBCP) ; Mr. DARIC, Vladimir (CNRS IBCP) ; Prof. DELEAGE, Gilbert (CNRS IBCP)

Presenter : Dr. BLANCHET, Christophe (CNRS IBCP)

Track classification : Interactivity and Portals

Contribution type : oral presentation

Submitted by : Dr. BLANCHET, Christophe

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Interactivity and Portals

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 185

Grids in the Clinic: The Hunt for the Radiotherapy Efficient Frontier

Description of the activity :

Modern radiotherapy delivers ionizing radiation energy to a tissue volume with the intent to disrupt cell proliferation, while avoiding as much as possible nearby non-involved tissues. A priori, it is not obvious the optimal radiotherapy treatment plan for a specific patient. We report on our experience of generating large number of competing treatment plans which are graded by humans in an attempt to construct an "efficient frontier island" in the relevant parameter phase space.

Grid added value :

Grid-like computational techniques have been previously reported in radiotherapy treatment planning either to speed-up traditional calculations or offer realistic access to monte-carlo simulations. In this work we use commercially available treatment planning algorithms to hunt for an efficient frontier in parameter space. For a 3-dim space we generate a 3, 4 or 5 lattice, with each point requiring between 15-20 min of single dedicated CPU time. The resulting plans (64, 125 or 216) are graded by seasoned practitioners and an efficient island is build in phase space. The computational load is spread over about 100 CPUs, physically located in one of our clinics. One or two radiotherapy cases can be completed during every workday, more during weekends. Such efficiency islands can provide starting points for additional computational efforts or offer a guide to the best available available treatment strategies.

Experience or proposed activity :

The size of USOncology clinics (up to 150 modern CPUs per physical location) offers an ideal match for this type of grid deployment. Physical security and excellent network connectivity circumvent certain known issues that prevent grid proliferation at 1994-95 web-server levels. Hardware age and standardization adds immensely in "silent worker CPU" deployment in the organization. We have also experience WAN-like deployment in metropolitan areas with very encouraging scaling results.
Still user credentialing, uniform level of authorization and similar administrative "trivia" continue to be issues, that we by-pass by silent installation. The cardinal issues of trust and security-especially when actual patient medical records are involved- are dealt at this stage by using the actual physical security of our clinics.

Future evolution :

The true potential of this work will be harvested from massive deployment of time-shifted grids, that is by definition, geographically separated nodes on a global scale, making the need for trusted and secure yet unobtrusive administration even more pressing. This appears to be consistently the major objection from the radiotherapy medical community in the limited feedback this work has elicited at this early stage.

Primary authors : Dr. DIMITROYANNIS, Dimitri (Kansas City Cancer Center, USOncology)

Co-authors :

Presenter : Dr. DIMITROYANNIS, Dimitri (Kansas City Cancer Center, USOncology)

Track classification : Workflow

Contribution type : oral presentation

Submitted by : Dr. DIMITROYANNIS, Dimitri

Submitted on Wednesday 14 February 2007

Last modified on : Wednesday 14 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 186

Approach to the remote instrumentation idea in the RINGrid project

Description of the activity :

A number of problems in science, industry and commerce may be addressed by using sophisticated equipment and top-level expertise, which is often locally unavailable. The answer for some of these problems is conception of Remote Instrumentation Services (RIS). RIS supports activities related with using rare equipment remotely e.g. workflows, post-processing, visualization, data management. This idea is especially attractive for: radio astronomy, chemistry, physics and medicine.

Grid added value :

The analysis of the wide implied RIS aspects are under of interest of the RINGrid (Remote Instrumentation in Next-generation Grids) project. This activity is part of 6th European Framework Programme and has been launched in October 2006. Briefly, the RINGrid project will provide systematically identification of instruments and corresponding user communities, the definition of their requirements as well as careful analysis of the remote instrumentation synergy with next-generation high-speed communications networks and grid infrastructure. These results will be the basis for the definition of recommendations for designing next-generation RIS. RINGrid associates partners coming from Europe and Latin America from 10 institutions. On the one hand it allows to achieve required level of generality and on the other hand gives desired impact by gathering scientists from different research domains. User communities are related with unique laboratory devices e.g. NMR spectrometers.

