EGEE'07

Monday 01 October 2007 - Friday 05 October 2007

Europe Congress Center

Book of Abstracts

Contents

Plenary / **0**

Towards Multi-Petascale Grids in the 2010s

Corresponding Author: matsu@is.titech.ac.jp

While there is general consensus that computing platforms underlying the grid infrastructures will continue to evolve, variance in the speed of technology acceleration in HPC is causing many of the assumptions made in the early days of grid to no longer hold. Such divergence
 in the metrics, as well as wider proliferation of related technologies such as Web2.0, will be changing the optimal design of the overall grid infrastructure towards more centralization. Based on our recent
 experiences with our TSUBAME supercomputer, which is currently Asia-Pac's fastest machine according to the Top500, and its next petascale generation design thereof, we will discuss the future design of multi-petascale grids with such machines being constituent massive resource
 nodes instead of vast distribution.

Technical Plenary / **1**

Sensible Cities: Digitally Augmented Urban Environments

The real-time city is now real! The increasing deployment of sensors and hand-held electronics in recent years is allowing a new approach to the study of the built environment. The way we describe and understand cities is being radically transformed - alongside the tools we use to design them and impact on their physical structure.

Studying these changes from a critical point of view and anticipating them is the goal of the MIT SENSEable City Laboratory.

At the Conference I will discuss the lab's long-term Wiki City project, which is based on last year's Real Time Rome project.
Here, the lab used analysis of aggregate cell phone traffic to generate traffic maps. With cell-phone locations mapped anonymously at regular intervals, Roman city planners could see the pulse of the city in real time.

Wiki City builds upon this idea, envisioning an interactive, digitally augmented city that would also integrate location-based services, WiFi-based wikis and environmental sensor input. This with the ultimate aim of leading to an overall increased efficiency and sustainability in making use of the city environment.

Technical Plenary / **3**

The National Grid Initiative D-Grid

Author: Wolfgang Gentzsch[1]

[1] *D-Grid, Duke, e-IRG, e-School, Father&Husband, GFSG, PCAST, and RENCI*

In 2005, the German D-Grid Initiative started with 7 projects in the areas grid infrastructure, astronomy, climate research, high energy physics, engineering, medical research, and humanities. Recently, 12 additional 'community grid' projects kicked off.

In this presentation, we will present an overview on D-Grid and summarize our experience in creating the grid infrastructure and the grid application communities, and the challenges we are facing. We will conclude with a set of lessons learned and recommendations.

Technical Plenary / 4

A Grand Challenge for the Information Age

Author: Francine Berman[1]

[1] *San Diego Supercomputer Center*

The information Age has created a paradigm shift for research, education, commerce, and modern life. Current estimates are that in 2006, 161 exabytes (10^18 bytes) of digital data were created, and by 2010, the torrent of digital data generated will reach a zettabyte (10^21 bytes). Critical to the success of efforts in the Information Age is the ability to use, access, manage, and preserve digital data, and a reliance on the assumption that our most valuable digital data will be there when we need it. Data is the foundation of the Information Age, and preserving digital data for the foreseeable future is an emerging Grand Challenge.

Best practices in digital data reliability involve the replication of valuable data collections. From the first, grid technologies have been a natural fit for the challenge of preserving multiple copies of digital data collections, and today, data grids provide a rich platform for data stewardship, management, and preservation services.

The San Diego Supercomputer Center is a pioneering U.S. Center leading the development and provision of data grid technologies and cyberinfrastructure for managing, storing, preserving, and using digital data. Leveraging collaborations in the U.S. with the National Science Foundation, the Library of Congress, and the National Archives and Records Administration, SDSC is providing innovative leadership in the area of data Cyberinfrastructure and data preservation. In this talk, SDSC Director Fran Berman describes SDSC's approach to digital preservation, and discusses the next generation of opportunities and challenges for the data that drives the Information Age.

6

Large-scale ATLAS simulated data on EGEE

Author: Xavier Espinal[1]

Co-authors: Rodney Walker [2]; Simone Campana [3]

[1] *PIC/IFAE*

[2] *TRIUMF*

[3] *CERN*

Corresponding Author: espinal@pic.es

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

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 (center of mass energy). ATLAS collaboration is composed of about 2000 scientists spread around the world. The experiment requirements for next year is of about 1,2PB of storage and 26 MskI2k of CPU, and is relying on GRID philosophy and EGEE infrastructure. Simulated events are produced and distributed over EGEE by the Atlas production system.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

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) while the

ATLAS Distributed Data Management system (DDM) stays on a top level and takes care of the asymmetric file movement on top of the FTS services.

Services which are causing problems are basically the Storage Elements, as 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.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

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 O(10)Pb.
The processing and storage need a distributed share of resources, spread worldwide and interconnected with GRID technologies as the requirements from the LHC are has no precedents.
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.The GRID allow this distributed infrastructure

Technical Plenary / **7**

EGEE Grid Operations – 3 years of progress

Corresponding Author: ian.bird@cern.ch

The EGEE grid infrastructure has grown over the 42 months of EGEE and EGEE-II from around 40 sites to over 250. It is now in daily use on a significant scale with continuous workloads of over 100,000 jobs per day, and is acting as the primary source of computing and storage for several application communities. Although this growth has been impressive, with the LHC coming online less than a year from now the scale of resources and workload is still expected to increase dramatically. This talk will summarise the progress that has been made in building this production infrastructure, and discuss the major challenges still to be faced both in the near future and in the longer term.

Plenary / **8**

Welcome from the local organiers - Károly Molnár, Rector of BME

Plenary / 9

Istvan Hiller, Minister of Education and Culture

Plenary / 10

Norbert Kroo, Vice-President of the Hungarian Academy of Science

Plenary / 11

Akos Detreköi, President of NCCIT

Plenary / 12

Ulf Dahlsten, Director of "Emerging Technologies and Infrastructures - Applications", EU

Plenary / 13

Gold sponsor - Oracle: Peter Fuzes, Oracle Hungary Managing Director

Title: Grid Adoption and Research
The Grid has proven itself as a viable solution for the Industry and Oracle Grid technology has been widely adopted by a large number of customers both in commercial and scientific cutting-edge environments. Research in Grid and related technologies however continues to be crucial to the evolution of large distributed infrastructures, data-driven information systems and services.

The talk will give an overview of directions and latest developments of Oracle Grid technology, outlining key R&D projects Oracle is leading in this domain and describing a few examples in Hungary, EGEE'07 hosting country.

Plenary / 14

Bob Jones, EGEE-II Project Director

Plenary / 15

Logistics Information

Technical Plenary / 16

The gLite middleware

Corresponding Author: claudio.grandi@cern.ch

The gLite 3 middleware has been in production for about one year and a half. Since its first release many improvements have been made in terms of performance and usability. Concurrently the gLite software process has been improved and the benefits of the single component release process are realised. gLite is now undergoing a multi-stage restructuring process with the aim to make it more maintainable, manageable and portable. Older legacy components will only be maintained until their successors are considered ready. The development effort will be focused on components that have a strategy to interoperate with other Grid infrastructures and to adhere to international standards.

COD-14 (closed meeting) / 17

Working groups proposed schedule 1/2

Track and Update issues:
- Tools Improvements for COD – includes SAM WG, SAMAP WG and gstat WG
- Best Practices for COD – includes Operational Procedure Manual editorial comittee

COD-14 (closed meeting) / 18

Working groups proposed schedule 2/2

- CIC integration –includes third party and permanent guests.
- Failover Process.

EGEE Grid Operations (SA1) - Site Monitoring and Visualization / 19

Grid Services Monitoring Working Group

Corresponding Authors: james.casey@cern.ch, ian.neilson@cern.ch

The presenter will provide an overview the Grid Services Monitoring WG activities and and introduction to the following presentations.

EGEE Grid Operations (SA1) - Site Monitoring and Visualization / 20

Visualizing the State of the Grid with GridMaps

Author: Max Böhm[1]

Co-author: Rolf Kubli [2]

[1] *CERN & EDS*

[2] *EDS*

A new approach to visualization of complex monitoring data of the entire Grid.

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

GridMaps are based on a novel use of Treemap visualizations in the context of the Grid. Sites or services of the Grid are represented by rectangles of different size and color allowing 2 dimensions of data to be visualized simultaneously. This representation of monitoring data requires much less space than conventional sorted tables or bar charts. One application is visualizing Service Availability Monitoring (SAM) data from different VO and geographical perspectives. Switching between different views help discover correlations and associated patterns in the monitoring data.
A prototype has been implemented and will be demonstrated online.

EGEE Operations (SA1) / 21

Report on SLA progress

Corresponding Author: iliaboti@grnet.gr

EGEE Operations (SA1) / 22

Evolution of the role of the PPS

EGEE Operations (SA1) / 23

How to deal with the media; and how to write articles

Tips from NA2 on working with the press/media, and writing articles for general consumption.

COD-14 (closed meeting) / 24

CIC integration

Report on portal integration : FR/CE –30'
- Update on GOCDB and scheduled downtimes : UK/FR
Tools Improvements for COD and feedback from sites: CE – 20'
- Report on SAM/FCR/COD status : CERN/FR
- Report on SAMAP : CE
- Report on GSTAT : TW
Failover procedures - Global: IT –30'
BEst PraCtices for COD – Global COD session : DE-CH –10'
- Update for Operations Procedure: SWE
AOB

EGEE Operations (SA1) / 25

ISSeG: Site Security Training

Corresponding Author: d.j.b.jackson@rl.ac.uk

Site security is important to us all but how can it be made easier to implement? This training session will outline the approaches to Integrated Site Security (ISS) and the general recommendations that have been identified following practical implementation at two European Grid sites. The emphasis is on practical advice on how to link technical, administrative and training issues to improve overall site security independent of site size, VO type or your role within the site. The session is not solely focused on technical issues so users and managers / administrators will be able to take away some advice for their sites.

COD-14 (closed meeting) / 26

Tools Improvements for CODs

EGEE Operations (SA1) / 27

VO manager - ROC manager topics

Corresponding Authors: charles.loomis@cern.ch, frederic.schaer@cea.fr, helene.cordier@in2p3.fr, rumler@in2p3.fr

- VO registration procedure explained (25') (Frédéric Schaer and Rolf Rumler)
  - Description of the process, states of a VO, security aspects, relation to other fields
  - Specific problems and actions:
  - Main problems encountered during registration
  - Site problems during VO deployment: results of a study (Frédéric)
  - Sites abandonning support for a specific VO, how-to, consequences (Frédéric)
  - VO deregistration
  - User Support (short; see next presentation)
- VO specific and grid specific user support (25') (Torsten Antoni)

The aim of this session is to bring together VO and Grid experts to discuss user support issues. With the LHC ramping up to full production the number of user on the grid with little experience in grid

computing will increase drastically. This will of course also increase the workload on the support staff from the VOs and from operations. Therefore it is vital to have a well defined, reliable and well working user support infrastructure, including operations and all VOs. The roles of all participants to this infrastructure and their responsibilities have to be clear and an overall process description should be worked out. The roles and responsibilities for the VOs in this process of course vary with the size and structure of the individual VOs and does the possible technical implementation.
In this session we aim to revive the discussion of this topic which is vital for building a sustainable production quality infrastructure.
We invite all VO managers and user support expert to join us, so that their requirements and needs will be properly taken into account.

- Resource allocation (25') (Rolf Rumler)
  - Current procedure (workflow, assessment/utilisation)
  - Ideas for EGEE-III
  - Discussion

COD-14 (closed meeting) / 28

Best Practices Session - incl GGUS and Operations Manual -

COD-14 (closed meeting) / 29

Failover procedures and tools topic

EGEE Operations and gLite Middleware (SA1 & JRA1) / 30

Overview of service management issues (checklist?)

Corresponding Author: alessandra.forti@cern.ch

30' presentation + 1h discussion

EGEE Operations and gLite Middleware (SA1 & JRA1) / 31

Summary of progress with improving logging

Corresponding Authors: ljocha@ics.muni.cz, s.m.fisher@rl.ac.uk

15' presentation + 15' discussion

EGEE Operations and gLite Middleware (SA1 & JRA1) / 32

Handling signals in the job wrapper

Corresponding Authors: francesco.giacomini@cern.ch, marco.cecchi@cnaf.infn.it

15' presentation + 15' discussion

KnowARC: Early results of the KnowARC Project / 33

Session opening

Author: Farid Ould-Saada[None]

KnowARC: Early results of the KnowARC Project / 34

The new web-service oriented Hosting Environment Daemon (HED)

Author: Ferenc Szalai[None]

This session will try to show the new HED, its language bindings and sample services (AREX).

KnowARC: Early results of the KnowARC Project / 35

Grid run-time environment solutions in KnowARC

KnowARC: Early results of the KnowARC Project / 36

The ARC -> gLite, gLite -> ARC interoperation

Author: Christian Soetrupp[None]

Co-author: Peter Stefan

Corresponding Author: stefan@niif.hu

Interoperability results in KnowARC

KnowARC: Early results of the KnowARC Project / 37

break

KnowARC: Early results of the KnowARC Project / 38

Standards and interoperations - the KnowARC approach

Corresponding Author: balazs.konya@hep.lu.se

KnowARC: Early results of the KnowARC Project / 39

ARC - KnowARC discussion

KnowARC: Early results of the KnowARC Project / 40

break

41

Top-level Grid Services Monitoring Visualization

Author: Max Böhm[1]

Co-author: Rolf Kubli [2]

[1] *CERN & EDS*

[2] *EDS*

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

This activity is a part of the CERN openlab / EDS collaboration undertaken together with the CERN IT Grid Deployment group. The purpose of the joint project is carrying out research and development in the field of monitoring, management and operation of grid services, currently with a focus on monitoring.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

Due to its large and distributed nature, it is very difficult to know and understand the current state of the grid. Currently, a multitude of historically grown tools and agents are used to capture availability data. Existing monitoring tools are based on conventional sorted tables and bar charts, which do not easily provide quick and action oriented oversight and insight into job failures and availability patterns of grid services.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

Reliability of the Grid infrastructure is very important. Top-level management visualizations of grid services monitoring data, which help faster understanding problems and thereby faster being able to take corrective action, help also increasing the reliability of the grid.

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.:

In this demo we show a prototype of a new Top-level Grid Services Monitoring visualization, which helps better understanding the state of the grid from different VO and geographical perspectives as well as correlations of the monitoring data. The visualization uses a 2 dimensional graphical representation of the monitoring data based on treemaps requiring much less space than conventional table oriented views.

42

Using hardware tokens to improve grid security

Author: Jan Just Keijser[1]

[1] *NIKHEF*

Corresponding Author: janjust@nikhef.nl

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

Hardware tokens offer the opportunity to store a grid certificate in a tamper-free environment. Grid security can be improved by using these hardware tokens to allow for 2-factor authentication. At Nikhef we have started using Aladdin eTokens to store and generated X.509 grid certificates and SSH public/private key pairs, on the Linux, Windows and Mac OS X platforms. We have also written a package to generate grid proxies directly from the token on all of these platforms.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

At Nikhef we have started using Aladdin eTokens to store and generated X.509 grid certificates and SSH public/private key pairs, on the Linux, Windows and Mac OS X platforms. We have also written a package to generate grid proxies directly from the token on all of these platforms.
Certificate Authorities in the UK and in the Czech republic are also in the process of using hardware tokens for grid authentication.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

Hardware tokens offer the opportunity to store a grid certificate in a tamper-free environment. Grid security can be improved by using these hardware tokens to allow for 2-factor authentication. By 2-factor authentication we mean that authentication is based on 2 things, e.g. what you know (a password) and what you possess (a hardware token). Hardware tokens offer a secure and tamper-free environment on which a grid certificate can be stored or, better yet, generated. The private key of such a certificate can never be copied off the token, making it an ideal place to store security-sensitive information. These tokens can be used for storing personal grid certificates, SSH public/private key pairs but also robot certificates for use by Grid Portal sites.

43

Service Availability Monitor for the LHC experiments

Authors: Alessandro Di Girolamo[1]; Andrea Sciaba'[1]; Elisa Lanciotti[1]; Nicolo Magini[1]; Patricia Mendez[1]; Roberto Santinelli[1]; Simone Campana[1]; Vincenzo Miccio[1]

[1] *CERN*

Corresponding Author: alessandro.di.girolamo@cern.ch

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

The four LHC experiment ALICE, ATLAS, CMS and LHCb depend on the EGEE grid to perform their scientific programme and all their computing activities preparing for the LHC collider startup (2008). The Service Availability Monitor (SAM) has been developed by SA1 primarily to test the EGEE infrastructure and to collect and to maintain the corresponding information.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

The integration activity has been slightly different for the four experiments, and we report on the following examples for each ones:
ALICE: VOBOX testing and installation
ATLAS and CMS: validation of the software installation and Storage Resource Manager (SRM) low-level testing
CMS: Calibration data base and local storage access from the worker nodes.
LHCb: software installation and validation integrated in the Dirac production and analysis system

We will report on the experiences from an application point of view of this integration, and on the improvements from the point of view of the operation of such large systems.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

The four LHC experiments rely on the EGEE infrastructure to perform their simulation, reconstruction, analysis activities. These large-scale activities require a stable environment, not only for the fundamental services, like the storage services, but also for experiment specific services like the Software and Data Distribution services.
The integration between the Service Availability Monitoring and the applications specific frameworks is essential to achieve high efficiency in large-scale activities and to provide dependable services for large users community. In fact the SAM system is widely used in the EGEE operations to identify malfunctions in grid services but it can be adapted to perform the same function on experiment-specific services. This tool is also capable to accommodate application dependent tests, which is of great interest for large-scale applications relying on the grid and on application specific services.

44

The Stellar Spectra Modeling as an example of data- and compute-intensive application running on the BalticGrid Project testbed

Authors: Bartek Palak[1]; Grazina Tautvaisiene[2]; Sarunas Mikolaitis[2]

[1] *Poznań Supercomputing and Networking Center*

[2] *Institute of Theoretical Physics and Astronomy of Vilnius University*

Corresponding Authors: bartek@man.poznan.pl, sarunas@itpa.lt

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

We present The Stellar Spectra Modeling package SYNTSPEC – the gridified tool for stellar spectra analysis – as an example of data- and compute-intensive application running on the testbed of the EU BalticGrid Project (http://www.balticgrid.org) – interoperable with EGEE resources and complementary with the EGEE infrastructure. The application brings the new quality to the research in the field of astrophysics in the Baltic States and accelerates the integration of science in EU.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

The SYNTSPEC is a very good example of an application that benefits usage of the BalticGrid testbed because of the need of powerful computing resources and high network throughput.
The application calculates normalized to the continuum stellar spectra that serve for determinations of e.g. chemical composition, effective temperatures and surface gravities of stars. Initially, it was used for modeling of small spectral regions, however the requirements increase rapidly, so the possibility of using grid infrastructure is a crucial way to accumulate enough resources for the analysis of massive data. This is very important in a preparation of infrastructure and procedures for analysis of large quantities of spectra that will be produced by the European Space Agency's GAIA space observatory to be launched in 2011.
The application for its submission, execution, monitoring and visualization of results exploits such EGEE services and components like WMS, LFC, CE, L&B, UI, SE, WN.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

The BalticGrid project is of high strategic importance for the Baltic States. The primary goal of the BalticGrid project is to establish a production-level, interoperable and complementary with the EGEE grid infrastructure, which enables scientists in the Baltic States to access critical resources, to form effective research collaborations and share efficiently unique instruments and data. It also rapidly increases the knowledge and use of grids in the Baltic States.
The usage of computing and networking grid infrastructure of the project has a significant impact on scientific communities, validated through everyday work of the scientists studying the Baltic Sea environment, bioinformatics, high energy physic, astrophysics and other sciences.
The real life example could be the SYNTSPEC application, which thanks to usage of project resources is able to calculate synthetic stellar spectra of significant wavelenght ranges and serves for the galactic and stellar research studies.

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.:

We intend to show a live-demo of the SYNTSPEC lifecycle, starting from defining of input files and computing parameters, submission of the job to the BalticGrid testbed, job monitoring, and ending with visualization of results. We will demonstrate how stellar spectra change because of the resolving power, because of stellar rotation and other parameters. Very important is a visual aspect of the presentation. The application will be run within "The Migrating Desktop" (MD) – advanced graphical user interface similar to a window-based operating system that hides the complexity of the grid middleware and makes access to the grid resources easy and transparent. We believe that the interesting scientific content of our demo combined with intuitive interface will attract the audience's attention. We will also present posters about the BalticGrid Project and MD, as well as several types of broshures. A single screen and a network connection is needed to present the live-demo.

45

Mathcell.Ru: Integrated Mathematical Model of Living Cell in GRID infrastructure

Author: Victor Lakhno[1]

Co-authors: Aleksandr Zaitsev [1]; Alexander Teplukhin [1]; Dmitriy Ustinin [1]; Gayane Tyulbasheva [1]; Mikhail Ustinin [1]; Nadezhda Fialko [1]; Nafisa Nazipova [1]; Sergey Filippov [1]

[1] *IMPB RAS, Pushchino, Russia*

Corresponding Author: nnn@impb.ru

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

Mathematical modeling of a living cell is a great challenge for modern science. Such modeling will allow ones to solve a number of practical problems, for example, acceleration of development of novel drugs and of prediction of their direct and mediated action, development of thin biochemical agents which will influence on metabolic reactions in the organism. Besides such model will be useful for development various nanostructures and nanomaterials.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

The MathCell portal includes interactive 3D cell model, encyclopedia on mathematical modeling of cell and software for modeling of some basic processes in living cell. Within the limits of the Project the interactive environment was developed, which allows to perform calculations of mathematical models in GRID infrastructure. The special Job Maintenance System was developed which automatically allows User Logging and Accounting, Job Submission, Job Status Monitoring, Job Queuing, Results Obtaining. At the present three models are deployed in GRID infrastructure:
- software for mathematical modeling of electron transfer in DNA molecule;
- simulation model of electron transfer on inner photosynthetic membrane in chloroplasts;
- software for calculation of dissolution energy of biomolecules in water by Monte Carlo method.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

The creation of integrated mathematical model of eukaryotic cell was the aim of the Mathematical Cell (MathCell) project (http://www.mathcell.ru) realized in the Institute of Mathematical Problems of Biology RAS during the EGEE Projects. Further development of the MathCell project is closely associated with advance of Grid infrastructure, it implies integration of individual components of the model into a program system which would simulate cell processes at different levels – from microscopic to macroscopic scales and from picoseconds to the cell lifetimes.

46

EGEE and Business

Business Track / 47

NESSI is about transforming the EU economy through Service Oriented business models

Businesses concentrate on activities where they can gain a competitive advantage. The main focus of NESSI is that of service. All definitions of service are based on the same principle: a service consumer does not own the service and therefore does not need to be concerned with all the aspects generally associated with ownership such as infrastructure, technology, integration and maintenance. What count are the business needs that are waiting to be met by the appropriate service.

Technology must be an enabler of service rather than an inhibitor. Enlargement of the European Union and the changing nature of trade with the rest of the world puts a strong emphasis on a multicultural approach to business. Furthermore, the tangible and intangible assets resulting from EU-funded R&D projects must reach the industry and be used by them.

