

3rd EGEE User Forum

Monday 11 February 2008 - Friday 15 February 2008

Le Polydôme, Clermont-Ferrand, FRANCE

Book of abstracts

Table of contents

The Porting of a Grid Software for Virtual Eye Surgery from Globus 4 to gLite	1
Distributed system for genetic linkage analysis using EGEE and BOINC	1
g-Eclipse - Easy access to Grid infrastructures	2
g-Eclipse - Grid in five minutes	2
Grid website vulnerabilities and the GridSite Security Framework	2
Grid Storage Interoperability Now!	3
DeployWare: A Framework for Automatic Deployment of Software Systems on Grids	3
A parallel data mining application for Gene Ontology term prediction	4
The National Grid Service User Accounting System	4
A simple SPAGO implementation enabling the connection of heterogeneous computational resources to a gLite based GRID	5
Stock Analysis Application	5
Modeling the EGEE latency to optimize job time-outs	6
Interfacing gLite services with the Kepler scientific workflow	6
Earth Science Application overview in EGEE infrastructure	6
TriAnnotPipelineGRID, a tool for the automated annotation of wheat BAC sequences	7
WISDOM	7
Exploitation path of Interactive European Grid on User Communities	8
Interactive European Grid: Advanced Applications Support Demo	8
Medical Data Manager: an interface between PACS and the gLite Data Management System	9
A Full Stokes Model for Large Scale Ice Dynamics Simulations	9
EGEETomo: 3D electron tomography on EGEE. A user-friendly and fault-tolerant grid application.	10
Strategies for experiment-specific monitoring in the Grid	10
Ganga - powerful job submission and management tool	11
New monitoring applications in the Experiment Dashboard	11
Increased productivity for emerging Grid applications: the application support system	12
The WLCG Common Computing Readiness Challenge: CCRC`08	12
Analysis of Metagenomes on the EGEE Grid	13
Optimizing a Grid workflow for the EGEE infrastructure: The case of Wien2k	13

Cosmological application in the Grid environment: Detection of SZ Clusters of Galaxies in data from the ESA Planck satellite mission	14
Portals for Earth Science	14
GRID BASED TELEMEDECINE APPLICATION FOR GATE MONTE CARLO DOSIMETRIC STUDIES	15
A telemedicine platform for information and image management on the grid	15
What can the CIC portal do for end users?	16
New developments on the LHCb Bookkeeping	16
ThIS on the Grid	17
A WS-DAIR Compatible Interface for gLite-AMGA	17
FUSION RESULTS WITHIN EGEE	18
Execution Time Prediction of Imperative Paradigm Tasks for Grid Scheduling Optimization	18
Mathcell.Ru: Integrated Mathematical Model of Living Cell in GRID Infrastructure	19
The EGEE user support infrastructure	19
Towards a statistical model of EGEE load	20
Optimisation applications support in RDIG Fusion VO	20
Service-Oriented Applications development support in gLite	20
WLCG-RUS: An Extensible Solution to Resource Usage Service	21
Grid enabled applications for modeling, simulation and optimization	21
OpenSAML extension library and API to support SAML2.0 - XACML protocol for interoperable authorisation infrastructure in distributed Grid applications	22
Supporting Statistical Semiconductor Device Analysis using EGEE and OMII-UK Middleware	22
GRelC DAIS: Towards Data Access and P2P Data Integration in gLite based Production Grids	23
A Network Monitoring Framework in the SCoPE Project	23
WISDOM: Grid enabled identification of active molecules against targets implicated in Malaria	24
EXPLOITING GRID FOR CONDENSED MATTER APPLICATIONS: QUANTUM DOTS AND BUILDING CLUSTERS ATOM-BY-ATOM	24
The Development of SRM interface for SRB	25
fMRI analysis on EGEE with the VBrower and MOTEUR	25
Fair Grid Scheduling	26
Smart adaptative identifier for a medical data retrieval system	26
Distributed Data Management on the petascale using heterogeneous grid infrastructures with DQ2	27
Charon GUI - Feature Rich Interface to Distinct Grid Niches	27
INTERCONNECTING GRID, DESKTOP GRID AND NETWORK	28
VO AUGER Large Scale Monte Carlo Simulations using the EGEE Grid Environment	28
Communication tools between Grid Virtual Organisations, middleware deployers and sites	29

Problem Solving Environment for stereological modeling of muscle cells	29
Design, implement and deploy Grid oriented applications in a cooperative way: a biomedical use case	30
Organising scientific data by dataflow optimisation on the petascale	30
RINGrid: conceptual design of the remote instrumentation systems	31
Enabling Distributed Access and Parallel Processing of Bioinformatics Data in Grid Environment: The EKTORAS platform	31
Workflow meta-scheduler for Grid Environments	32
Advantages of Pre-Production WLCG/EGEE services for VOs and Users	32
GRid-aware Optimal data Warehouse design (GROW)	33
Dissemination and exploitation of Grids in Earth Science	33
A service oriented framework to create, manage and update metadata for earth system science	34
Grid computing for wildfire danger assessment: porting RISICO in glite	34
R-GMA: Now With Added Authorization	35
Simple, fault tolerant, lightweight grid computing approach for bag-of-tasks applications	35
Numerical modeling of electrodynamic aggregation of magnetized dust in electric discharges	36
Development and adaptation of a web enabled in silico oncology application in grid environment	36
gCube Grid Services	37
SAGA API for gLite Service Discovery	37
All-in-one graphical tool for grid middleware management	37
Early failure detection: a method and some applications	38
Extension of DIRAC to enable distributed computing using Windows resources	38
The Experiment Dashboard for medical applications	39
Interactive Workflow Management in int.eu.grid	39
AIDA: Atlas graphical Interface for Distributed Analysis	40
The G-DSE (Grid-Data Source Engine)	40
Making the Grid and the Virtual Observatory mutually interoperable	40
Astronomy and Astrophysics applications on EGEE from Paris, Grenoble, Lyon and Strasbourg observatories	41
GEOSCOPE applications on EGEE	41
Implementation of geospatial services in gLite: the RISICO case study	42
The gCube Metadata Framework: integrated environment for managing Metadata Objects and relationships on top of Grid-enabled Storage Systems	42
LOFAR@EGEE	43
IV Grid Plugtests: composing dedicated tools to run an application efficiently on Grid'5000	43
Decentralized access to medical images in Research and Enterprise PACS Applications	43

Interaction of a 3D finite-difference application for computing synthetic waveforms with the Grid infrastructure	44
Evaluating meta data access strategies through implementing the GOME test suite	44
CRAB, the CMS tool to allow data analysis in a distributed environment	45
Towards a WBEM-based Implementation of the OGF GLUE Information Model	45
Performance Analysis and Optimization of AMGA for the WISDOM environment	46
The Dutch Life Science Grid	46
Evaluation of EGEE-grid security.	47
Distributed Task Scheduling for Physics Fusion Applications	47
Utilization of EGEE Grid as a Computational Platform for Chemistry	48
Parallel Execution of Chemical Software on EGEE Grid	48
Using grid technologies for the mapping out of taxation policy	49
Partnership for Advanced Computing in Europe (PRACE)	49
High-throughput GRID application for Life Sciences: The Gene Analogue Finder	50
Testing 65536 parallel pseudo-random number streams	50
SALUTE: New results for the inhomogenous case	50
EDGeS: Integrating EGEE with Desktop Grids	51
Long range air pollution transport over Europe, studied on the Grid	51
GridAE: A Grid-based Framework for Artificial Evolution Applications	52
Grid Solving a Bioinformatics Challenge: a First Step to Anchoring the Nucleosome	52
High Performance Computing on the GRID infrastructure of COMETA	52
Configuring and enabling Condor for LHC Computing Grid	53
Batch Service Management Tool within g-Eclipse	53
MPI Support on the Grid	53
Genome Wide Association Studies of human complex diseases with EGEE	54
XtreemOS: A Grid Operating System Providing Native Virtual Organization Support	54
GridVideo: a grid-based multimedia application	54
The ZEN Project	55
gUSE: grid User Support Environment	55
Experiences from porting the astrophysical simulation “The unified theory of Kuiper-belt and Oort-cloud formation” to EGEE grid	56
An Application of ETICS Co-Scheduling Mechanism to Interoperability and Compliance Validation of Grid Services	56
European Grid Initiative Design Study (EGI_DS)	57
Bioinformatics portal on Grid: the GPSA - Grid Protein Sequence Analysis - case.	57

EnginFrame Genius Grid Portal and VOMS Proxy creation.	58
Non-parametric parallel GRID Harris-affine detector	58
Enhancement and Breast Hierarchical Segmentation on Digital Mammograms	59
The gLite Workload Management System	59
GRID distribution supporting chaotic map clustering on large mixed microarray data sets	60
Solving Data Transfer Level Grid Interoperability among SRM, SRB and OGSA-DAI data resources	60
Expo : an experiment framework for dedicated platforms	61
The study of Cytochrome C oxidase on EGEE Grid	61
gLibrary/DRI: A grid-based platform to host multiple repositories for digital content	62
European Space Agency astronomy projects using Grid technology : XMM-Newton and Integral	62
The Grid Application Platform: Development and Applications	63
The Health-e-Child (HeC) Gateway and Case Reasoner, a Concrete Implementation and Use-Case of a gLite-based Healthgrid for European Paediatrics	63
New results on a comparative evaluation of software providing access to different relational databases interfaced to the grid	64
Molecular dynamics refinement and rescoring in WISDOM virtual screenings	64
BioinfoGRID: Bioinformatics Grid applications for life science	65
Deploying e-Infrastructures for new User Communities in DORII project	65
Panel Discussion	65
All-in-one graphical tool for grid middleware management	65
Introduction	65
Gold Sponsor: SHARE	66
Welcome from EGEE Project Director	66
Welcome from Local Organisation	66
Computer Science Grids and research in Grid Computing	66
Application Porting in EGEE	66
Challenges and success of the HEP GRID	66
ArchaeoGRID, a Laboratory for the Past on e-Infrastructures	67
Be Elegant on the Grid	67
VO-level Application Support in EGEE	67
Computational Chemistry & Material Science Summary	67
Earth Science Summary	68
Finance & Multimedia Summary	68
Fusion Summary	68
Astronomy & Astrophysics Summary	68

Interaction with the Grid'5000 initiative Summary	68
Life Sciences Summary	68
Monitoring, Accounting & Support Summary	68
Data Management Summary	68
Grid Access Summary	69
Interoperability and Resource Utilisation Summary	69
Workflow and Parallelism Summary	69
Demo/Poster Session Summary and Awards	69
User Forum Closing Talk	69
Fusion Demonstration: Ion Kinetic Transport and Stellarator Optimization	69
Workflow and Parallelism Overview	70

Posters - Board P01 / 8

The Porting of a Grid Software for Virtual Eye Surgery from Globus 4 to gLite

Dr. BOSA, Karoly Jozsef¹; Dr. SCHREINER, Wolfgang¹

¹ JKU/RISC

Corresponding Author: kbosa@risc.uni-linz.ac.at

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

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

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

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

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

Life Sciences / 9

Distributed system for genetic linkage analysis using EGEE and BOINC

Mr. SILBERSTEIN, Mark¹; Mr. SHAROV, Artyom¹; Prof. SCHUSTER, Assaf¹; Prof. GEIGER, Dan¹

¹ Technion - Israel Institute of Technology

Corresponding Author: marks@cs.technion.ac.il

Tasks are submitted via web and parallelized into thousands or even millions of CPU-bound jobs ranging from a few seconds to a few minutes long. Efficient and reliable execution is complicated due to unbounded queuing times, high execution and scheduling overheads, high job failure rates and insufficient scalability of the EGEE middleware.

Our solution is to first submit lightweight clients which, when started on remote resources, fetch the actual jobs from the central job server and execute them. For this purpose we adopt open-source BOINC platform, used in the last few years for large-scale cycle-stealing such as SETI@HOME and many others. Built for volatile desktop environments, BOINC is capable of efficiently managing billions of jobs and millions of unreliable clients, yielding high performance through sophisticated scheduling mechanisms to overcome network, hardware and software faults. Furthermore, BOINC is firewall friendly and has a built-in accounting functionality.

Grid Access / 10

g-Eclipse - Easy access to Grid infrastructures

Dr. KORNMAYER, Harald ¹; Dr. STUEMPERT, Mathias ²

¹ *NEC Laboratories Europe*

² *FORSCHUNGSZENTRUM KARLSRUHE (FZK)*

Corresponding Author: harald.kornmayer@iwr.fzk.de

The usage of a common and reliable tool eco system will help the developers from different domains to port their legacy applications to Grids. But not only developers will benefit from a general tool Eco System, but also Grid users and Grid resource operators can integrate their use cases in such a general Grid Tool Eco System. The g-Eclipse project built such an general, middleware independent tooling framework for Grid infrastructures on top of the well known Eclipse Eco System. The first release of the framework is available and exemplary support for the gLite middleware is available. The GRIA middleware support is currently be implemented.

The g-Eclipse framework requires stable and reliable basic Grid services like information systems, data replication systems and resource brokers. g-Eclipse is a JAVA application and requires either JAVA APIs or well defined WS descriptions for the basic Grid services, which are independent of the Grid operation system.

Demonstrations / 11

g-Eclipse - Grid in five minutes

Dr. KORNMAYER, Harald ¹; Dr. STUEMPERT, Mathias ²

¹ *NEC Laboratories Europe*

² *FORSCHUNGSZENTRUM KARLSRUHE (FZK)*

Corresponding Author: harald.kornmayer@iwr.fzk.de

The g-Eclipse framework requires stable and reliable basic Grid services like information systems, data replication systems and resource brokers. g-Eclipse is a JAVA application and requires either JAVA APIs or well defined WS descriptions for the basic Grid services, which are independent of the Grid operation system.

For the demo, only a computer with JAVA and an arbitrary OS (Windows, Linux, MacOS) is needed.

Grid Access / 12

Grid website vulnerabilities and the GridSite Security Framework

Dr. MCNAB, Andrew ¹

¹ *UNIVERSITY OF MANCHESTER*

Corresponding Author: andrew.mcnab@man.ac.uk

CSRF and XSS attacks have been used against major public websites, such as Google's GMail, for several years, and generally involve "confused deputy" scenarios in which an authenticated user's web browser is deceived into carrying out an action desire by the attacker. Due to the support for Javascript functions such as XMLHttpRequest in browsers, it can be possible for an attacker's script to communicate with a website using the user's credentials without their knowledge. The credentials involved have typically been HTTP cookies issued by websites to legitimate users, and the attacks have relied on users being logged-in at the time of the attack.

However, in Grid environments many websites authenticate users with their X.509 user certificates, and so users are always logged-in from the point of view of an attacker's script.

Data Management / 14

Grid Storage Interoperability Now!

Dr. JENSEN, Jens ¹; Mr. ROSS, Derek ¹; Mr. DOWNING, Roger ²; Dr. HODGES, Matt ¹

¹ STFC-RAL

² STFC-DL

Corresponding Author: j.jensen@rl.ac.uk

SRM and SRB are traditionally the two "islands" in Grid data, achieving interoperability only amongst themselves. We now show data being transferred between SRMs and SRBs, effectively making SRBs available (with some restrictions) as a Storage Element to gLite-based Grids. The main use case is to enable data sharing between such Grids - files are copied from one to the other and can be registered in transfer catalogues. Rather than using simple tools, the use case calls for using gLite's advanced data management tools.

This work has been done as a contribution to the OGF GIN (Grid Interoperability Now) activities, and as it builds heavily on gLite, it is a suitable activity for the EGEE user forum.

From research to production grids: interaction with the Grid'5000 initiative / 15

DeployWare: A Framework for Automatic Deployment of Software Systems on Grids

FLISSI, Areski ¹; Dr. MERLE, Philippe ²

¹ LIFL - UMR 8022 CNRS

² INRIA

Corresponding Author: areski.flissi@lifl.fr

Deployment, which can be defined as a set of tasks to orchestrate such as installation/uninstallation of software on remote nodes, configuration of nodes and software, starting/stopping of application servers or data collecting, is a nightmare for Grid Computing users. A first challenge is complexity of the orchestration of the several deployment tasks and software dependencies, and the administration of such large distributed software systems. A second challenge is heterogeneity of:

- the software systems to deploy, which use different paradigms (parallel or object-oriented programming, component-based or services-oriented approaches) but also a plethora of runtime platforms/middleware (e.g. MPI, Globus, GridCCM, ProActive, SOA-based systems, etc.)
- the targeted physical infrastructures in terms of hardware, operating systems, network, protocols.

A third challenge is scalability: a typical scenario is to automatically perform the deployment of a software system on thousands of nodes.

Posters - Board P02 / 16

A parallel data mining application for Gene Ontology term prediction

Mr. PSOMOPOULOS, Fotis¹; Mr. GKEKAS, Christos¹; Prof. MITKAS, Pericles¹

¹ *Aristotle University of Thessaloniki*

Corresponding Author: fpsom@danae.ee.auth.gr

Gene ontology can be thought of as a database of expert-based terms. The application presented utilizes the motifs that exist in already annotated protein sequences in order to model the corresponding GO terms. The input data set is created in a semi-automatic way, using the unique (UNIPROT) code of each protein and the InterProScan tool so that all available sequence databases (such as PRODOM, PFAM etc) will be taken under consideration. For each GO term that appears in the original protein set, a new training set is created, which contains all the protein sequences that have been annotated with the specific GO term. Based on the motifs present in the new data sets, a finite state automaton model is created for each GO term. In order to predict the annotation of an unknown protein, its motif sequence is run through each GO model thus producing similarity scores for every term. Results have shown that the algorithm is both efficient and accurate in predicting the correct GO term.