Experience or proposed activity :

All RINGrid effects will be practically verified in the last stage of the project. Prototype installations will be set up, by taking into consideration user communities and instruments as well as used software. One of the systems which will be used in validation process is PSNC Virtual Laboratory (VLab). VLab (vlab.psnc.pl) project is developed by Poznań Supercomputing and Networking Center in collaboration with the Institute of Bioorganic Chemistry since 2002.
The main research goal of the VLab is definition of a framework for building many different types of laboratory. It will facilitate and automate building new laboratories using existing modules with their functionality. The PSNC Virtual Laboratory system should not be comprehended solely as a set of mechanisms to submit, monitor and execute jobs. It is also a possibility to give access to the resources of the digital library, communication, and e-Learning systems.

Future evolution :

Basing on the demands and requirements and taking into account the state of the art, future needs and trends will be analyzed in respect of RIS. Guidelines concerning the design, development and use of next-generation RIS will be provided. Special attention will be paid to present and on-going research activities (e.g. EGEE, gLite), enabling a cooperative and integrated use of Grid technologies and self-organizing, self-configuring, self-optimizing, self-healing networks with QoS support.

Primary authors : Dr. LAWENDA, Marcin (Poznan Supercomputing and Netwrking Center)

Co-authors :

Presenter : Dr. LAWENDA, Marcin (Poznan Supercomputing and Netwrking Center)

Track classification : Interactivity and Portals

Contribution type : oral presentation

Submitted by : Dr. LAWENDA, Marcin

Submitted on Wednesday 14 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Interactivity and Portals

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 187

The Knowledge-based Workflow System for Grid Applications

Description of the activity :

The Knowledge-based Workflow System for Grid Applications (K-Wf Grid) project addresses the need for a better infrastructure for the future Grid environment.

Grid added value :

The K-WfGrid project introduces several new ideas and innovations in the area of Grid workflow application development, application of knowledge-based technology and performance monitoring services.
The main innovation of the K-WfGrid workflow management technology includes a novel approach of supporting knowledge-based workflow orchestration and Grid Service selection by means of an expert system.
K-WfGrid will develop an ontological app

Experience or proposed activity :

The K-Wf Grid system consists of several layers. The user interacts with the top layer, can access the service workflows using an interactive workflow GUI, and can access collaborative knowledge base, existing workflows, and results using the User Assistant. The top layer also provides access to results (intermediate as well as final ones). All these tools are integrated in one web-based portal. The actual middleware layer consists of a workflow management system, a knowledge store, and modules for monitoring, performance analysis, and performance prediction. The filtered monitoring information is processed in the performance prediction module, where new facts are created and stored in the knowledge store. This knowledge store is then queried during workflow construction/execution for best available services.

Future evolution :

K-Wf Grid is originally developed based on Globus Toolkit 4. The work is beging ported to gLite middleware using WSRF job submission service adapter, that receives jobs from the knowledge-based workflow management system and submits it into gLite.

Primary authors : TRAN, Viet (Institute of Informatics, Slovakia) ; HLUCHY, Ladislav (Institute of Informatics, Slovakia)

Co-authors :

Presenter : HLUCHY, Ladislav (Institute of Informatics, Slovakia)

Track classification : Workflow

Contribution type : --not specified--

Submitted by : TRAN, Viet

Submitted on Wednesday 14 February 2007

Last modified on : Thursday 15 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Workflow

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 189

PSNC Virtual Laboratory system as implementation of remote instrumentation idea

Description of the activity :

A number of problems in science, industry and commerce may be addressed by using sophisticated equipment and top-level expertise, which is often locally unavailable. The answer for some of these problems is conception of Remote Instrumentation Services (RIS). RIS supports activities related with using rare equipment remotely e.g. workflows, post-processing, visualization, data management. This idea is especially attractive for: radio astronomy, chemistry, physics and medicine.