NESSI can connect eInfrastructures with industrial users. Companies can contribute to the sustainability of eInfrastructures in different ways and use them in order to improve performance, to gain efficiency, effectiveness and competitive advantage.

eInfrastructures could be part of the ecosystem that NESSI is determined to create. The resulting self-sustained network could operate in a cost-effective way producing even profit and following a kaizen approach without the umbrella of the European Commission.

Business Track / 48

An update on HP's Tycoon at CERN: integration with EGEE

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

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.
One of our problems will be to keep intact 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 leak nor modify security constraints.

Business Track / 49

Opportunities for Technology Transfer in Grid Computing

The rapid development of virtualisation and distributed computing technologies, including grids, means that the paradigm shift towards utility computing could happen sooner rather than later. Scientific Grids such as WLCG could contribute to this process. gLite middleware is a good example: developed to work with geographically distributed heterogeneous resources which often belong to different owners it fits perfectly with the requirements of future inter-enterprise Grids. The Science and Technology Facilities Council (STFC) supports several Grid commercialisation projects that showcase market opportunities for Grid technologies.

Business Track / 50

Grids in Industry: Lost in Transition?

The journey from proof-of-concepts and operational Grid research infrastructures to credible Enterprise technologies has significant technical, economical and cultural challenges. For the last 10 years IT Innovation has been working with industry and commerce to address these challenges from the early days of meta-computing experiments to the Grid and now Service-Oriented Infrastructures.

This presentation revisits the evolving vision and proposition presented by the Grid research community to industry over the last decade. The barriers for adoption of successful operational research Grids in commercial context are presented and success stories that show how these barriers can be overcome.

Business Track / 51

Implementation of an industrial-strength pharmaceutical workflow into a Grid environment

The goal of the SIMDAT Pharma application activity is to demonstrate the usefulness of Grid technology in the area of life sciences. We implement an industrial strength pharmaceutical workflow across a Grid test-bed comprised of both academic and industrial partners. Naturally, these different types of partners are operating their Grid node with different security and quality of services policies and obviously have different commercial interests.

The existing Grid prototype employs all infrastructure elements required to establish trusted relationships between academic and industrial partners. The infrastructure allows an academic service provider to accept outsourcing of high value applications or databases from a research group in order to give controlled and managed access to this application for other partners, both academic and industrial. The underlying model guarantees the preservation of intellectual property rights among the partners.

In the presentation we will concentrate on the Grid-specific developments of the current Pharma prototype and discuss the impact of Grid technology in the life science sector.

EELA: E-Infrastructure shared between Europe and Latin America / 52

Project overview

Corresponding Author: marechal@if.ufrj.br

EELA: E-Infrastructure shared between Europe and Latin America / 53

EELA Grid infrastructure

Corresponding Author: roberto.barbera@ct.infn.it

EELA: E-Infrastructure shared between Europe and Latin America / 54

e-Science Applications in the EELA Grid

Corresponding Author: rafael.mayo@ciemat.es

EELA: E-Infrastructure shared between Europe and Latin America / 55

Dissemination and training activities

Corresponding Author: leandro.ciuffo@ct.infn.it

This talk will present the dissemination activities done by the EELA project highlighting its collaboration with EGEE.

Business Track / 56

Inter-Enterprise CAE Design with Secure Controlled Sharing of Data and Analysis Services

The use of CAE analysis within the design process is common place in the Aerospace industry but is very focused on using a limited set of capabilities within the boundaries of the enterprise. To enable wider access to physical modelling and to enable the aerospace industry to take advantage of more cost effective provisioning models there is a requirement to enable the exploitation of CAE analysis services that are supplied beyond the enterprise boundaries. This introduces challenges in terms of managing Intellectual Property especially as this is encapsulated in the process in which the analysis codes are applied hence a focus on workflows, data and associated embedded knowledge.

This talk will focus on what has been achieved within the Aerospace activity of SIMDAT and present some of the issues addressed with the Aerospace sector.

Business Track / 57

SIMDAT Grid-based Solutions for the Virtual Product Design Process in the Automotive Industry and Lessons Learned.

The growing competition in the automotive industry requires continuous reduction of development and innovation cycles. On the other hand the demands on quality, safety and comfort are increasing. During the past years, advances in the area of CAD (Computer Aided Design), CAE (Computer Aided Engineering) and CAT (Computer Aided Testing) technologies and processes have contributed significantly to the ability of the automotive industry to keep up with these requirements.

Today, the data resulting from the CAD-, CAE-, and CAT-processes are stored in separate databases without common interfaces. In addition, the development teams often work at distributed locations. It is due to the ongoing transition from many small manufacturers to a few large conglomerates. This transition requires the integration of previously separate teams into single virtual development teams. They need to work on common car platforms to reduce development costs. SIMDAT aims at to provide solutions for these problems.

The presentation will give an overview of industrial use cases SIMDAT project deals with in the automotive area and of solutions intended to ease the collaboration between engineering organizations and departments based on Grid Technologies employed and extended in the SIMDAT project.

Business Track / 58

EGEE & Platform Technical Collaboration: Milestones & Challenges

Yearlong working relations between EGEE consortium members and Platform Computing lead to systematic collaboration since 2005, officially manifested as an EGEE Business Associate at EGEE'06. EGEE as the worldwide leading research Grid and Platform Computing as the leading provider of commercial Grid software underwent intensive joint investigations in order to achieve best possible performance for HEP and other communities job profiles and requirements.

This presentation will highlight the joint success story from the original aim to "learn, how to better exploit LSF", to the most recent developments and future plans.

Business Track / 59

Avanade Grid achievements and EGEE collaboration benefits

Avanade solution offering includes a proprietary Grid environment, based on .NET but supporting other platforms. From the project launch, three years ago, we built various projects on the platform, including mission critical computations and heavy load financial computations. On this speech, Avanade will present a short overview of his achievements on the Grid space, business requirements to further evolve of the platform, and our vision about benefits and opportunity to leverage gLite standards in future releases and actively collaborate with EGEE.

Business Track / 60

Industrial Applications of Grids

For Philips Research an IT infrastructure geared towards the needs of researchers is an indispensable element. In recent years Philips Research has undergone a number of major changes. As a result of the focus shift at Philips, Philips Research has also changed its programme considerably. The adoption of the open innovation concept has resulted in an open organisation that enjoys strategic cooperation with companies, universities and other institutes. Facilities are shared with partners in order to improve cost-effectiveness and efficiency.

The Research ICT Department has modified its strategy so as to facilitate Research's new wishes. An example of this is the expected increase in High Performance Computing in application areas such as medical imaging, bio-informatics and simulation of physical devices. This presentation deals with the above-mentioned changes, the way we are responding to them and the need for intensive support.

Business Track / 61

NICE EnginFrame and A-WARE - An easy Way to Access GRID Resources

While Grids can dramatically innovative business models and marketplaces, for end-users the term "Grid" is frequently associated with highly complex infrastructures dissimilar to the everyday tools and technologies they are
used to. The A-WARE project tackles this problem and develops a stable, supported, commercially exploitable, high quality technology able to give easy access to Grid resources.

A-WARE technology will merge two complementary experiences, capable of easing the complexity of Grid technology and complex job management. A-WARE will enhance mature European Grid middleware technologies – through workflows – and a deliver improvements of accessibility, visibility - through Web 2.0 technology.

The A-WARE project is based on the EnginFrame Grid portal framework developed by NICE and adopted by a number of important companies worldwide. The EnginFrame Grid portal, delivers user-friendly, highly customisable access to Grid-enabled applications and infrastructures. It completes the Grid solution by increasing usability and user-friendliness, without sacrificing flexibility and control for the most advanced computing scenarios, like Oil & Gas, Automotive, Aerospace, HEP, Electronics, Telecommunications etc. EnginFrame is also the core technology of the GENIUS portal adopted by EGEE.

Technical Network Liaison Committee - Closed / **62**

Operational interface status

Author: Mathieu GOUTELLE[1]

[1] *CNRS/UREC*

Corresponding Author: mathieu.goutelle@urec.cnrs.fr

This session will give an update on the ENOC and the Operational Interface that have been built between EGEE and the NRENs.

Technical Network Liaison Committee - Closed / **63**

Network Service Level Agreement (SLA) Implementation

Authors: Maria Grammatikou[None]; Vasiliki Pouli[None]

Co-author: Chrysostomos TZIOUVARAS [1]

[1] *GRNET*

Corresponding Authors: tziou@grnet.gr, vpouli@netmode.ntua.gr, mary@netmode.ntua.gr

Update on the use of advanced network services by the EGEE community

Technical Network Liaison Committee - Closed / **64**

Network trouble ticket standardisation for exchange

This talk will present the recent progress on the field on network Trouble ticket standardisation in order to exchange these operational data more easily.

Business Track / **65**

gLite industry adoption barriers

Industrial adoption of any Grid middleware depends on its ability to satisfy the needs of many different communities, starting from technology providers through developers to the end users. For this reason the EGEE Industry Forum Steering Committee is starting a special focus group on gLite adoption. This group will focus on analysis of the technical barriers to gLite adoption in industry and provide maximum feedback to developers in order to make the gLite middleware fit industrial needs. This presentation is in part an invitation to start a discussion on these topics and lists a set of potential issues taken from previous discussions and presentations.

66

BioinfoGRID: Bioinformatics Grid Application for life science / **67**

BioinfoGRID Project Introduction

Author: Luciano Milanesi[1]

[1] *National Research Council - Institute of Biomedical Technologies*

Corresponding Author: luciano.milanesi@itb.cnr.it

BioinfoGRID: Bioinformatics Grid Application for life science / **68**

GRID computing applications in genomics

Corresponding Author: s.suhai@dkfz-heidelberg.de

BioinfoGRID: Bioinformatics Grid Application for life science / **69**

Transcriptomics and Phylogenetics Applications in GRID

Author: Pietro Liò[None]

Corresponding Author: pl219@cam.ac.uk

BioinfoGRID: Bioinformatics Grid Application for life science / **70**

Databases and Applications

Corresponding Author: gabriele.trombetti@itb.cnr.it

BioinfoGRID: Bioinformatics Grid Application for life science / **71**

Comparative evaluation of tools providing access to different types of data resources exposed on the Grid

Corresponding Author: giacinto.donvito@ba.infn.it

BioinfoGRID: Bioinformatics Grid Application for life science / **72**

Molecular Dynamics Applications - WISDOM

Corresponding Author: kasam@clermont.in2p3.fr

73

BioinfoGRID summary and future prospective

Corresponding Author: luciano.milanesi@itb.cnr.it

Security Policy and Vulnerability Handling / **74**

Grid Security Vulnerability Group

Author: Linda Ann Cornwall[1]

[1] *RAL*

Security Policy and Vulnerability Handling / **75**

Joint Security Policy Group

Author: David Kelsey[1]

[1] *RAL*

Technical Plenary / **76**

Current Use of the EGEE Infrastructure and Evolution of NA4 into EGEE-III

Corresponding Author: charles.loomis@cern.ch

This presentation will summarize the current use of the EGEE production infrastructure by the EGEE user community and provide an overview of the functionalities required, both present and future, by the users' applications. The talk will include an overview of the services currently provided by NA4 to the user community and discuss how the provision and organization of those services will evolve into EGEE-III.

int.eu.grid: Support to MPI and interactivity on glite infrastructures / **77**

Introduction and Session overview

78

Structuring Medical Images with the Grid

Authors: Damian Segrelles[1]; Ignacio Blanquer[1]; Vicente Hernández[1]

[1] *UPV*

Corresponding Author: iblanque@dsic.upv.es

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.:

The most challenging issue in the project has been deployment. The current infrastructure requires dealing with the restrictions of private networks and firewalls of modern hospitals. GridFTP protocols, for example, needed to be carefully configured with port maps to enable the communication outside from the hospitals.

The key services in the infrastructure, inherited from EGEE, have been the authorisation service and the file transfer. The authorisation service has been extended to deal with encryption and multiple CAs to guarantee the security of multiple administrative domains. File transfer is crucial when exchanging large data files (on the order of several tens up to few hundreds of Megabytes) in short time.

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

Sharing and organising medical imaging knowledge is a key issue in medical research and training. Evidence-based medicine is also demanding high-quality well-organised knowledge bases to check for second opinion and drive diagnosis.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

The first advantage lies on the organisation of data. The availability of a Grid platform to securely share cases will enable increasing the significance of the studies, through the enlargement of the study sample, and the support to learning through representative cases. Currently, data is organised by administrative and demographic keys, which prevents from searching for specific diagnosis.

In order to implement this functionality, an authorisation architecture has been implemented to define workgroups, case access permission and relations. Encryption and key share management is necessary to prevent from unauthorised access from users with administrative privileges.

The relevant studies should be explicitly selected and carefully documented through the structured report. In order to ease this process, advanced post-processing tools are provided in the form of WSRF services. This Grid services provide the users with the access to advanced algorithms and computer resources.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

However, sharing and organising medical imaging data is not straightforward. Technological and legal problems on exchanging data make it difficult or even impossible with the current infrastructures. On the other side, the index criteria used in clinical practice are inefficient when searching for knowledge.

CVIMO (Valencian Cyber-infrastructure for Medical Imaging in Oncology) is a platform developed to share and organise medical studies and reports based on ontologies constructed upon the fields of structured reports. It is based on a Grid Software Architecture of WSRF services that organise coding, access rights and data location for different studies and reports.

This platform enables the users to submit new cases which are automatically organised according to the semantic criteria defined through the Virtual Organisations and groups. The platform provides a virtual data catalogue based on the metadata coming from the evaluation report of the radiologists.

Operational Security Coordination Team Training / 79

Introduction: Grid and security

Author: Remi Mollon[1]

[1] *CERN*

Corresponding Author: remi.mollon@cern.ch

Operational Security Coordination Team Training / 80

Grid systems installation and configuration

Author: Louis PONCET[1]

[1] *CERN*

Corresponding Author: louis.poncet@cern.ch

Operational Security Coordination Team Training / 81

Centralised logging

Author: Eddie Aronovich[1]

[1] *Tel-Aviv University*

Corresponding Author: eddie.aronovich@cs.tau.ac.il

82

Discussion / Questions

Author: Romain Wartel[1]

[1] *CERN*

Operational Security Coordination Team Training / 83

Discussion / Questions

Author: Romain Wartel[1]

[1] *CERN*

Corresponding Author: romain.wartel@cern.ch

Operational Security Coordination Team Training / 84

Protecting administrative credentials

Author: Mingchao Ma[1]

[1] *Unknown*

Operational Security Coordination Team Training / 85

Testing and monitoring Grid systems

Author: Michal Prochazka[1]

[1] *CESNET*

Operational Security Coordination Team Training / 86

Incident response (policies and procedures)

Author: Carlos Fuentes Bermejo[1]

[1] *Unknown*

Corresponding Author: carlos.fuentes@rediris.es

Operational Security Coordination Team Training / 87

Discussion / Questions

Author: Romain Wartel[1]

[1] *CERN*

Corresponding Author: romain.wartel@cern.ch

Business Track / 88

EGEE's vision and roadmap for commercial uptake of Grids

EGEE (Enabling Grids for E-sciencE) is Europe's flagship Grid infrastructure project now in its second 2-year phase. The EGEE Grid now combines the processing power of more than 30.000 CPUs, a storage capacity of about 5 million GB on a global network established by GEANT and its extensions connecting more than 200 sites in 48 countries. This Grid is already providing an essential daily service to more than 100 applications in such areas as astrophysics, medical imaging, bioinformatics, climate studies, oil and gas exploration, pharmaceutical research, and financial forecasting. It's now the world's largest general-purpose scientific computing Grid, and is getting bigger every month.

Envisaged as an infrastructure for the support of scientific research, EGEE has attracted interest from numerous business sectors that see it as either a service that can be used as a virtual computer centre, as a means of selling their services and products to the diverse communities it supports, or as an example of how to build partner grids with their supply chains.

This talk will give an overview of the range of businesses with which it is working, explain the mechanisms it places to allow commercial organisations to get involved and suggest some directions in which it could develop on a time-scale of 2 to 5 years.

89

Statistical analysis of the workload and modeling of the time a job spends at different states in the EGEE environment

Author: Konstantinos Christodoulopoulos[1]

Co-author: Emmanuel Varvarigos [1]

[1] *University of Patras*

Corresponding Author: kchristodou@ceid.upatras.gr

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

In this study we analyze the workload of LCG/EGEE production Grid and the time the job spends at different states. The probabilistic models obtained for the job submission process and the delay components introduced at different stages is important for modeling production Grids such as the EGEE.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

Our results indicate that if we consider the LCG/EGEE Grid as the level of our observation, jobs are submitted without any specific weekly or daily patterns. The job interarrivals match very well with a rounded exponential distribution. We define and model four delay components that comprise the overall job processing. We observe that the total time of a job in the EGEE Grid is dominated by the CE's Register and Queuing time (D3) and the WN's Execution time (D4), while the other two components, namely, the time spent in the Pending, Submitted and Waiting states (D1), and the time in the Ready state (D2), are small.
We also evaluate the efficiency of the EGEE Grid and (indirectly) of the employed super-scheduling algorithm by comparing the total delay experienced by a job with that of a hypothetical ideal super-cluster. We found that we would have similar performance if we submitted the same workload to a super-cluster having 34% of the total average number of CPUs participating in EGEE

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

We have considered the overall LCG/EGEE infrastructure as a single entity and observed the general properties of job submission and execution, by collecting and analyzing the daily reports of the Real Time Monitor tool. Based on the EGEE job flow diagram we distinguish four delay components of the job processing, each corresponding to the time spent at different states in the EGEE environment. We analyze and model each delay component separately and we also model the job arrival process. The existence of good probabilistic models is important for the improved understanding of the Grid Computing concept, the prediction of their performance, the design of the middleware they use, and the evaluation of new scheduling and quality of service policies. We also examine the efficiency of the EGEE environment and (indirectly) of the employed super-scheduling algorithm.

int.eu.grid: Support to MPI and interactivity on glite infrastructures / 90

Support to MPI

int.eu.grid: Support to MPI and interactivity on glite infrastructures / 91

How to run MPI applications

int.eu.grid: Support to MPI and interactivity on glite infrastructures / 92

Resources for MPI Jobs: an interoperable framework

int.eu.grid: Support to MPI and interactivity on glite infrastructures / 93

Interactivity and fast job allocation at the level of Resource Brokers

int.eu.grid: Support to MPI and interactivity on glite infrastructures / 94

Visualization on the Grid

int.eu.grid: Support to MPI and interactivity on glite infrastructures / 95

Using the Migrating Desktop

int.eu.grid: Support to MPI and interactivity on glite infrastructures / 96

Pannel: interoperability, operation and business models

int.eu.grid: Support to MPI and interactivity on glite infrastructures / 97

Discussion Panel: feeback from site managers

98

Grid based telemedecine application for Monte Carlo dosimetric studies

Authors: Lydia Maigne[1]; Matteo Diarena[1]; Simon Nowak[1]

Co-author: Vincent Breton [1]

[1] *CNRS IN2P3 LPC*

Corresponding Authors: diarena@clermont.in2p3.fr, maigne@clermont.in2p3.fr, novak@clermont.in2p3.fr

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

Nowadays, grid technology offers new possibilities in terms of storage and data sharing capabilities. To overcome medical data heaviness and confidentiality, new tools have been developed enabling the secure storage and sharing of medical information. Based on EGEE middleware (gLite), our application is a medical information system allowing physicians to easily exploit grid capabilities to store and securely share medical data and images, coming from hospital's PACS systems.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

The Medical Information Platform (MIP) is mainly based on Web Service technology that implements, through gLite Grid Services, the interaction with various EGEE Grid components. This platform creates an abstraction layer between users and the Grid tacking care of all the processes involving interaction with Grid or sensitive information management and sharing. MIP users access the platform using a web portal developed on top of GridSphere Portal Framework taking advantage from the flexibility and modularity offered by portlet technology. Patients information are stored in an AMGA server installed in each hospital while the images are stored encrypted in Grid SEs. Information privacy and security constraints avoid replication of information between AMGA servers installed in different hospitals. For this reason we implemented on-request data exchange, between users and the platform, uses SOAP protocol over an SSL secured connection.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

Today medical informations are digitalized and this opens new opportunities for physicians in terms of information exchange but the resources needed inside hospitals to manage these medical images is a growing issue. Using the EGEE middleware and Web Service technology we developed a distributed medical platform that allows physicians to manage and exchange patient information (clinical information and medical images) securely between different hospitals through the EGEE Grid. In a near future hospitals will find in the Grid the solution to store and manage their medical data. Moreover we also plan to use grid computational resources to run Monte Carlo simulations on medical images to estimate "a priori" the radiation dosimetry in cancer treatment. This is a very complex and CPU intensive job that can be parallelized over several grid nodes to reduce computation time.

DILIGENT - Dynamic Digital Libraries and Virtual Research Environments on the Grid / 99

A Service-Composed Information Retrieval Engine

Author: George Kakaletris[1]

[1] *University of Athens*

Corresponding Author: gkakas@di.uoa.gr

This talk presents the gCube Search Framework that is the orchestrator of the Information Retrieval process, distributed in its nature. It consolidates query and environment information, prepares and plans retrieval execution and ultimately offers the gluing elements that bring together all the independent components into a meaningful and performant collaboration.
The gCube Search Framework supports Full Text Search, Content Based Search, Geospatial Search, Search over semi structured (XML) and structured (tabular) data; it is totally independent from underlying schemas and organisation of information; it exploits advanced Distributed Information Retrieval and Indexing facilities provided by external services; it builds in execution planning and optimization avoiding misutilization and overloading failures; and finally it handles efficiently the plurality and dynamicity of resources available on a gCube-based grid infrastructure.

EGEE and Collaborating Projects / **100**

Technical Collaboration with Related Projects

Corresponding Author: erwin.laure@cern.ch

The EGEE Technical Director will give an overview of the technical aspects of EGEE that are of interest to other projects.

EGEE and Collaborating Projects / **101**

Training Synergies with Collaborating Projects

Corresponding Author: sipos@sztaki.hu

The leader of the EGEE Training and Induction ctivity (NA3) will present how Collaborating Projects can work together with EGEE to enhance their training initiatives and leverage common ground.

EGEE and Collaborating Projects / **102**

The EGEE Ecosystem and FP7

Corresponding Author: louridas@grnet.gr

The leader of the EGEE Policy and International Cooperation Activity (NA5) will present the current outlook of the EGEE Ecosystem in FP7.

EGEE and Collaborating Projects / **103**

Panel Q & A

Corresponding Author: frederic.hemmer@cern.ch

The participants will be able to field their questions to high-level EGEE representatives (Activity Leaders or Deputy Activity Leaders) from activities such as SA1 (Grid Operations, Support and Management), SA2 (Networking Support), JRA1 (Middleware Reengineering), NA2 (Dissemination, Out-

reach and Communication), NA3 (Training and Induction), NA4 (Applications, Identification and Support), NA5 (Policy and International Cooperation).