Monitoring, Accounting & Support / 17

The National Grid Service User Accounting System

Ms. WEEKS, Katie¹; Mr. HAINES, Kevin¹; Dr. RICHARDS, Andrew¹

¹ *STFC*

Corresponding Author: k.m.weeks@rl.ac.uk

The User Accounting system has simplified administration of users on the NGS. Users who go over their allocated CPU hours find themselves warned when they reach 90% of their allocation, and automatically locked out when they reach 100%. They can apply for further resources using the online form. Since the system entered production in October 2006, it has locked 74 users out, of which 80% successfully reapplied for more time. 183 new users have applied using the system, enabling the NGS to collect more valuable information on users than before. This automatic collection of data and the use of the Oracle Apex system mean statistics can be produced daily. Figures on how much CPU time is allocated and used, where our users are located, how they found out about the NGS and who funded them are available. The users themselves have benefited from the Accounting System, using the online interface to update their contact data, see their allocations or apply for more resources.

Interoperability and Resource Utilisation / 19

A simple SPAGO implementation enabling the connection of heterogeneous computational resources to a gLite based GRID

Dr. BRACCO, Giovanni ¹; Dr. SANTORO, Andrea ¹; Dr. MIGLIORI, Silvio ¹; Dr. SCIO', Carlo ¹; Dr. QUINTILIANI, Andrea ¹; Dr. ROCCHI, Alessio ¹

¹ ENEA FIM, 00044 Frascati, Roma, Italy

Corresponding Author: bracco@frascati.enea.it

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 operating systems which are not currently supported by gLite middle-ware. SPAGO concept has been developed in the implementation of the ENEA Gateway which now provides access from EGEE to the ENEA AIX SP systems. Although the ENEA Gateway implementation requires a solution for the interoperability between ENEA-GRID and EGEE (due to the different authentication mechanisms, AFS and Kerberos 5 vs. X509), a much simpler solution has been found for standard UNIX/Posix systems where NFS and ssh can be adopted as the base for the proxy implementation. This simplified solution is the object of this presentation.

Finance & Multimedia / 20

Stock Analysis Application

Dr. MICHALAREAS, Giorgios ¹; Dr. COZZINI, Stefano ²; Mr. CORSO, Ezio ²; Prof. SKOURAS, Spyros ³

¹ Department of Electronics and Computer Science, University of Southampton, UK

² ICTP / EU-Indiagrid team

³ Athens University of Economics and Business

Corresponding Author: ecorso@ictp.it

The proposed application will automatically manage the analysis of a large mass of financial data. For each financial instrument there is a zip file: its content is one text file per trading day containing high frequency time-series information for that instrument. Overall there are 4 TB of unzipped data: compression reduces it to roughly 100 GB. One analysis run consists in launching one job for each stock; for each instrument about 150 time-series are constructed and analysed; about 700 instruments will be analysed in each run. Many runs are expected, as both the analysis and the time intervals of interest will change during open-ended research on the properties of the data. For a reasonably exhaustive analysis on all the data, about 200 GB of zipped output files are expected.

From research to production grids: interaction with the Grid'5000 initiative / 22

Modeling the EGEE latency to optimize job time-outs

Dr. MONTAGNAT, Johan ¹; Dr. LINGRAND, Diane ²; Dr. GLATARD, Tristan ¹

¹ CNRS

² UNSA

Corresponding Author: johan@i3s.unice.fr

Jobs submitted to a production grid infrastructure are impacted by a variable delay resulting from the grid submission cost (middleware overhead and queuing time). The actual execution time of a job will depend on the process execution time, which can be known through benchmarks, and the variable grid latency duration, which is difficult to anticipate due to the complexity of the grid infrastructure and the variable load patterns it is enduring. We aim at estimating the grid latency through a probabilistic approach that is well adapted to complex system modeling. We derived a model of the expected execution time of a job function of the time-out value in a time-outing and resubmission setting. To follow on the variable load conditions, a monitoring service sends regular probe jobs to the infrastructure and measures their latency duration. This information is injected into the model and a numeric minimization provides the time-out value that minimizes the expected execution time.

Fusion / 23

Interfacing gLite services with the Kepler scientific workflow

Mr. METERY, Rene ¹; Dr. GUILLERMINET, Bernard ²

¹ CS

² CEA

Corresponding Author: rene.metery@c-s.fr

The interfacing process has been realized with the different operations made by an end-user in mind, when he wants to authenticate and submit a job on the grid. First the creation of an actor making the voms-proxy-init operation was needed to create a temporary proxy as well as returning informations concerning his role in his Virtual Organization. Then the creation of several separate actors making all the operations of the job lifecycle was mandatory to build a complete and modular workflow. After having validated all the actors on the Pre Production Service infrastructure, the creation of a workflow representing the complete job lifecycle has been possible.

Earth Science / 24

Earth Science Application overview in EGEE infrastructure

Dr. PETITDIDIER, Monique ¹; Mr. WEISSENBACH, David ¹

¹ IPSL

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

Due to the large variety of ES applications it is not possible to describe all the results obtained. However some ES applications, already ported, provide scientific 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. The ES applications that already obtained results could be used to point out the kind of problems very well suited to Grid infrastructure like statistical approaches (monte Carlo method, ensemble of jobs..), sharing data or algorithm, performing a very large number of independent jobs that permits to have a rapid solution.

Life Sciences / 25

TriAnnotPipelineGRID, a tool for the automated annotation of wheat BAC sequences

Mr. LEROY, Philippe ¹; Mr. REICHSTADT, Matthieu ²; Mr. GIACOMONI, Franck ¹; Mr. CHARPENTIER, Christian ³; Mr. ALAUX, Michael ⁴; Mr. REBOUX, Sebastien ⁴; Mr. CLAUDE, Alexandre ⁵; Mrs. LIAUZU, Marianne ⁵; Mr. TANAKA, Tsuyoshi ⁶; Mr. ITOH, Takeshi ⁶

¹ INRA Clermont-Ferrand

² Auvergrid

³ INRA Clermont-Ferrand

⁴ INRA URGI

⁵ Lifegrid

⁶ National Institut of Agrobiological Sciences

Corresponding Author: reichma@clermont.in2p3.fr

A long term project of the IWGSC is to sequence the wheat genome to decipher the chromosomal location and biological function of all genes. This knowledge should enhance the understanding of the biology of the wheat plant and create a new paradigm for the improvement of this major crop. Because of the genetic and metabolic conservation among species in the grass family, efforts to decipher gene function in wheat and its close relatives will work synergistically with similar efforts in maize, rice, sorghum, barley and other grasses, for a global understanding of the function, structure, and evolution of the grass genomes.

The aim of the project is to provide a wheat automated annotation system to annotate new BAC sequences and regularly updating previous BAC annotations. It also provides a GBrowse graphical viewer. The pipeline integrates programs for prediction and analysis of protein-coding gene structure, as well as the identification of transposable elements and repeats in BAC sequences.

Demonstrations / 26

WISDOM

Mr. SALZEMANN, Jean ¹; Mr. BLOCH, Vincent ¹; Mr. KASAM, Vinod ²

¹ IN2P3/CNRS

² SCAI Fraunhofer

Corresponding Author: jean.salzemann@clermont.in2p3.fr

The environment evolved throughout the development: It was first made of a set of scripts that generate the jobs, submit the files and check regularly their status while they are on the Grid. Through this abstract we want to present the new environment that is based on the AMGA metadata catalog for more flexibility and on a Java environment that can be used through web services. The environment is very flexible and can be used for any type of bioinformatics application. In this new environment we control the job distribution, maintaining the choice of directly submitting the executable (push mode) or implementing a two-way submission where the system submits generic wrapper jobs which request their payload (the executable and the input data) only when they start to be executed (pull mode).

Grid Access / 27

Exploitation path of Interactive European Grid on User Communities

Dr. CAMPOS PLASENCIA, Isabel ¹; Dr. MARCO DE LUCAS, Jesus ¹

¹ *Instituto de Fisica de Cantabria CSIC*

Corresponding Author: isabel.campos@cern.ch

The Interactive European Grid project (int.eu.grid) aims to deploy and operate a production quality Grid infrastructure oriented to service research communities with specific needs regarding parallel MPI support and interactive access to grid resources. Over the past user forums (Geneva 2006, Manchester 2007) it has been observed that there is a clear need and interest in the scientific community for the services being developed by int.eu.grid on top of gLite as well as for an infrastructure that deploys these services for the user communities.

Demonstrations / 28

Interactive European Grid: Advanced Applications Support Demo

Dr. CAMPOS PLASENCIA, Isabel ¹; Mr. HARDT, Marcus ²; Mr. PLOCIENNIK, Marcin ³; Dr. GARCIA, Lino ⁴

¹ *Instituto de Fisica de Cantabria CSIC*

² *Forschungszentrum Karlsruhe*

³ *Poznan Supercomputing and Networking Center*

⁴ *CESGA*

Corresponding Author: isabel.campos@cern.ch

The package includes an enhanced version of fusion application that won the last User Forum Best Demo award in Manchester. In environmental sciences we will show the integration of an open source Geographical Information Systems on the Grid, GRASS, which is used for environmental analysis of water in reservoirs. Also in the environment sector we will show the application to analyze the evolution of pollution clouds. In medical science we will show how to optimize radiotherapy plans by computing the amount of radiation absorbed by the human body organs in cancer treatments. The total computing time goes down from a few days to few minutes when MPI job submission to several sites is used. In particular we will show the submission of MPI jobs distributed between different clusters using PACX-MPI. A second application of the field of medical science will be shown. It presents a prototype of running Matlab Applications for Ultrasound Computer Tomography (USCT) on interactive grids.

Data Management / 29

Medical Data Manager: an interface between PACS and the gLite Data Management System

Dr. MONTAGNAT, Johan¹; Mr. FROHNER, Akos²; Dr. TEXIER, Romain¹; Mr. NIENARTOWICZ, Krzysztof²; Dr. BAUD, Jean-Philippe²

¹ CNRS

² CERN

Corresponding Author: johan@i3s.unice.fr

Hospitals continuously produce tremendous amounts of image data that is managed by local PACS (Picture Archiving and Communication Systems). These systems are often limited to a local network access although the community experiences a growing interest for data sharing and remote processing. Indeed, patient data is often spread out different medical data acquisition centers. Furthermore, researchers in the area often need to analyze large populations whose data can be gathered through federations of PACS. Opening PACS to the outer Internet is challenging though, due to the stringent security requirements applying to medical data manipulation. The gLite Data Management System provides the distribution, user identification, data access control and secured transportation core services needed to envisaged wide scale deployment of the medical imaging applications. The MDM provides an upper layer to interface to PACS and manipulate medical data with the required level of security.

Posters - Board P04 / 30

A Full Stokes Model for Large Scale Ice Dynamics Simulations

Dr. ZWINGER, Thomas¹; Dr. RÅBACK, Peter¹; Dr. GRÖHN, Matti¹; Dr. LYLÿ, Mikko¹

¹ CSC - Scientific Computing Ltd.

Corresponding Author: matti.grohn@csc.fi

In contrary to scaled equations, Elmer applies Full Stokes (FS) simulations, where horizontal scales of the mesh are of similar size than vertical, leading to a scale-up of the problem size by a factor 100. Models earlier run on a single workstation consequently occupy 100 and more processors if FS is applied, demanding parallel computations on clusters or Grid environments. The main focus of the work presented here is to make the needed modules for FS ice-dynamics modeling within Elmer available on the EGEE environment. With the increased capacity of the EGEE infrastructure, we attempt to obtain an enhanced resolution down to a horizontal scale in the size of a few kilometers, resulting in computations containing millions of degrees of freedom. At these scales, details such as ice streams, which were below the resolution of a standard SIA run can be investigated. This provides an enhanced insight into the mechanics and thermodynamics of ice sheets.

Posters - Board P05 / 31

EGEETomo: 3D electron tomography on EGEE. A user-friendly and fault-tolerant grid application.

Dr. BILBAO-CASTRO, J.R.¹; Dr. GARCÍA, I.¹; Dr. FERNÁNDEZ, J.J.¹

¹ Dept. Computer Architecture and Electronics, University of Almería, 04120, Almería, Spain

Corresponding Author: jrbcast@ace.ual.es

3D Electron Tomography is a key imaging technique when studying large biological complexes like cells organelles or even whole cells structures. Projection images of the specimen are taken through electron microscopes. Nevertheless, technical limitations reduce the number and quality of the projections that can be obtained from the specimen under study. Because of this, the commonly used reconstruction algorithms, like WBP, which are relatively fast, present some limitations in terms of the reconstruction quality they provide. On the other hand, the iterative reconstruction techniques provide better reconstruction quality but at a much higher computational cost. Fortunately, the whole reconstruction task can be divided into smaller, independent, reconstruction subtasks. This makes the grid a perfect place to run tomographic reconstructions, with hundred of long-lasting independent tasks. Nevertheless, making such application usable implies making it user-friendly and fault-tolerant.

Monitoring, Accounting & Support / 32

Strategies for experiment-specific monitoring in the Grid

Dr. MENDEZ LORENZO, Patricia¹; Dr. SCIABA, Andrea¹; Dr. CAMPANA, Simone¹; Dr. SANTINELLI, Roberto¹; Dr. LANCIOTTI, Elisa¹; Dr. MICCIO, Enzo¹; Dr. MAGINI, Nicolo¹; Dr. DI GIROLAMO, Alessandro¹

¹ CERN IT

Corresponding Author: patricia.mendez@cern.ch

The LHC experiments perform most, if not all, of their computing activities on Grid resources. This requires an accurate and updated picture of the status of the Grid services used by them, and of the services which are specific to the experiment. To achieve this, a common method is to periodically execute tests on the services, where the functionalities tested may be different from a VO to another. The SAM framework, developed for the EGEE operations, can be easily used to run and publish the results of arbitrary tests, from basic functionality tests, to high-level operations from real production activities. This contribution describes in detail how the monitoring system of each LHC experiment has taken advantage of SAM

Demonstrations / 33

Ganga - powerful job submission and management tool

Dr. MAIER, Andrew ¹; Dr. LEE, Hurng-Chun ¹; Dr. MENDEZ LORENZO, Patricia ¹; Dr. MOSCICKI, Jakub ¹; Dr. LAMANNA, Massimo ¹; Dr. MURARU, Adrian ¹

¹ CERN IT

Corresponding Author: patricia.mendez@cern.ch

Ganga has already gained widespread use, the incomplete list of applications using Ganga include: Imaging processing and classification (developed by Cambridge Ontology Ltd.), Theoretical physics (Lattice QCD, Feynman-loop evaluation), Bio-informatics (Avian Flu Data Challenge), 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 without the underlying knowledge of each one: EGEE gLite and NorduGrid ARC middlewares, Condor and Cronus (Condor/G), various batch systems, etc

Monitoring, Accounting & Support / 34

New monitoring applications in the Experiment Dashboard

ANDREEVA, Julia ¹; GAIDIOZ, Benjamin ¹; MAIER, Gerhild ¹; RICARDO, Rocha ¹; SAIZ, Pablo ¹; CATALIN, Cirstoui ¹; SIDOROVA, Irina ²; WAKEFIELD, Stuart ³; HERRALA, Juha ¹

¹ CERN

² CERN/JINR

³ IC

Corresponding Author: julia.andreeva@cern.ch

The LHC experiments ALICE, ATLAS, CMS and LHCb are preparing for data acquisition planned to start in 2008. The LHC experiments are relying on several GRID infrastructures (LCG/EGEE, OSG, NDGF). Providing the reliable monitoring system which enables the transparent view of the experiment activities across different middleware platforms and combines the Grid monitoring data with information which is specific for the experiment/activity/application is a vital and challenging task. The Experiment Dashboard is used by all four LHC experiments to follow their activities on the Grid. There are multiple monitoring applications of the Experiment Dashboard which are in production and are widely used by the LHC VOs. At the same time the project is in active development phase. Existing applications are evolving and new applications are developed following the suggestions of the user community.

Monitoring, Accounting & Support / 35

Increased productivity for emerging Grid applications: the application support system

Dr. MAIER, Andrew¹; Dr. LEE, Hurng-Chun¹; Dr. MENDEZ LORENZO, Patricia²; Dr. MOSCICKI, Jakub¹; Dr. LAMANNA, Massimo¹; Dr. MURARU, Adrian¹

¹ CERN IT

² CERN IT/GD

Corresponding Author: patricia.mendez@cern.ch

The CERN Grid application support team has been working with the following real-life applications: medical and particle physics simulation (Geant4, Garfield), satellite imaging and geographic information for humanitarian relief operations (UNOSAT), telecommunications (ITU), theoretical physics (Lattice QCD, Feynman-loop evaluation), Bio-informatics (Avian Flu Data Challenge), commercial imaging processing and classification (Imense Ltd.) and physics experiments (ATLAS, LHCb, HARP).