Grid added value :

The PSNC Virtual Laboratory (VLab) is an exemplary implementation of the remote instrumentation system. The main goal of VLab is definition of a framework for building many different types of laboratory. It will facilitate and automate building new laboratories using existing modules with their functionality. Definition of all accessible remote facilities as simple resources in the Grid infrastructure allows treating jobs submitted to the real laboratory devices as any other Grid task. Dynamic measurement scenarios allow flexible defining the process of experiment, from pre-processing, through executing the experiment, to the post-processing and visualization tasks. Users are also allowed to add their own module as a part of the scenario. The PSNC Virtual Laboratory system should not be solely comprehended as a set of mechanisms to submit, monitor and execute jobs. It is also a possibility to give access to the resources of the digital library, communication, and e-Learning systems.

Experience or proposed activity :

We are going to demonstrate two implementations of the VLab. First demo – the Virtual
Laboratory of Nuclear Magnetic Resonance Spectroscopy (VLab-NMR), includes two parts: measurement scenario presentation and demonstration of the Digital Science Library functionality. The aforementioned measurement scenario contains: real experiment performed on the NMR, post processing and visualization. The experiment will be executed on Varian Unity 300 spectrometer placed in the Institute of Bioorganic Chemistry in Poznan (Poland). The output data of real experiment are automatically sent to the Data Management System (DMS). The next step covers the post processing analysis, which is performed on the application server using TopSpin application – third party software. The second implementation of the VLab is Virtual Laboratory of Radio Astronomy. It is possible to demonstrate how very large (32 m diameter) radio telescope, situated in Piwnice in Poland, can be remotely accessed and controlled.

Future evolution :

PSNC Virtual Laboratory is still under development and much new functionality is anticipated to implement. Due to consuming a lot of resources by virtual laboratory applications beginning from network bandwidth and finishing on storage for experiment results, integration with e-Infrastructure (like EGEE infrastructure) is essential. Development of virtual laboratory claims also functionality which is neither available yet nor well defined e.g. real bandwidth on demand.

Primary authors : Dr. LAWENDA, Marcin (Poznan Supercomputing and Netwrking Center)

Co-authors :

Presenter : Dr. LAWENDA, Marcin (Poznan Supercomputing and Netwrking Center)

Track classification : On-line Demonstrations

Contribution type : demo presentation

Submitted by : Dr. LAWENDA, Marcin

Submitted on Thursday 15 February 2007

Last modified on : Friday 16 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : On-line Demonstrations

Judgment :
Judged by :
Date :
Comments : ""

Abstract ID : 190

Grid Interoperation on Data Movement between NAREGI and EGEE gLite

Description of the activity :

GIN (Grid Interoperation Now) is an activity of OGF for interoperation among production grids. Major grid projects are participating, such as EGEE, NAREGI, UK National Grid Service, NorduGrid, OSG, PRAGMA, TeraGrid, etc.). They are trying to identify islands of the interoperation between production grids and grow those islands.

Grid added value :

We have implemented a Grid interoperability system on data movement (GIN-data) between NAREGI and EGEE. We have ported the srmcp (SRM Copy) module of gLite into the beta-2 version of the NAREGI software. For the file transfer interface of the NAREGI Grid filesystem (AIST Gfarm), we have developed a GridFTP-DSI (Data Storage Interface) component for the Gfarm filesystem.

Experience or proposed activity :

We have succeeded bilateral file transfer between NAREGI Gfarm and EGEE SRM (DPM) servers.

Future evolution :

For the future plan, the bilateral retrieval of file catalog information is needed for seamless file exchange between NAREGI and EGEE gLite.

Primary authors : Prof. MATSUDA, Hideo (NAREGI/Osaka University)

Co-authors : Mr. KIDO, Yoshiyuki (Mitsui Knowledge Industry Co.,Ltd./Osaka University) ; Mr. WAKATSUKI, Kentaro (Hitachi Software Engineering Co.,Ltd)

Presenter : Prof. MATSUDA, Hideo (NAREGI/Osaka University)

Track classification : Related Projects

Contribution type : oral presentation

Submitted by : Prof. MATSUDA, Hideo

Submitted on Thursday 15 February 2007

Last modified on : Sunday 18 February 2007

Comments :

Status : SUBMITTED

Track judgments :

Track : Related Projects
Judgment :
Judged by :
Date :
Comments : ""