DILIGENT - Dynamic Digital Libraries and Virtual Research Environments on the Grid / 104

Welcome

Author: Donatella Castelli[1]

[1] *CNR - ISTI*

105

DILIGENT: Managing Virtual Research Environments on the Grid

Author: Pasquale Pagano[1]

[1] *CNR-ISTI*

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

The DILIGENT project (http://www.diligentproject.org) provides an infrastructure for supporting the creation of on-demand, transient Digital Libraries (DL). It relies on the EGEE middleware and infrastructure to deliver advanced DLs functionality to two user communities: ImpECt from the environmental e-science domain and ARTE from the cultural heritage domain.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

The gCube system offers a wide range of DL services. Some of these services, due to their characteristics, are specially tailored to be executed in a computing infrastructure as the one provided by EGEE. Several DILIGENT applications like watermarking, feature extraction and image processing can largely profit from the distributed computation handled by gLite. One concrete example is the currently ongoing experimentation with feature extraction on images. This experimentation started in the third week of July and was able to process so far around 25 million images, something completely impossible without the exploitation of the EGEE PPS infrastructure.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

The DILIGENT vision has been put in place with the implementation of gCube. The gCube software reflects the three-dimensional Grid vision of resources sharing (computation resources, data/information and application services), and implements a SOA over a Grid middleware. gCube adopts the WS-Resource Framework, WS-Addressing, WS-Notification, and WS-Security set of conventions and relies on the EGEE project middleware (gLite) and infrastructure. By merging a service oriented approach with Grid technology, gCube exploits the advantages of both.
The demand for DL has steadily increased during the last years. Distributed communities and organizations, regardless domain, need to share their multi-type data and information, collaborate and produce knowledge by processing data, in a secure and efficient manner. DILIGENT provides an infrastructure that addresses these scenarios providing Virtual Research Environments (VRE) to these communities.

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a

successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.:

The proposed demonstration will illustrate the DILIGENT infrastructure and the capabilities behind its configuration and management. This includes the creation, monitoring and management of VREs and the users and resources (computers, services, running instances, collections, gLite resources, etc.) associated to it.
Several implemented gCube DL services will be demonstrated. It will be explained how these novel functionality for autonomic management, controlled sharing, storage, indexing, etc., has been developed by exploiting the WS-*/WSRF specifications, the gLite middleware and the EGEE infrastructure.
From the DILIGENT user communities point of view, the functionality will be exemplified by demonstrating two VRE: one to support courses and teaching activities in the culture-heritage domain and another to support the automatic generation of reports in the environmental sector. This will highlight the added value of the DL services being exploited.

DILIGENT - Dynamic Digital Libraries and Virtual Research Environments on the Grid / 106

gCube overview: exploiting the grid to create and manage VREs

Author: Pasquale Pagano[1]

[1] *CNR - ISTI*

Corresponding Author: pasquale.pagano@isti.cnr.it

A brief overview of the gCube framework is provided. In particular, this talk presents the way it embeds and exploits gLite by offering a feature full platform for distributed hosting and management of web services; storage, retrieval, and access of raw data and information; and a framework for extending state-of-the-art indexing, selection, fusion, extraction, description, annotation, transformation, and presentation of digital information.
Finally, it presents how gCube eliminates manual deployment overheads, guarantees automatic placement of services within the infrastructure and opens opportunities for outsourcing state-of-the-art implementations.

DILIGENT - Dynamic Digital Libraries and Virtual Research Environments on the Grid / 107

Identity management issues in the gCube framework

Author: Paolo Roccetti[1]

[1] *Engineering Ingegneria Informatica*

In Virtual Research Environments, the management of service identities must face with issues related to the dynamic services deployment in the infrastructure. This talk briefly presents how these issues have been solved in the gCube framework.

DILIGENT - Dynamic Digital Libraries and Virtual Research Environments on the Grid / 108

Content and Storage Management in gCube

Author: Heiko Schuldt[1]

Co-author: Pasquale Pagano [2]

[1] *Universität Basel Switzerland*

[2] *CNR-ISTI*

This talk introduces the gCube Content and Storage Management that provides means for persistently storing and physically structuring any content in a Grid-enabled VRE. In addition, it presents novel approach to advanced replication in the Grid that allows for increased availability of content in a VRE.

DILIGENT - Dynamic Digital Libraries and Virtual Research Environments on the Grid / 109

Process Management in gCube: Supporting Distributed Grid Applications by Service Composition

Author: Heiko Schuldt[1]

Co-author: Evangelos Floros [2]

[1] *Universität Basel Switzerland*

[2] *Unknown*

This talk presents gCube's approach to define, run, and optimize complex process-based applications n the Grid. Following the idea of programming-in-the-large, gCube first supports a powerful graphical user interface for the composition of both Grid services and jobs into processes. Second, gCube provides a fully distributed approach to process (workflow) management that avoids any single point of failure and allows to dynamically adjust to changing environments. Third, it includes support for structural process modifications to maximise parallelism during process optimization.

EGEE and Collaborating Projects / 110

Networking with Collaborating Projects

Corresponding Author: sarah.purcell@cern.ch

The Dissemination, Outreach and Communication (NA2) Activity Leader will present to the participants EGEE's networking efforts, synergies in outreach, what EGEE can do for other projects and what other projects can do for EGEE.

DILIGENT - Dynamic Digital Libraries and Virtual Research Environments on the Grid / 111

The DILIGENT infrastructure: running gCube and gLite

Author: Pedro Andrade[1]

[1] *CERN*

Corresponding Author: pedro.andrade@cern.ch

This talk gives on overview of the DILIGENT infrastructure highlighting its different constituents. It is explained how gCube and gLite are made available to enable the creation of Virtual Research

Environments. In particular it is explained the importance of the DILIGENT Hosting Node (DHN) to host the gCube services and how the gLite services are associated with the EGEE PPS infrastructure. This talk also presents the status of the DILIGENT infrastructure currently deployed for the different DILIGENT applications.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

The DILIGENT infrastructure brings together gCube and gLite services. By building on top of gLite, gCube not only extends the gLite functionality but brings more computing and storage power to the EGEE infrastructure.

DILIGENT - Dynamic Digital Libraries and Virtual Research Environments on the Grid / 112

ImpECt - Implementation of Environmental Conventions

Author: Veronica Guidetti[1]

[1] *ESA*

The goal of the ImpECt scenario is to improve accessibility, interoperability and usability of environmental data, models, tools, algorithms and instruments integrating the distributed data sources with specialized data handling services. In particular, the DILIGENT infrastructure is used as a means for supporting the preparation of projects and periodical reports on specific environmental topics of concern. This experience is reported.

DILIGENT - Dynamic Digital Libraries and Virtual Research Environments on the Grid / 113

SAPIR - Search In Audio Visual Content

Author: Matteo Mordacchini[1]

[1] *CNR - ISTI*

SAPIR is an EU project aimed to extend the power of web searches beyond centralized text and metadata searches to include distributed audio-visual content.
By exploiting the DILIGENT infrastructure, a large collection of images, counting 100 millions of objects, is under processing with the aim to generate the bigger data set of image features than at any time before. The result of this data challenge is presented.

DILIGENT - Dynamic Digital Libraries and Virtual Research Environments on the Grid / 114

Panel - Exploiting the Grid: approaches, experiences and possible synergies

115

GridWay Interfaces for On-demand Access to EGEE

Authors: Eduardo Huedo Cuesta[1]; Ignacio Martín Llorente[1]; Javier Fontán Muiños[1]; José Herrera Sanz[1]; José Luis Vázquez Poletti[1]; Rubén Santiago Montero[1]; Tino Vázquez Blanco[1]

[1] *Universidad Complutense de Madrid*

Corresponding Authors: jfontan@fdi.ucm.es, ehuedo@fdi.ucm.es

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

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 transfer and information services. The demonstration will mainly show the interfaces provided by GridWay allowing on-demand access to EGEE resources.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

Currently, there are four interfaces in GridWay that allow on-demand access to grid resources:
- "GridGateWay" encapsulates a grid behind standard Globus services, mainly GridFTP, MDS and GRAM (the latter two, both WS and pre-WS), enabling any application using Globus services to access a whole grid infrastructure as if it were a single Globus resource.
- "SGE Transfer Queue to Globus and GridWay" allows the forwarding of jobs from SGE to Globus and GridWay in a transparent way under some load conditions.
- "Excel plug-in for GridWay" allows Excel spreadsheets to divert compute-demanding tasks to grid resources managed by GridWay.
- "C and Java DRMAA bindings for GridWay" enable developers and ISVs to write portable applications using DRM systems.

The above interfaces are deployed on the client side, not needing the deployment of central services. Moreover, GridWay can be currently deployed on an EGEE UI.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

GridWay interfaces for on-demand access to grid infrastructures allow end-users to obtain computing power where needed. For instance, computing resources can be used on-demand by overloaded LRMS (like SGE), ISV applications using DRMAA (like gridMathematica), Excel spreadsheets (e.g. to immediately visualize results) or tools using standard Globus CLIs and APIs (like a workflow engine or a computing portal).
Moreover, since GridWay is able to simultaneously harness resources from the most important grid infrastructures (EGEE, TeraGrid, OSG...), both directly and through GridGateWays, end-users don't need to worry about the underlying resources and middleware their jobs are using.

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.:

After a short introduction (supported by some slides) is given, a live demo of the chosen interface will be performed. For this demo, a grid infrastructure will be set up, involving a GridWay instance located at UCM giving access to some resources from EGEE (and possibly from TeraGrid and OSG), a GridGateWay wrapping that GridWay instance, a SGE cluster with a transfer queue to GridWay configured, an Excel spreadsheet submitting tasks to GridWay and some DRMAA applications using both GridWay and SGE depending on the DRMAA library they are dynamically linked. This infrastructure will accessed using SSH, therefore a wired network connection would be appreciated.

Operational Security Coordination Team / 116

Introduction: status, progress and issues

Operational Security Coordination Team / 117

Security Service Challenges

SSC3: update and discussions

Operational Security Coordination Team / 118

Best practice and training

RSS feed topics, training and other dissemination activities

Operational Security Coordination Team / 119

Security Monitoring

Progress on SAM security monitoring, monitoring tools supported by the OSCT and discussion on possible gLite-based security monitoring.

Operational Security Coordination Team / 120

Web applications security

Operational Security Coordination Team / 121

Technical session (CLOSED)

TRANSITS Technical module

Operational Security Coordination Team / 122

AOB

AOB and next meeting

123

Productive End-user System for Large-scale Drug Discovery

Author: Hurng-Chun Lee[1]

Co-authors: Chon-Chen Lee [2]; Hsin-Yen Chen [3]; Li-Yung Ho [3]; Pei-Ying Yao [3]; Simon Lin [3]; Ying-Ta Wu [2]; Yu-Hsuan Chen [3]

[1] *ASGC, CERN, U. Innsbruck*

[2] *Genomic Research Center, Academia Sinica*

[3] *ASGC*

Corresponding Author: hurng-chun.lee@cern.ch

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

Efficient computer-based drug discovery for neglected and emergent disease is based on the molecular docking simulation where the potential drug candidates are selected from a huge amount of chemical compounds (˜ 10ˆ6). The previous grid challenge preparing for fighting avian flu mutations has demonstrated that the biomedical communities can benefit from the EGEE infrastructure in terms of the speed and the reaction time of screening over a full spectrum of the compound libraries.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

The system uses DIANE to distribute docking simulations on the grid. DIANE features an agent-based task pulling model with high-level failure recovery mechanism to ensure a steady job throughput. The system could also utilize as many as available grid resources by running multiple DIANE instances.

The distributed DIANE instances are organized by a Virtual Queuing System, part of the Grid Application Platform developed by ASGC. Through it, users can manage the distributed DIANE instances as controlling jobs in a job queuing system.

Essential information from the simulation results are stored in the AMGA catalogue system as the metadata. Aggregative data analysis could be done easily by AMGA queries rather than looking into the results widely distributed on the grid storage elements.

The thin client consists of a set of Java APIs and a command shell. It can be launched in any Java-enabled desktop environment, providing an opportunity of integratigring the grid with desktop utilities.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

The grid is an ideal environment providing "on-demand" resources for the docking simulation. The objective of our work is to deliver a productive system enabling biologists to run docking simulations and to manage the docking results on the grid as simple as using a desktop utility in the daily research.

Based on our experience in running previous computational challenge, the system has integrated several existing technologies in order to improve the usability, scalability and stability of running molecular docking simulations on the grid.

During the second grid challenge started in this August for avian flu drug analysis, the system has been used by biologists to run large scale docking simulations on the EGEE infrastructure. Within the same system, biologists also performed a first-level analysis on the distributed results for planning a refinement simulation. A significant part of the simulation works was done by non-grid experts on Windows desktop.

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.:

The demo will show a real production system used by biology users to run docking simulations on the EGEE infrastructure.

In the demo, a real-life job containing about 1000 docking simulations (˜6000 minutes CPU time) will be prepared and submitted from the thin client running on a laptop. The job preparation is based on answering few application-level questions. Few hundred worker nodes on the EGEE infrastructure will be requested to process the simulations.

When the job is running, a real-time progress monitor will show a steady job throughput even under the condition that unexpected errors may occur on some grid worker node. We will also demonstrate how user can visualize the results produced by the completed simulations and perform an aggregate analysis among the simulations (e.g. an binding-energy histogram).

To have better demo performance, a wired internet connection would be preferred.

RINGrid: Remote instrumentation in contemporary science / 124

User requirements for interactive controlling and monitoring of applications in grid environments

Corresponding Author: isabel.campos@cern.ch

RINGrid: Remote instrumentation in contemporary science / 125

RINGrid: Evaluation of Remote Instrumentation Infrastructures

Corresponding Author: mpolak@gup.jku.at

RINGrid: Remote instrumentation in contemporary science / 126

Workflow management in remote instrumentation systems - e-VLBI experiences

RINGrid: Remote instrumentation in contemporary science / 127

On the evaluation of platforms for remote instrumentation on the Grid

Corresponding Author: ckotso@grnet.gr

RINGrid: Remote instrumentation in contemporary science / 128

Discussion

Closing Plenary / 129

Presentation of the next EGEE event: EGEE User Forum 11-14 Feb 2008

Closing Plenary / 130

Wrap up and closure of the conference

Corresponding Author: robert.jones@cern.ch

131

Astronomy and Astrophysics applications on EGEE from Paris, Grenoble, Lyon and Strasbourg Observatories

Author: marie-lise dubernet[1]

Co-authors: A-L. Melchior [1]; A. Schaaf [2]; Alexandre Faure [3]; B. Semelin [1]; C. Pinte [3]; F. Le Petit [1]; F. Menard [3]; F. Roy [1]; Francoise Combes [1]; H. Wozniak [4]; Igor Chilingarian [1]; J. Lecubin [1]; J. Marchand [1]; Jean-Michel Alimi [1]; L. Wiesenfeld [3]; O. Tiret [1]; P. Le Sidaner [1]; P. Valiron [3]; P. di Matteo [1]; S. Baek [1]; V. Lainey [1]; W. Thuillot [1]; Y. Revaz [1]

[1] *Paris Observatory, France*

[2] *Strasbourg Observatory, France*

[3] *Grenoble Observatory, France*

[4] *Lyon Observatory, France*

Corresponding Author: marie-lise.dubernet@obspm.fr

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

We plan to post-process theoretical astronomical data contained in databases, in particular results of numerical simulations of the formation and evolution of galaxies in cosmological context (HORIZON project) and orbits computation in galactic potentials, as well as porting numerical simulations related to astronomy, astrophysics, cosmology, astrochemistry, molecular physics and dynamics of small bodies of planetary systems. We wish to test workflows for astronomical image processing.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

We are not experienced yet with EGEE, and will adapt our simulations and processing while learning. Nevertheless we have experience on two other systems: Grid'5000 (HORIZON collaboration) and CIMENT (Astrochemistry and Radiative transfer). Grid'5000 is a research effort developing a large scale nation wide infrastructure for Grid research. The main purpose of this platform is to serve as an experimental testbed for IT research in Grid Computing. Clusters in Grenoble are regrouped within the CIMENT initiative (https://ciment.ujf-grenoble.fr/) which is part of the regional grid CiGri (http://cigri.gforge.inria.fr/). CNRS researchers in Grenoble have been involved in the specifications and testing of CIGRI, and have demonstrated its ability to tackle large campaigns of millions of jobs.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

A&A applications require the deployment of codes on the Grid on the fly and in a transparent way. The use of a common infrastructure will foster a set of good practices and encourage evaluation, development and distribution of tools and services able to guarantee such a transparent code deployment. It will encourage development and deployment of standardized codes useful for each given A&A community. Lastly it will favor the fast return of scientific results needing millions of runs. As examples this concerns the exploitation of the theory in the virtual observatory, collaborative projects such as HORIZON, scientific preparation and exploitation of observational space and ground missions such as HERSCHEL/ALMA.

EELA: E-Infrastructure shared between Europe and Latin America / **132**

BELIEF - The Project & the Digital Library

133

MOTEUR manager for workflow-based data-intensive applications

Authors: Diane Lingrand[1]; Johan Montagnat[2]; Tristan Glatard[3]

[1] *UNSA*

[2] *CNRS*

[3] *CNRS / INRIA*

Corresponding Author: johan@i3s.unice.fr

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

The MOTEUR workflow manager is a generic workflow enactment service for grids. It targets a wide user community. It was designed inside the biomed community to fulfill the needs for efficient enactment of data-intensive workflow-based applications on the EGEE grid infrastructure, although it is not limited to this application area. MOTEUR is interfaced to the gLite workload manager and the code is open source.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

MOTEUR has been successfully used to efficiently enact data parallel applications to medical images processing. It has been exploited in production for more than one year and it has been evolving to include various execution time optimization strategies. The workflow engines shields the end user from the details of the application parallel execution and from the grid interface. It ensures consistent parallel execution and interface to the gLite middleware through its command line interface. Large scale experiments, involving thousands of medical image processings have been conducted with very good execution time as compared to other comparable workflow managers. Extensions to larger scale applications to test the scale-up capability of MOTEUR are planed in a near future.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

MOTEUR is interpreting and enacting the Scufl workflow language developed in the context of the Taverna project. Scufl is a pure data flow language which eases the description of data-parallel applications in a workflow framework. It provides a large flexibility in data composition through the use of iteration strategies. MOTEUR defines a strict semantics for this data composition, relying on the implicit or explicit user input describing the connections between data fragments. While preserving this semantics,

MOTEUR provides completely asynchronous parallelization of the workflow tasks and data segments to be processed, thus optimizing the degree of parallelism that can be achieved. It is also providing a legacy code wrapper enabling the execution of non-instrumented codes in a service-oriented framework, interfaced with the grid infrastructure.

134

A Framework for Providing Hard Delay Guarantees in the EGEE

Author: Panagiotis Kokkinos[1]

Co-authors: Emmanouel Varvarigos [2]; Kostas Koumantaros [3]

[1] *University of Patras*

[2] *Research Academic Computer Technology Institute - RACTI*

[3] *Greek Research & Technology Network - GRNET*

Corresponding Author: kokkinop@ceid.upatras.gr

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

Future Grid Networks should be able to provide Quality of Service (QoS) guarantees, in order to support real world commercial applications and complex scientific simulations and computations. In this work we present a framework that provides hard delay guarantees to its Guaranteed Service (GS) users and examine its applicability to the EGEE infrastructure.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

We believe that the proposed QoS framework can be incorporated in the EGEE environment. The GS users can in fact be Virtual Organizations (VO), single applications or just different User Interface (UI) machines, using the EGEE's infrastructure. Furthermore the resources, Computing Elements (CE) capable of serving the GS users should be defined during their installation and configuration by the site's administrator. These CE will publish to the Information Service of EGEE (Berkely Database Information Index - BDII) not only data regarding their current load but also information needed for the operation of the proposed QoS framework. When a user wishes hard delay guarantees (GS user) for the scheduling of his tasks, then he can ask such a service from the Workload Management System (WMS). The WMS will be responsible for the registration of the user to GS resources and for the scheduling of his tasks to registered resources, capable of executing them before their deadline expires.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

In the proposed framework no quantitative resource reservation is performed. Instead, the users and the resources simply agree upon the task load the former will generate and the latter will serve. Specifically the GS users are leaky bucket constrained, and so they follow a (ρ, σ) constrained task generation pattern, which is agreed separately with each resource during a registration phase. So using the proposed framework a GS user can choose a resource that will execute his task before its deadline expires with absolute certainty. Also based on this framework various resources types can be defined, based on whether they serve GS users or BE (best effort) users, or both.

135

The ENEA gateway approach to provide EGEE/gLite access to unsupported platforms and operating systems

Author: Giovanni Bracco[1]

Co-authors: Alessio Rocchi [2]; Andrea Quintiliani [1]; Andrea Santoro [1]; Carlo Scio' [3]; Silvio Migliori [1]

[1] *ENEA*

[2] *CASPUR*

[3] *Esse3Esse*

Corresponding Author: bracco@frascati.enea.it

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

The success of the GRID depends also on its flexibility in accommodating the computational resources available over the network. A big effort is underway to develop accepted GRID standards but in the meanwhile solutions have to be found to include into EGEE infrastructure resources based on platforms or operation systems which are not currently supported by gLite middleware. The ENEA gateway approach provides a working solution to this issue, now enabling GRID access to its AIX SP systems.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

The poster will describe the architecture and the implementation of the gateway solution built on the main components of ENEA-GRID middle-ware, which is based on very mature and reliable software, namely the AFS distributed file system and LSF Multicluster. The key element of the architecture is a set of Linux proxy machines, running standard gLite middle-ware, which support the communication between the non standard worker nodes and the EGEE infrastructure. EGEE technical reports have also been prepared to document the implementation.
In the last year the ENEA-INFO EGEE site has been certified in the gateway configuration for AIX resources and it is open to production jobs. The site supports at present several VOs (COMPCHEM, EGRID, FUSION) and the experimentation with applications is underway. Proposals by other interested VO are also well accepted.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

ENEA, the Italian agency for the energy, environment and new technologies, has a substantial experience in GRID technologies and its multi-platform HPC resources are integrated in the ENEA-GRID infrastructure. ENEA participation in EGEE has focused on the interoperability between EGEE and ENEA-GRID and resulted in the development of a gateway architecture. The gateway provides a flexible and affordable solution for the access in principle to all the platforms and operating systems available in ENEA-GRID and has been finalized to the case of the AIX SP system, but tests have also been performed for Altix IA64, IRIX, MacOS X and Solaris.
This result can be used to expand the EGEE GRID capability by including a wider range of resources but also, on the other hand, to take advantage on the maturity of the gLite grid services to offer a working GRID solution to communities have been up to now discouraged by the middle-ware rigidity.