Using the EGEE Grid we created a standard infrastructure - set of services and tools - customized for the emerging applications. This includes creation of a generic Virtual Organization easily accessible by small communities and adding resources and services to it. We provide the consultancy service to help the porting of the applications to the Grid using the Ganga and DIANE tools. The system may be operated with only small maintenance and support overhead and is easily accessible by new applications.

Posters - Board P06 / 36

The WLCG Common Computing Readiness Challenge: CCRC`08

Dr. MENDEZ LORENZO, Patricia¹; Dr. SCIABA, Andrea²; Dr. RENSALL, Harry²; Dr. CAMPANA, Simone²; Dr. SANTINELLI, Roberto²; Dr. LAMANNA, Massimo²; Dr. LANCIOTTI, Elisa²; Dr. DI GIROLAMO, Alessandro²; Dr. MAGINI, Nicolo²; Dr. MICCIO, Enzo²; Dr. SHIERS, Jamie²

¹ CERN IT/GD

² CERN IT

Corresponding Author: patricia.mendez@cern.ch

The LHC machine will produce some 15PB of data per year. The management and the analysis of these data relies on a worldwide production Grid service involving hundreds of sites from EGEE and collaborating Grids. One significant challenge remains: to demonstrate that these computing facilities can be used to satisfy simultaneously the needs of the 4 major experiments of the LHC at full 2008 rates. During the CCRC`08 we will demonstrate precisely this. Given the importance of the challenge, two phases are foreseen: an initial run in February, when not all sites will have their full 2008 resources in place and a second period in May, when the full 2008 capacity is required to be in place

Life Sciences / 37

Analysis of Metagenomes on the EGEE Grid

Mr. APARICIO, Gabriel¹; Dr. BLANQUER, Ignacio¹; Prof. HERNÁNDEZ, Vicente¹

¹ UPV

Corresponding Author: iblanque@dsic.upv.es

Metagenomic analysis requires several iterations of alignment and phylogenic classification steps. Source samples reach several millions of sequences. These sequences are compared to the eukaryotic species of the "Non-redundant" database.

The deployment process involves three stages: First, public databases are copied in relevant SEs to reduce the access time by increase the geographic replication of the data. Second, the available resources are tested through short test jobs that check the different operations. Finally, the experiment is performed.

The sequences of the source sample are split into different jobs. Each job is submitted through the RBs to the CEs that have been selected in the test phase. Jobs copy all the relevant databases from close SEs to the local storage, install locally the BLAST and clustalW software and execute the scripts. After the completion of the job, results are copied back through the SEs and GridFTP as a backup solution.

Workflow and Parallelism / 38

Optimizing a Grid workflow for the EGEE infrastructure: The case of Wien2k

Dr. BERGER, Maximilian¹; Prof. FAHRINGER, Thomas¹

¹ University of Innsbruck

Corresponding Author: maximilian.berger@uibk.ac.at

In real production grids the time between submitting a grid activity and its execution ranges from 10 to 60 minutes. When porting a complex workflow to the Grid, such as Wien2K, this overhead does not only appear once, but repeatedly, and increases the execution time of workflows largely.

In our previous presentation we showed aggregation of grid activities. Since then we have also experimented with other means of reducing overhead through scheduling:

Worker nodes may be scheduled without an active task, which they request through a pull model from a coordinator. These workers would be submitted once, and thus the overhead of scheduling them would only appear once. However, this mechanism is unfair towards other users, as it occupies resources for a longer period than one task. We will show results of our experiments, and classify them according to speed and fairness.

Astronomy & Astrophysics / 39

Cosmological application in the Grid environment: Detection of SZ Clusters of Galaxies in data from the ESA Planck satellite mission

Dr. LOPEZ-CANIEGO, Marcos¹; Dr. HERRANZ, Diego¹; Dr. MARTINEZ-GONZALEZ, Enrique¹

¹ *Instituto de Fisica de Cantabria, CSIC-UC, Santander, Spain*

Corresponding Author: caniego@ifca.unican.es

Several techniques have been proposed to detect the SZ clusters. Most of them are based on linear filters that try to take into account the frequency dependence of the SZ effect to combine information from different channels to produce a single map where the clusters can be detected with a higher SNR than in the individual frequency maps (Planck will image the microwave sky in nine frequencies ranging from 30 GHz to 857 GHz). Within the Planck Collaboration an exercise to detect SZ clusters using realistic simulations has been proposed. The purpose of this exercise is to compare the performance of different algorithms to detect the SZ clusters, give an estimation of their integrated flux (and error) and an estimation of their size (and error). We have tested our implementation of the Matched Multifilter (MMF) developed by Herranz et al. 2002. We have analyzed nine full resolution Planck sky maps and detected approx. 1100 clusters above the 5 sigma level.

Grid Access / 40

Portals for Earth Science

Mr. LINFORD, Julian¹; Mr. PLEVIER, Camiel²; Dr. LECCA, Giuditta³; Mr. POLIAKOV, Andrey⁴; SIMO, Branislav⁵; Mr. GOTAB, Benoit⁶

¹ *ESA*

² *DutchSpace*

³ *CRS4*

⁴ *GCRAS*

⁵ *UISAV*

⁶ *IPGP*

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

Large sets of ES data are available and distributed all over the world. The data come from satellites, ground-based network and sensors aboard balloons, aircrafts, and/or sounding rockets. A critical requirement is the organisation of the data, their accessibility and in some cases tools to define the workflow of the application.

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, permits to define requirements for implementation on gLite and for development of new services.

Life Sciences / 41

GRID BASED TELEMEDICINE APPLICATION FOR GATE MONTE CARLO DOSIMETRIC STUDIES

Dr. MAIGNE, Lydia¹; Mr. DIARENA, Matteo¹; Mr. NOWAK, Simon¹; Dr. DONNARIEIX, Denise²; Dr. THIAM, Cheick¹; Dr. BRETON, Vincent¹

¹ CNRS/IN2P3, LPC Clermont-Ferrand

² Centre Jean Perrin

Corresponding Author: maigne@clermont.in2p3.fr

The secured web portal has been designed to be used by physicians and medical physicists to perform Monte Carlo calculations :

to optimize the acquisition and data processing protocols of medical scans,
to ensure accurate treatment plannings for some specific radiotherapy applications.

In that way, developments focus on the creation of a secured web platform to access grid computing to split the GATE simulations.

Functionalities of this platform enable :

A secure authentication to assess grid computing

The retrieving of medical data from a PACS server, this service contains the anonymization of data, encryption and extraction of metadata stored in a base on the grid.

The secured and parallelized computing using medical images on the grid.

The monitoring and resubmission of calculations in case of failure.

The visualization of results (dosimetry map, sinograms...) as images, directly from the client machine of the use.

Demonstrations / 42

A telemedicine platform for information and image management on the grid

Mr. DIARENA, Matteo¹; Mr. NOWAK, Simon¹; Dr. BRETON, Vincent¹; Mr. LEGRE, Yannick¹; Dr. MAIGNE, Lydia¹; Mr. DAO, Van Tuyet²; Prof. BOIRE, Jean-Yves³; Dr. BILBAO, José Roman⁴

¹ CNRS IN2P3 LPC Clermont-Ferrand

² Ho Chi Minh City Institute of Information Technology

³ ERIM INSERM ER14

⁴ University of Almeria

Corresponding Author: matteo.diarena@cern.ch

The medical field offers a wide and challenging scenario in which new grid applications can be developed to improve collaborative work between scientists. The development of grid-based medical applications needs to take into account some key factors such as the need to conform to strict legal constraints in terms of data privacy and security. Moreover physicians are quite reluctant to use new applications that change their way of working, for this reason applications developed on this context need to be as intuitive and user friendly as possible.

Demonstrations / 43

What can the CIC portal do for end users?

CORDIER, Helene ¹; AIDEL, Osman ¹; MATHIEU, Gilles ¹; L'ORPHELIN, Cyril ¹

¹ *CNRS/IN2P3*

Corresponding Author: helene.cordier@in2p3.fr

The CIC portal added value to the Grid infrastructure is for all the EGEE actors, whether scientist, VO or site manager, or grid operator. Every type of group has its own entry point to the portal. The information on the operational state of the grid is filtered and presented according to the usefulness for a particular group.

The tools presented to a given group are those which could be useful for it, 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.

Monitoring, Accounting & Support / 44

New developments on the LHCb Bookkeeping

Dr. LANCIOTTI, Elisa ¹; Dr. MAIER, Andrew ¹; Dr. SANTINELLI, Roberto ¹; Dr. KOBLITZ, Birger ¹

¹ *CERN IT*

Corresponding Author: patricia.mendez@cern.ch

The Bookkeeping (Bkk) is a crucial component in the LHCb software infrastructure, both for the production, as it registers and uploads to the database all newly produced files, as well as for the data analysis, since it is the tool which allows physicists to retrieve datasets and their metadata.

The motivation for this activity on the Bkk arises from requirements of the physicists, who outlined a lack of efficiency of the service. Issues raised include the current user interface, implemented as a web page, is not flexible enough and has broken functionality. Furthermore, the service does not provide exhaustive information on the metadata and returns the output to the user in a rather cumbersome way. The objective now is to provide a new client to allow physicists to search for data and relative metadata in the most flexible and efficient way possible. The new client is implemented in Python, for consistency with the rest of the LHCb software infrastructure

Life Sciences / 45

ThIS on the Grid

Mrs. CAMARASU, Sorina ¹; Mr. GUIGUES, Laurent ¹; Mr. BENOIT-CATTIN, Hugues ¹; Mr. SARRUT, David ²

¹ CNRS - CREATIS LRMN

² CNRS - CREATIS LRMN, Léon Bérard Cancer Center

Corresponding Author: sorina.camarasu@creatis.insa-lyon.fr

Our main requirements concern computing resources and data management.

The simulation is split into sub-jobs. Each sub-job uses a different random seed number, allowing to be statistically independent and to be run concurrently. By dividing one simulation into hundreds of different sub-jobs, computation time can be reduced from more than one day to less than 1 hour if computing resources on the grid are rapidly available.

Data requirements are also important. ThIS needs about 50Mo of input data and can produce between 15 to 150Mo of output data per each sub-job.

ThIS is based on the Geant4 toolkit and consequently uses Geant4 and CLHEP libraries. In order to keep our application independent on the installed software on the grid, we decided to compile it statically. Thus, each time a complete simulation is run, we provide the executable and all the needed input files as a tar ball stored on a storage element (SE) of the grid. Once completed, output data are retrieved and merge on a SE.

Data Management / 46

A WS-DAIR Compatible Interface for gLite-AMGA

Mr. JAVADZADEH BOLOORI, Ali ¹; Dr. KOBLITZ, Birger ²

¹ Royal Institute of Technology (KTH)

² CERN

Corresponding Author: ali.javadzadeh.boloori@cern.ch

The addition of a WS-DAIR interface to the gLite AMGA metadata service will greatly improve the extensibility and interoperability with other Data access services based on the Open Grid Service Architecture. As the standard also defines the interaction of relational database services among each other, it will allow to integrate data access services of different types.

We will present as an example the Avian Flue Drug Discovery application implemented by Academia Sinica Grid Computing (ASGC), which has been used as a test case for validation and evaluating the new interface, compared to the older TCP-Socket based of AMGA with respect to performance, scalability, fault tolerance, interoperability and ease of use for Grid applications.

The result of the evaluation has also been presented at SC '07.

As AMGA is in fact the first metadata service to adapt the WS-DAIR standard, we will present our findings on the usability of this standard as well as on its overall design.

Fusion / 47

FUSION RESULTS WITHIN EGEE

Mr. VELASCO, José Luis ¹; Dr. TARANCÓN, Alfonso ¹; Mr. VÁZQUEZ-POLETI, José Luis ²; Dr. VOZNESENSKY, Vladimir ³; Dr. TERESHCHENKO, Max ⁴; Dr. CASATEJÓN, Francisco ⁵; Dr. CAPPA, Álvaro ⁵

¹ *BIFI*

² *Universidad Complutense*

³ *Kurchatov Institute*

⁴ *General Physics Institut, Russian Academy of Sciences*

⁵ *CIEMAT*

Corresponding Author: francisco.castejon@ciemat.es

Ion kinetic transport application allowed the estimation of ion collisional transport both in tokamaks and stellarators, showing properties that could not be found by the customary methods: transport is not diffusive and that there exist important asymmetries, oppositely to what was thought. Further improvements of the applications are ongoing.

MaRaTra calculations have allowed to estimate the heating of plasmas confined in complex geometries, like that of TJ-II stellarator, by quasi-electrostatic waves. This is especially challenging, since it is necessary to consider a huge number of rays. The optimization of wave launching requires a large number of runs with 10^4 - 10^5 rays. The heating system for TJ-II stellarator has been designed using these results. Gridway metascheduler is used in MaRaTra.

Stellarator Optimization. The application is running and several stellarator configurations are explored. Future activities will involve the optimization of TJ-II stellarator.

Kepler

Workflow and Parallelism / 48

Execution Time Prediction of Imperative Paradigm Tasks for Grid Scheduling Optimization

Ms. LIM, Mei Kuan ¹; Dr. YAP, Yee Jiun ¹; Ms. KIRAN, Maleeha ¹; Dr. LAI, Weng Kin ¹

¹ *Mimos Berhad*

Corresponding Author: lim.mkuan@mimos.my

The proposed system is intended to be implemented in the KnowledgeGRID Malaysia to improve the efficiency of the scheduling system. The work is focused on imperative paradigm tasks since they are commonly used in the aforementioned grid. Imperative paradigm refers to a sequence of commands for the computer to perform and the normally used imperative paradigm programming languages are R!, Fortran and C. In the current phase, all testing and evaluation is done via a web-based wrapper which is developed specially for this purpose. The testing and evaluation involves test cases sampled from jobs submitted to the aforesaid grid.

Demonstrations / 49

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

Dr. USTININ, Mikhail¹; Dr. NAZIPOVA, Nafisa¹; Mr. FILIPPOV, Sergey¹; Mr. ZAITSEV, Aleksandr¹; Dr. FIALKO, Nadezhda¹; Mrs. TYULBASHEVA, Gayane¹; Dr. TEPLUKHIN, Alexander¹; Dr. USTININ, Dmitriy¹; Prof. LAKHNO, Victor¹

¹ IMPB RAS

Corresponding Author: ustinin@impb.psn.ru

The MathCell Project includes 3D interactive living cell model, encyclopedia on mathematical modeling of cell and software for modeling of basic processes in living cell. Within the limits of the Project the interactive environment was developed, which allows to perform calculations of mathematical models using GRID infrastructure. The special Job Maintenance System was developed which automatically allows User Logging & 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.

Demonstrations / 50

The EGEE user support infrastructure

Dr. ANTONI, Torsten¹; Dr. MILLS, Alistair²

¹ GGUS, INSTITUT FÜR WISSENSCHAFTLICHES RECHNEN, FORSCHUNGSZENTRUM KARLSRUHE

² CERN

Corresponding Author: torsten.antonni@iwr.fzk.de

The grid user support model in EGEE can be captioned "regional support with central coordination". This model is realised through a support process which is clearly defined and involves all the parties that are needed to run a project-wide support service. This process is sustained by a help desk system which consists of a central platform integrated with several satellite systems belonging to the Regional Operations Centres (ROCs) and the Virtual Organisations (VOs). The central system (Global Grid User Support, GGUS) interconnects the ROCs and VOs and the other project wide groups like middleware, network, and service groups. Additionally the central system acts as a portal for users, offering the possibility to document problems and requests in the form of trouble tickets. Since all trouble tickets pass through the GGUS system it is the perfect place to store information on problems and of course also their solution in a knowledge base, available to users and support staff.

Workflow and Parallelism / 51

Towards a statistical model of EGEE load

GERMAIN-RENAUD, Cecile; VAZQUEZ, Emmanuel¹; COLLING, David²

¹ *Supelec*

² *Imperial College London*

Corresponding Author: germain@lal.in2p3.fr

Preliminary results indicate that EGEE job traffic shares some properties with the Internet traffic: the distributions of the inter-arrival times seem to be heavy-tailed and the time series of the loads indicate long-range power-law correlations. Precise characterizations are currently investigated on two aspects.

a) Marginal distributions. Modeling the distributions at different spatial and temporal scales will provide an insight into the way the flow of jobs is actually dispatched on the resources. Relevant statistical approaches are parametric modeling of the nominal behavior as well as the tail behavior of the distributions, and specifically, extreme value theory.

b) Time-dependent structures in the time series. We explore two kinds of well-known stochastic models: Poisson processes, with a possibly non homogeneous or stochastic intensity; and self-similar stochastic models such as fractional Gaussian noises (FGN) and fractional ARIMA processes (ARFIMA).

Fusion / 52

Optimisation applications support in RDIG Fusion VO

Mr. VOZNESENSKY, Vladimir¹; Dr. SEMENOV, Igor²; Dr. DOBRETISOV, Vladimir¹; Mr. RYABINKIN, Eugene¹; Mr. LARIONOV, Andrey¹; Mr. MIKHAILOV, Mikhail²

¹ *Inst. of Information Systems, RRC "Kurchatov Inst."*

² *Nuclear Fusion Inst., RRC "Kurchatov Inst."*

Corresponding Author: vovic@nfi.kiae.ru

Two examples of CPU-intensive numerical optimisation applications were found in Russian nuclear fusion and plasma science: optimisation of stellarator magnetic field configuration and optimisation of the reflectometry radiowave shape in ITER plasma.