136

Grid-DB Management in gLite based production Grids with the GRelC Data Access Service

Authors: Giovanni Aloisio[1]; Roberto Barbera[2]; Sandro Fiore[1]

Co-authors: Alessandro Negro [1]; Emidio Giorgio [2]; Maria Mirto [1]; Massimo Cafaro [1]; Salvatore Vadacca [1]

[1] *SPACI Consortium & University of Salento*

[2] *INFN Sez. di Catania, via S. Sofia, 64*

Corresponding Author: sfiore@mailing.unile.it

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

The activity aims at providing an advanced grid database management service (GRelC DAS) tightly coupled with the gLite middleware, the EGEE architecture and the GILDA t-Infrastructure. The target community includes scientists/VOs belonging to different domains within the EGEE collaboration that need for their applications to transparently, efficiently and securely access huge and distributed DBs. The GRelC DAS is a GSI and VOMS enabled grid service providing advanced database management features

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

The GRelC DAS, is a data grid access service developed by the GRelC Team (SPACI Consortium and Univ. of Salento). Currently, it is tested on the GILDA t-Infrastructure and used for training activities on grid-database management. The GRelC DAS is a GSI/VOMS enabled web service addressing extreme performance, interoperability and security. It provides a uniform access interface to relational and non-relational (i.e. XML db) 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 consolidated grid standards and specifications as well as production grid middleware (gLite). Currently it has been extended with additional functionalities such as asynchronous queries, it delivers resultsets by streaming and compressing data, it can be managed and accessed via a web interface (GRelC Portal) or a graphical console management (XGRelC).

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

The Grid could allow intensive data mining applications on databases (exploiting and leveraging the GRelC DAS) on hundreds/thousands of machines. Working intensively in a grid environment with databases new communities could be attracted due to the high number of computational resources. Moreover, solutions based on database replication could allow workload management and distribution, able to serve a huge load of user requests (queries) with regard to classical systems/approaches/environments. Due to (i) the transversal role of the GRelC DAS, (ii) extreme performance and (iii) advanced delivery mechanisms implemented (streaming, chunking, compression and prefetching), many experiments and Virtual Organizations within the EGEE project could benefit from it (bioinformatics, astrophysics, financial, etc.). The grid could also become a big "repository" containing many databases to be accessed, integrated, federated, etc.

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.:

The demo will focus on several parts and it will show how the GRelC DAS provides a strong data virtualization layer compliant with gLite. We will show how a grid-database can be imported, managed and accessed in grid with the GRelC DAS by using: (i) the command line interface, (ii) a graphical console named XGRelC and finally (iii) the GRelC Portal. Network connection is required in order to remotely manage databases that are currently deployed on the GILDA t-Infrastructure. We will show how authentication can leverage VOMS extensions and how query submission can be easily carried out on widely distributed databases. During the demo we will also show some examples of query submission through the GRelC Portal, a Web interface including all of the key features provided by this grid service. Query demo will include: synchronous and asynchronous query submission and examples of queries submitted (as a job) to a Broker through a GRelC DAS enabled User Interface.

Bridging gLite with the Volunteer Computing middleware "XtremWEb-CH"

Author: Nabil Abdennadher[1]

[1] *University of Applied Sciences Geneva. (HES-SO)*

Corresponding Author: nabil.abdennadher@hesge.ch

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

To design a "bridge" which allows the migration of jobs from XWCH (www.xtremwebch.net) to gLite and vice versa. A particular interest will be given to the security aspect. Indeed, the two softwares have different security and safety policies: gLite is rather a GRID middleware whereas XWCH is rather a Volunteer Computing platform.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

No comment

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

Two applications are being "gridified" on XWCH. The first one is a package of programs for inferring phylogenies: PHYLIP. PHYLIP (the PHYLogeny Inference Package) is a package of programs for inferring phylogenies (evolutionary trees). Developed during 1980s, PHYLIP is one of the most widely-distributed phylogeny packages. It has been used to build the largest number of published trees. PHYLIP has over 15,000 registered users. The application IS USED to build trees of HIV viruses.
The second one is a medical application used to construct a functional neural map of the humain brain. The system to build is used to detect the areas responsible of epileptic crises.

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.:

XtremWeb-CH (XWCH: www.xtremwebch.net) is a software system that makes it easy for scientists and industrials to deploy and execute their parallel and distributed applications on a public-resource computing infrastructure (Volunteer Computing). XWCH supports diverse High Performance applications, including those with large storage or communication requirements. Universities, research centres and private companies can create their own XWCH platform while anonymous PC owners can participate to these platforms. They can specify how and when their resources could be used.
Two applications are being "gridified" on XWCH. The first one is a package of programs for inferring phylogenies: PHYLIP. PHYLIP (the PHYLogeny Inference Package) is a package of programs for inferring phylogenies (evolutionary trees).
The second one is a medical application used to construct a functional neural map of the humain brain. The system to build is used to detect the areas responsible of epileptic crises.

138

What is the CIC portal for, how can I use it and why is it useful ?

Author: Helene Cordier[1]

Co-authors: Cyril L'Orphelin [1]; Gilles Mathieu [1]; Osman Aidel [1]

[1] *CNRS/IN2P3*

Corresponding Authors: gilles.mathieu@in2p3.fr, cyril.lorphelin@cc.in2p3.fr, osman.aidel@in2p3.fr, helene.cordier@in2p3.fr

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

The CIC Portal, or Operations Portal, is conceived to be useful to any grid user, but also to VO-, ROC-, and site managers. Namely, information about available resources, existing VOs and their requirements is available. Moreover, grid wide communication tools are there to ease up operations and broadcast relevant info in a transversal way to the different communities involved in EGEE operations. The demo shows how to access and use all this; it will also adress questions and record feedback.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

The range of tools and information proposed by the CIC portal over the last 3 years has allowed major improvements in daily work and procedures for various actors. The best example of this being the work of "Grid Operators on Duty" (COD) who use the CIC portal as their central operational tool. Tools and procedures established to support their work have proven to be stable and scalable, as the number of sites they have been taking care of has been multiplied by 5 in less than 3 years. Key services essential to our activity (knowing that the CIC portal is in itself a key service for various grid activities) are numerous as the CIC portal is an integration platform. Indeed, it is interfaced to basic operational tools like GGUS, GOCDB, gstat, the grid's information system, FCR, and SAM. Finally, in order to ensure the High-Availability of their service, the "Grid Operators on Duty (COD)" have set-up internal working groups to elaborate, namely, failover processes of the tools they use.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

The CIC portal added value to the Grid infrastructure is aimed at all the EGEE actors, whether scientist, VO or site manager, or grid operator. Indeed, each community has its own entry point to the portal.
The information on the operational state of the grid is filtered out and presented according to its usefulness to a particular community. The tools we will present to a given group of customers of the portal are those which could be useful to them, like VO ID card updates for VO managers, EGEE broadcasts for various communities, dashboards for grid operators and so on.
For instance, any new VO can immediately, directly and simply benefit from the portal: a given VO manager defines indeed its "VO ID card" in the portal, which is the starting point for any site administrator to configure and allow VOs' access to his sites' resources.
The scale of the activity is tied up to the regular EGEE operations and any new application or user community will find their interest in consulting it.

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.:

Interacting with the demonstrators, different actors of the grid will learn:
- what operational info is available to them and demonstrators will then get the feedback on the ones they actually find the most useful.
- how to access this info;
- which tools are available, what benefit they get out of them;
- which procedures and workflows are implemented to ease daily work in operations. In particular, VO managers will see how to use the VO ID card for registering a new VO or updating its information; site administrators may discover tools like alert notification for SAM tests or the user tracking tool; users may grasp what resources are actually available for their VO.

Moreover, posters show screenshots on the scenariis we elaborated to illustrate functionalities that each specific group of users may access, as well as schemas explaining the basic workflows implemented in the operations portal.
Logistic specifications are: network, flat screen, 2 tables, 4 chairs, support for posters.

139

AWare - An easy Way to Access GRID REsources

Author: Nicola Venuti[1]

[1] *NICE srl*

Corresponding Author: nicola.venuti@nice-italy.com

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

AWare is a European Grid technology project aimed at the development of an opensource, userfriendly, solution, to create,store and manage complex job workflows.
There are many communities and related Grid projects within EGEE (eg 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

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

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.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

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

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.:

The A-WARE, is a demonstration will show how its possible to create and manage the complete the life-cycle of a grid workflow through a web application and standard web browser.

The demo consists of four principle phase. The first; the design phase is where the workflow is graphically designed using an applet.
The second; the service binding phase, connects the A-WARE services to the workflow designed in phase one.
The next phase consists of data mapping, where the input and output data are mapped to the A-WARE services. The final deployment phase, generates the actual workflow which is then memorized in the system.
After the four phases are complete the workflow can then be submitted to the Grid. To run the demonstration a standard screen (or projector) and internet connection is needed.

140

BalticGrid Special Interest Groups

Authors: Algimantas Juozapavicius[1]; Margarita Kazakeviciute[1]

[1] *Vilnius University*

Corresponding Author: margarita.kazakeviciute@mif.vu.lt

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

BalticGrid project extends European Grid by integrating new partners from the Baltic States in the European Grid research community. BalticGrid offers for users an activity context - Special Interest Groups (SIG). The aim of SIG is to bring people together working in the same research/scientific project. Such activity includes a common effort in the design, implementation and evaluation of research task. SIG provides efficient tools to exchange information, data, applications and ideas.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

Each BalticGrid SIG is provided with the following functionality:
- Gridcom - a user friendly, intuitive web interface for launching applications on a Grid.
- Repository - directory for storing and sharing data, files, results of calculations among members of a SIG.
- Online audio/video conferencing – easy and fast way to organize meetings of SIG community, exchange news and ideas.
- Forum – online discussion board of BalticGrid users for experience and news sharing.
- Mailing list - every user can join mailing list of a corresponding SIG to get news and send messages for SIG community.

These functions make SIGs quite popular among BalticGrid users because it facilitates a work in research groups.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

The functionality of SIG offers an easy, well-formed, intuitive web interface for groups of users to launch common applications, exchange data and ideas. This gives a possibility to expand user communities as well as to emerge the new ones.

There are eight SIGs formed in BalticGrid project with research areas in physics, astronomy, computer modeling, computational and applied linguistics, modeling of chemical, physical and biological processes, multidimensional scaling. One more SIG is planned to be formed this year.
According to the fields of research of scientists in the Baltic States and technical analysis of the developed applications the following BalticGrid Special Interest Groups are formed:
-Baltic Sea Eco-System Modeling,
-Text-to-Speech, Text Annotation,
-Stellar Spectra,

-Atomic and Nuclear,
-Computer Modeling,
-Multidimensional Scaling,
-GAMESS.

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.:

Demo presents all the functionality which is provided for BalticGrid Special Interest Groups (http://sig.balticgrid.org): Gridcom, repository, online audio/video conferencing, forums, mailing lists.
The main part of demo is dedicated to the possibilities of Gridcom - a user friendly, intuitive web interface for launching applications on grid. It shows how user may:
-upload/download a grid application, view the list of grid applications, launch a work for an application;
-view the list of launched works with indicated states, relaunch a work or delete it, view a log file of a work, download output of each work;
- view the list of submitted jobs of the selected work with indicated states, resubmit job, download output of each job;
- use the interface to manage data and files repository (output of applications is placed in the repository).
Gridcom is a secure interface with authentication based on a secure low level Linux authentication.

141

Genetic Algorithm for Solving Chess Endgames in Grid Environments

Author: Antonio Gómez-Iglesias[1]

Co-authors: Juan A. Gómez-Pulido [2]; Juan M. Sánchez-Pérez [2]; Manuel Rubio del Solar [1]; Miguel A. Vega-Rodríguez [2]; Miguel Cárdenas Montes [1]

[1] *CIEMAT*

[2] *UNEX*

Corresponding Author: antonio.gomez@ciemat.es

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

Chess game needs a high computational cost in order to be played correctly by a machine. Our proposal relies on the use of grid computing and genetic algorithms (GAs). On the one hand, grid computing offers us the potential for deeper game tree analysis. On the other hand, genetic algorithms can reduce significantly the computational cost of a brute force search, obtaining a good solution in a very lower execution time. Combining these two approaches (grid + GAs) we can obtain a very good chess.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

Using gLite as middleware in our grid environment we can achieve good results with GAs applied to chess game. Executing and managing of the jobs submitted to the grid can be easily developed. Jobs are submitted using JDL (without WSDL).

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

This is our first step in GA with grid computing and we hope to apply the acquired knowledge in future developments in scientific areas. Grid computing is a key issue in our work because it allows to access to many resources and to use the computational elements of the grid environments.

ETICS/EGEE: Quality assurance workshop / 142

Introduction to QA Workshop

Corresponding Authors: gabriel.zaquine@cern.ch, marc-elian.begin@cern.ch

ETICS/EGEE: Quality assurance workshop / 143

Introduction to ITIL - Best Practices for IT Service Management

ITIL has become the de facto standard for IT Service Management. The ITIL Framework defines Roles, Processes, Key Performance Indicators and outlines Critical Success Factors. These definitions helps every kind of IT organisation WHAT has to be implemented for IT service management in order to achieve better customer orientation, customer satisfaction and measurable service quality. In this talk one will be informed about the basics of the IT Infrastructure Library - which is since Version 3 now called - ITIL, Best Practices for IT Service Management.

ETICS/EGEE: Quality assurance workshop / 144

The ITIL pilot project and further investigations at FZK

To implement IT service management processes and the therefore necessary organisational structures, is a bigger project that has to be carefully evaluated. This talk will present the essential steps during the ITIL pilot project in the FZK. The main attention lies here on the service support processes, which are assisting directly the IT personnel in the process of IT service production and provisioning to the customers.

ETICS/EGEE: Quality assurance workshop / 145

EGEE ITIL Group first works - ITIL evaluation for ENOC

The presentation will describe a first study conduced within the EGEE ITIL group to evaluate the way to reach consistency between the work of the ENOC and the ITIL way of work.
ENOC is the EGEE Network Operation Center, part of EGEE SA2 (Network activity).
The presentation will raise which issues for the ENOC to reach ITIL?

ETICS/EGEE: Quality assurance workshop / 146

ETICS Quality Assurance Certification Model

ETICS/EGEE: Quality assurance workshop / 147

ETICS Service overview

ETICS/EGEE: Quality assurance workshop / 148

Panel discussion

149

Interactive Air Pollution Simulation

Author: Branislav Simo[1]

[1] *Institute of Informatics, Slovak Acedmy of Sciences*

Corresponding Author: branislav.simo@savba.sk

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

The radioactive air pollution simulation is a computational core of the commercial IMS Model Suite - a software system for environmental pollution assessment and prediction of consequences of nuclear accident or radiological emergence. It provides the target audience (scientists, emergency response authorities, etc.) with the ability to predict the spread of radioactive air pollutants given the pollutant release data (location, amount) and the meteorological conditions in the area.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

The simulation will require the operational grid with standard services like CE, SE, resource broker. The interactivity requires the usage of the Migrating Desktop UI which is used for job submission, monitoring and as a container of application specific plug-in handling the data flowing to/from the interactive channel. For the setup of the interactive channel the ability to execute the gLogin tool is required.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

The application core simulates a number (thousands) of particles that have been released into the atmosphere using the Lagrangian trajectory model. The computational requirements increase with added number of particles, which are necessary for increased precision of the simulation. There was a clear need for using parallel computation to decrease the computational time. As the individual particles are independent of each other, a very efficient parallelization is achieved. Using the clusters from the grid and the special interactivity feature enables the users to perform the simulation faster and with the ability to monitor and steer the simulation as it proceeds. Even several clusters might be used to run it in a parameter study-like scenario.

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.:

The demo presents an interactive steering of a simulation running on the grid. The Migrating Desktop UI is used as a platform for submitting the application to the grid and launching application specific

interaction plug-in.
In the demo, the user submits the application using the predefined parameters stored as an icon on the desktop. Then he monitors the status of the job in the job monitoring dialog. When the job enters running state the user starts the application plug-in, which connects via interactive channel to the application. As the computation proceeds, images showing map with the particles are sent to the user for each hour of simulation time (i.e. every few seconds). Images are buffered locally for the sake of viewing past images and for playing simple animation. Each simulation step, the user can send a command to split particles, thus doubling particle count simulated. There is also a command for terminating the simulation correctly instead of aborting it the usual way.

150

Earth Science requirements tracing using test suites

Authors: Monique Petitdidier[1]; Wim Som de Cerff[2]

Co-authors: Horst Schwichtenberg [3]; Viet Tran [4]

[1] *CNRS/IPSL*

[2] *KNMI*

[3] *Fraunhofer SCAI*

[4] *UI SAV*

Corresponding Author: sdecerff@knmi.nl

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

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, geology, and geophysics. The DEGREE project is a consortium of ES partners aiming at promoting the uptake of Grid technology in ES.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

The test suite is organized in 3 parts: the test suite organization document, the test suite specification documents and the application software. The test suite organization document provides a high level overview of the test suite as well as templates. The test suite specification documents are written using the common template specifying the test cases and test procedures for each application. Each ES application consists of the application software, data, database schemas, documentation and contact points for technical and scientific support. The test specification for each ES application consists of one or more test cases; each test case is designed to address one or more specific ES requirements. The results of performing the specified tests are documented in the test reports. The reports are usable for the Grid developers to find weaknesses and requirement gaps in their software. A total of 7 test suite specifications is available and can be downloaded from the DEGREE site.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

The ES community has two main IT related challenges: Modeling, which requires vast amounts of computational resources, and exploration and production of large shared data sets. In both Grid can play an important role. ES requirements for Grid are based on several requirements studies. Both ES and Grid have evolved and at the present state it is not known to what extent requirements have been implemented. A method is needed in order to provide an unbiased way to monitor progress of Grid and trace ES requirements. It is not sufficient to provide only the requirements, as they might be misunderstand or misinterpreted by possible lack of domain knowledge. The test suites aim at providing well-documented test specifications and real applications plus data to the Grid developers. They can use

the test suites to check whether the developed Grid software matches the ES requirements. Also, showing working application examples are more convincing than a statement certain requirements are met.

151

Promoting e Science at Minority Serving Institutions (MSI)-CyberInfrastructu (CI)

Author: Dr. Richard Antonio Alo'[1]

Co-authors: Al Kuslikis [2]; Alex Ramirez [3]; Diane Baxter [4]; Geoffrey Fox [5]; Karl Barnes [6]

[1] *university of houston downtown*

[2] *American Indian Higher Education Consortium*

[3] *Hispanic Association of Colleges and Universities*

[4] *San Diego Supercomputer Center*

[5] *Indiana University*

[6] *National Association for Equal Opportunity in Education*

Corresponding Author: ralo@uh.edu

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

Three organizations that comprise the Alliance for Equity in Higher Education (AEHE) – American Indian Higher Education Consortium (AIHEC), the Hispanic Association of Colleges and Universities (HACU), National Association for Equal Opportunity in Higher Education (NAFEO), have established the Minority Serving Institutions (MSI)-Cyberinfrastructure (CI) Empowerment Coalition (MSI-CIEC). Coalition will foster CI-enabled distributed education and research network providing e-science education

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

Through a novel collaboration of MSIs and key leaders in CI, this project provides a model for building the CI and e-science knowledge and skills of MSI faculty and growing an inclusive national CI-enabled research infrastructure. The project will advance the knowledge associated with building technology-mediated multi-institutional and cross ethnic research and educational collaborations critical for CI development. It provides the needed human "middleware" to foster CI-enabled MSI and Research I university research and education collaborations, accelerating the inclusion of MSIs and their students into the growing national CI activities while helping to prepare the current and next-generation scientists at MSIs to use, support, deploy, develop or use CI. It is designed to develop the MSI capacity to prepare underrepresented minority students for a future in CI-facilitated science and the knowledge-based economy.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

Based in large part on lessons learned from the MSI CI Institute pilot project (NSF project #0537498), MSI-CIEC will provide the "human middleware" – the social and technological mechanisms facilitating the necessary communication and support linkages between MSI faculty and students, and researchers associated with e-science and CI initiatives. This will ensure that MSI institutions emerge as full partners in the national CI-enabled research and education infrastructure. Specifically, the project will involve: Expansion of the coalition of CI leaders and MSIs and resources and expertise brought into the collaboration by such leaders in emerging CI as TeraGrid, SDSC, NCSA, TACC, RENCI, and leading e-scientists in computational biology at majority and minority institutions; Implementation of CI-supported collaborative research and education projects using e-Science and e-Learning within and across our respective communities.

152

LiveWN: CPU scavenging in the grid era

Authors: Fotis Georgatos[1]; Giannis Kouretis[1]

[1] *National Technical University of Athens*

Corresponding Author: fotis.georgatos@cern.ch

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

The goal of this research is to introduce a bootable CD and DVD as an easy and versatile way to provide and use gLite-based Grid resources without the need of any OS installation or middleware configuration. At the same time it provides an excellent training tool for newer Grid users and a flexible playground for people that want to experiment, without enforcing any installation or expert knowledge. We have been testing it thoroughly under different circumstances with firm success.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

LiveWN was prototyped in 2006 as a grid-bootable CD supporting diskless, easy-to-deploy worker nodes and requiring virtually zero administration upon deployment.

The system mixes key technologies including LiveCD (a self-booting Linux CD), gLite (the grid middleware stack), OpenVPN (virtual private network) and OpenAFS (open distributed filesystem), proving that the initial concept was correct and workable and efficient enough to dynamically provision for worker nodes and user interfaces, without having to resort to a hard-disk installation. This "grid distribution" was simple enough to run even from a USB stick. During testing, it became clear that there was a need for two solutions: a CD format called LiveWN; and a DVD format called gLiteDVD.

gLiteDVD has a much larger footprint and it is a superset of LiveWN. On top of LiveWN, it includes a full desktop environment and some extra scientific software, including tools from high energy physics project ATLAS, Wine, PovRay and others.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

What if grid nodes were as simple to run as music on a compact disk player?
What if a new cluster involved zero installation and only minimal—if any at all—configuration?
What if setting up that new cluster took only a few minutes?

If all this were true, we would able to readily exploit the idle time of publicly funded computers, such as those in libraries and universities. We could share resources among arbitrary users, Internet Service Provider members or Open Source Software projects. We could rapidly multiply computing power for urgent tasks. We could quickly and easily train novice users.

LiveWN can address all these features through its adaptable scavenging mechanism.

It will perhaps become apparent once the large LHC experiments start processing their huge datasets during 2008 and other research fields enter the E-Science realm:
all embarrasingly parallel problems should ideally use the resources of dormant computing resources in public institutes and volunteers' systems!

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.:

The demo activity will include distributing a number of LiveWN CDs, DVDs or even USB sticks to interested individuals with a laptop, and show them how easy it can be to become part of the grid within a few minutes, only by booting with that media.

The solution should be expected to work fairly well with most laptops; with the exception of wireless network cards, of whom a fraction might present driver compatibility problems. For this reason we hereby suggest that a switch with network cables will be necessary, so that everybody can enjoy the LiveWN grid experience regardless of his hardware. Needless to say, there is a need for upstream internet connection, a NAT'd private address space won't be a problem though.