The first application has been successfully ported from a supercomputer to the EGEE infrastructure. This porting has revealed two issues. First, the grid dictates use of mathematical methods that allow asynchronous completion of parallel jobs without loss of efficiency. For optimisation, that means shifting from iterative (gradient-based) methods to results-based spool-driven (genetic, stochastic interpolation) ones.

Second, tools for such application-specific jobs and data management are required.

The second application is in its development phase. The functional requirements are the same, but it requires orders of magnitude more CPU than the first one. Supercomputers cannot fulfill its demands.

Posters - Board P07 / 53

Service-Oriented Applications development support in gLite

Ms. GORANOVA, Radoslava; Prof. DIMITROV, Vladimir

Corresponding Author: radoslava.goranova@cern.ch

Basic characteristics of service-oriented architectures are self-describing interfaces in platform-independent XML documents and interoperability between different systems and programming languages. gLite provides a lot of grid services but very few of them have self-describing interfaces with endpoints and WSDL documents. The lack of WSDL files for main gLite grid services make them very hard to use in service-oriented grid application. And moreover, the end user often needs more complex functionality, which can be achieved as composition of grid services or grid processes.

Interoperability and Resource Utilisation / 54**WLCG-RUS: An Extensible Solution to Resource Usage Service**Mr. CHEN, Xiaoyu ¹; Mr. AKRAM, Khan ¹¹ *Brunel University***Corresponding Author:** xiaoyu.chen@brunel.ac.uk

The project originated from providing a RUS compliant solution for WLCG accounting, which requires collection of usage data from three operational Grids, the Open Science Grid (OSG), EGEE and NorduGrid. The collection of usage data are to be stored centrally and summarized for usage reporting on per site, per VO, per month basis. These collected usage data are persistent in relational database based on WLCG accounting schema. At present, usage providers from each operational Grid simply email SQL statements for insertion of summary usage records to project manager who runs a simple script to populate usage data into storage. The WLCG-RUS project is therefore proposed to automate and standardize usage data sharing and reporting processes with interoperability to RUS implementations available or being developed in operational Grid projects.

Posters - Board P08 / 56**Grid enabled applications for modeling, simulation and optimization**Dr. NEAGU, Gabriel ¹; Dr. ANDREI, Neculai ²; Dr. SUCIU, Alin ³; Dr. ZAHARIE, Daniela ⁴; Dr. SIMA, Vasile ²; Prof. CRISTEA, Valentin ⁵; Dr. NAE, Catalin ⁶; Prof. POTOLEA, Rodica ⁷; Prof. PETCU, Dana ⁸; Mr. STANCIU, Alexandru ²; Mr. POP, Florin ⁹; Mr. PRICOP, Victor ¹⁰¹ *National Institute for R in Informatics - ICI Bucharest, Department of Research*² *ICI Bucharest, Department of Research*³ *UTCN, Computer Science Department*⁴ *WUT, Computer Science Department*⁵ *University "Politehnica" of Bucharest - UPB, National Center for Information Technology - NCIT*⁶ *National Institute for Aerospace Research - INCAS Bucharest, Subsonic Wind Tunnel Laboratory*⁷ *Technical University of Cluj-Napoca - UTCN, Computer Science Department*⁸ *West University of Timisoara - WUT, Computer Science Department*⁹ *UPB, NCIT*¹⁰ *INCAS Bucharest***Corresponding Author:** gneagu@ici.ro

The candidate applications to be ported on grid were selected based on the following criteria: mature implementation in classic mode, extensive memory and/or computational requirements, grid as innovative approach for the given MOSI domain, expected user community in research and academic area, potential interest from industry. The list of selected applications includes: GridModRed - Model Order Reduction, GridIdent - System Identification, CGALLP and BIBFR - Unconstrained optimization based on conjugate gradient algorithms and Related high performance library (developed by ICI Bucharest), OPT-GT and MFCC - Optimizer based on Grid Technology and Application cluster for CFD-FEM oriented simulation (INCAS Bucharest), DEMO/G - Distributed Evolutionary Multiobjective Optimization on Grid (WUT), CryptoGrid - Cryptographic and Cryptanalytic Algorithms for the Grid (UTCN), DIOGENES - Application oriented task scheduling using genetic algorithms (UPB).

Data Management / 57

OpenSAML extension library and API to support SAML2.0 - XACML protocol for interoperable authorisation infrastructure in distributed Grid applications

SAGEHAUG, Hakon Tuvin; DEMCHENKO, Yuri; VENTURI, Valerio; FORTI, Alberto

Corresponding Author: hakon.sagehaug@bccs.uib.no

Authorisation is an important component of the Grid security infrastructure. AuthZ decision is typically based on the AuthZ policy that contains a set of access control rules depending on user credentials or attributes. In many cases AuthZ service is a part of the application platform and uses a policy specific to application. Consistency of the access control enforcement can be achieved by introducing the Site Central AuthZ Service (SCAS) that will allow applying common policies centrally and leave a possibility for applying local policies and enforcement mechanisms.

The proposed SAML-XACML library and API provide all necessary functionality for the PEP (bound to the Grid resource or application) to call out to external SCAS. The API provides the helper classes to create and parse SAML-XACML messages and also extendible functionality for policy Obligations handling. The proposed functionality is specifically targeted to support pool account management when submitting Grid job to WNs

Computational Chemistry & Material Science / 58

Supporting Statistical Semiconductor Device Analysis using EGEE and OMII-UK Middleware

Mr. REID, Dave ¹; Dr. ROY, Scott ¹; Dr. MILLAR, Campbell ¹; Dr. ROY, Gareth ¹; Prof. SINNOTT, Richard ²; Dr. STEWART, Graeme ¹; Prof. ASENOV, Asen ¹

¹ *University of Glasgow*

² *NeSC*

Corresponding Author: seabass@elec.gla.ac.uk

The only method by which variability in nano-CMOS transistor characteristics (and thus the variability and yield of the circuits making use of them) can be predicted, understood and designed around is through large scale numerical simulation. To capture in detail the statistical parameter distributions in real device architectures requires vast computational resources generating extensive device ensembles. Using resources such as ScotGrid (www.scotgrid.ac.uk), it has been possible, for the first time, to generate an ensemble of more than 100,000 microscopically different 35nm gate length devices – comparable with state-of-art production devices. Early results indicate that designers must assume much larger parameter variations than are currently considered. It is also clear that some of the assumptions underlying present design techniques are no longer valid, due to significant, non-analytic deviations at the extremes of the statistical distribution.

Demonstrations / 60

GRelC DAIS: Towards Data Access and P2P Data Integration in gLite based Production Grids

Dr. FIORE, Sandro ¹; Prof. ALOISIO, Giovanni ¹; Prof. BARBERA, Roberto ²; Mr. VADACCA, Salvatore ³; Mr. NEGRO, Alessandro ⁴; Dr. CAFARO, Massimo ¹; Mr. GIORGIO, Emidio ²

¹ SPACI Consortium University of Salento

² INFN Sez. di Catania, via S. Sofia, 64 - 95123 Catania

³ Euro-Mediterranean Centre for Climate Change

⁴ ISUFI Settore Nanoscienze e Grid Computing

Corresponding Author: sfiore@mailing.unile.it

Decoupling routing aspects from data access, the GRelC DAIS can be used in several data integration scenarios. Some examples are: integration of information stored within several distributed metadata DBs for Earth Science to perform queries across distributed collections described by a common metadata model; distributed queries on accounting services deployed at several sites (e.g. APEL or DGAS within EGEE) to infer “global reports” on the grid infrastructure usage (ordered list - per job, cputime, etc. - of VOs/GridUsers which are mostly using the grid). The architecture is very modular, query language independent and easily extensible, so the integration can be related to a set of XML DBs (using XPath), distributed Relational DBs (using SQL), distributed BDII (using LDAP), distributed flat files (using GFAL libraries), distributed Monitoring of the GRelC DAIS P2P Network (retrieving a KML files for client-side Google-Earth based visualization), etc.

Posters - Board P09 / 61

A Network Monitoring Framework in the SCoPE Project

Dr. PARDI, Silvio ¹; Dr. SERIO, Fulvio ¹; Prof. RUSSO, Guido ¹; Dr. CIUFFOLETTI, Augusto ²

¹ Dipartimento di Scienze Fisiche - Università degli Studi di Napoli Federico II - Italia

² Univ. di Pisa - Largo B. Pontecorvo - Pisa

Corresponding Author: spardi@na.infn.it

The network reliability appear to be a crucial factor to guarantee the distributed services availability and the proper system functioning of a Grid infrastructure.

The network performances can affect dramatically the job computation time in those applications that processing bulk of dataset and is obviously crucial during data replication activities.

Currently in the main Grid deployments, the network resource is considered as pure facility and the middleware components act agnostically to respect the network parameters, so that the Grid infrastructure working globally below the best effort threshold, if we considered the Grid, in the first approximation, as an integration of computational, storage and network resources.

In this work we present GlueDomain, a network monitoring framework created to support the middleware services by offering a set of measurement useful for general operations, bandwidth performance previsions and our deploy in the SCoPE infrastructure.

Life Sciences / 62

WISDOM: Grid enabled identification of active molecules against targets implicated in Malaria

Mrs. DA COSTA, Ana Lucia¹; Mr. KASAM, Vinod²

¹ CNRS-IN2P3, LPC Clermont-Ferrand

² Fraunhofer-SCAI

Corresponding Author: dacosta@clermont.in2p3.fr

The in silico workflow which we employed starts with docking to evaluate the binding energy between a target and a ligand, then selected compounds are refined by Molecular Dynamics (MD). In 2005, against Plasmeprin target, the WISDOM initiative achieved 41 million dockings, using FlexX, in 45 days on 1700 computers which is equivalent to 80 CPU years on one machine. The best 5000 compounds identified were reranked by MD with Amber9 in 7 days equivalent to 124 CPU days on one machine. In 2006, this success led to a second assault against 4 other malaria targets. During 90 days, ~140 million dockings were achieved which is equivalent to 413 CPU years, representing an average throughput of 80,000 dockings per hour. MD simulations were performed on 15000 docking poses against wild type Dihydrofolate reductase target and 5000 docking conformations against Glutathione-S-transferase target respectively. The total 25 000 simulations lasted for 25 days equivalent to 347 CPU days in one machine.

Computational Chemistry & Material Science / 63

EXPLOITING GRID FOR CONDENSED MATTER APPLICATIONS: QUANTUM DOTS AND BUILDING CLUSTERS ATOM-BY-ATOM

Prof. KANHERE, Dilip¹; Mr. PUJARI, Bhalchandra¹; Mr. ZORRIASATEIN, Shahab¹; Mr. GHAZI, Mohammad¹; Mr. KAWARE, Vaibhav¹; Dr. JOSHI, Kavita¹

¹ Pune University

Corresponding Author: bspujari@unipune.ernet.in

QUANTUM DOTS: A fully self consistent real space DFT approach has been implemented.

We calculate the ground state charge and spin density, magnetic state, eigenvalue spectrum and investigate effects of impurity over a wide range of sizes of quantum dots (50-300 nm with 2-20 electrons). The results reveal Wigner localized state (Wigner molecule). The impurity induces the novel magnetic states and anti ferromagnetic spin distribution and enhances the localization.

CLUSTERS: We have obtained equilibrium geometries (~100) of clusters of Sodium (Sizes 20-150) by a combinations of simulated annealing and local minimization. We calculate stability, binding energies, HOMO-LUMO gap etc. We also investigate growth patterns and find that the growth shows order- disorder cycles. The shape analysis of the ground states is shown to be correlated with the shapes of heat capacities. Thus nature of the ground states and the isomers spectrum dictates the behavior of clusters at finite temperature.

Data Management / 64

The Development of SRM interface for SRB

Mr. TSAI, Fu-Ming¹; Mr. WEI-LONG, Ueng¹; Mr. SHUN-CHE, Hsiung¹; Mr. TSUNG-HSIEN, Lin¹; Mr. ERIC, Yen¹; Mr. SIMON C., Lin¹

¹ *Academia Sinica Grid Computing*

Corresponding Author: sary357@twgrid.org

The standard SRM services for SRB were developed to make the popular SRB data grid system interoperable with the gLite e-infrastructure. AMGA is used to implement the File catalog and to provide uniform interface for replication and to the backend database. Currently the development is under standard SRM functional testing and validation, and establishment of full SRM 2.2 functionalities will be finished by early 2008. In the first phase, targeted use cases are to: 1) Make SRB an archival system of gLite-based e-Infrastructure, 2) Support Lifetime policy for files - volatile, durable, and permanent, and 3) Impose the same VO and security control to SRB as the Grid infrastructure. Other than the basic directory, permission and data access functions, user authorization, web service interface, gridftp deployment, and SRB-DSI had been supplemented and demonstrated at the SC07 in Nov. 2007 as well. Next, the interoperation between SRB and DPM, dCache, Castor, etc.

Demonstrations / 65

fMRI analysis on EGEE with the VBrower and MOTEUR

Mr. DE BOER, Piter¹; Dr. OLABARRIAGA, Silvia²; Dr. GLATARD, Tristan²; Mr. BOULEBIAR, Kamel²

¹ *Institute of Informatics, University of Amsterdam*

² *Institute of Informatics / Academic Medical Center, University of Amsterdam*

Corresponding Author: glatard@essi.fr

* Data management requirements

The data has to be handled directly by end-users. It is thus mandatory to set up a high-level data management tool providing a uniform view of distributed storage. A user-level file access control is also required to prevent users from ruining someone else's experiment because of wrong data manipulations.

Although the Logical File Catalog provides a uniform view of the SEs, a layer is missing to make them usable by end-users. In an experiment, several file transfers need to be performed between local and Grid storage, which must be hidden. Moreover, file access control is based on the VO membership and not sufficiently fine-grained.

* Jobs management requirements

The needs of the application in this area are:

- Intuitive parameter sweep specification
- Intermediate status check-pointing
- Fault tolerance mechanisms

The first point is addressed by a dedicated GUI. The two last points are expected to benefit from the use of a workflow management system.

Posters - Board P10 / 66

Fair Grid Scheduling

Mr. MEDERNACH, Emmanuel ¹

¹ LPC - Clermont, France

Corresponding Author: medernac@clermont.in2p3.fr

Jobs durations vary a lot among groups or users depending of the kind of applications. For instance a group dedicated to test the middleware have 90% of their jobs of duration less than 5 minutes. Biomed jobs have a quarter of jobs running less than 2 minutes, another quarter between 2 minutes and 8 hours, another quarter between 8h and one day and the last quarter between one day and 3 days. Other groups have many jobs running up to 3 days.

Each site have to schedule jobs coming from the grid in an efficient and fair way. Site needs sometimes to be able to justify part of resources granted to groups or users.

Consequences of unfairness are possible leaving of groups or users. These differences in the way jobs are scheduled could lead to unfair treatment of users or groups. We propose to take fairness into account when dealing with multi-users scheduling problems and to seek how scheduling could be improved in order to be fair.

Posters - Board P11 / 67

Smart adaptative identifier for a medical data retrieval system

Mr. DE VLIENER, Paul ¹; Prof. BOIRE, Jean-Yves ²

¹ ERIM / LPC

² ERIM INSERM ER14

Corresponding Author: vliener@clermont.in2p3.fr

Grids provide a lot of possibilities in terms of data storage and exchange. By using the telemedicine application developed at LPC Clermont-Ferrand we plan to add some functionalities to manage patient data throughout medical centres.

To get back all data around Europe concerning a patient we need to be able to identify these data ownership while certifying that alone, it must be impossible to regain the patient id by using data encryption technologies.

For that we decided to consider the patient as the center of the medical data retrieval system. The idea is to store a scalable and dynamic identification number, stored in the patient' smart card and this identifier should be the key to regain all data regarding this patient.

Each part of this identifier match a accurate medical act in the patient life. So the data retrieval system consists in an analysis of these parts, a download of associated data and a decryption.

Data Management / 68

Distributed Data Management on the petascale using heterogeneous grid infrastructures with DQ2

Mr. BRANCO, Miguel¹; Dr. GARONNE, Vincent¹; Mr. LASSNIG, Mario²; Mr. SALGADO, Pedro¹

¹ CERN

² CERN University of Innsbruck, Austria

Corresponding Author: mario.lassnig@cern.ch

DQ2 is specifically designed to support the access and management of large scientific datasets produced by the ATLAS experiment using heterogeneous grid infrastructures. The DQ2 middleware manages those datasets with global services, local site services and enduser interfaces. The global services, or central catalogues, are responsible for the mapping of individual files onto DQ2 datasets. The local site services are responsible for tracking files available on-site, managing data movement and guaranteeing consistency of available data. The enduser interfaces provide users with the ability to query, manipulate and monitor datasets and its transfers. The distinction between global and local services is a core design decision as it clearly separates site-specific information, e.g. local site storage management, from global information. With this separation, any change within site infrastructures does not affect global reliability of the system and QoS requirements can be guaranteed.