Poster, newsletter, website and papers all exist, so outreach material shouldn't be a problem: http://gridathome.sf.net

153

Enabling Distributed Processing of Bioinformatics Data in Hellas Grid Environment

Author: Ilias Maglogiannis[1]

Co-authors: Aris Chatzioannou [2]; Ioannis Kanaris [1]; Ioannis Soldatos [3]; Vassilis Mylonakis [1]

[1] *University of Aegean*

[2] *NHRF*

[3] *AIT*

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

In this work we present a Web based portal, which enables intelligent processing of biological data in Grid environments. The deployed software aims at creation of tools for processing data from microarray experiments over the Hellenic Grid infrastructure. This work serves as a starting point for building a more complete and integrated Grid enabled microarray experimentation environment.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

The developed microarray processing algorithms have been initially provided by as a set of MATLAB libraries. However, no nodes of the Hellenic Grid Infrastructure provide support for MATLAB execution. Furthermore MATLAB is a commercial product, which raises intricate licensing issues when it comes to installing it in the Grid and makes it unlikely to become available in the near future. Therefore, we have investigated possible alternatives, the most prominent being the use of Octave Forge, which is the GNU open-source alternative to MATLAB. Accordingly, we dealt with the task of parallelizing the (Octave Forge) microarray application and making it appropriate for use over the Grid. During grid programming we introduced a hierarchical system of sub-jobs. It is likely that a job could utilize the services of the grid environment to launch one or more sub-jobs.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

Users accessing the Web Interface through the implemented portal are given the ability to submit their experiments, retrieve their results and also compare them with formerly submitted experiments. Since the portal is set up on Hellas Grid User Interface (HG-UI), users have the ability to actually access the whole Grid infrastructure, consisting of many grid nodes. Access to services is enabled by parsing input files and accordingly activating the 'gridified' algorithms for processing the microarray experiments. Both data parsing operations and launching of experiments are specified as Grid jobs, using the Job Description Language (JDL). The provided microarray input files, which are usually structured according to formats that are standard for the microarray bioinformatics community are pre-processed so as so to

be usable by the range of algorithms available.

154

An Agent Based Framework for Complex Systems Simulations on the Grid

Authors: George Kampis[1]; Laszlo Gulyas[1]

Co-authors: Gábor Szemes [2]; Peter Fazekas [1]; Stephen Robinson [3]

[1] *Collegium Budapest*

[2] *Aitia International Inc.*

[3] *University of Ulster*

Corresponding Author: kampis.george@gmail.com

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

As members of the EC-funded QosCosGrid consortium (www.qoscosgrid.eu) we are using the Globus 4.0 toolkit for developing grid middleware for grid enabled Complex Systems applications.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

Because of the universal view taken in this project, a large set of applications, covering the whole spectra of CS types, from a totally opportunistic to an ideally parallel implementation, is needed. Such use cases are not only a set of demonstrations of the developing technology, but the driving force to design grid tools that may be taken as standard in any CS simulation one may encounter.

We describe a general framework for dealing with Complex Systems Simulations (CSS) in a Grid environment. The framework is based on a unifying agent based modeling (ABM) approach. We consider the following problems: (1) basic definitions (2) communication topologies (3) parallelization.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

Many real-world systems involve large numbers of highly interconnected heterogeneous elements. Such structures, known as complex systems (CS), typically exhibit non-linear behavior and emergence. The methodologies used to understand their properties involve modeling, simulation and often require considerable computational resources such that only supercomputers and/or grids can deliver. Currently there is no grid technology with the capability to harness the available grid resources and provide an environment computationally equivalent to a supercomputer service.

This paper is part of a project the aim of which is to develop core grid technology capable of providing quasi-opportunistic supercomputing grid services and technology for CSS-s. The test bed used to evaluate this quasi-supercomputing grid technology will consist of a number of diverse CSS applications operating as part of complex problem-solving environment.

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.:

We present results and demonstrations of a parameter sweep application for CS written in RePast / Java featuring a Windows-based GUI and running under ProActive and Globus on the European grid spanned by CosCosGrid machines.

155

Searching a solution for Mlalign2D / Mlrefine3D on the Grid

Author: Germán Carrera[1]

Co-authors: Alfredo Solano [1]; Jose R. Valverde [1]; Jose-María Carazo [1]

[1] *Centro Nacional de Biotecnología CSIC*

Corresponding Author: gcarrera@cnb.uam.es

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

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.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

We have tried diverse ways to carry out the port of our applications within the Grid infrastructure, generating a good body of experience.
- As a first approach, we tried to send independent jobs with our executables to the Grid to observe the results.
- Next we tried partitioning the data sets which provided the bases for subsequent works..
- Thanks to the group ARDA inside CERN, we used a script integrated in the excellent DIANE framework, but in the end we had some problems relative to the operational structure of our programs and other issues.
- Since we already have MPI-optimized versions of MLaligng2D and MLrefine3D, we tried to use them on MPI clusters within EGEE. However this approach encountered many practical problems related to MPI adoption in EGEE.
- Currently, we are developing our own solution for the jobs management with satisfactory results. We hope to improve its performance in the next future.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

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.

Each of the EM images can be regarded as a projection image of the specimen 3D structure from an unknown projection direction. A key task is, therefore, to determine a posteriori the projection direction. Furthermore, in many cases there is a mixture of different conformations of the same macromolecule, and a "structural class sorting" has to be accomplished at the same time that the orientation search. Probably the most advanced methods is the one REF, refered as ML2d/ML3d, included in the package of xmipp. Typical runs are several single CPU months or years, making it a good target for parallelization over t

156

New results on a comparative evaluation of software providing access to different relational databases exposed on the Grid

Author: Giacinto Donvito[1]

Co-authors: Alessandro NEGRO [2]; Andrea BARISANI [3]; Andreas GISEL [4]; Antonia GHISELLI [5]; Antonio PIERRO [5]; Atul JAIN [6]; Claudio VUERLI [3]; Cristina AIFTIMIEI [5]; Emidio GIORGIO [5]; F. Manna [3]; Fabio PASIAN [3]; Federico DEL FREO [7]; Giorgio MAGGI [6]; Giovanni ALOISIO [2]; Giuliano TAFFONI [3]; Giuseppe LA ROCCA [5]; Luciana CAROTA [5]; Marco VERLATO [5]; Massimo CAFARO [2]; Mirco MAZZUCATO [5]; Roberto BARBERA [8]; Salvatore VADACCA [2]; Sandro FIORE [2]

[1] *INFN-Bari*

[2] *Università Lecce*

[3] *INAF*

[4] *CNR-ITB*

[5] *INFN*

[6] *INFN+Politecnico Bari*

[7] *EGG project*

[8] *INFN+Università Catania*

Corresponding Author: giacinto.donvito@ba.infn.it

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

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 some updates on the comparative evaluation of four tools to access different types of data resources exposed onto Grids: G-DSE, GRelC, OGSA-DAI and AMGA. A special attention is focused on the test in order to prove the scalability of each tool under high load.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

A test bed, which includes several sites belonging to the INFN and the SPACI Grid Infrastructure, has been set up. In each site the instances of the G-DSE, GRelC, OGSA-DAI and AMGA servers and client have been installed.

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. Some tests are based also on very complex query coming from Public Administration requirements in order to test the flexibility of the tools.

The tool comparisons has been extended to the provided authentication mechanism for granting authorization.
The ability of the tools to allow conditional access depending on the user group, role and capability as derived from Virtual Organization Membership Service has also been investigated.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

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).

EUChinaGRID/EGEE/ETICS: Grids and IPv6 / 157

Introduction: Why the needs for IPv6

Corresponding Author: mathieu.goutelle@urec.cnrs.fr

EUChinaGRID/EGEE/ETICS: Grids and IPv6 / 158

IPv6 programming: The example of the BDII

Corresponding Author: xavier.jeannin@urec.cnrs.fr

EUChinaGRID/EGEE/ETICS: Grids and IPv6 / 159

IPv6 support in ETICS

Corresponding Author: marian.zurek@cern.ch

EUChinaGRID/EGEE/ETICS: Grids and IPv6 / 160

IPv6 in EGEE related projects: EUChinaGrid

Corresponding Author: gabriella.paolini@garr.it

161

IPv6 in other Grid projects: NAREGI (tbc)

162

A service oriented infrastructure to integrate earthsystem databases into the grid

Author: Kerstin Ronneberger[1]

Co-author: Stephan Kindermann [1]

[1] *DKRZ*

Corresponding Author: ronneberger@dkrz.de

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

Climate research is data- and collaboration-intensive. Climate and earthsystem models are calibrated and driven by data of different scientific and technical sources. Model results describe several spheres and are needed and analysed by various scientists of diverse disciplines. Moreover, multi model comparisons gain in importance to evaluate uncertainty of models and results. Yet, most of the data is stored in large archives and central databases and analysis is done locally and individually.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

A prototype, developed in collaboration with the German, community driven Grid initiative C3-Grid (http://www.c3-grid.de/), was setup to demonstrate the feasibility of the developed system. Data of different German earthsystem science data centres can be discovered, browsed and uploaded to the EGEE infrastructure via a central Web portal. Via the same Web portal an example workflow can be triggered to run on EGEE; the results are automatically described in ISO 19115 and republished to a central metadata catalogue.
The administration, update and republishing of processed data is based on the data management services of EGEE, such the lfc-tools, the lfn catalogue and the storage elements. To find and retrieve the data of different data centres, currently tools of the C3Grid are used. The system can be easily expanded by further international data providers. Respective collaborations with the British NERC datagrid and the US-American earth system grid are ongoing.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

The Grid offers a common platform to share data, tools and resources, which could be useful to the entire climate community. Even though most large climate and earthsystem models are designed for specialized computer architectures, the pre- and post-processing of input and output data could be done on the grid, once data and tools were accessible. But still some effort is required to seamlessly and efficiently integrate data - described by complex metadata and e.g. stored in databases - into the grid.
We develop a service-oriented architecture to integrate external data sources with complex metadata into a grid infrastructure. The system is built modular and based on common standards such as webservice technology and gridftp for data access, ISO 19115 for data description and OAI protocol for metadata harvesting. This makes the system easily adaptable and expandable and thus potentially beneficial also to other communities that wish to integrate databases and their describing metadata

163

Earth Science Application overview in EGEE infrasctructure

Author: Monique Petitdidier[1]

Co-author: David Weissenbach [1]

[1] *CNRS/IPSL*

Corresponding Author: monique.petitdidier@cetp.ipsl.fr

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

Earth Science (ES) is an all-embracing term for sciences related to the planet earth covering a large and diverse user community, Academy, organisation and industry. Since 2000 within DataGrid and CrossGrid ES applications from various domains have been ported on a Grid infrastructure. Examples of thematic area are atmospheric chemistry by satellite and simulation, climate, seismology, hydrology, geology, geophysics…

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

Many ES applications have been ported on EGEE. It would be important to provide some overview and describe the requirements related to each application, the solution adopted and the interest to use the Grid. The examples shown concerns the ozone chemistry, climate, earth quakes, seismology simulation, exploration of the geoscope database, hydrology (flood and water management), Geocluster, a company software….and many others
In conclusion it would mention the key Grid services needed for those applications and the ES expectation for other key services.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

The interest to use a Grid infrastructure for the ES community is related to problems difficult to solve on a local or national computing infrastructure even high-power one. The typical applications for which Grid presents a large interest are related to the use of statistical approaches (monte Carlo method, ensemble of jobs..), to sharing data or algorithm, to performing a very large number of independant jobs that permits to have a rapid solution and also to take advantage of a large number of CPUs a needed at one moment and not on a routine base.
T ES applications, already ported, provide some results published in international journal and conference proceedings and included in PhD report. Those results are a mean to convince the ES community of the potentiality of the Grid infrastructure like EGEE.

164

Job management and control in Earth Science

Author: Viet Tran[1]

[1] *Institute of Informatics, Slovakia*

Corresponding Author: viet.ui@savba.sk

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

Earth Science (ES) covers a large range of topics related to the solid earth, atmosphere, ocean and their interfaces as well as planet atmospheres and cores. Examples of ES research areas are meteorology, hydrology, geology and geophysics. DEGREE is a Specific Support Action (SSA) project which aims to promote GRID throughout a large and diverse Earth Science community, in order to increase the awareness and uptake of GRID technology and infrastructure by EU Earth Science Industry and Res

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

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. Most important missing technologies are near-realtime job executions, that are required by applications dealing with risk managements, and QoS for applications running in full operation modes. The test suites have been created with two applications: Centroid Seismic Moment Tensor (CMT) and Flood Forecasting Simulation Cascade (FFSC). Each ES application consists of the application software, data, database schemas, documentation and contact points for technical and scientific support. The test suites can help developers to check if the their Grid middleware and tools can satisfy the ES requirements.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

ES applications deal with numerical modeling, which requires vast amounts of computational powers. Grid would be the key technology for providing such computational power for ES applications. How-

ever, despite the enthusiasm of the Earth Science pioneers, exploitation of Grid technologies in ES is not a trivial task. Between the Grid and the ES applications, there persists a significant gap due to complex computing protocols in Earth Science. In the DEGREE project, the requirements of ES applications have been collected and analyzed, and the missing technologies required for full operation of ES applications in Grid infrastructures are identified. We also provide suggestions for improvements and test suites with well-documented test specifications and real applications plus data to the Grid developers.

165

Integrating Windows Compute Cluster Server 2003 in GRID Environment

Author: Philipp Fuchs[1]

[1] *Instituto de Fisica de Cantabria CSIC*

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

In 2006 Microsoft entered the High Performance Market with Windows CCS 2003 (Comp. Cluster Server) and until now it has been rather separated from the Grid. But what if Grid users want to run Windows Applications on the cluster? The prototype, which will be described below, shows a way of how to integrate CCS into the Grid and how to provide a convenient way to the users to access both Clusters. Then, scientists will be able to submit jobs to both Grid and Windows CCS 2003 from the same UI.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

Key services of this prototype are Globus-Gatekeeper and the modified job Scheduler on Linux side. They will run on a special Computing Element which has a dedicated job-manager for every Windows Application that shall be executed. Jobs for the Windows Cluster are submitted through the globus-job-run command and this makes sure that the job gets to the correct CE.
The job scheduler will need to submit information to a Web Interface which runs on a Web Server (Internet Information Server) on the Windows Head Node. From there the scheduling mechanism of Windows CCS takes care of the execution.
However, at the moment there are still issues about the polling for job status and cancelling jobs. In addition to that, this version of the prototype only forwards command line input. Future versions of this prototype will improve this and provide a way to get the output of Windows Application to the User Interface.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

The added value is that scientists will now be able to submit jobs also to a Windows Cluster, in case they need to run simulations on a Windows platform. In addition to that they don't have to leave their User Interface, to which they are already accustomed. Everything will be accessible from a single point of entry.

166

Grid-enabled approach to design and execution of virtual experiments through workflows

Authors: Ivan Porro[1]; Livia Torterolo[1]; Luca Corradi[1]; Marco Fato[1]

[1] *Bio-Lab, DIST, University of Genoa*

Corresponding Author: pivan@dist.unige.it

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

According to the Virtual Physiological Human paradigm, integration is often required among different levels and disciplines.Due to such an integrated approach, clinical and genomic studies tend to converge in more complex multiscale biomedical experiments to open the way to new research goals. Virtual experiments on the grid can be used to assess and validate models describing human or animal biological subsystems, even if they require large computing or storage resources.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

Starting from these reasons, a Grid based platform, the BMPortal (Bio-Medical Portal), has been developed as a core block that can rely on different specific Grid environments able to integrate data storage, retrieval, and analysis in several fields of biomedicine. Currently available middleware is gLite but a native Globus and a Nordugrid plugin are planned.Using BMPortal, three different approaches can be followed: first, developers can implement entire plugin for a given middleware or only interfaces/services on available infrastructures. Then, "power users" may implement application and research tools based upon the atomic component available, i.e. combining workflow of services. Finally, end users (clinical, research) only connect to the portal to use the available high-level interfaces.This approach -based on a strong work on data integration- allows contributions as SOAP services from third party developers improving interoperability.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

Both for storage of distributed biomedical data and metadata, and for access to distributed analysis tools, a Grid based approach may provide a shared, standardized and reliable solution. Using certificates and the security infrastructure of the Grid, applications can expose data and resources according to fine grained user permissions. This allows developers to implement authentication and authorization schemas easily and on top of industry standard tools (GSI). This is useful in three main situations: (a) constrain access to biomedical data according to strict security and privacy policies, (b) constrain access to storage and computing resources according to given policies and (c) allow for accounting, in the perspective of a service based environment.
The complexity and heterogeneity of data and services should be hidden from unpractised users: the adoption of a Grid portal and workflow execution tools can help developers to implement really user friendly applications.

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.:

Several use cases have already been implemented and they could be demonstrated through the connection to a Grid portal available on Internet. A standard web browser and Internet connection are the only tools required to show the demo.
As example of atomic service, we are able to show a tool for the analysis of gene expression microarray data performed by a grid version of the dChip software ported on the gLite, using GFAL APIs for remote data access through the Grid.
Then the demo could show how atomic services like dChip can be combined in more complex virtual experiments: workflow will be executed and monitored on the portal using client tool Taverna and BMPortal authentication. Workflows on the portal can be also exposed as normal atomic services themselves thus allowing an unprecedented flexibility in combining available tools recursively.

Cosmological applications in the GRID environment: detection of compact sources and non-gaussian signatures in data from the ESA Planck satellite mission

Author: Marcos Lopez-Caniego[1]

Co-authors: Diego Herranz [1]; Enrique Martinez-Gonzalez [1]; Patricio Vielva [1]; Rita Belen Barreiro [1]

[1] *IFCA-CSIC*

Corresponding Author: caniego@ifca.unican.es

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

In 2008 ESA's Planck satellite will be launch. The main objective of this mission is to produce a map of the cosmic microwave background radiation (CMB), a relic radiation from the Big Bang. To study this map, the compact source emission from distant galaxies and clusters of galaxies must be detected and extracted. Similarly, the study of non-gaussian signatures provides additional information about the underlying sistematic effects or even a primordial departure from gaussianity of the data.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

We have been using part of Astronomy & Astrophysics cluster to run this kind of tests. We are testing the performance of two different filters, the Matched Filter and the Mexican Hat Wavelet 2, both operating on flat patches. The input images to analyse are aprox. 200 MB and are stored in the local storage element. Then in every test we modify one of the parameters and run the application for the nine input simulated maps, corresponding to the nine frequencies of Planck. The process takes about 90 minutes to run in a 3.5 GHz intel Xeon processor. The output is a catalogue of detected objects. This catalogue of objects is also converted into a map in sperical coordinates and displayed to compare it with the input map. Depending on the output, the input parameters may be modified and then the application is run again. A paralelization of this application will help decrease the time it needs to run from 90 minutes to just a few minutes. This will increase the interactivity of the proces

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

There are several techniques used to detect point sources and non-gaussian signatures in maps of the CMB. In particular, for the case of point sources, most of these techniques are based on linear filters, matched filter, wavelets, etc. Within the Planck satellite collaboration a set of realistic simulations of the sky at microwave frequencies has been produced. This sky model can be used to test our detection techniques extensively. The techniques that we use operate on flat patches of the sky, whereas the output maps of Planck will be in spherical coordinates. Therefore, each map has to be proyected into several hundreds of patches, each of which needs to be filtered to increase the signal-to-noise ratio of the point sources, which in turn increases the number of detected objects. There are several parameters that may be modified such as the image size, pixel size, overlapping, etc, and only using a GRID infrastructure these techniques can be thoroughly tested in a reasonable time

168

On low energy nuclear activities in EGEE

Author: Faustin Laurentiu Roman[1]

[1] *IFIN-HH*

Corresponding Author: froman@tandem.nipne.ro

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

High accuracy nuclear data is a strong requirement for reliable, safe and cost effective modern nuclear facilities and the only method for obtaining all the required quantities unitary is through nuclear data evaluations by nuclear model calculations. Recent nuclear codes developments employ microscopical approaches combined with large-scale global calculations that are inherently computational intensive and thus being suited for Grid environment.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

The outcome was firsthand the reduced computational time but also served as a practical experience of porting low-energy computer codes to Grids.
The proven success should stimulate the adoption of Grid technologies for other low energy nuclear activities, e.g. uncertainty estimation, transport and reactor design computer codes, and therefore an eventual Virtual Organization to support such activities in EGEE would prove suitable.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

A modified version of SCAT-2 nuclear code that searches 'best fit' nuclear optical model parameters - SCAT2MIN - was ported to the grid environment using the EGEE Fusion VO resources. This work has showed that it is rather easy to integrate low-energy nuclear computer codes into modern computational environments and perform large scale computations with good performance. There are still some issues regarding the success rate of the jobs, a solution investigated being the use of DIANE framework.

169

The CMS Data Transfer System: scaling up to a reliable LHC scenario

Author: José Flix[1]

Co-authors: Artem Trunov [2]; Brian Bockelman [3]; Daniele Bonacorsi [4]; Derek Feichtinger [5]; Jens Rehn [2]; Lassi Tuura [6]; Pavel Goglov [2]; Ricky EGELAND [7]; Simon Metson [8]; Tony Wildish [2]

[1] *PIC, Barcelona, Spain*

[2] *CERN, Geneva, Switzerland*

[3] *University of Nebraska–Lincoln, Lincoln, NE, USA*

[4] *INFN–CNAF, Bologna, Italy*

[5] *PSI, Villigen, Switzerland*

[6] *Northeastern University, Boston, MA, USA*

[7] *University of Minnesota – Twin Cities, Minneapolis, MN, USA*

[8] *University of Bristol, Bristol, UK*

Corresponding Author: jose.flix.molina@cern.ch

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

The CMS experiment needs to sustain uninterrupted high reliability, high
throughput and very diverse data transfer activities as the LHC operations start. PhEDEx, the CMS data transfer system, is responsible for the full range of the transfer needs of the experiment. At the infraestructural level, PhEDEx interacts with the local storage, grid file transfer services, the CMS dataset bookkeeping system, the dataset location system and a site-local file catalogue.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

As said, we have sustained production transfers exceeding 1 PB/month for several months and have demonstrated core system capacity several orders of magnitude above expected LHC levels.

The Phedex application is build on top of key services like, for example, the grid file transfer service (FTS). In the poster we will also cover some of the new modifications included in Phedex in order to accomodate to recent FTS 2.0 release.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

Phedex provides an interface for CMS and site data managers to manage and monitor data placement decisions, schedules transfer requests for execution, and dispatches file transfers to underlying grid file and storage management services.

In order to produce the system with confirmed capability to meet the objectives, the PhEDEx data transfer system has undergone rigourous development and numerous demanding scale tests. We have sustained production transfers exceeding 1 PB/month for several months and have demonstrated core system capacity several orders of magnitude above expected LHC levels.