Posters / 69

Charon GUI - Feature Rich Interface to Distinct Grid Niches

KMUNICEK, Jan¹; PLSEK, Vitezslav¹; KUBA, Martin¹

¹ CESNET

Corresponding Author: kmunicek@ics.muni.cz

Charon GUI is Java-based application currently running at specific server that functions as a dedicated frontend/user interface to individual virtual organization on EGEE or national Grid environment. The one and only prerequisite on the server side is to have Java Runtime Environment installed. Charon GUI displays on a remote X-server that can be either Linux OS or MS Windows with X-Window emulator. The full list of Charon GUI features includes key Charon Extension Layer functionality (job submission, monitoring, results retrieval as well as exploration of available application modules) enhanced by the graphical representation. Charon GUI functions as a laboratory book to keep track of end user's research projects and computational jobs allowing full project and/or jobs manipulation. Secondly, the exhaustive job overview and filtering functionality is ready to provide overview of the individual research project progress.

Interoperability and Resource Utilisation / 70

INTERCONNECTING GRID, DESKTOP GRID AND NETWORK

Mr. GEORGAKOPOULOS, Kostas ¹; Mr. STEFANIDIS, Vassilis ²; Prof. MARGARITIS, Kostas ²

¹ *University of Macedonia*

² *University Of Macedonia*

Corresponding Author: kostas.georgakopoulos@gmail.com

We are researching and evaluating systems that can provide a bridge to non-dedicated resources for the grid. We focused on the Condor system because it is a technology that has been around for many years. Furthermore, computing elements (CE) that use the LCG or gLite middleware can be configured to interact with Condor pools and forward jobs to be executed. Our aim was to test the functionality of this bridge and also to research issues like security, reliability and network functionality. Our testbed for this research was a Condor pool we set up and the gLite Pre-Production site we administer as part of the EGEE project. Additionally we researched, to some extent, other systems that provide a non-dedicated resources computing model like B.O.I.N.C and the LiveWN project. Our end goal is to provide some case studies that document the possible solutions for expanding the grid with non-dedicated resources and also to investigate the restrictions and boundaries imposed by these solutions.

Astronomy & Astrophysics / 71

VO AUGER Large Scale Monte Carlo Simulations using the EGEE Grid Environment

Ms. SCHOVANCOVA, Jaroslava ¹; Dr. CHUDOBA, Jiri ²; Mr. MULAC, Milos ¹; Mr. RUDA, Miroslav ¹; Mr. SALVET, Zdenek ¹; Mr. SITERA, Jiri ¹; Mr. SUSTR, Zdenek ¹; Dr. TRAVNICEK, Petr ³; Mr. DVORAK, Frantisek ¹; Mr. FILIPOVIC, Jiri ¹; Dr. KMUNICEK, Jan ¹; Mr. KRENEK, Ales ¹; Dr. MATYSKA, Ludek ¹

¹ *CESNET*

² *CESNET and Institute of Physics, ASCR, Prague*

³ *Institute of Physics, ASCR, Prague*

Corresponding Author: jaroslava.schovancova@cern.ch

VO AUGER simulations made use of many CPUs connected in the EGEE Grid, which enabled us to simulate events with higher precision. The results of simulations were uploaded and stored on Storage Elements and registered in LFC Catalogue, therefore they can be accessed globally by all the VO AUGER members.

Monitoring, Accounting & Support / 72

Communication tools between Grid Virtual Organisations, middleware deployers and sites

DIMOU, Maria ¹

¹ CERN

Corresponding Author: maria.dimou@cern.ch

Today's Grid usage is still very far from the simplicity and functionality of the web.

While pressing for middleware usability, we try to turn the Global Grid User Support (GGUS) into the central tool for identifying areas in the support environment that need attention. To do this, we exploit GGUS' capacity to expand, by including new Support Units that follow the project's operational structure. Using tailored GGUS database searches we obtain concrete results that prove where we need to improve procedures, Service Level Agreements and deployment methods. These are reliable indicators of the health status of all Grid services. They are also useful to anticipate and plan changes of direction in the required strategy and procedures.

It is via regular reporting to the ROCs and the users within the VOs that we show the value of using GGUS.

It is also by using user input for GGUS improvement that we try to make this a trully useful tool.

Posters / 73

Problem Solving Environment for stereological modeling of muscle cells

Mr. CIGLAN, Marek ¹; Mr. PARULEK, Julius ²; Mr. SIMO, Branislav ³; Dr. SRAMEK, Milos ⁴; Dr. HLUCHY, Ladislav ³; Dr. ZAHRADNIK, Ivan ²

¹ *Institute of Informatics, Slovakia*

² *Institute of Molecular Physiology and Genetics, Slovak Academy of Sciences*

³ *Institute of Informatics, Slovak Academy of Sciences*

⁴ *Austrian Academy of Sciences, Austria*

Corresponding Author: viet.ui@savba.sk

From the viewpoint of grid computing, SM-PSE requires:

Computing Intensive Parametric Studies: The created cell tools enable users to repetitively create hundreds of huge models, which, consecutively, will be stereologically verified. The computation of volume and surface densities of a single model can take up to hours; i. e., computation of all the models would last for days.

Heterogeneous Computing Platform Support: Visualization rendering platform is limited to Windows OS. Therefore, combining model solving with model visualization on Windows platform has to be provided. Moreover, rendering requirements are not the only ones that define the platform demands. Several existing tools, capable of working with stereological data that are considered to extend our PSE are available also solely on the Windows platform.

Demonstrations / 75

Design, implement and deploy Grid oriented applications in a cooperative way: a biomedical use case

Dr. PORRO, Ivan ¹; Dr. SCIFO, Salvatore ²; Dr. SCUDERI, Giordano ³; Dr. GATTI, Stefano ⁴; Dr. TORTEROLO, Livia ¹; Dr. CORRADI, Luca ¹; Dr. FATO, Marco ¹; Dr. BARBERA, Roberto ⁵; Dr. PARISI, Salvo ⁶; Dr. VENUTI, Nicola ⁷; Dr. SCARDACI, Diego ⁸; Dr. CALANDUCCI, Antonio ⁸

¹ *Department of Communication, Computer and System Sciences, University of Genoa*

² *COMETA Consortium*

³ *Unico Informatica s.r.l*

⁴ *Aitek S.p.A.*

⁵ *University of Catania and INFN*

⁶ *IR T engineering s.r.l.*

⁷ *NICE s.r.l.*

⁸ *INFN Catania*

Corresponding Author: pivan@dist.unige.it

Nowadays, many biomedicine studies are dealing with large, distributed, and heterogeneous repositories as well as with computationally demanding analyses. Complex integration techniques are more often required to handle this complexity, even if for small sized applications, when they are intrinsically distributed: this particular scenario is frequently found in medical informatics applications, where the health care provider is not a single institution but a collection of actors that play different roles in the territory.

The BMPortal is a platform thought to promote collaboration and cooperation among scientists and healthcare research groups, enabling the remote use of resources integrated in complex software platform services forming a virtual laboratory. It is designed to host several medical use cases and it is able to deploy several analysis that can be combined in large applications using a workflow strategy: here, the engineering of BIOLAB SPM Alzheimer application is presented

Posters / 76

Organising scientific data by dataflow optimisation on the petascale

Mr. LASSNIG, Mario ¹

¹ *CERN University of Innsbruck, Austria*

Corresponding Author: mario.lassnig@cern.ch

We analyse the Distributed Data Management system Don Quijote 2 (DQ2) of the High-Energy Physics experiment ATLAS at CERN. ATLAS presents unprecedented data transfer and data storage requirements on the petascale and DQ2 was built to fulfill these requirements. DQ2 is built upon the EGEE infrastructure, while seamlessly enabling interoperability with the American OSG and the Scandinavian NorduGrid infrastructures. Thus it serves as a relevant production-quality system to analyse aspects of dataflow behaviour in the petascale. Controlled data transfers are analysed using the central DQ2 bookkeeping service and an external monitoring dashboard, provided by ARDA. However monitoring dynamic data transfers of jobs and enduser data transfers cannot happen centrally because there is no single point of reference. Therefore we provide opportunistic clients tools for all scientists to access, query and modify data. Those tools report the needed usage information in a non-intrusive, scalable way.

Grid Access / 78

RINGrid: conceptual design of the remote instrumentation systems

Dr. LAWENDA, Marcin ¹; Dr. MEYER, Norbert ¹; Mr. KOTSOKALIS, Constantinos ²; Prof. DAVOLI, Franco ³; Mr. PROKOSCH, Thomas ⁴; Dr. KALVACHEV, Yuri ⁵

¹ *Poznan Supercomputing and Networking Center*

² *GRNET*

³ *Italian National Consortium for Telecommunications*

⁴ *GUP*

⁵ *CLMC*

Corresponding Author: lawenda@man.poznan.pl

The analysis of the wide implied RIS aspects are under of interest of the RINGrid (Remote Instrumentation in Next-generation Grids) project. This activity is part of 6th European Framework Programme and has been launched in October 2006. Briefly, the RINGrid project will provide systematically identification of instruments and corresponding user communities, the definition of their requirements as well as careful analysis of the remote instrumentation synergy with next-generation high-speed communications networks and grid infrastructure. These results will be the basis for the definition of recommendations for designing next-generation RIS. RINGrid associates partners coming from Europe and Latin America from 10 institutions. On the one hand it allows to achieve required level of generality and on the other hand gives desired impact by gathering scientists from different research domains. User communities are related with unique laboratory devices e.g. NMR spectrometers.

Posters / 79

Enabling Distributed Access and Parallel Processing of Bioinformatics Data in Grid Environment: The EKTORAS platform

Prof. MAGLOGIANNIS, Ilias ¹

¹ *University of Aegean*

Corresponding Author: imaglo@hol.gr

Users, accessing the platform's 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. 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 to be usable by the range of algorithms available. The results of this pre-processing step are directed to the Grid's storage elements (SE). Then the data are being processed by parallel applications distributing the parallel chunks & jobs to various nodes-processors of the Grid.

Posters / 80

Workflow meta-scheduler for Grid Environments

Mr. POP, Florin¹; Prof. CRISTEA, Valentin¹

¹ *University "Politehnica" of Bucharest*

Corresponding Author: florinpop@cs.pub.ro

Based on our contribution to the 2nd EGEE User Forum, we extend the DIOGENES (DIstributed Optimal GENetic algorithm for grid application Scheduling) project that provides a solution to the Grid scheduling problem at application level. The extension consists of a new algorithm that aims to achieve a distributed, fault-tolerant, scalable and efficient method for dependable task (DAG) assignment in Grid environments. Several metrics including scheduling time, scheduling length and load balancing are used to highlight improved behavior of our proposal as compared with other existing scheduling strategies (HLFET – Highest Level First with Estimated Times, ETF – Earlier Time First, MCP – Modified Critical Path). The scheduling priority is represented by the ALAP (As Late As Possible) parameter. The improvements are made in the node sorting step: defining an additional criteria for nodes with the same ALAP and use of descendant prediction to improve the critical path.

Posters / 81

Advantages of Pre-Production WLCG/EGEE services for VOs and Users

Mr. THACKRAY, Nicholas¹; Dr. FERNANDEZ SANCHEZ, Carlos²; Mr. FREIRE GARCIA, Esteban²; Mr. SIMON GARCIA, Alvaro²; Dr. LOPEZ CACHEIRO, Javier²; Mr. RETICO, Antonio¹

¹ *CERN*

² *CESGA*

Corresponding Author: esfreire@cesga.es

Currently a well-defined set of procedures is followed in PPS, beginning with the pre-deployment activity and finishing with the approval of a new middleware release that can go into PROD. The PPS Coordination Team takes care of supervising all these steps, mainly trying to spot possible bugs in the middleware before it goes into PROD.

Unfortunately, VO contribution in this part of the deployment is currently very limited and the PPS team does not have a full feedback from these important groups. One of the problems is that PPS is not very well-known for many VOs and users and this situation should change. When a VO has a new application or there is an upcoming middleware upgrade, they should be aware that they can test it in PPS, helping to discover possible problems before going into PROD. Indeed, PPS sites have a great quantity of free resources that can and should be used. Finally, VO and users should remember that PROD is a critical service not to be used for testing purposes.

Workflow and Parallelism / 82

GRid-aware Optimal data Warehouse design (GROW)

Mr. VELINOV, Goran¹; Mr. CEREPNALKOSKI, Darko¹; Mr. JAKIMOVSKI, Boro¹

¹ *University of Sts. Cyril and Methodius*

Corresponding Author: boroj@ii.edu.mk

The application is implemented as a Java framework for executing genetic algorithms in a distributed fashion using a Grid. The framework consists of two parts: genetic algorithm framework and grid tools. The first part enables researchers to easily implement new optimization problems by simply extending several classes. The second part enables researchers to make their application Grid-aware. In other words it enables easy Grid job submission, job status and output retrieval. The GROW application is a VIS optimization. The chromosomes are bit sequences, each bit representing weather particular view or index is materialized in the database. The chromosomes are evaluated on a set of database queries, where for each query we estimate the time and memory usage for its execution. The parameters for the GA optimization influence both per population GA execution and grid execution workflow. Some parameters are: mutation and crossover probability, islands, epochs, seasons, migration width.

Earth Science / 83

Dissemination and exploitation of Grids in Earth Science

Dr. HLUCHY, Ladislav¹; Dr. TRAN, Viet¹; Dr. PETITDIDIER, Monique²; Mr. SOM DE CERFF, Wim³; Dr. SCHWICHTENBERG, Horst⁴; Dr. LINFORD, Julian⁵; Dr. FUSCO, Luigi⁵

¹ *Institute of Informatics, Slovakia*

² *Institut Pierre Simon Laplace*

³ *Koninklijk Nederlands Meteorologisch Instituut*

⁴ *Fraunhofer Institute for Algorithms and Scientific Computing*

⁵ *European Space Agency*

Corresponding Author: viet.ui@savba.sk

The key requirements of ES applications on data management, job management and portal technology have been identified and analyzed in five ways: (1) A panel of representative ES applications have been analysed; (2) Existing data management tools and policies that are being used in ES applications are surveyed in order to find common required features of ES community; (3) The existing Grid data management technologies have been analyzed in order to provide the solutions for the requirements; (4) The available Grid middleware and tools for job submission and workflow management have been analyzed and several missing features required by ES applications are identified; (5) Existing Grid portals for ES applications are surveyed for creating common Grid based portal and service oriented architectures for ES applications.

Data Management / 84

A service oriented framework to create, manage and update metadata for earth system science

Dr. RONNEBERGER, Kerstin ¹; Dr. KINDERMANN, Stephan ¹; Dr. BIERCAMP, Joachim ¹

¹ DKRZ

Corresponding Author: ronneberger@dkrz.de

The system is built modular and service oriented to be expandable and easily maintainable. All Xml metadata instances, as well as the service layer composed of XSL stylesheets, XQuery/XUpdate modules and XML templates and property files are stored in a native XML database (eXist) and are accessible via different interfaces, depending on the interest in the metadata.

For users of the data, described by the metadata, a detailed view on the information contained, is offered; users of the metadata can explore the structure and content of the metadata format via HTML pages; additional interfaces to manually (for existing data) and automatically (during processing) create, update and parse metadata files are offered to potential providers of data and metadata. For the automatic update, a request file can be submitted to the eXist database via a java interface, which sets on OGSADAI. Manual update is realized via an interactive GUI based on XForm technology.

Earth Science / 85

Grid computing for wildfire danger assessment: porting RISICO in glite

Dr. D'ANDREA, Mirko ¹; Dr. DAL PRA, Stefano ²; Dr. VERLATO, Marco ²; Dr. FIORUCCI, Paolo ¹; Dr. GAETANI, Francesco ¹; Dr.

ANGELINI, Valerio ³

¹ CIMA Foundation

² INFN

³ CNR-IMAA

Corresponding Author: stefano.dalpra@pd.infn.it

RISICO works on a set partitioned on identical and algorithmically independent squared cells. Computing a cell requires "cell status" and meteo data and produces a "next status" as also the wanted output of the simulation. A RISICO run on the Italian territory with cells of 1 km² requires approximatively the computing of 330.000 cells and 150Mb of input data, producing 1Gb of output data in a 20 min of time on a common workstation. Finer simulations using 0.01 km² sized cells, leads to quadratically increased input and output data size as also needed computation time. We can afford these higher needs through Grid. We divide a run into a few tens of jobs, each using Storage Elements for input and output data, and the requirement that every job successfully ends in a given maximum time. This gets achieved thanks to a "job status monitor" program running at UI level which polls about jobs termination, retrieve outputs and resubmits failed or late to finish jobs.

Monitoring, Accounting & Support / 86

R-GMA: Now With Added Authorization

Mr. CHIDAMBARAM, Arun ¹; Mr. DUNCAN, Alastair ¹; Dr. JIANG, Ming ¹; Mr. KUSEJU, Adebisi ¹; Dr. WILSON, Antony ¹; Ms. BHATTI, Parminder ¹; Dr. FISHER, Steve ¹

¹ RAL

Corresponding Author: s.m.fisher@rl.ac.uk

R-GMA is currently being used by APEL, the ARDA Dashboard and Service Discovery on the LCG grid and by Grid Ireland. APEL uses a producer at each site to publish accounting data. A consumer is used to accumulate all of the data in a central location where it is migrated off line for later analysis. The ARDA dashboard has a consumer that pulls in monitoring data about the status of grid jobs as published by resource brokers. This enables real time monitoring of the progress of jobs on the grid. The Service Discovery API has an R-GMA plugin. A producer at each site publishes information about available services and their current status. This enables other middleware components to select required services. Grid Ireland are using R-GMA to monitor TCP logs. With the increased robustness of the code and the fine grained authorization we expect that users will find new applications for R-GMA.

From research to production grids: interaction with the Grid'5000 initiative / 87

Simple, fault tolerant, lightweight grid computing approach for bag-of-tasks applications

GEORGIU, Yiannis ¹; CAPIT, Nicolas ¹; BZEZNIK, Bruno ²; RICHARD, Olivier ¹

¹ LIG Laboratory

² Projet CIMENT

Corresponding Author: yiannis.georgiou@imag.fr

Our approach consists by a (RMS) resource management system OAR, responsible for the efficient allocation of local cluster resources and a grid lightweight service CIGRI that uses only the idle cluster resources by not interfering to the normal functionality of the interconnected clusters.

The approach is based on the concept of "best effort" tasks, introduced by OAR. This type of jobs have the minimum execution priority and are submitted only if there is an idle resource. However, if during their execution the resource is requested by a local cluster user, the grid "best-effort" job is killed by the local RMS. The CIGIR grid fault-treatment mechanism can resubmit the killed jobs and thus guarantee a successful completion of the whole calculation.

Features like web portal for grid monitoring, checkpoint/restart, results collection, support of diskless PCs environment (ComputeMode) and application data transfer are implemented and provide ease of use and quality of service to the user.

Posters / 88

Numerical modeling of electrodynamic aggregation of magnetized dust in electric discharges

Dr. KUKUSHKIN, Alexander¹; Mr. MARUSOV, Nikolay¹; Dr. MINASHIN, P.V.²; Mr. NEVEROV, V.S.³; Mr. CHEREPANOV, K.V.¹

¹ *RRC "Kurchatov Inst."*

² *Moscow Engineering Physics Institute, Russia*

³ *Moscow Engineering Physics Institute*

Corresponding Author: isemenov@nfi.kiae.ru

We examine the hypothesis for a fractal condensed matter composed of magnetized nanodust capable of forming a skeleton of filamentary structures observed in various laboratory electric discharges, severe weather phenomena and space [2], suggested for explaining the unexpected longevity of these filaments and their unexpected (sometimes transverse) direction with respect to that of main electric current. A 3-D numerical model [3] of many-body system of basic blocks (magnetized, electrically conducting thin rods) managed to describe the following processes:

- self-assembling of a quasi-linear filament from a random ensemble of basic blocks and the capability of such filaments to close the electric circuit,
- self-assembling of coaxial tubular skeleton in a system of initially-linear electric current filaments, composed of above basic blocks and linked to the biased electrodes,
- the trend towards self-similarity of structuring during these self-assembling processes.

Posters / 89

Development and adaptation of a web enabled in silico oncology application in grid environment

Mr. ATHANAILEAS, Theodoros¹; Mr. MENYCHTAS, Andreas¹; Mr. KOUSIOURIS, Georgios¹; Ms. DIONYSIOU, Dimitra¹; Mr. STAMATAKOS, Georgios¹; Prof. KAKLAMANI, Dimitra¹; Prof. VARVARIGOU, Theodora¹; Prof. UZUNOGLU, Nikolaos¹

¹ *NTUA*

Corresponding Author: a_menyctas@telecom.ntua.gr

This tool was developed following a multi-tier architectural approach in order to provide access to the core grid services through a state of the art web interface. Initially the in silico oncology source code was modified to facilitate the execution of simulations to grid nodes using parameter files that are automatically created from the end users using the tool. Additionally, grid specific wrapper-scripts were developed for setting up the simulation and for gathering useful statistics for the QoS mechanisms. The end users exploiting these mechanisms' functionalities were able to create dynamically simulation specific JDL files based on the user requirements and on the status of the grid infrastructure. Finally a web portal was designed and developed that simplified the access to the grid resources and automated the job submission and monitoring. This portal enabled additional services to the framework, such as user management and job scheduling based on QoS criteria.

Interoperability and Resource Utilisation / 90

gCube Grid Services

Mr. ANDRADE, Pedro ¹; Dr. PAGANO, Pasquale ²

¹ CERN

² CNR-ISTI

Corresponding Author: pedro.andrade@cern.ch

gCube reflects within its name a three-sided interpretation of the Grid vision of resource sharing: sharing of computational resources, sharing of structured data, and sharing of application services. As such, gCube embodies the defining characteristics of computational grids, data grids, and virtual data grids. Precisely, it builds on gLite middleware for managing distributed computations and unstructured data, includes dedicated services for managing data and metadata, provides services for distributed information retrieval, allows the orchestration of workflows, and offers a novel approach for managing these services. Rather than interfacing the infrastructure, the gCube services are transparently deployed across its constituent nodes. This is genuinely ambitious and entirely novel: like computational resources and data before, application logic in gCube becomes a pervasive commodity within an infrastructure which abstracts over its physical location at any point in time.

Interoperability and Resource Utilisation / 91

SAGA API for gLite Service Discovery

Dr. FISHER, Steve ¹; Dr. PAVENTHAN, A ¹

¹ RAL

Corresponding Author: s.m.fisher@rl.ac.uk

The API returns a list of service descriptors matching search criteria. A random choice can then be made from the URLs returned. Information about the individual services can be obtained from the descriptors if it is desired to rank the services returned or produce a web page of some subset of services. The search criteria are specified by means of three filters – service, VO and data. The service filter selects on the basis of some GLUE attributes such as the “type” of service. The VO filter allows the user to select from those services he is allowed to use and the data filter makes use of a GLUE feature of key/value pairs associated with each service. SAGA components have a “plugin” architecture. In the case of Service Discovery, a plugin is required for each underlying information system; so far we have R-GMA and BDII support in C++.

Demonstrations / 93

All-in-one graphical tool for grid middleware management

Dr. CARON, Eddy ¹; Dr. AMAR, Abdelkader ¹; Mr. LOUREIRO, David ¹

¹ ENS-Lyon / INRIA / CNRS / UCBL

Corresponding Author: eddy.caron@ens-lyon.fr

When dealing with grid environments, grid middleware are powerful tools for the development of computational servers able to exploit the available resources. But managing a grid middleware, and a fortiori the grid environment itself can be a hard task when no dedicated tools exist. Some are usable through nice graphical interfaces, but they are all dedicated to one or some limited tasks and do not fulfilled all the needs of a grid end-user wanting to deploy grid applications easily and fastly. The aim of this paper is to present an all-in-one software, designed for the management of grid middleware gathering user-friendly graphical interfaces answering to the various needs of a end-user. The software moreover eases the use of the grid by avoiding the scripting layer under a nice GUI enabling the user a faster and more efficient use of the grid environment. By this way they demonstrate how the DIET Dashboard fulfilled all the needs of an unified tool for the grid management.

Posters / 94**Early failure detection: a method and some applications**GERMAIN-RENAUD, Cecile; KRENEK, Ales¹¹ *Masaryk Univeristy***Corresponding Author:** germain@lal.in2p3.fr

The complexity of the hardware/software components, and the intricacy of their interactions, defeat attempts to build fault models only from a-priori knowledge. A black-box approach, where we observe the events to spot outliers, is appealing by its simplicity, and large body of experience in quality control. The general challenge is to detect anomalies as soon as possible. Much better solutions than simple thresholding are routinely used in e.g. clinical trials and the supervision of production lines. In the case of abrupt changes, the Page-Hinkley statistics provides a provably efficient method, which minimizes the time to detection for a prescribed false alarm rate. We have applied this method to quantities (e.g. number of arrived and served jobs per unit of time) that are easily computed from the output of existing services. The main result is that we are able to efficiently detect failures of very different origins (e.g. some software bugs, blackholes) without human tuning.

Workflow and Parallelism / 95**Extension of DIRAC to enable distributed computing using Windows resources**TSAREGORODTSEV, Andrei¹; COLES, Jeremy²; HARRISON, Karl²; LI, Ying Ying²; LYUTSAREV, Vassily³; PARKER, Andy²¹ *Centre de Physique des Particules de Marseille*² *University of Cambridge*³ *Microsoft Research***Corresponding Author:** karl.harrison@cern.ch

The LHCb experiment, designed for high-precision studies of matter-antimatter asymmetries in the decays of b-hadrons, is one of the four main experiments at the CERN Large Hadron Collider (LHC). DIRAC has been developed to meet the experiment's need for processing petabytes of data per year, using globally distributed resources. It can be used either as a standalone Grid implementation, or as an optimisation layer on top of another system, such as the EGEE Grid, and has performed impressively in data challenges held since 2002. Although mostly written in Python, which is largely platform-independent, various features of the implementation have previously limited use of DIRAC to Linux machines. We have extended DIRAC to allow its use also on Windows platforms, making the core code more generic in a number of places, integrating Windows-specific solutions for certificate-based authentication and secure file transfers, and enabling interaction with Microsoft Windows Compute Clusters.

Posters / 96

The Experiment Dashboard for medical applications

ANDREEVA, Julia ¹; BOULEBIAR, Kamel ²; LAMANNA, Massimo ¹; MAIER, Gerhild ¹; OLABARRIAGA, Silvia D. ³; SAIZ, Pablo ¹; SIDOROVA, Irina ¹; TEMPLON, Jeff ⁴; CIRSTOIU, Catalin ¹; DA ROCHA, Ricardo ¹; GAIDIOZ, Benjamin ¹; HERRALA, Juha ¹; KAN, Andrevan ⁴

¹ CERN

² Informatics Institute, University of Amsterdam

³ Academic Medical Center, University of Amsterdam

⁴ NIKHEF

Corresponding Author: julia.andreeva@cern.ch

Functional magnetic resonance imaging (fMRI) is a popular tool used in neuroscience research to study brain function.

The Virtual Lab for fMRI (VL-fMRI) is developed as one of the activities of the "Medical Diagnosis and Imaging" subprogram of the Virtual Laboratory for e-Sciences Project. VL-fMRI has taken steps to enable data management and analysis tasks for fMRI studies on the Grid infrastructure. Since spring 2006 the Experiment Dashboard is used for job processing monitoring of the VL-fMRI activities. The Experiment Dashboard provides an easy way to users to follow their jobs on the distributed infrastructure. Furthermore, the system allows to detect problems or inefficiencies of Grid sites or services and to understand the underlying problem. This functionality is important for site administrators and VO support teams.

Workflow and Parallelism / 97

Interactive Workflow Management in int.eu.grid

Mr. SIMO, Branislav ¹; Mr. HABALA, Ondrej ¹; Dr. HLUCHY, Ladislav ¹

¹ Institute of Informatics, Slovak Academy of Sciences

Corresponding Author: branislav.simo@savba.sk

Grid computing is a useful tool for complex scientific applications, enabling their execution over a large pool of resources. Many of the deployed applications are a complex workflow composed of many smaller parts. However, most of these applications appear to their users as a monolithic black box, usually driven by a complicated and finely tuned shell script. Once the job starts executing, the user has no finer control over it than being able to abort it or to wait until it finishes. The described tool is able to visualize the inner workflow of the application. The user can completely control the job during execution, can see partial results, and can even alter it while it is still running. This allows not only to associate the produced data to the job workflow, to extend it, or to shorten it, but also to interactively debug and tune the job—something that would otherwise be possible only for a domain expert, and would be more time-consuming.

Posters / 98**AIDA: Atlas graphical Interface for Distributed Analysis**

Dr. LAMPOUDIS, Christos¹; Mr. HOUIRIS, Anastasios¹; Prof. SAMPSONIDIS, Dimitrios¹; Prof. PETRIDOU, Charikleia¹; Dr. CHRISTIDI, Ilektra¹; Mr. KANELLOPOULOS, Christos²; Mr. KOROSGLOU, Paschalis²; Mr. ZILASKOS, Dimitrios²

¹ Aristotle University of Thessaloniki - Department of Physics

² Aristotle University Grid Operations Center

Corresponding Author: christos.lampoudis@cern.ch

AIDA specifically addresses the needs of the ATLAS high energy physics experiment, performing large scale data processing on globally distributed resources. AIDA is an easy-to-use strong tool implemented in java. User faces a single Graphical User Interface, instead of having a set of different applications. In such a way, the application offers the opportunity to avoid the complexity of GRID's command line interaction. This tool assists the user in all steps of a job's life cycle, starting from job creation, data set mining, multiple job submission, job monitoring and result collection.

Demonstrations / 99**The G-DSE (Grid-Data Source Engine)**

Dr. VUERLI, Claudio¹; Dr. TAFFONI, Giuliano¹; Mr. BARISANI, Andrea¹; Dr. PASIAN, Fabio¹

¹ INAF-OA Trieste

Corresponding Author: vuerli@oats.inaf.it

The I/O of Astronomical Applications almost always involve one or more databases. Grids unable to directly access databases force users to access databases off-line and transfer data of interest in classical SEs before the execution of the applications. Similarly, output will be stored in classical SEs and transferred to a database off-line after the termination of the application. This way of working is extremely uncomfortable and discourage users to choose the Grid as a collaborative tool. The G-DSE is one of the proposed solutions; with G-DSE a database becomes an embedded resource of the Grid. Grids extended through the G-DSE allow users to submit special jobs containing instructions to get data from databases, process them and store the processing results in one or more databases. In this way users can intermix database-related and processing related directives in their jobs. A databases in fact is one of the shared resources in Grid as a CPU, a disk space storage and so on.

Astronomy & Astrophysics / 100**Making the Grid and the Virtual Observatory mutually interoperable**

Dr. TAFFONI, Giuliano¹; Dr. PASIAN, Fabio¹; Dr. VUERLI, Claudio¹

¹ INAF-OA Trieste

Corresponding Author: vuerli@oats.inaf.it

Astrophysical applications handle simulated, theoretical and observed data and the amount of data requested by a single application is not negligible. Astronomical data are usually kept in databases most of them are now federated in the VObs. Users accessing these data usually expect some key capabilities like: a) find data by specifying their characteristics; b) retrieve them whatever is their physical location; c) generate them on-the-fly if not found; d) permanently store them in some place for the benefit of future users; e) apply further processing to them; f) save the results somewhere to be subsequently exploited by the whole community. The Grid allows the sharing of resources of different nature (hardware, software, data, and so on) so its tight synergy with the VObs is of strategic importance. A set of standards, tools and services are currently in preparation to make possible the necessary interoperability of these two technologies.

Posters / 101

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

ROUEFF, Evelyne¹; ROY, Fabrice¹; THUILLOT, william²; VALIRON, pierre³; VALLS-GABAUD, David⁴; WIESENFELD, Laurent³; WOZNIAK, Herve⁵; LECUBIN, julien²; DUBERNET, Marie-Lise⁶; LE-PETIT, Franck¹; LE-SIDANER, Pierre⁷; BERTHIER, Jerome²; SCHAAF, Andre⁸; ALIM, Jean-Michel¹; BALANCA, Christian⁶; COMBES, Françoise⁶; CHILINGARIAN, Igor⁶; CORASANITI, Pier-Stefano¹; DI MATTER, Paola⁶; FAURE, alexandre³; FEAUTRIER, Nicole⁶; FOUCHARD, Marc²; FROUARD, julien²; SEMELIN, Benoit⁶; GODARD, Benjamin⁶; LAINEY, Valery²; LE BOURLLOT, Jacques¹; LENAIN, Jean-Philippe¹; LESAFFRE, Pierre⁶; MENARD, Francois³; RAZERA, Yann¹; REVAZ, Yves⁶; SHIH, Albert⁷; SLEZAK, Eric⁹; SOL, Helene¹; SPIELFIEDEL, Annie⁶; STEHLE, chantal⁶; TABURET, sylvestre⁴

¹ LUTH - Observatoire de Paris

² IMCCE - Observatoire de Paris

³ LAOG

⁴ GEPI - Observatoire de Paris

⁵ CRAL

⁶ LERMA - Observatoire de Paris

⁷ UMS - Observatoire de Paris

⁸ CDS - Observatoire de Strasbourg

⁹ Observatoire de Nice

Corresponding Author: julien.lecubin@imcce.fr

Numerical simulations, data analysis, as well as design study for new instruments and telescopes, often require important computing time. It is not uncommon that the analysis of a single observation requires to run a huge number of times the same simulation code to explore the parameters space. On the other hand, the same reduction pipeline has to be used several times for data reduction of a set of observations. On the other hand, the physics introduced in simulation codes may be limited due to computing time restrictions. By sharing computing facilities thanks to grid technology, we can expect to work faster and to go further in the detail of the physics of simulations

Posters / 102

GEOSCOPE applications on EGEE

Mr. WEISSENBACH, David¹; Dr. CLÉVÉDÉ, Eric¹; Dr. STUTZMANN, Eleonore¹

¹ CNRS

Corresponding Author: weissenb@ccr.jussieu.fr

For now, applications dealing with seismic noise (signal collected between earthquakes, more than 99% of the data) or major earthquakes are taking advantage of the grid facilities:

- average seismic noise level per day over the past years is computed

- polarized noise and source determination also over the past years.

Although much more demanding on CPU time, this was achieved faster than the corresponding raw noise calculation as GEOSCOPE data was already transferred on EGEE.

- source and mechanism determination for large earthquakes. It is still one of ESR's most successful application on EGEE as hundreds hours of computing time are completed during a very short period, delivering the results almost right after the data is available.