We describe the level of scalability reached, and how we got there, with focus on the main insights into developing a robust, lock-free and scalable distributed database application, the validation stress test methods we have used, and the development and testing tools we found practically useful. We expect sharing this to be useful for developers building robust applications on Data management

170

Grid-enabling a Problem Solving Environment: Implementation and Everyday Use

Authors: Efstratios Gallopoulos[1]; Giorgos Kollias[1]; Konstantinos Georgiou[1]

[1] *University of Patras*

Corresponding Author: gdk@hpclab.ceid.upatras.gr

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

We have conveniently integrated Grid access capabilities within Jylab, a Problem Solving Environment (PSE) we designed. This PSE already provides interactive access to a suite of Java libraries targeting the applied scientist in general however also serving his need to compute with data and peer machines available over the Internet. Jylab can be dynamically installed at Grid nodes, thus being the execution environment for the scenaria we tested(e.g. Web search engine functionality over the Grid)

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

Grid-enabling and Grid-ifying (dynamic installation) of our PSE are mature enough for submitting jobs to Grid nodes (used mainly HellasGrid infrastructure). The applications tested are coded in Python (with the extra benefit of calling into Java libraries). E.g. we have suitably packaged Nutch search engine (http://lucene.apache.org/nutch/) with the scripts needed to crawl and index predefined small sets of URLs down to some linking depth from within Grid nodes; all indices remotely constructed are currently downloaded to the submission machine -not necessarily the User Interface machine, ssh tunnels can be transparently employed- and merged for subsequent visualization using Jung library (http://jung.sourceforge.net/) or queries using ready-to-use, servlet-based Web interface, all without leaving the comfort of the PSE environment. Grid interaction scripts are light-weight Python scripts wrapping the CLI (Command Line Interface) of standard available Grid toolkits (mostly gLite).

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

We consider the simplification of access to Grid resources for the applied scientist as a priority task. By Grid-enabling our PSE and also providing for its dynamic installation at Grid nodes we also propose a simple example of the implementation path that could be taken by more established PSEs (like Matlab, Scilab, etc) for actually exposing the Grid to applied scientists lacking the skills of a computer engineer, however in need of aggregate computational resources for their applications. In particular our Grid applications are taken from the field of Internet algorithmics - computing with data and metadata (e.g. link structure) - collected directly from the Web. We particularly follow the interesting pattern of exploiting the network bandwidth available to Grid nodes for collecting this info in a distributed manner (by interactively submitting a set of Grid jobs from within our PSE) and then merging it for further processing (e.g. with numerical linear algebra kernels)

171

Deploying Job Provenance: First Application Experience

Authors: Ales Krenek[1]; Frantisek Dvorak[1]; Jan Kmunicek[1]; Jiri Filipovic[1]; Jiri Sitera[1]; Ludek Matyska[1]; Milos Mulac[1]; Miroslav Ruda[1]; Zdenek Salvet[1]; Zdenek Sustr[1]

[1] *CESNET*

Corresponding Author: kmunicek@ics.muni.cz

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

Here we demonstrate a generic gLite middleware service - Job Provenance - providing a backbone for custom application solutions requiring job cataloque capabilities. The target communities vary from small research groups trying to set up their own solution fulfilling their specific needs up to potential usage in well established communities like the high energy particle experiments.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

We demonstrate JP usage in two cases. In the first one, management of large parametric studies in computational chemistry (molecular docking) JP, together with a thin graphical front-end, was used to build the job catalogue from scratch. It allowed the researchers to easily manipulate computational jobs (input modification, jobs resubmission), to search and selected desired (finished/non-finished, aborted) jobs and finally utilize specific plugins for results presentation (e.g. visualization). In the second case, we augmented production jobs of the Atlas experiment to interact with JP, yielding functionality similar to Atlas ProdDB but with emphasis in job history. We routed part of the Atlas production traffic to JP (approx. 1100 jobs/day) as well as performed stress tests on snapshot of these jobs in order to demonstrate JP readiness for production deployment.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

gLite Job Provenance (JP) is a generic job catalogue service keeping long-term track of execution of Grid jobs. It provides a sophisticated machinery to support application and user annotations of the Grid computational jobs. Furthermore it provides data mining over the raw data and annotations. Being a standard part of gLite middleware stack it

offers continuous and guaranteed service to store all the primary information. On top of storage facilities, Job Provenance Index Servers allow for efficiently looking for expected and unexpected patterns within the stored information through user queries. While JP can be used directly by all gLite middleware users, specialized job catalogues can be built with moderate effort compared to custom solutions (custom job catalogue development) taking considerable effort.

172

File Transfer Service and CMS data transfer optimizations at PIC Tier-1 center

Author: José Flix[1]

Co-authors: Artem Trunov [2]; Carlos Borrego [1]; Christian Neissner [1]; Daniele Bonacorsi [3]; Gonzalo Merino [1]; José Hernández [4]; Simon Metson [5]

[1] *PIC, Barcelona, Spain*

[2] *CERN, Geneva, Switzerland*

[3] *INFN–CNAF, Bologna, Italy*

[4] *CIEMAT, Madrid, Spain*

[5] *University of Bristol, Bristol, UK*

Corresponding Author: jose.flix.molina@cern.ch

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

The CMS experiment needs to sustain uninterrupted high reliability, high throughput and very diverse data transfer activities as the LHC operations start. PhEDEx, the CMS data transfer system, is responsible for the full range of the transfer needs. At the infraestructural level, PhEDEx interacts with the grid file transfer services. Optimisations on SWE region and plans for new FTS-interaction are intended to be shown in this poster.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

CMS has demonstrated to have sustained production transfers exceeding 1 PB/month for several months and a core system capacity several orders of magnitude above expected LHC levels.

This year 2007, the PIC Tier-1 center has succesfully transferred about 1 PB of data. The Phedex application is build on top of key services like the grid file transfer service (FTS). The modifications and tuning of the Phedex system to the FTS service have helped to improve the quality, stability and throughput from/to PIC Tier-1 center. In the poster we will also cover some of the new modifications to be included in Phedex in order to accomodate to recent FTS 2.0 release to boost the Phedex performance.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

Phedex provides an interface for CMS and site data managers to manage and monitor data placement decisions, schedules transfer requests for execution, and dispatches file transfers to underlying grid file and storage management services.

In order to produce the system with confirmed capability to meet the objectives, the PhEDEx data transfer system has undergone rigourous development and numerous demanding scale tests. While being involved in these global CMS tests, the PIC Tier1 center has carried out several tests in order to improve and optimise the Phedex interactivity to site FTS server. This has improved the

quality, stability and throughput between PIC and SWE Tier2 sites, CERN, Tier1 and non-regional Tier2 centers.

The impact of these tests has collapse to new ideas and plans to develop a new FTS backend for the Phedex Data System, that will be reviewed as well in the poster. We expect sharing this to be useful for developers of data management applications.

173

Portals for Earth Science applications

Author: Camiel Plevier[1]

Co-authors: Julian Linford [2]; Mathieu Lonjaret [3]

[1] *Dutchspace*

[2] *SERCO*

[3] *CNRS/IPGP*

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

Earth Science (ES) is an all-embracing term for sciences related to the planet Earth, covering a large and diverse user community. Since several years the applications need more and more to have access to different large and heterogenous sets of data, in general via web portals, and to intensive computing facilities.
DEGREE is a consortium of ES partners aiming at promoting the uptake of Grid technology in ES. DEGREE has in charge to define requirements for future ES portals on Grid.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

From a very large number of existing ES portals, a survey was done to focus on, analyze and document those of particular interest and relevance. The focus is on ES portals which are employing, to a greater or lesser extent, some combination of the following relevant technologies and methodologies, Grid, e-collaboration, Service oriented architecture, semantic web and Ontology.
This survey provides a clear picture of wide range of emerging technologies in ES portals. The high-level of web-based portal services, provided to end-users, permit to define requirements to implement them on gLite and develop new ones.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

Large sets of ES data are available and distributed all over the world. The data come from satellite, also from ground-based network and sensors aboard balloon, aircraft, and/or rocket. An urgent need has been the organisation of the data, their access with respect to the data policy, and in some cases tools to define the workflow of the application. Some web-based ES portals have appeared with different tools for discovery, download, and local computation . Grid infrastructure offers the capability to explore those large sets that was not done previously due to the computing power limitation and to deploy complex simulation based on a combination of various large sets of data.
The combination of data web services and Grid via a portal will open new fields and discovery by the full exploration of the data. It is not limited to Earth science

174

The Health-e-Child Gateway and Case Reasoner, a Concrete Implementation and Use Case of a gLite-based Healthgrid for European Paediatrics

Authors: David Manset[1]; Martin Huber[2]

[1] *Maat Gknowledge*

[2] *Siemens*

Corresponding Author: dmanset@maat-g.com

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

Health-e-Child (HeC) endeavours to respond to the increasingly pressing demand to fully integrate and exploit heterogeneous biomedical information for improved clinical practice, medical research and personalized health care. As an integrated project of the Sixth Framework Programme of the European Commission, HeC brings together three major paediatric medical centres with several European companies, university groups and research institutions specialized in grid-based (bio)medical technologies.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

Early January, HeC has demonstrated at the Hospital Necker (APHP - Necker Enfants Malades in Paris), its first prototype of the so-called "Health-e-Child Gateway" software, a secure portal of online medical services for storing and manipulating patient data from clinical records to medical images. Based on leading edge technologies such as the gLite grid middleware, this advanced prototype highlights the maturity of such complex technologies and their relevance when applied to health care. The HeC prototype is, at the time of writing, the result of nearly two years of active research and development in collaboration with an important community of clinicians and domain experts, which has been developed and deployed inside the HeC private gLite-based grid infrastructure. Amongst the first technological contributions of the project, a security prototype was delivered as well as innovative domain specific client applications.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

Following the grid vision and aiming at enriching it, the HeC project is developing an integrated health care platform for European Paediatrics that makes use of the underlying grid infrastructure through a series of biomedical domain specific services. Given the objectives and the scope of the project, the grid has been naturally selected as a low cost and cost-efficient facility which can appropriately support the end-users needs in terms of computational power (e.g. for executing medical image processing and data mining algorithms), storage capacity (e.g. for storing thousands of patients' scans and associated clinical reports) and security (e.g. for guarantying a satisfactory security level when accessing patient data). The grid is also used as a technological glue to interconnect the several partners of the project, in particular for dealing with their highly heterogeneous Information Systems. Beyond connectivity, the grid also serves as a collaboration vehicle.

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.:

HeC proposes a highly graphical demo of its platform and gLite-based infrastructure, which requires two workstations (to be provided by HeC), a video-projector and a simple Internet connection (without proxy). Demonstration flow: user logs in the system thanks to a physical USB Key, which he connects to his workstation. The Key contains his credentials and basic portable applications to enter the HeC VO. Immediately after authentication, the so-called Peer-to-Peer Patient Privacy network (P2P3) is established in the local network to which the clinician's workstation is connected, which enables him to store and share patient private information. The clinician then accesses patients' medical records from various clinical centres. In a second time the clinician starts the HeC Case Reasoner application, which enables to explore and compare the records related to several patients, and to visualize their place in

the corresponding distribution.

175

Using Grid for Micro-Architecture Research

Author: He Liqiang[1]

Co-authors: Christiana Ioannou [1]; Marios Kleanthous [1]; Yiannakis Sazeides [1]

[1] *University of Cyprus*

Corresponding Author: liqiang@cs.ucy.ac.cy

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

The Xi Computer Architecture Lab (University of Cyprus) research projects are memory hierarchy optimizations, speculation techniques, reliability methods for hard errors and mechanisms to reduce temperature problems on multi-cores. A routine activity in our group, and most architectural research teams, is the running of simulation experiments to investigate the potential of new techniques we design.The EGEE provides a high throughput powerful computing platform that matches our simulation needs.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

A good interface for job submission, management (resubmit, cancel e.t.c) and retrieving results is necessary. We developed scripts for this but a complete suite of such services can be very useful since the grid intend for high throughput that assumes 1000 of jobs running simultaneously by each user.Using the Storage Element to save the results has the advantage of not being affected by any user interface machine failures but it requires the user to delete any unused files to avoid flooding the SE. Initializing the proxy on every submission avoids any unexpected proxy expiration and job failures. Specifying different requirements for different job categories can increase throughput if slow jobs go to faster but smaller clusters and faster jobs to slower but bigger clusters (more nodes).A more detailed report on job failures will be very useful to understand and deal with it.Sometimes grid seemed very unreliable.Usually only 80% of the jobs submitted were finished in less than 12 hours.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

The grid provides a powerful a computing resource to perform more quickly and comprehensively design space exploration that is crucial for determining both timely and good solution points in the design space we explore. In general, every set of experiments in our projects may require several hundreds of simulations, due to a plethora of interacting parameters, with individual runs requiring several anything from several hours to few days.

There are jobs that require certain hardware configurations such as large memory size and number of processors. With grid functionality, we can define criteria for the resources that best match our needs and then run in parallel as many jobs as available resources to achieve a high throughput simulation methodology. The usual quick turnaround from submission to completion enables us to become more efficient to
determine good solution points.

176

Experience in testing the gLite workload management system and the CREAM computing element

Author: Vincenzo Miccio[1]

Co-authors: Andrea Sciabà [1]; Simone Campana [1]

[1] *CERN-PSS*

Corresponding Author: vincenzo.miccio@cern.ch

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

The LHC experiments will study the physics of p-p interactions at a center-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. They use the EGEE infrastructure to perform their offline computing activities (data reconstruction and analysis, Monte Carlo simulation, calibration and alignment, data replication).

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

This report describes the experience of the EIS team in testing the gLite Workload Management System and the CREAM computing element. 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. Tests aimed at determining if the WMS and CREAM met the acceptance criteria defined by WLCG were successfully performed, demonstrating a level of performance compatible to the LHC experiments requirements.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

The LHC experiments are undoubtedly 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 2008. It is expected that each experiment will need to run hundreds of thousands of jobs per day. The Grid is expected to fulfill the experiment needs in terms of available resources and middleware.

177

New Developments in the gLite AMGA Metadata Catalogue

Authors: Ali Javadzadeh Boloori[1]; Birger Koblitz[2]

[1] *CERN/KTH*

[2] *CERN-IT*

Corresponding Author: birger.koblitz@cern.ch

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

The gLite-AMGA metadata catalogue is developed by NA4 to provide simple relational metadata access for the EGEE community. AMGA is extensively used by the biomedical community to store medical images metadata, digital libraries, in HEP for logging and bookkeeping data and the climate community. This presentation focuses on presenting new developments in the catalogue, like metadata replication, a new WS-DAIR compatible interface, making existing databases accessible and the gLite 3.1 integration

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

We expect that the new developments in AMGA will allow the biomedical community to evolve their current much centralized setup to a truly distributed environment, without making compromises on the security. In a replicate environment security 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. In particular the experiences gained by the Health-e-Child project should be very valuable.

The WS-DAIR compatible interface to AMGA, together with the new import feature, will make access to large existing databases and integration with other data sources much simpler and should make it much simpler to port existing applications to the Grid, when they need relational data access.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

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. We will give an update on the latest additions to the replication features of AMGA and also report on how AMGA's capabilities have been validated and used by the Health-e-Child project.

Another ongoing development in AMGA is the addition of a WS-DAIR standard compliant web-service interface, which should allow improved interoperability with other WS-DAI grid services, in particular in the field of earth observation with its heterogeneous data sources. Here, also the addition of an importation feature should be of great significance as it allows to directly access pre-existing databases without copying data.

Finally we will give an overview of the ongoing integration of AMGA into the gLite 3.1, which will further eas

178

Ganga – or how I lost the fear of running my app on the grid

Authors: Adrian Muraru[1]; Andrew Maier[1]; Anton Lechner[2]; Hurng-Chun Lee[3]; Jakub Moscicki[1]; Patricia Mendez[1]

[1] *CERN*

[2] *Atominstitut der Österreichischen Universitäten and CERN*

[3] *ASGC, CERN and University of Innsbruck*

Corresponding Authors: patricia.mendez@cern.ch, jakub.moscicki@cern.ch, andrew.maier@cern.ch

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

The computational and storage capability of the Grid are attracting several research communities, also beyond HEP. Ganga is a lightweight Grid job management tool developed at CERN. It is a key component in the Distributed Data Analysis for ATLAS and LHCb. Ganga's open and general framework allows to plug-in applications, which has attracted users from other domains outside HEP. In addition, Ganga interfaces to a variety of Grid and non-Grid backends using the same, simple end-user interface.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

From January to August 2007 Ganga has been used by around 850 users (500 ATLAS, 150 LHCb, 200 other user communities) and has been installed locally in more than 50 sites around the world. Recently also the educative aspect of Ganga has been recognized and Ganga has become a part of the official EGEE

tutorials. Contrary to other portals or tools, Ganga is not limited to specific VOs or infrastructures allowing the new user to immediately use the full EGEE infrastructure.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

Ganga has already gained widespread use, the incomplete list of applications using Ganga include:
- Image processing and classification (developed by Cambridge Ontology Ltd.)
- Theoretical physics (Lattice QCD, Feynman-loop evaluation),
- Bio-informatics (Avian Flu Data Challange)
- Geant4 (Monte Carlo package)
- HEP data analysis (ATLAS, LHCb)

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 "shielding" the applications from the details of the Grid. 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.

Ganga supports a large number of backends to which jobs be submitted too without the underlying knowledge of each of these backends. This allows users to use more than one computational resource available to them.

The incomplete list of supported backends includes:
- EGEE gLite middleware
- NorduGrid ARC middleware
- Condor and Cronus (Condor/G)
- Various batch systems (LSF,PBS,SGE)
- DIRAC (LHCb production system)
- PANDA (ATLAS production system)

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.:

We will show how an application is gridified and how a user can immediately profit from the Grid using Ganga. We demonstrate application cases outside of the initial scope of High Energy Physics in which the
tool has been developed. The Grid added values include:
- easier transition to the Grid environment for the end-users
- integration of Grid and local resources which is required in many scientific activities
- education and dissemination
- technology exchange across different application domains

The demo and poster will show the ease of this transition from the traditional submission, running on single machines or a local batch cluster to running on the full EGEE infrastructure. Using a concrete example from Lattice QCD, we will show how the application was girdified and how subsequently 30 CPU year have been harvested in 1 week of running the application on the Grid.

179

Genome Wide Association Studies of human complex diseases with EGEE

Author: Alexandru Munteanu[1]

Co-authors: Cecile Germain-Renaud [2]; David Tregouet [1]; François Cambien [1]

[1] *INSERM UMR S 525*

[2] *LRI*

Corresponding Author: cecile.germain@lri.fr

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

As part of the research conducted at INSERM U525, the THESIAS software was created in order to analyze statistically, associations between gene polymorphisms and diseases. Given a data set containing the genotypes of case and control individuals, THESIAS measures haplotype frequencies combining several polymorphisms and associations with the disease. This research can lead to the identification of new causes and mechanisms of disease of potential therapeutic and preventive interest.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

For whole genome haplotype analysis we decided to test the EGEE grid. EGEE provides enough computational power to perform these analyses. The EGEE infrastructure was used in our project and proved to be very effective. As a proof of principle, we have analyzed thousands of SNPs for their association with cardiovascular disease in thousands of individuals, and found that the performance improvement relative to a modern PC is about 300. Nevertheless, better reporting and analysis of failures would be very desirable.
A user interface on top of the gLite user interface has been created to simplify batch job submissions, monitoring and automatic resubmission of failed jobs. We have created this interface, which is very easy to use, and allows non computer scientists to use the EGEE grid ressources. We plan to analyse the entire genome as soon as we will have the data.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

Until now this kind of analysis was limited to single genes and a few polymorphisms (<25). The recent availability of DNA chips allowing to genotype hundreds of thousands of polymorphisms implies a change in scale in the necessary computations.
The complexity of identifying which DNA sequences variations(SNPs) are associated to a disease on the entire human genome increases exponentially with the number of SNPs. Frequencies of combinations of multiple SNPs must be estimated. Ideally all the possibilities would be analyzed. With at least 10 millions SNPs on the human genome, calculating all the combinations is hardly imaginable. Fortunately, SNPs located close to each other are frequently tightly correlated (linkage disequilibrium); they define haplotype blocks that can be tagged by a limited number of marker-SNPs. The most recent genotyping arrays with 1 million marker- SNPs are highly informative. The computation may be further reduced by investigating haplotypes in a sliding window

180

Implementation of Geospatial services on Grid platform for Civil Protection applications

Authors: Marco Verlato[1]; Paolo Mazzetti[2]; Stefano Nativi[2]; Valerio Angelini[2]

[1] *INFN-Padova*

[2] *CNR-IMAA*

Corresponding Author: marco.verlato@pd.infn.it

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

The CYCLOPS (CYber-Infrastructure for CiviL protection Operative ProcedureS) EU project started the 1st June 2006, with the main goal to bridge the gap between Grid and Global Monitoring for Environment

and Security (GMES) communities making Civil Protection (CP) people be aware of the services provided by Grid infrastructures, and, at the same time, letting Grid researcher to be aware of Civil Protection specific requirements and service enhancement needs.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

In the geoscience community the Web Coverage Service (WCS) OGC standard is used to share and retrieve geospatial data as coverages,digital geospatial informations representing space-varying phenomena.
In this context a Grid-enabled WCS prototype has been developed:it exposes a standard WCS interface to the web,while it is able to process a user request in a distributed Grid environment. It evaluates the request, splits it in an arbitrary number of sub-requests, generates a JDL le 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. This approach allows to handle a huge set of geographically distributed datasets and to process an arbitrary number of high demanding requests. In the same way will be implemented other OGC services,such as WMS,WPS,WFS.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

Civil Protection procedures, as well as most of GMES applications, require a strict integration with research infrastructures providing heterogeneous and distributed resources useful in the full cycle of emergency situations,from forecasting to post-emergency assessment. Moreover this kind of activities typically involves many different actors who need to share resources in a coordinated and effective way.
Consequently the adoption of a Grid-based infrastructure seems a natural choice. Current Grid platforms are mainly designed to support research and applications requiring intensive processing and data management.
The CYCLOPS project aims to define a set of services that are essential for GMES and in general for the Earth Science community.
On top of the Grid platform an intermediate layer of services will be defined. Such services will make use of well known solid standards and will be designed to access and process data using the advanced grid capabilities required by CP.

181

Large scale Bioinformatics portal on EGEE Grid : the GPSA example

Author: Christophe Blanchet[1]

Co-authors: Alexis Michon [1]; Christophe Combet [1]; Gilbert Deléage [1]

[1] *CNRS IBCP*

Corresponding Author: christophe.blanchet@ibcp.fr

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

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 predefined and connected workflows.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

The bioinformatics portal GPSA was improved with the adding of MPI capabilities for some of its applications: BLAST and CLUSTALW. This work has been done with the help of the MPI group , and according to the recommendations produced during the activity of this group. CNRS IBCP has followed these recommendations, and some time adapted them, to be able to run MPI jobs on the EGEE productin platform. The bioinformatics applications that can now be ran on EGEE grid are mpiBLAST and mpi-CLUSTALW. There are both famous applications in the Bioinformatics field and are ran daily by lots of scientists. These MPI tools have been validated on the EGEE production platform mainly on the current most usable MPI-site at IN2P3 LAL in Orsay. Some other tests has been done on few others MPI-sites but not with the same success rate.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

In GPS@, we simplify the grid analysis query: GPS@ Web portal runs its own EGEE low-level interface and provides biologists with the same interface that they are using daily in NPS@. They only have to paste their protein sequences or patterns into the corresponding field of the usual submission Web pages. Then simply pressing the "submit" button, they get the results of executing these jobs on the EGEE grid platform. All the EGEE workload management is encapsulated into the GPS@ low-level layers: submitting, monitoring and getting the results of the bioinformatics jobs.
Running MPI-parallelized bioinformatics applications on EGEE production platform was still a challenge, and especially MPI-BLAST that has raised in the past several tricky points from MPI and EGEE sites configuration. One other applications is the CLUSTALW multiple alignment engine. Both programs are C/C++ writen, and could be compiled with gcc. (GCC release available from SL3.0.9)

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.:

GPS@ grid web portal (Grid Protein Sequence Analysis, http://gpsa-pbil.ibcp.fr) is the port of our Bioinformatics integrated portal, the NPS@ protein analysis portal, and would provide the biologist with a user-friendly interface for the GRID resources (computing and storage) made available by the project EGEE (2004-2008).
This genomic grid user interface hides the mechanisms involved for the execution of Bioinformatics analyses on the grid infrastructure. The bioinformatics algorithms and databanks have been distributed and registered on the EGEE grid and GPS@ runs its own EGEE interface to the grid. In this way, GPS@ portal simplify the Bioinformatics grid submission, and provide biologist with the benefit of the EGEE grid infrastructure to analyze large biological datasets with parallelized bioinformatics software: for example doing a large multiple alignment with CLUSTALW, or analysing large protein sequence sets, with BLAST, on the Grid.