Posters / 103

Implementation of geospatial services in gLite: the RISICO case study

Dr. ANGELINI, Valerio ¹; Dr. MAZZETTI, Paolo ¹; Dr. NATIVI, Stefano ¹; Dr. DAL PRA, Stefano ²; Dr. D'ANDREA, Mirko ³

¹ CNR-IMAA

² INFN

³ CIMA Foundation

Corresponding Author: angelini@imaa.cnr.it

RISICO presently runs in gLite accessing input data stored in a SE using various proprietary formats. Our aim is to integrate this application with a framework of standard geospatial web services to gain more flexibility. In this case the data will be stored in standard formats (GRIB) and will be accessed through standard interfaces.

The workflow will be as follows:

- The CP user selects an area in which the model should be run, selects the input data URIs and indicates an appropriate priority for the action;
- A Web Processing Service (WPS) receives the request, evaluates the input size and the priority, and then activates various independent data access services (WCS/WFS) which split the input data;
- When the various inputs have been set up, the WPS spawns and distributes an adequate number of jobs on the grid. These are responsible for the execution of the core algorithm;
- When all the jobs have correctly run, the WPS takes care of merging the results and publishes them to the CP user.

Data Management / 104

The gCube Metadata Framework: integrated environment for managing Metadata Objects and relationships on top of Grid-enabled Storage Systems

SIMI, Manuele ¹; Dr. KAKALETRIS, George ²; Dr. PAGANO, Pasquale ¹

¹ CNR-ISTI

² University of Athens

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

The framework allows to: i) store, update, validate, manipulate, and retrieve metadata through the Metadata Catalog; ii) arbitrarily transform metadata through the Metadata Broker; iii) index metadata through the XML Indexer and discover them through XQuery and XPath expressions; iv) manage annotations through the Annotation Management stack.

The granularity of each operation varies from a single metadata entity, to bulk, passed by-reference, entities that allow managing entire collections. The outputs of the operations can be a static or dynamic, continually updated products of their inputs.

Each component is a well defined Web Service. The framework itself has been factored to support inclusion of new services at any (even run-) time. Moreover, apart the service that manages the upload and the relationships among the metadata items and the objects they describe, the rest of the services can be omitted, if the provided functionalities are not desired.

Astronomy & Astrophysics / 105

LOFAR@EGEE

Dr. BELIKOV, Andrey¹

¹ *Kapteyn Astronomical Institute*

Corresponding Author: a.n.belikov@astro.rug.nl

The data volume produced by LOFAR will reach 4 PB per year. Data processing of Lofar requires an online access to the most of this data volume. We will need to implement a distributed data storage in a multicomponent environment, and the multicomponent environment (from the point of view of hardware, software and grid concepts) will become a key feature of the project. We will need to integrate EGEE storage elements as a part of storage space for LOFAR information system.

From research to production grids: interaction with the Grid'5000 initiative / 107

IV Grid Plugtests: composing dedicated tools to run an application efficiently on Grid'5000

BESSERON, Xavier¹; DANJEAN, Vincent¹; GAUTIER, Thierry¹; GUELTON, Serge¹; HUARD, Guillaume¹; WAGNER, Frédéric¹

¹ *LIG - MOAIS project*

Corresponding Author: xavier.besson@imag.fr

Exploiting efficiently the resources of whole Grid'5000 with the same application requires to solve several issues:

- 1) resources reservation;
- 2) application's processes deployment;
- 3) application's tasks scheduling.

For the IV Grid Plugtests, we used a dedicated tool for each issue to solve.

The N-Queens contest rules imposed ProActive for the resources reservations (issue 1).

Issue 2 was solved using TakTuk which allows to deploy a large set of remote nodes. Deployed nodes take part in the deployment using an adaptive algorithm that makes it very efficient.

For the 3rd issue, we wrote our application with Athapascan API whose model is based on the concepts of tasks and shared data. The application is described as a data-flow graph using the Shared and Fork keywords. This high level abstraction of hardware gives us an efficient execution with the Kaapi runtime engine using a work-stealing scheduling algorithm to balance the workload between all the distributed processes.

Posters / 108

Decentralized access to medical images in Research and Enterprise PACS Applications

Mr. KULHANEK, Tomas¹; Dr. SAREK, Milan¹

¹ *CESNET*

Corresponding Author: tomaton@centrum.cz

A significant activity in the classic distributed solution with computer servers and long-term storage devices of high capacity is the MeDiMed project (Metropolitan Digital Imaging System in Medicine), where the Masaryk University in Brno cooperates with a range of university and city hospitals. To be interconnected they take advantage of the CESNET2 high speed computer network. The goal of the MeDiMed project is to create shared outsourced medical multimedia data archiving and communication center (Metropolitan PACS). The presented pilot project is a parallel project next to MeDiMed. It provides an interface able to operate in DICOM standard (Digital Imaging and Communications in Medicine). It allows interchange medical images with modalities (e.g. medical devices or software application) and with existing PACS (e.g. with the mentioned MeDiMed).

Posters / 109

Interaction of a 3D finite-difference application for computing synthetic waveforms with the Grid infrastructure

Mr. SKARLATOUDIS, Andreas¹; Mr. KOROSOGLOU, Paschalis²; Mr. KANELLOPOULOS, Christos²; Prof. PAPAACHOS, Constantinos¹

¹ *Geophysical Laboratory, Univ. Thessaloniki*

² *Grid Operations Centre, A.U.Th.*

Corresponding Author: pkoro@physics.auth.gr

The aim of this study is to shed some light into the ground motion properties of Thessaloniki in 3 dimensions. Using a computer code that implements a 3D - 4th order staggered-grid velocity-stress finite-difference (FD) scheme (Moczo et al., 2002) full 3-dimensional synthetics of ground motion have been computed. The studying grid covers an area of 63 km² with a depth of 12 km, which is translated in approximately 47x108 nodes. The execution of the 3D FD code is very demanding in terms of CPU power and computer memory and for the previous grid the memory demands reach the 20 GB and the time of computations is approximately 30 hours in a 4-processor machine.

Data Management / 110

Evaluating meta data access strategies through implementing the GOME test suite

Mr. GEMUEND, Andre¹; Mr. SCHWICHTENBERG, Horst¹

¹ *FhG/SCAI*

Corresponding Author: gemuend@scai.fraunhofer.de

With the realisation of the GOME validation test suite we tested the capabilities of the three underlying data access services as well as their integration into the gLite middleware. For the validation process data from a satellite and data from ground measuring stations have to be assigned by their spatial coordinates. For this we used GIS features available in modern databases when applicable and the services allowed it, e.g. with OGSA-DAI using the PostGIS extension for the PostgreSQL database. We additionally reviewed some of the advanced features of the underlying services.

Grid Access / 111

CRAB, the CMS tool to allow data analysis in a distributed environment

FANZAGO, Federica ¹; CINQUILLI, mattia ²; CODISPOTI, giuseppe ³; CORVO, marco ¹; FANFANI, alessandra ³; FARINA, fabio ⁴; KAVKA, carlos ⁵; LACAPRARA, stefano ⁶; SPIGA, daniele ⁷; VAANDERING, Eric ⁸

¹ *CERN-CNAF*

² *INFN-Perugia*

³ *university bologna*

⁴ *Univesity Milano Bicocca*

⁵ *INFN-TS*

⁶ *INFN-LNL*

⁷ *INFN-PG*

⁸ *FNAL*

Corresponding Author: fanzago@pd.infn.it

The CMS experiment will produce few PBytes of data each year to distribute and store in many computing centres spread in the countries participating to the CMS collaboration and made available for analysis to world-wide distributed physicists. CMS will use a distributed architecture based on Grid infrastructure to analyze data stored at remote sites, to assure data access only to authorized users and to ensure remote resources availability. Data analysis in a distributed environment is a task that assume to know which data are available, where data are stored and how to access them. To simplify analysis job creation and management the CMS collaboration is developing CRAB (CMS Remote Analysis Builder) a tool to allow users with no specific Grid knowledge to be able to run analysis in the distributed environment as data were in their local farm. CRAB is developed as tool standalone and client-server to improve the throughput, the scalability and to automatize most of CRAB functionalities

Interoperability and Resource Utilisation / 112

Towards a WBEM-based Implementation of the OGF GLUE Information Model

Dr. ANDREOZZI, Sergio ¹; Mr. CANAPARO, Marco ¹; Mr. CARPENÈ, Michele ¹

¹ *INFN-CNAF*

Corresponding Author: sergio.andreozzi@cnafe.infn.it

The current gLite middleware relies on the GLUE 1.3 information model and its implementation in LDAP in order to advertise the available resources and their characteristics in the EGEE infrastructure. In the context of the Open Grid Forum, the GLUE Working Group is defining the evolution of this information model to improve the current design and to unify a number of existing approaches in a community standard. The OMII-Europe project is engaged in this activity and is developing a modular framework for managing the information providers based on WBEM (Web-Based Enterprise Management) technologies. A client supporting multiple renderings is being developed in order to be useful for different consumers (e.g. LDAP for BDII, XML for Web Services, SQL for R-GMA).

Life Sciences / 113

Performance Analysis and Optimization of AMGA for the WISDOM environment

Mr. AHN, Sunil¹; Mr. KIM, Namgyu¹; Mr. LEE, Seehoon¹; Dr. HWANG, Soonwook¹; Dr. KOBLITZ, Birger²; Dr. BRETON, Vincent³

¹ KISTI

² CERN

³ CNRS

Corresponding Author: siahn@kisti.re.kr

In the WISDOM environment, thousands of job agents distributed on the Grid may have access to an AMGA server simultaneously (1) to take docking tasks out of the AMGA server to execute on the machine that they are sitting, (2) to get the related ligand and target information, and (3) to store the docking results. The docking tasks take about 10 to 30 minutes to finish depending on the machine that they run and the docking configuration.

We have carried out some performance analysis on the current AMGA implementation. Due to the overhead required to handle GSI/SSL connection on the Grid, it showed about 350% poorer throughput compared with a direct DB access. In the current version of WISDOM, AMGA is used as a placeholder for a task distribution table where docking tasks are stored and maintained. We have found a serious performance degrade due to the overhead caused by the need to lock the whole table to prevent different agents from taking the same task.

Posters / 114

The Dutch Life Science Grid

Dr. BOUWHUIS, Maurice¹; Dr. JANSEN, Machiel¹; Mr. VAN BEEK, Pieter¹

¹ SARA

Corresponding Author: maurice.bouwhuis@sara.nl

Today's Life Scientists need to have advanced High Performance Computing facilities at their disposal. For that reason, and commissioned by the Dutch BioInformatics Centre (NBIC) and the Dutch grid infrastructure project (BiGGrid), SARA places, maintains and supports small but powerful computer clusters at the local sites of academic medical hospitals and universities. These clusters are interconnected by high speed network connections and can be used simultaneously by the use of GLITE Grid middleware.

A number of Use Cases have been formulated and development of a number of biological applications running on this infrastructure is in progress. Among the areas which are involved are metabolomics, proteomics and micro array analysis. The use cases describe several biological pipelines which will be realized by Grid and web services, interconnected by workflow descriptions.

Posters / 115

Evaluation of EGEE-grid security.

Ms. FOUOSSONG, Dorine ¹; Dr. BRETON, VINCENT ¹; Mr. REICHSTADT, PIERRE LOUIS ¹; Mr. LIAUZU, STEPHANE ¹

¹ IN2P3-LPC

Corresponding Author: dorine@clermont.in2p3.fr

Our first goal is an evaluation of the overall security.

Using the grid introduces in the information system of a laboratory or a company new infrastructure and processes that have to be taken into account in the management. This is specially true in the security management, where relevant subsystem should be introduced in the trust chain. The trust chain is the subset of the information system that is "secure" in terms of the security strategy plan of this specific firm. The everyday work of the security manager is to deploy and maintain that trust chain.

A good management strategy in our terms should follow the Deming Wheel: plan, do, check, act.

Our work is mainly concerned by the third point which is "check" as it evaluates the security level of the new trust chain (modified to take in account the EGEE-grid resources).

Thus, this activity should contribute significantly to the security risk management of the EGEE-grid environment.

Fusion / 116

Distributed Task Scheduling for Physics Fusion Applications

Dr. MONTERO, Ruben Santiago ¹; Prof. LLORENTE, Ignacio Martín ²; Mr. HERRERA, Jose ³; Dr. HUEDO, Eduardo ¹

¹ Associate Professor

² Full Professor

³ Teaching Assistant

Corresponding Author: jherrera@fdi.ucm.es

There are two kinds of parallel loop schedulers to distribute program loops among the processors of a parallel architecture: static and dynamic scheduling. In this work, we will focus on dynamic schedulers because they are more suitable for heterogeneous environments such as a Grid. In general, in these algorithms a central node dynamically distributes a fraction of the computational workload (chunk) to the rest of the worker nodes. Depending on how the chunks are calculated, different simple self-scheduling schemes can be devised. An alternative to these schemes is the distributed self-schedulers. In this case, the scheduler takes into account the characteristics of the different components of the system (e.g. the cpu speed or the system load) to optimize the chunk assigned to each node. This work presents a new distributed self-scheduler scheme that takes into account all Grid characteristics: a high degree of heterogeneity, high fault rate, dynamic resource availability, etc.

Demonstrations / 117

Utilization of EGEE Grid as a Computational Platform for Chemistry

Dr. STERZEL, Mariusz¹; Dr. LAGANA, Antonio²; Dr. GERVASI, Osvaldo²; Dr. KMUNICEK, Jan³; Dr. BERGER, Maximilian⁴

¹ ACC CYFRONET AGH

² University of Perugia

³ CESNET

⁴ University of Innsbruck

Corresponding Author: mariusz.sterzel@cern.ch

A huge amount of work has been devoted to satisfy chemical community requirements on the grid. The activity of CompChem, Gaussian and VOCE VOs has been mainly focussed on the grid ports of chemical software packages and on the development of grid tools that simplify job manipulation and workflows, automating complex data management tasks. The ports targeted commercial software packages like Gaussian, Turbomole or Wien2k, in particular as the community is highly accustomed them. The main difficulty concerned licenses for which grid solutions had to be developed. In parallel, grid ports of other packages for ab initio, molecular dynamics and quantum dynamics (including time) has been developed mainly for the GEMS project. The development in job management resulted in the Charon Extension Layer, the latest version of which allows easy job manipulation via a web browser. Workflows are mainly available to the Wien2k and GEMS communities to solve their complex data submission issues.

Workflow and Parallelism / 118

Parallel Execution of Chemical Software on EGEE Grid

Dr. STERZEL, Mariusz¹

¹ CERN

Corresponding Author: mariusz.sterzel@cern.ch

The current state of development of grid middleware allows easy parallel execution in case of software using any of MPI flavour. Unfortunately many chemical packages do not use MPI for parallelization therefore special treatment is needed. Gaussian can be executed in parallel on SMP architecture or via Linda. These require reservation of certain number of processors/cores on a given WN and the equal number of processors/cores on each WN, respectively. The current implementation of EGEE middleware does not offer such functionality therefore the only solution is to enforce required configuration via Maui scheduler. The solution we present does not require Linda for parallel execution. It allows utilization of maximum number processors/cores on a given WN. Taking in to consideration the WNs supporting Gaussian VO parallel execution on maximum 8 processors/cores is possible. The main disadvantage of our solution is necessity of local Maui configuration on each site by an administrator.

Finance & Multimedia / 119

Using grid technologies for the mapping out of taxation policy

Mr. PLOSKAS, Nikolaos¹; Mr. GLAVELIS, Themistoklis¹; Prof. MARGARITIS, Konstantinos¹; Prof. SAMARAS, Nikolaos¹

¹ *University of Macedonia*

Corresponding Author: mai084@uom.gr

The regression that we used to appreciate the tax policy of Greek government is the following:

$$Sgovt = a_0 + a_1 * T + a_2 * TR + a_3 * INT + a_4 * G$$

where a_i , $i=1, \dots, 4$ are the coefficients of the regression and a_0 is the constant term, Sgovt the Government budget deficit/surplus, TR is the Transfer Payments, INT is the Net Interest Payments and G is the Government Purchases. Due to the lack of real elements of many years, the application creates a lot of instances of data. Sample of elements for the past fifteen years were taken from the databases of OASA, Eurostat and National Statistical Service of Greece. The application exports a report that includes all the statistic and econometric results of the model with the most adequate data. Using such a kind of report the government could forecast its budget deficit or surplus setting up various scripts. Obviously, this is only a tool for examining different solutions of the taxation policy and cannot substitute the theoretical approach of the problem.

Interoperability and Resource Utilisation / 120

Partnership for Advanced Computing in Europe (PRACE)

Ms. VÄRTTÖ, Saara¹; Dr. KOSKI, Kimmo¹

¹ *CSC - Finnish IT center for science*

Corresponding Author: saara.vartto@csc.fi

The objectives of PRACE are to:

- Create and implement by 2009/2010, a persistent, sustainable pan-European HPC service with several HPC leadership systems of petaflop/s performance.
- Define and establish a legal and organizational structure involving HPC centers, national funding agencies, and scientific user communities.
- Prepare for the deployment of petaflop/s systems in 2009/2010 under the responsibility of European supercomputing centers having the expertise, competency, and required infrastructure to provide a comprehensive service to industry and academia user groups.
- Collect requirements and demands from the user community about future challenging applications.

The infrastructure will be complemented with network and grid access, and the services required to enable applications. These include development of parallel application software expertise, packages, and data handling. It will use concepts and services from EC-funded projects, such as EGEE, GÉANT2 and DEISA.