182

Development of Neuraminidase Inhibitor by grid-enabled virtual screening

Author: Doman Kim[1]

Co-author: Ying-Ta Wu [2]

[1] *Chonnam National University*

[2] *Academia Sinica Taiwan*

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

The community targeted is the drug discovery community. The WISDOM initiative aims at lowering the cost for finding new drugs by developing the use of grid technologies in the drug discovery process. Its first success was to demonstrate how grids can speed up and reduce cost for large scale in vitro screening.

But the scientific impact of this approach depends on its biological relevance which is addressed in this abstract.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

We report here on the in vitro tests which were carried on to validate the results obtained in silico. The recombinant NA from H5N1 influenza virus strain A/Vietnam/1203/04 was successfully expressed and purified in this experiment. Neuraminidase activity was determined using (MU-Neu5Ac) as a fluorogenic substrate. Inhibition activity of NA was determined by incubating enzyme solution with 40 mM sodium phosphate buffer (pH 7.2), MU-Neu5Ac [2'-(4-methylumberlliferyl)-□-D-N-acetylneuraminic acid], and with or without target compounds as following the measuring the fluorescence using excitation at 365 nm and emission at 450 nm. Compared with Grid score, target compounds were ranked by the degree of inhibition of NA. The results of in vitro analysis demonstrate the relevance of the approach adopted in silico and the potential grid impact to reduce the cost and time of structure-based drug design.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

Virtual screening is about selecting and ranking in silico the best molecules which could impact the target biochemical activity. Screening millions of compounds which are made available due to advances in the combinatorial chemistry takes years and terabytes of storage.
Grid infrastructures are solutions for such large scale experiments.
In April 2006, data challenge for influenza neuraminidase (H5N1) was carried out by Grid-enabled high throughput in-silico screening based on AutoDock and Python programs against 308,585 compounds. As a result, a subset of compounds was identified with putative inhibition activity on neuraminidase.

Plenary / **183**

Ilona Vass, Vice-President of National Office for Reserach and Technology

What the Network can do for the Grid / **184**

GÉANT2 services for users

Author: Emma Apted[1]

[1] *Dante*

Corresponding Author: emma.apted@dante.org.uk

The basic outline is:

PERT (support team for e2e investigations on perf. issues)
perfSONAR/MDM (e2e perf. monitoring)
point to point services (dedicated links)
Premium IP/AMPS (network resource reservations).

What the Network can do for the Grid / **185**

[CANCELLED] Usage of premium network services: the GridCC example

Author: Maria Grammatikou[1]

[1] *Unknown*

Corresponding Author: mary@netmode.ntua.gr

What the Network can do for the Grid / **186**

What the ENOC can do for you

Author: Guillaume Cessieux[1]

[1] *CNRS / CC-IN2P3*

Corresponding Author: guillaume.cessieux@cc.in2p3.fr

Presentation of the services provided by the ENOC to the Grid community

Application Track (Grid Observatory) / **187**

The Grid Observatory: goals and challenges

Author: Cecile Germain-Renaud[1]

[1] *LRI*

Application Track (Grid Observatory) / **188**

Endogenous versus exogenous dynamics and scaling laws in YouTube, Open Source Softwares and Cyber-risks

Author: Didier SORNETTE[1]

[1] *Dept of Management, technology and Economics ETH Zurich*

Grid Observatory / **189**

L&B and Job Provenance services: What do we know about Grid job

Author: Ales Krenek[1]

[1] *MASARYK UNIVERSITY, BRNO, CZECH REPUBLIC*

This talk will present some details on the internal design of two major tools that are relevant for the Grid Observatory purpose.

Grid Observatory / 190

Summary of requirements and contributions

Author: Cecile Germain-Renaud[1]

[1] *LRI*

Grid Observatory / 191

Requirements and contributions roundtable

192

193

194

195

196

197

Business Track / 198

Interactive Media & e-Infrastructures: building a successful future

Network evolution, parallel computing, emerging technologies and distributed computational power are keys to understand the future of interactive media applications. While physical formats will fade away and network will stand stronger, online services will need innovative technology and business models to replace existing mechanisms. Study in those fields and active projects' research work are tremendous opportunities for Europe to get ready for next years' competition.

199

Application Track (Grid Observatory) / 200

Analyzing the Grid using the Grid - case study and possible applications

Business Track / 201

The COMETA consortium and its activities for Grid adoption by Industry in the context of the PI2S2 Project

The presentation will outline the activities of the Consorzio COMETA to foster the adoption of grid technologies by Industry. Special care will be devoted to the porting of the gLite middleware to Windows platforms carried out in collaboration with Microsoft."

GEMS: Grid Empowered Molecular Simulators and CompChem Tools / 202

Ab initio photodynamics calculations on the Grid: approaches and applications

Author: Hans Lischka[1]

[1] *University of Vienna*

Medical imaging / 203

Introduction

Author: Johan Montagnat[1]

[1] *CNRS*

Medical imaging / 204

A Grid Implementation of SPM Analysis for Early Diagnosis of Alzheimer's Disease

Authors: Andrea Schenone[None]; Ivan Porro[None]; Livia Torterolo[1]

[1] *Bio-Lab, DIST, University of Genoa*

Medical imaging / 205

Virtual reservations for Short Deadline Jobs

Author: Romain Texier[1]

[1] *CNRS*

Medical imaging / 206

Structuring Medical Images with the Grid

Author: Ignacio Blanquer[1]

[1] *UPV*

Medical imaging / 207

Workflow-based medical image analysis algorithm assessment

Authors: Diane Lingrand[None]; Johan Montagnat[1]; Tristan Glatard[None]

[1] *CNRS*

Medical imaging / 208

Grid-enabling ThIS (radiotherapy) and CAVIAR (cardiovascular) applications

Author: Hugues Benoit-Cattin[None]

Business Track / 209

European Commission views and instruments on Grid industry uptake

Business Track / 210

Changing intentions - Opportunities for gLite in finance and related industries

gLite has been exceedingly successful as an enabling infrastructure, and has been a massive success in bringing together scientific and technical communities to provide the compute to address previously incomputable problems. Not so in the finance industry.

In its current form gLite would be a business disabler!! There are other middleware tools that solve the finance communities compute problems much better. Things have moved on, gLite had a purpose on first inception – to harness as much compute as possible - and there are moves afoot in the open source community to evolve the technology to address other, more sophisticated needs such as utility and interactive computing. Excelian is well placed as a provider of Grid consultancy services for the Finance community and through its relationship to the EGEE to identify and help exploit opportunities as the research and business worlds converge.

Because of the strong third party presence in the finance industry, such opportunities are far and few between, but they are there, especially as we expand sideways into related verticals such as the smaller hedge funds and energy companies. This talk will give an overview of the barriers to adoption in the finance industry and highlight some of the opportunities offered in this and related industries as the ideas around Grid mature.

Business Track / 211

Interactive Panel Discussion – Conclusions, Wrap-up and Recommendations

Grid Observatory / 212

Usage Patterns

213

VGRID: Virtual Machine Life Cycle Management Portal

Author: Omer Khalid[1]

[1] *CERN*

Corresponding Author: omer.khalid@cern.ch

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).:

VGrid platform's build upon Xen Virtualization to provide secure, isolated and fine-grained environments which could be deploy and terminated on demand. They are very useful in testing/certification of the middle ware where new software patches and releases have to be regularly tested on clean machine. Another activity domain could be where root access is required for the testing/certification on the host machines and making sure that the host machine always stays in a consistent state.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.:

Following projects have been so far using VGRID portal in their activities:

1) gLite certification team members deploy virtual machines via vgrid portal on demand for testing the new software patches and releases. Its been almost a year since its been deployed and rigorously upgraded with new features.

2) gLite data management team is using vgrid portal and virtual machine in their testing and certification activity.

3) ICEAGE Grid Training school have used virtualization for their approach to set-up a grid in a box. They deployed on-demand virtual machines with pre-configured gLite software components and made them available to students to get hands on experience with out any risk of disturbing the host machine or environment.

Presently VGrid portal is been deployed on test environments and later on could be deployed production systems.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications:

Modern day virtualization is enabling consolidation of hardware resources for better utilization, over all resource saving as less number machines required which further lowers administration and maintainance costs. The added advantage of having VGrid portal in test environments would be that productivity could be significantly increased as the man hours and efforts lost previously required for machine re-installations would not be necessary due to the fact that clean, isolated and secure virtual machine environments could be deployed from the VGrid's browser interface on demand.

Such an approach reduces inconsistency on the host systems, and is extremely useful for the communities and activities where hardware and human resources are restricted but requires multiple operating system distributions.

Presently 6 different GNU/Linux distributions are available to VGrid users with host machines running Scientific Linux 4.

Business Track / 214

Best practices in Grid infrastructure testing - Bringing industry-class reliability and quality of service

As Grid installations worldwide enter production phase, they become an attractive target for industry looking to either replicate the technology internally, or outsource their processing and data operations externally. Reliability in Grids is not a trivial issue. On one hand we face a degree of uncertainty due to distribution, while on the other, applications and users require certain quality of service.

GridwiseTech works with customers building scalable infrastructure internally or using external Grid resources. Our recent research concerns state-of-the-art practices in testing scalable and distributed resource management infrastructures.

Application Track (Workflow) / **215**

Workflow Tools on the EGEE Grid infrastructure. Overview and Perspective.

Workflows have become a daily tool for applications on the Grid. In essence every non trivial application requires the composition and orchestration of data and functionality to produce useful scientific results. This talk will give an overview of the tools and technologies that currently utilize the EGEE Grid and gLite, in order to define, run and monitor workflows. The talk will outline the current status, potential problems, missing functionality and the road ahead, setting the landscape for the discussion that will follow.

Application Track (Workflow) / **216**

The Taverna Workflow Workbench and Tool Suite.

The OMII-UK myGrid project (http://www.mygrid.org.uk) has developed the popular Taverna workflow workbench, used for a wide range of data-centric Life Science problems and increasing within other domains such as astronomy and music. Taverna was designed to suit the work-a-day bioinformatician in a normal (not especially well-resourced) research laboratory, in order to ease and automate the routine burden of plumbing together the myriad of data resources and analytical tools publicly available and privately developed. It has been widely adopted by over 200 projects and labs, with many thousands of everyday users, and stands at nearly 40,000 downloads (1,500 per month).

The Taverna Workflow Workbench is an open-source workflow tool which provides a data-centric workflow language (Scufl) and graphical interface to: facilitate the building of workflows over distributed services hosted on remote and local machines; run these workflows on their own data; and visualise the results. Workflows can be enacted as part of the workbench or from third party applications. The enactor and the workbench have been designed with a range of extensibility points to make them open. A suite of plug-ins and components provide: service ingest and management; the cataloguing and discovery of services and workflows using semantic descriptions; and the recording of the provenance of the data outcomes and the execution log of the workflow runs. Recent work has focused on better support for third party applications, including a Web browser workflow client that enables Taverna workflows to be launched from a web browser, and the publishing of workflows as Google gadgets.

Complementing Taverna, and capitalising on the increasing numbers of workflows and users, the myExperiment initiative
(http://myexperiment.org) is creating a social networking environment and workflow bazaar for workflow workers. This recognizes that workflows are scientific assets in their own right, to be exchanged, traded and reused.
The talk will introduce the Taverna Workbench and its associated suite of components, its current

status and plans, some success stories and some possible exploitation plans and points of collaboration for EGEE.
In particular, the Taverna enactor is currently undergoing a major revision that will retain all the functionality of the current Taverna, but with better scalability, data streaming, data management, grid services integration and security management.

Application Track (Workflow) / 217

Discussion

The preceding session talks are expected to bring forward important issues and topics for analysis and discussion. Research teams that currently develop workflow tools for Grids and application teams that use them will have an opportunity for an open discussion regarding the current status but also with an eye to future advanced issues like knowledge-based workflow execution and cross-team collaborative workflow technologies.

GEMS: Grid Empowered Molecular Simulators and CompChem Tools / 218

About handling on the Grid quantum molecular knowledge related to molecular simulators

Author: Antonio Laganà[1]

[1] *University of Perugia, Dept. of Chemistry*

GEMS: Grid Empowered Molecular Simulators and CompChem Tools / 219

Activities of the COST D37 Computational Chemistry Workflow Group

Author: Thomas Steinke[1]

[1] *Zuse Institute Berlin*

GEMS: Grid Empowered Molecular Simulators and CompChem Tools / 220

The CompChem Virtual Organization

Author: Osvaldo Gervasi[1]

[1] *University of Perugia, Dept of Mathematics and Computer Science*

GEMS: Grid Empowered Molecular Simulators and CompChem Tools / 221

A proposal for a data format to help program interchange in the Quantum Chemistry domain

Author: Elda Rossi[1]

[1] *CINECA*

GEMS: Grid Empowered Molecular Simulators and CompChem Tools / 222

Chemistry visualization tools in an integrated discovery cycle

Author: Mario Valle[1]

[1] *CSCS*

SA3 Partner Review (Closed) / 223

PSNC

Review of Posznan's SA3 activities

SA3 Partner Review (Closed) / 224

GRNET

Review of GRNET's SA3 activity

SA3 Partner Review (Closed) / 225

INFN

OMII-Europe/EGEE: What can we do for each other? / 226

Introduction

Corresponding Author: erwin.laure@cern.ch

OMII-Europe/EGEE: What can we do for each other? / 227

OMII-Europe - Overview and interoperable components

OMII-Europe/EGEE: What can we do for each other? / 228

EGEE roadmap and requirements

OMII-Europe/EGEE: What can we do for each other? / 229

Discussion

Application Track (Uniform Grid Access) / 230

User view and requirements: example of Bioinformatics applications

Application Track (Uniform Grid Access) / 231

GridSphere Portal : an High-level User Interface to Grid

Application Track (Uniform Grid Access) / 232

The GridWay approach for job Submission and Management on Grids

Application Track (Uniform Grid Access) / 233

Discussion

234

Author: Jonathan Silverstein[1]

[1] *University of Chicago*

Health-e-Child/SHARE/HealthGrid: Applications for Health, from Data Acquisition to its Integration and Use in Grids / 235

HealthGrid activities in the United States

Author: Jonathan Silverstein[1]

[1] *University of Chicago*

Health-e-Child/SHARE/HealthGrid: Applications for Health, from Data Acquisition to its Integration and Use in Grids / 236

Scientific results of WISDOM first data challenges on malaria and avian flu

Author: Doman Kim[1]

[1] *Chonnam National University*

The talk will present the results of the in vitro tests performed on the best compounds selected in silico during the first data challenges on malaria and avian flu

Health-e-Child/SHARE/HealthGrid: Applications for Health, from Data Acquisition to its Integration and Use in Grids / 237

Health-e-Child

Author: Tamas Hauer[1]

[1] *University of the West of England*

Health-e-Child/SHARE/HealthGrid: Applications for Health, from Data Acquisition to its Integration and Use in Grids / 238

CoreGRID

Author: Carmela Comito[1]

[1] *University of Calabria*

Health-e-Child/SHARE/HealthGrid: Applications for Health, from Data Acquisition to its Integration and Use in Grids / 239

Grid enabled epidemiology

Author: Ignacio Blanquer[1]

[1] *UPV*

Health-e-Child/SHARE/HealthGrid: Applications for Health, from Data Acquisition to its Integration and Use in Grids / 240

SHARE: a roadmap for healthgrids

Health-e-Child/SHARE/HealthGrid: Applications for Health, from Data Acquisition to its Integration and Use in Grids / 241

Round table

Author: David Manset[1]

Co-authors: Ignacio Blanquer [2]; Luciano Milanesi [3]; Siegfried Benkner [4]

[1] *University of the West of England*

[2] *UPV*

[3] *National Research Council - Institute of Biomedical Technologies*

[4] *University of Vienna*

The round table will focus on 3 major topics (10' each) and then if extra time, questions from the audience will be taken by the panelists.

Topic1: Identifying ways forward in Europe for the convergence toward a community healthgrid platform and infrastructure
Topic2: Security and privacy in healthgrids, Common practices in
European projects. Emergence of a healthgrid European regulation
Topic3: Medical data integration and exploitation in grids.
European Technologies and their Integration in the gLite Grid Middleware

ACGT: the GRID to fight cancer / 242

Post-genomic Clinical Trials: Needs and Requirements (including legal issues)

243

ACGT Scenarios: Indicative post-genomic scenarios

ACGT: the GRID to fight cancer / 244

The ACGT architecture and its key services

ACGT: the GRID to fight cancer / **245**

The ACGT Data Integration Architecture

246

The ACGT Security Architecture

ACGT: the GRID to fight cancer / **247**

The ACGT Knowledge discovery services

ACGT: the GRID to fight cancer / **248**

Joining ACGT: Measures for the implementation of the open access policies of ACGT

CE Federation meeting (Closed) / **249**

NA3

Author: Gergely Sipos[1]

[1] *Mr.*

MTA SZTAKI (Gergely Sipos)

CE Federation meeting (Closed) / **250**

NA4

Author: Jan Kmunicek[1]

[1] *CESNET*

CESNET

CE Federation meeting (Closed) / **251**

NA2

Each partner

CE Federation meeting (Closed) / 252

SA1

Author: Tomas Szepieniec[1]

[1] *CYFRONET*

CYFRONET

CE Federation meeting (Closed) / 253

SA3

Author: Norbert Meyer[1]

[1] *PSNC*

PSNC

CE Federation meeting (Closed) / 254

JRA1

Author: Ales Krenek[1]

[1] *CESNET*

CESNET

CE Federation meeting (Closed) / 255

CE JRUs

Author: Martin Polak[1]

[1] *GUP*

Current situation, status in EGEE III, perspectives (NGIs)

CE Federation meeting (Closed) / 256

Info from PMB

Author: Ludek Matyska[1]

[1] *CESNET*

Information related to the current EGEE II project.
Experiences from the preparation of the EGEE III

CE Federation meeting (Closed) / 257

AOB

General discussion

CE Federation meeting (Closed) / 258

Welcome

Author: Ludek Matyska[1]

[1] *CESNET*

Business Track / 259

The Health-e-Child Project

Health-e-Child (pronounced healthy child) is an attempt to respond to the increasingly pressing demand to fully integrate and exploit heterogeneous biomedical information for improved clinical practice, medical research, and personalized healthcare. As an integrated project of the Sixth Framework Programme of the European Commission, the project brings together three major paediatric centres with several European companies, university groups and research centres specialized in information-based medical technology.

Pre-Production Service: All Sites Meeting / 260

Introduction and Agenda

Authors: Antonio Retico[1]; Nicholas Thackray[1]

[1] *CERN*

Quick overview of PPS activity and infrastructure. Presentation of the agenda

Pre-Production Service: All Sites Meeting / 261

Operations

Authors: Antonio Retico[1]; Nicholas Thackray[1]

[1] *CERN*

Business Track / 262

BEinGRID - Business Experiments in Grid

Middleware Security Group (MWSG) / 263

OSG Security

Author: Robert Cowles[1]

[1] *OSG*

OSG security update

Astronomy and Astrophysics cluster meeting / 264

Current status of the A&A cluster

Authors: Claudio VUERLI[1]; Claudio Vuerli[2]

[1] *INAF*

[2] *INAF-OA Trieste*

Middleware Security Group (MWSG) / 265

AuthZ mechanisms in gLite

Author: Christopher Witzig[1]

[1] *ETH*

266

AuthZ mechanisms in gLite

Astronomy and Astrophysics cluster meeting / 267

French projects and current status

Middleware Security Group (MWSG) / 268

VOMS Migration to SSL: Roadmap and Issues

Author: Vincenzo Ciaschini[1]

[1] *INFN*

Motivation: TCG required us to remove all globus dependencies from VOMS. Especially for the APIs, this require some coordination among JRA1 developers, so some things needs to be discussed.

269

Interoperations/Interoperability

Authors: Antonio Retico[1]; Nicholas Thackray[1]

[1] *CERN*

270

Interoperations/Interoperability

Authors: Andreas Unterkircher[1]; Antonio Retico[1]; Nicholas Thackray[1]

[1] *CERN*

Pre-Production Service: All Sites Meeting / 271

PPS Users

Authors: Antonio Retico[1]; Nicholas Thackray[1]

[1] *CERN*

Pre-Production Service: All Sites Meeting / 272

Meeting sum-up (decisions + input for SA1 session)

Middleware Security Group (MWSG) / 273

The SAML-XACML protocol in EGEE & OSG Authorization and resource enforcement tools

Author: Oscar Koeroo[1]

[1] *NIKHEF*

This presentation will explain and discuss the work done to get the SAML-XACML protocol usable in OSG and EGEE authorization and resource enforcement tools for better inter-operation between our commonly used tools.

It will explain the big picture on which problems we'd like to solve and how the existing tools will be used on a site architecturally. The involved tools are the commonly used grid tools like the gatekeeper, gridftpd, glexec, gjaf, dCache when deployed with lcas & lcmaps and/or prima/gplazma & gums on both EGEE and OSG sites.

Middleware Security Group (MWSG) / 274

Security concerns from GridPP and GSVG

Author: Linda Ann Cornwall[1]

[1] *RAL*

To briefly describe the most common security concerns from the GridPP meeting discussions and from the GSVG.

Middleware Security Group (MWSG) / 275

Pseudonymity-related work in EGEE-II

Author: Henri Mikkonen[1]

[1] *HIP/EGEE*

A need for anonymous access to grid resources has been identified. It prevents the resource owners and other external parties from tracing the users and their actions. The talk gives an overview and status of the related work in the EGEE-II project.

Middleware Security Group (MWSG) / 276

Proxy Restrictions for the WN

Author: Gergely Debreczeni[1]

[1] *CERN, EGEE*

There have been demonstrations on how to steal a proxy from the WN and there is an ongoing work to add policies as restrictions. What we would like to be discussed is the status of this work.

Middleware Security Group (MWSG) / 277

gLExec – gluing grid computing jobs to the Unix world

Author: David Groep[1]

[1] *NIKHEF*

What is the purpose for gLExec and how and what does it do. What is the relation of gLExec with regard to pilot jobs and what can gLExec do for you on the WN or CE for both users and site administrators.

Middleware Security Group (MWSG) / 278

Authorization interoperability

Author: Gabriel Garzoglio[1]

[1] *FNAL*

Astronomy and Astrophysics cluster meeting / 279

Support of Astrophysical applications in Slovakia

Astronomy and Astrophysics cluster meeting / 280

Activities in progress ad IFCA and examples of applications

Astronomy and Astrophysics cluster meeting / 281

Italian Projects and current Status

Astronomy and Astrophysics cluster meeting / 282

A&A cluster in EGEE: Future developments and perspectives

This is a general discussion about the AA cluster in EGEE and future perspectives concerning the participation of the astronomical community to EGEE. All people who attend the session are invited to contribute.

Business Track / **283**

Distributed Computing with MATLAB in Grids

MATLAB is a popular programming environment with over 1,000,000 users worldwide, primarily scientists and engineers engaged in technical computing. In 2004, The MathWorks introduced a product for distributed and parallel computing that enables MATLAB users to take advantage of high-performance environments to solve computationally or data intensive problems.

Engineers and scientists can now develop parallel applications independently of the resources that are available for execution; they can prototype applications on their desktops and then scale to a computer cluster without any code changes. This presentation will discuss the additional technical and licensing challenges that are being resolved so that scientist can take advantage of the grid for their collaborative research activities.

Business Track / **284**

Sun Technology advances in HPC

Business Track / **285**

Applying Grid Technology on Business Support Systems - A paradigm in the Telecom sector

Business Track / **286**

The WISDOM Project

Business Track / **287**

A Grid-based Collaborative Environment for new Space Systems design

Nowadays, space activities are characterised by increased constraints in terms of cost and schedule combined often with a higher and higher technical and programmatic complexity.

To answer this challenge, the European Space Agency has set up the Concurrent Design Facility starting in 1998. This has widely demonstrated the advantages of applying the Concurrent Engineering approach to the assessment and conceptual design of future space missions and has raised

an enormous interest among the European partners (academia, scientific communities, industry, other agencies) in the space sector.

At the same time, starting from mid 90's, a remarkable increase in computing power has been achieved by designing and prototyping technologies, most notably the Grid, to support distributing tasks and data on distributed computing centres linked with high-speed networks. Grid technology can, therefore, provide the means for secure connectivity of design environments as well as integrate multiple heterogeneous systems into a powerful virtual "single" system.

Within this framework, the European Space Agency, at beginning of 2006, awarded a project called Grid-based Distributed Concurrent Design (GDCD) to study how to allow geographically distributed facilities to interact each other in real-time over wide area networks adopting the Grid technology for the purpose of space projects, to make the structure deployment reliable, cheap and compatible with Concurrent Facilities.

SEE-GRID-2: Regional Grid projects concertation workshop / 288

EELA Applications

Author: RAFAEL MAYO[1]

[1] *CIEMAT*

SEE-GRID-2: Regional Grid projects concertation workshop / 289

EUMEDGRID Applications

Author: Giuseppe Andronico[1]

[1] *INFN SEZIONE DI CATANIA*

SEE-GRID-2: Regional Grid projects concertation workshop / 290

SEEGRID Applications

Author: Asli Zengin[1]

[1] *TUBITAK ULAKBIM*

SEE-GRID-2: Regional Grid projects concertation workshop / 291

BalticGrid Applications

Author: Algimantas Juozapavicius[1]

[1] *associate professor*

292

EUCHINAGRID Applications

Author: Marco Verlato[1]

[1] *Dipartimento di Fisica Galileo Galilei*

SEE-GRID-2: Regional Grid projects concertation workshop / 293

EUINDIAGRID Applications

Author: Marco Verlato[1]

[1] *Dipartimento di Fisica Galileo Galilei*

SEE-GRID-2: Regional Grid projects concertation workshop / 294

EUCHINAGRID Applications

295

DSA1.6

Corresponding Author: ian.bird@cern.ch

SEE-GRID-2: Regional Grid projects concertation workshop / 296

Applications Panel

Author: Roberto Barbera[1]

[1] *UNIV. CATANIA AND INFN*

EGEE Grid Operations (SA1) - Site Monitoring and Visualization / 297

Using Nagios for Grid Service Monitoring

Corresponding Author: emir.imamagic@cern.ch

EGEE Grid Operations (SA1) - Site Monitoring and Visualization / 298

Experience at NIKHEF deploying Nagios for monitoring grid services

Corresponding Author: ronalds@nikhef.nl

EGEE Grid Operations (SA1) - Site Monitoring and Visualization / 299

Report on status of Open Science Grid monitoring

Corresponding Author: rquick@iupui.edu

ROC Managers Meeting / 300

Review of GGUS tickets

How are we doing with clearing all tickets in the GGUS Ticket Escalation report? Click on "more information" to see the latest reports.

ROC Managers Meeting / 301

Cleaning up the LCGAdmin role for dteam

The people who will retain the LCGAdmin role are:

CERN
* Maarten Litmaath
* Andreas Unterkirche

Italy
* Mirco Ciriello
* Simone Dalla Fina
* alessandro paolini

ROC Managers Meeting / 302

Status of new features in CIC Portal

 YAIM Configuration Tool:
 <u>Update from Gilles</u>: The tool is fully integrated to the CIC portal since the release of version 4.5, i.e. on August, 22nd. The Yaim Tool database back-end is integrated to the CICDB schema, thus ensuring data coherency about all officially registered VOs. Managers of such VOs don't need to go to the tool to update their data: all that is needed is taken from the VO ID card. These 2 points solve the main issue of having 2 entry points to update this info. The web interface for the tool is now also hosted at IN2P3, and available as follows: <ul> As a page of the RC section in the portal: https://cic.gridops.org/index.php?section=rcpage=yaim As a page of the VO section

in the portal: https://cic.gridops.org/index.php?section=vopage=yaim As a standalone version: https://cic.gridops.org/yaimtool/yaimtool.py YAIM configurator code is, in CIC portal code structure, an independent module. Development and maintenance is assured by Dimitar, which has access to this module in our cvs repository.
 Failover procedures have not been established yet. For that, we need to check python configuration on the replica instance of the portal at CNAF, test the application there, and validate the process in the same way we did with SAM Admin's Page.

Tool for announcing scheduled downtimes <u>Update from Osman</u>: The schedule downtime process is composed of 6 steps: Step 1 : Integration of BDII data in the CIC db: This part determines the impacted VO by a downtime 100
Step 2 : Notification from GOC portal to the CIC portal 90
Step 3 : The implementation of a Supervisor componant was implemented: this allows the management of the whole process Step 4 : The downtime broadcasting by email - 90
Step 5 : The downtime broadcasting by RSS Feed is OK - 100
Step 6 : The subscription form has not implemented yet but we are working on it: this form will enable to any user to be notified by email or RSS Feed according to his need. In summary, I think I will need 2 weeks (after the EGEE conference) to finish and test the whole process

ROC Managers Meeting / 303

Use of benchmarking tool for SpecInt2000

 Extract from the minutes of the WLCG MB, 5 June <small>(https://twiki.cern.ch/twiki/pub/LCG/MbMeetingsMinutes/LCG
$/small >:< br >< br >< i >< b > Decision of the WLCG Management Board :< /b >$
$The MB agreed that, until the HEPiX benchmarks are defined, the proposal distributed by L. Robertson is accepted. <$
$br >< br > This proposal is contained in the note distributed to the MB "Proposal for an interim CPU capacity metric for W$
$. This uses the methodology of GridKA (i.e. run one copy of the SPECCPU2000 code per core), uses the compiler flags as spe$
$pthread - fPIC), with a scaling factor of 1.5 (reflecting the lower level of compiler optimization compared with the curren$
$br >< br > A HOWTO section with benchmark script and configuration files is available on the HEPiX pages :$
$see http : //hepix.caspur.it/processors/. < br >< br > * < small > (https : //mmm.cern.ch/public/archive -$
$list/w/worldwide - lcg - management - board/CPU$

SEE-GRID-2: Regional Grid projects concertation workshop / 304

EUCHINAGRID Infrastructure

Author: Andrea Chierici[1]

[1] *Unknown*

SEE-GRID-2: Regional Grid projects concertation workshop / 305

EELA Infrastructure

Author: Roberto Barbera[1]

[1] *UNIV. CATANIA AND INFN*

SEE-GRID-2: Regional Grid projects concertation workshop / 306

EUMEDGRID Infrastructure

Author: Konstantinos Koumantaros[1]

[1] *Unknown*

SEE-GRID-2: Regional Grid projects concertation workshop / 307

SEEGRID Infrastructure

Author: Antun Balaz[1]

[1] *Institute of Physics, Belgrade*

SEE-GRID-2: Regional Grid projects concertation workshop / 308

BalticGrid Infrastructure

Author: Margarita Kazakeviciute[1]

[1] *Vilnius University*

SEE-GRID-2: Regional Grid projects concertation workshop / 309

EUINDIAGRID Infrastructure

Author: Sergio Fantinel[1]

[1] *Unknown*

SEE-GRID-2: Regional Grid projects concertation workshop / 310

Infrastrucure Panel

Author: Ake Edlund[1]

[1] *Unknown*

SEE-GRID-2: Regional Grid projects concertation workshop / 311

SEEGRID NGIs

Author: Gabriel Neagu[1]

[1] *Unknown*

SEE-GRID-2: Regional Grid projects concertation workshop / 312

BalticGrid NGIs

Author: Algimantas Juozapavicius[1]

[1] *associate professor*

SEE-GRID-2: Regional Grid projects concertation workshop / 313

EELA NGIs

Author: Bernard Marechal[1]

[1] *Instituto de Fisica*

SEE-GRID-2: Regional Grid projects concertation workshop / 314

EUMEDGRID NGIs

Author: Federico Ruggieri[1]

[1] *INFN*

SEE-GRID-2: Regional Grid projects concertation workshop / 315

NGIs Panel and wrap-up

Author: Aleksandar Belic[None]

ROC Managers Meeting / 316

Abbreviations for Site-Names in GridView

Corresponding Author: antonio.retico@cern.ch

Examples of the GridView graphs with long and short site names are in the attached material.
The conversion table is here: http://gridview.cern.ch/GV/

More:
http://gvdev.cern.ch/GVVAI/
now you are in the uncharted territory, this is (one of) the development
displays used by gridview development.
So what you see there is not yet in production, but will be soon.

Click on "Job Status".
In the left pane you'll see two radio buttons "Use Site Abbreviation" and "Use Full Site Names".
So far only the "Use Site Abbreviation" works (remember that's a development portal), but you get the idea, roll down a bit to see the effect in the graph.

Same as above: both are available, everyone can pick and choose to his liking.

ROC Managers Meeting / 317

Viewing contact details in Operations Tools

ROC Managers Meeting / 318

Incentives/penalties for sites under the SLD

Why will sites stick to the SLD? Considering that the SLD will be aiming to set a minimum "standard" for sites, what incentive is there for sites to do better?
One suggestion is to set up a "league table" showing the relative 'goodness' of sites.
Any other suggestions?

ROC Managers Meeting / 319

Integration of CE ROC's first line support within the COD operations

http://goc.grid.sinica.edu.tw/gocwiki/TIC_1st_line_support_integration

ROC Managers Meeting / 320

GGUS 6.0 upgrade

ROC Managers Meeting / 321

ARM and COD meetings in Lyon in February

Corresponding Author: alistair.mills@cern.ch

ROC Managers Meeting / 322

ROC Manager access to SA1 PPT sheets of their federation partners

At least to a summary page every 3 months

ROC Managers Meeting / 323

Problems with the process for the update of root certificates?

Corresponding Author: sven.hermann@iwr.fzk.de

The update of root certificates still is an issue in EGEE. The CERN (EGEE) repository http://grid-deployment.web.cern.ch/grid-deployment/lcg2CAlist.html isn't updated quick enough after announcement of the new certs on https://dist.eugridpma.info/distribution/igtf/ Then problems arise when the SAM tests check for the new certs before the official EGEE broadcast, and before the official EGEE repository is updated. IMHO a better work flow is needed here.

The wiki pages explaining the process are here:
http://goc.grid.sinica.edu.tw/gocwiki/Procedure_for_new_CA_release

ROC Managers Meeting / 324

Issues raised by ROC DECH: Request for comments from other ROCs

1) gLite deployment has to be better agreed with VO requirements before being certified (see e.g. "prd-/pool-account" discussions after rollout to production, VOs still put high pressure on sites because there's still no common, agreed solution; instead VOs and the gLite development team should communicate directly).

2) the gLite community is not organised transparently enough, e.g. there aren't links to corresponding site administration/MW configuration wikis (GOC Wiki etc.) or further practical information about usage of gLite
from the official gLite web page. This concern has also been expressed by users and admins at GridKa School 2007.

325

Introduction

326

Data management in gLite

327

test

Author: Liabotis Ioannis[None]

EUChinaGRID/EGEE/ETICS: Grids and IPv6 / **328**

Short report on basic IPv6 testing

Author: Mario Reale[1]

[1] *GARR*

Corresponding Author: mario.reale@garr.it

ROC Managers Meeting / **329**

WLCG / EGI issues

Author: Jamie Shiers[1]

[1] *CERN*

Corresponding Author: jamies.shiers@cern.ch

Further information:
EGI workshop agenda (part of EGEE'07): http://www.eu-egi.org/workshop/oct07/
WLCG Service Reliability workshop: http://indico.cern.ch/conferenceDisplay.py?confId=20080

Pre-Production Service: All Sites Meeting / **330**

Interoperability/Interoperations

Authors: Antonio Retico[1]; Nicholas Thackray[1]

Co-author: Andreas Unterkircher [1]

[1] *CERN*

331

Example in medical data

332

Discussion

Application Track (Data Mgt.) / 333

Accessing and Managing Databases in Grid with G-DSE

Application Track (Data Mgt.) / 334

Grid database management in gLite based production Grid with the GReIC DAS

Application Track (Data Mgt.) / 335

Discussion

EGEE Operations and gLite Middleware (SA1 & JRA1) / 336

Managing a CREAM CE

Corresponding Author: zangrando@pd.infn.it

20' + 10' discussion

EUMEDGRID: Grids and their role in sustaining development / 337

eInfrastructures in the Mediterranean: the Eumedconnect network success story

EUMEDGRID: Grids and their role in sustaining development / 338

Grids and their role towards developement

EUMEDGRID: Grids and their role in sustaining development / 339

The EUMEDGRID project (with some examples of applications)

EUMEDGRID: Grids and their role in sustaining development / 340

Can Grid Technology reverse brain drain to brain gain in Africa? (NO SHOW)

EUMEDGRID: Grids and their role in sustaining development / 341

Biological applications in GRID - the EuchinaGRID experience

342

…….

343

Panel disccussion

Business Track / 344

Aerospace research on the Grid

345

Introduction

346

Data Management in gLite

347

Data management in gLite

348

Data management in gLite

349

.....

Application Track (Data Mgt.) / 350

Example in Biomedical image management

Application Track (Data Mgt.) / 351

Introduction

Application Track (Data Mgt.) / 352

Data management in gLite

EUMEDGRID: Grids and their role in sustaining development / 353

Towards grid-enabled Telemedicine in Africa

EUMEDGRID: Grids and their role in sustaining development / 354

Panel discussion

EUMEDGRID: Grids and their role in sustaining development / 355

Scientific Data Sonification and the ASTRA project for ancient instruments reconstruction within EGEE

Application Track (Data Mgt.) / 356

Discussion

Business Track / 357

EGEODE business model, based on gLite

Business Track / 358

NESSI-Grid Vision and Strategic Research Agenda

The Networked European Software and Services Initiative (NESSI) aims to create a strategic research agenda for European research in services and their foundations. NESSI-Grid forms part of that activity by defining a vision and strategic research agenda (SRA) for grid infrastructures used in business environments and in particular in NESSI scenarios.

A first version of both, vision and SRA, focusing on more short-term challenges is presented. It addresses all stakeholders in this field, in particular researchers, industry and policy makers, to get a precise understanding about envisioned scenarios, state-of-the-art, challenges, business impact and actual roadmaps and recommendations for realizing the vision. The SRA is open for contribution.

GEMS: Grid Empowered Molecular Simulators and CompChem Tools / 360

Adaptation of chemical dynamics codes to the GRID

Author: George Lendvay[1]

[1] *Institute of Chemistry, Hungarian Academy of Science*

SW Federation meeting (Closed) / 361

Welcome

Author: Kai Neuffer[1]

[1] *PIC*

SW Federation meeting (Closed) / 362

SWE SA1 status

Author: Kai Neuffer[1]

[1] *PIC*

SW Federation meeting (Closed) / 363

SWE SA3 status

SW Federation meeting (Closed) / 364

SWE NA2 status

SW Federation meeting (Closed) / 365

SWE NA3 Status

Author: Antonio Fuentes[1]

[1] *Unknown*

SW Federation meeting (Closed) / 366

SWE NA4 status

SW Federation meeting (Closed) / 367

SWE EGEE-III proposal

Author: Kai Neuffer[1]

[1] *PIC*

SW Federation meeting (Closed) / 368

A.O.B.

PORTAL Working group / 369

Introduction and usecase: the GPSA Bioinformatics portal

PORTAL Working group / 370

Portals and Authentication

PORTAL Working group / 371

JSPG contribution

PORTAL Working group / 372

Web application security

PORTAL Working group / 373

Shibboleth and Grid Portals

PORTAL Working group / 374

Discussion

FR Federation meeting and SA1-FR meeting (Closed) / 375

ROC report

Author: Pierre Girard[1]

[1] *Unknown*

Corresponding Author: pierre.girard@in2p3.fr

FR Federation meeting and SA1-FR meeting (Closed) / 376

OAG overview and state

Author: Rolf Rumler[1]

[1] *Unknown*

Corresponding Author: rumler@in2p3.fr

Mission, achievements, and status

FR Federation meeting and SA1-FR meeting (Closed) / 377

COD and CIC

Report on the grid operators' work (COD), status of the CIC portal work

FR Federation meeting and SA1-FR meeting (Closed) / 378

French Federation meeting

Author: guy wormser[1]

[1] *LAL Orsay*

Corresponding Author: wormser@lal.in2p3.fr

1) EGEE-III
Présentation générale (GW)
Tour de table des partenaires
2) Institut des grilles:
Présentation et perspectives (GW)
3) EGEE-II : questions diverses
4) AOB

KnowARC: Early results of the KnowARC Project / 379

Towards the next generation ARC middleware

Early results of the KnowARC project

The gLite Release Process and Porting / 380

The release process, gLite 3.1 status, moving towards SL4

The gLite Release Process and Porting / 381

Certification Process

The gLite Release Process and Porting / 382

developer driven builds

The gLite Release Process and Porting / 383

source distribution and RPM signing

The gLite Release Process and Porting / 384

Partner contribution process

EUChinaGRID/EU-IndiaGrid: Grids Interoperability between Europe and Asia, status and future strategies / 385

OGF/GIN Status and perspectives

Author: Erwin Laure[None]

EUChinaGRID/EU-IndiaGrid: Grids Interoperability between Europe and Asia, status and future strategies / 386

The gLite-GOS gateway

Author: Marco Pappalardo[1]

[1] *INFN Catania*

EUChinaGRID/EU-IndiaGrid: Grids Interoperability between Europe and Asia, status and future strategies / 387

Garuda

Author: Subrata Chattopadhyay[None]

388

Are Web Services the answer to interoperability? the OMII Europe view

Author: Morris Riedel[None]

EUChinaGRID/EU-IndiaGrid: Grids Interoperability between Europe and Asia, status and future strategies / 389

Are Web Services the answer to interoperability? the OMII Europe view

Author: Morris Riedel[None]

EUChinaGRID/EU-IndiaGrid: Grids Interoperability between Europe and Asia, status and future strategies / 390

Panel session

Authors: Erwin Laure[None]; Marco Verlato[None]; Morris Riedel[None]; Subrata Chattopadhyay[None]

The topics addressed will be:
- Are the Grid middlewares converging or diverging?
- Are standards really used (and useful)?
- Web services: do they have a role in the interoperability arena?
- A Grid of Grids: just like Internet?

Middleware Security Group (MWSG) / 391

authN Interoperability: The INFN View

gLite Configuration Management (SA3) / 392

Session Overview

gLite Configuration Management (SA3) / 393

YAIM Overview and Tutorial

gLite Configuration Management (SA3) / 394

Developer Experience (SGE and FTM)

gLite Configuration Management (SA3) / 395

YAIM and QUATTOR, (shaving a bald Yak ?)

gLite Configuration Management (SA3) / 396

Discussion

Application Track (Data Mgt.) / 397

OGSA-DAI -version3

gLite Interoperability: ARC and UNICORE (SA3) / 398

Unicore interoperability, status and outlook

gLite Interoperability: ARC and UNICORE (SA3) / 399

ARC interoperability, status and outlook

gLite Interoperability: ARC and UNICORE (SA3) / 400

Multi platform support

P-GRADE Grid Portal and Developer Alliance / 401

P-GRADE Portal and Developer Alliance

P-GRADE Grid Portal and Developer Alliance / 402

Exposing Legacy Applications through the P-GRADE Portal

P-GRADE Grid Portal and Developer Alliance / 403

Achienvements of Turkish partners of P-GRADE Portal Alliance

P-GRADE Grid Portal and Developer Alliance / **404**

Discussion

Topics:
- Open source P-GRADE Portal
- Open source licences
- Future releases and features
- Support services

int.eu.grid: Support to MPI and interactivity on glite infrastructures / **405**

MPI support on EGEE

Author: Stephen Childs[1]

[1] *TCD*

Status of the MPI support in the EGEE infrastructure

NA3 Partner review (Closed) / **406**

NA3 status, news, actions

Technical Plenary / **407**

Demo and Poster Awards

NA3 Partner review (Closed) / **408**

NA3 activities in German/Swiss federation

NA3 Partner review (Closed) / **409**

UCY Activities

NA3 Partner review (Closed) / **410**

GRNET activities

NA3 Partner review (Closed) / 411

IPB activitites

NA3 Partner review (Closed) / 412

RoGrid JRU

NA3 Partner review (Closed) / 413

ICM UW activities

NA3 Partner review (Closed) / 414

IPP-BAS activities

NA3 Partner review (Closed) / 415

BEgrid

NA3 Partner review (Closed) / 416

II-SAS activities

NA3 Partner review (Closed) / 417

NA3 Activity in Russia

NA3 Partner review (Closed) / 418

NA3 TUBITAK

NA3 Partner review (Closed) / 419

TAU activities

NA3 Partner review (Closed) / 420

UK - Ireland federation

NA3 Partner review (Closed) / 421

CESNET activities

NA3 Partner review (Closed) / 422

Italian Federation

NA3 Partner review (Closed) / 423

JKU activities

NA3 Partner review (Closed) / 424

PSNC activities

NA3 Partner review (Closed) / 425

UIBK activities

NA3 Partner review (Closed) / 426

South West Europe Federation

NA3 Partner review (Closed) / 427

SZTAKI activities

NA3 Partner review (Closed) / 428

NA3 tasks at CSCS