Life Sciences / 121

High-throughput GRID application for Life Sciences: The Gene Analogue Finder

DE SARIO, Giulia¹; DONVITO, Giacinto²; TULIPANO, Angelica²; Prof. MAGGI, Giorgio²; Dr. GISEL, Andreas¹

¹ *Istituto di Tecnologie Biomediche, CNR*

² *INFN Bari*

Corresponding Author: andreas.gisel@ba.itb.cnr.it

The algorithm is a very high data and data-access intensive application. The results of the functional analogous search demonstrates that the information contained by the GO is adequate to run such analysis using the gene production description. For example most of the homologous gene products of most of the model organisms were assigned as functional analogues, although the annotations were done independently. This result also proves that the algorithm assigns functional analogues in the right way. More important, the algorithm finds significant functional analogous gene products within the same or within different organisms, also non-model organisms, which have such a low sequence similarity so that with conventional methods those assignments would not have been found.

Functional analogous associations is a very important information for scientists in the laboratory which are able to find new information and hints about the functionality of the gene product they are working on.

Posters / 122

Testing 65536 parallel pseudo-random number streams

Mr. REUILLON, Romain¹; Prof. HILL, David¹

¹ *LIMOS*

Corresponding Author: reuillon@isima.fr

Monte Carlo simulations are typical Grid applications, they are considered as naturally parallel because many replications of the same experiment can be distributed on multiple execution units to reduce the global simulation time. However, one needs to take care of the underlying random number streams and ensure that the generated streams do not show intra or inter-correlations.

TestU01 is a well known stringent sequential statistical tests battery that aims to detect defaults on pseudo-random number sequences. Matsumoto designed a parallel version of a very good and famous pseudo-random generation algorithm called Mersenne Twister. With a parameterization technique, we have generated independent parallel Mersenne Twisters that have to be tested for statistical deficiencies using TestU01. The best generators can then be safely used in parallel for nuclear medicine Monte Carlo simulations.

Computational Chemistry & Material Science / 123

SALUTE: New results for the inhomogeneous case

Dr. KARAIIVANOVA, Aneta¹; Dr. ATANASSOV, Emanouil¹; Dr. GUROV, Todor¹

¹ *IPP-BAS*

Corresponding Author: anet@parallel.bas.bg

SALUTE integrates a set of advanced Monte Carlo and quasi-Monte Carlo algorithms developed by the application team. In our recent work we studied the inhomogeneous case in the presence of electric field. We obtained new results for the distribution density, energy distribution and Wigner function, which give insight into the quantum effects that occur in this case. The understanding of the physics of the inhomogeneous case give more realistic picture of the intra-colisional field effect and are important for improving the simulation process for new semiconductor devices.

Posters / 125

EDGeS: Integrating EGEE with Desktop Grids

Dr. LOVAS, Robert ¹; Prof. KACSUK, Peter ¹; Mr. LODYGENSKY, Oleg ²; Mr. CARDENAS MONTES, Miguel ³; Dr. GILLES, Fedak ⁴; Mr. KISS, Tamas ⁵; Mr. KELLEY, Ian ⁶; Dr. ARAUJO, Filipe ⁷; Mr. EMMEN, Ad ⁸; Dr. VAN LEEWEN, Manon ⁹

¹ *MTA SZTAKI*

² *CNRS LAL IN2P3*

³ *CETA-CIEMAT*

⁴ *INRIA*

⁵ *Univ. of Westminster*

⁶ *Univ. of Cardiff*

⁷ *Univ. of Coimbra*

⁸ *Stichting AlmereGrid*

⁹ *Fundecyt*

Corresponding Author: rlovas@sztaki.hu

For the EGEE users, having the option to enhance jobs with the ability to seamlessly migrate them to a DG environment will open up new possibilities and enable more widespread and frequent use of data challenge problems that require massive resources. Large DGs far outstrip the performance of even the largest supercomputers, and although DGs are certainly not intended, nor suited, for every application, especially those that are tightly coupled and require inter-processor communication, they have proven to be very beneficial for a wide range of applications, including many that are currently being run on expensive supercomputers and clusters. E.g. most parameter sweep applications of EGEE can easily be migrated into the connected DGs. The EGEE-DG Bridge will provide a mechanism to migrate these applications to a Desktop Grid environment when needed, thereby freeing up resources for use by tightly coupled MPI applications that require low latency inter-processor communication.

Earth Science / 126

Long range air pollution transport over Europe, studied on the Grid

Dr. OSTROMSKY, Tzvetan ¹; Prof. DIMOV, Ivan ¹; Dr. ZLATEV, Zahari ²

¹ *IPP - BAS, Sofia, Bulgaria*

² *NERI, Roskilde, Denmark*

Corresponding Author: ceco@parallel.bas.bg

The Danish Eulerian Model (DEM) is a powerful air pollution model, designed to calculate the concentrations of various dangerous species over a large geographical region (e.g. Europe). This is a huge computational task and requires significant resources of storage and CPU time. Parallel computing is essential for the efficient practical use of the model. However, it is not sufficient when a large number of experiments have to be done. A number of experiments for one year period were necessary to carry out in order to detect the set of the most important parameters, which have influence on certain specific pollutant (e.g. the ozone). The computational grid was used for this purpose. Promising results have been obtained within a reasonable time by using EGEE infrastructure. Overview of these results will be presented in this talk.

Posters - Board P35 / 128

GridAE: A Grid-based Framework for Artificial Evolution Applications

Dr. SENER, Cevat¹; Mr. KETENCI, Ahmet¹; Dr. SAHIN, Erol¹; Mr. SIVRI, Umit¹

¹ *Middle East Technical University*

Corresponding Author: sener@ceng.metu.edu.tr

The candidate solutions found in AE are used to generate "offsprings" based on their fitness values; fitter candidates generating more offsprings into the next generation. This makes the entire computational requirements tend to be proportional to that of evaluating a single individual. Hence, the fitness evaluations of candidate solutions need to be spread over a large number of processors to make the whole process viable.

The development of GridAE (supported under the SEE-GRID2 project) aims to create a Grid-based framework for AE applications by porting the idea and experience of our earlier study, Parallelized Evolution System (developed as a part of our Swarm-bots project to be run on clusters), onto the Grid.

This framework should create a transparent interface for the AE user (similar to BEAGLE or GALib), which would manage the execution of the evolution on the Grid, to be achieved both as a command line interface and a GUI through a portlet on the TR-Grid P-GRADE portal.

Life Sciences / 129

Grid Solving a Bioinformatics Challenge: a First Step to Anchoring the Nucleosome

Dr. BLANCHET, Christophe¹; Mr. MICHON, Alexis¹; Dr. ZAKRZEWSKA, Krystyna¹; Dr. LAVERY, Richard¹

¹ *CNRS IBCP*

Corresponding Author: christophe.blanchet@ibcp.fr

The nucleosome involves a complex of eight proteins (histones) binding to 147 base-pairs of DNA. Simulating a nucleosome core bound to a single DNA sequence would require treatment of roughly 250,000 atoms and many months of computer time. To understand selective binding we need to compare many potential binding sequences and hence perform many such simulations. Given that any of the four nucleic acid bases can occupy each position within the bound DNA, there are roughly 10^{86} potential sequences to test. We have been able to reduce this task by dividing the DNA into overlapping fragments containing four nucleotide pairs ($4^4=256$ sequences for each pair). By minimizing each sequence in turn for each fragment, and then moving one step along the nucleosome-bound DNA, we can reconstruct the binding energies of all possible sequences with approximately 36,000 optimizations using the JUMNA program (developed in our team). The whole task would take roughly four years on a single processor.

Astronomy & Astrophysics / 130

High Performance Computing on the GRID infrastructure of COMETA

Dr. SACCO, Germano¹; COMETA, team²; Dr. ORLANDO, Salvatore³; Prof. PERES, Giovanni⁴; Prof. REALE, Fabio⁴; Dr. BOCCHINO, Fabrizio³

¹ *COMETA consortium, INAF - Osservatorio Astronomico di Palermo*

² *COMETA consortium*

³ *INAF - Osservatorio Astronomico di Palermo*

⁴ *Dip.S.F.A. - University of Palermo*

Corresponding Author: orlando@astropa.inaf.it

FLASH is a parallel MHD code based on Message-Passing Interface (MPI) library and designed to be executed on HPC systems. The simulations performed required a substantial amount of distributed computational resources made available through the GRID infrastructure of COMETA.

Posters / 131

Configuring and enabling Condor for LHC Computing Grid

DAS, Santanu

Corresponding Author: santanu.das@cern.ch

Condor is a specialized workload management system for compute-intensive jobs, which can effectively manage a variety of clusters of dedicated compute nodes. Today, there are grid schedulers, resource managers, and workload management systems available that can provide the functionality of the traditional batch queuing system e.g. Torque/PBS or provide the ability to harness cycles from idle desktop workstations. Condor addresses both of these areas by providing a single tool. In Grid-style computing environment, Condor's "flocking" technology allows multiple Condor compute installations to work together and opens a wide range of possible options for resource sharing. Although Condor, as a batch system, is officially supported by gLite/EGEE, various part of the middleware still limited to the PBS/Torque in terms of transparent integrity. We have extended the support to allow middleware to work seamlessly with Condor and enable interaction with University Compute Clusters.

Posters / 132

Batch Service Management Tool within g-EclipseDr. GJERMUNDROD, Harald¹; Prof. DIKAIKAKOS, Marios¹¹ *University of Cyprus***Corresponding Author:** harald@cs.ucy.ac.cy

The Batch Service editor within the g-Eclipse framework allows the administrator of a Grid site to manage her site(s). The editor presents the administrator with a color-coded representation of the current state of the computing element, queues, and worker nodes as well as their properties. For large sites, the administrator can zoom in/out to view all the elements of the site. Using this editor the administrator can choose from the context menu of the specific item(s) to start/stop/enable/disable/delete queues, enable/disable worker nodes, hold/release/move/delete batch jobs, and initiate wizards to create new queues. If multiple items are selected, then only the actions that can be applied to all the items are available.

In addition to the editor there is also a Batch Job view which presents a table of the batch jobs that are currently present on the selected element (worker node, queue) in the editor. The columns in the table present the properties associated with the batch jobs.

Workflow and Parallelism / 133

MPI Support on the GridMr. DICHEV, Kiril¹; Mr. KELLER, Rainer¹¹ *High Performance Computing Center Stuttgart***Corresponding Author:** dichev@hlrs.de

MPI-Start was developed for the Interactive European Grid project in order to improve MPI support for its infrastructure. MPI-Start supports different MPI implementations (currently Open MPI, MPICH, MPICH2, LAM-MPI). Also, it offers support to different batch systems (currently PBS, SGE, LSF). In addition, support for MPI tools like Marmot is already integrated into MPI-Start.

PACX-MPI supports any implementation of the MPI 1.2 standard and delivers the support for seamlessly running one large MPI application on heterogeneous clusters or supercomputers.

Marmot can be useful for different MPI correctness checks at runtime like using correct data types, deadlocks etc.

Life Sciences / 134

Genome Wide Association Studies of human complex diseases with EGEE

Mr. MUNTEANU, Alexandru Ionut¹; Mr. CAMBIEN, François¹; Mr. TRÉGOUËT, David¹

¹ INSERM, UMR S 525, Faculté de Médecine Pitié-Salpêtrière, Paris, France

Corresponding Author: munteanu@chups.jussieu.fr

As part of the research conducted at the INSERM U525 laboratory, 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. Until now this kind of analysis was restricted to single genes and a few polymorphisms (<25). The recent availability of DNA chips allowing to genotype hundreds of thousands of polymorphisms across the genome implies a change in scale in the necessary computations. For whole genome haplotype analysis we decided to use the EGEE grid.

Grid Access / 135

XtreemOS: A Grid Operating System Providing Native Virtual Organization Support

Dr. MORIN, Christine¹; Dr. JÉGOU, Yvon¹; Mr. SANCHEZ, Oscar David¹

¹ INRIA

Corresponding Author: oscar.sanchez@irisa.fr

While much has been done to build Grid middleware on top of existing operating systems, little has been done to extend the underlying operating systems for enabling and facilitating Grid computing, for example by embedding important functionalities directly into the operating system. XtreemOS project aims at investigating and proposing new services that should be added to current operating systems to build a Grid infrastructure in a simple way.

This approach can be seen to have some advantages over conventional Grid middleware toolkits, which may have different programming interfaces and lack of a unifying model. A common interface can be provided to simplify the task of the application developer on the Grid by making the Grid support native to the operating system, and also by removing layers of abstraction, leading to higher dependability of services.

Finance & Multimedia / 136

GridVideo: a grid-based multimedia application

Mr. IELLAMO, Giuseppe¹; Mr. MINUTOLI, Giuseppe¹; Dr. BRUNEO, Dario¹; Prof. PULIAFITO, Antonio¹

¹ Università di Messina

Corresponding Author: giellamo@unime.it

The application is divided into two different activities:

- the Multimedia Upload activity during which service providers make multimedia objects available to their customers by uploading them to the Grid Storage Elements.
- the Multimedia Streaming activity where the media are requested by end-users through a GUI. Upon these requests the media chunks are recovered, tailored and finally streamed towards user device. This activity calls for stringent time requirements between different jobs.

Using the Grid allows for both seamless data dissemination during the Upload activity(through the use of Storage Elements and file catalogs) and performance scalability during the Streaming activity(by adapting the amount of resource used to the number of users).

Astronomy & Astrophysics / 137

The ZEN Project

Dr. TILQUIN, André ¹

¹ CPPM/IN2P3

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

Evidence for the accelerated expansion of the Universe has been observed in the last decade with the many cosmological observations. The origin of accelerated expansion remains one of the most challenging research activities today. Progress in this field requires both theoretical innovations and many accurate observational probes with controlled systematic error estimates. The difficulty in performing combinations of different observations is to manage in a global analysis a large number of cosmological and astrophysical parameters (14 or more). As correlations are large and cannot be ignored at the percent accuracy level, the statistical method and the construction of an efficient numerical tool represent an important step of the ZEN project. We promote a frequentist approach which is commonly used by the High Energy Physics community and well under control. Our results are in agreement with complementary methods (mainly Bayesian using Monte Carlo Markoff chain).

Monitoring, Accounting & Support / 138

gUSE: grid User Support Environment

Dr. LOVAS, Robert ¹; Prof. KACSUK, Peter ¹; Mr. KAROCZKAI, Krisztian ¹; Mr. SCHNAUTIGEL, Andras ¹; Mr. MARTON, Istvan ¹; Mr. SIPOS, Gergely ¹; Mr. HERMANN, Gabor ¹

¹ MTA SZTAKI

Corresponding Author: rlovas@sztaki.hu

Users of gUSE can be either grid application developers or end-users. Application developers can develop sophisticated workflow applications where workflows can be embedded into each other at any depth. Even recursive workflows allowed. gUSE enables to embed workflows derived from other workflow systems (e.g. Taverna, Triana, Kepler, etc.). gUSE supports the concept of workflow templates, abstract workflows, concrete workflows and workflow instances. All of them can be published in a workflow repository. Members of a developer community can import workflows from the repository and can continue the work on them. End-users can import completed workflow applications from the repository and can execute them based on a simple user interface that hides grid details from them. Grid is exposed only for the application developers. Parametric sweep nodes and normal nodes can be used in a mixed way in the workflows enabling very complex applications to develop in gUSE.

Astronomy & Astrophysics / 139

Experiences from porting the astrophysical simulation “The unified theory of Kuiper-belt and Oort-cloud formation” to EGEE grid

Mr. ASTALOS, Jan¹; Dr. HLUCHY, Ladislav¹; Mr. DOBRUCKY, Miroslav¹

¹ *Institute of Informatics, Slovak Academy of Sciences*

Corresponding Author: astalos.ui@savba.sk

The experiment was performed in the scope of collaboration between Astronomical Institute of Slovak Academy of Sciences, Catania Observatory and Adam Mickiewicz University in Poznan. The simulation was ported to EGEE by Institute of Informatics Slovak Academy of Sciences and it ran in EGEE and TriGrid from February to October 2007.

The simulation consists of a sequence of sub-simulations with many independent tasks within each sub-simulation. The necessary requirement is to finish all the tasks of a given sub-simulation before starting the next sub-simulation.

The main problem when running the large number of jobs in grid was the reliability of grid infrastructure. Job management was rather time consuming due to the time spent on the analysis of the failed jobs and their resubmission. Moreover, the jobs that were waiting at some sites in a queue for a long time were blocking whole simulation.

Interoperability and Resource Utilisation / 140

An Application of ETICS Co-Scheduling Mechanism to Interoperability and Compliance Validation of Grid Services

Mrs. RONCHIERI, Elisabetta¹; Mr. AGUADO SANCHEZ, Carlos²; Mr. DIEZ-ANDINO SANCHO, Guillermo²; Dr. DI MEGLIO, Alberto²; Dr. MARZOLLA, Moreno³

¹ *INFN CNAF*

² *CERN*

³ *INFN*

Corresponding Author: elisabetta.ronchieri@cnafl.infn.it

Interoperability and compliance to standards are important quality attributes of software developed for Grid environments where many different parts of an interconnected system have to interact. Compliance to standard is one of the major factors in making sure that interoperating parts of a distributed system can actually interconnect and exchange information. Taking the case of the Grid environment (Foster and Kesselman, 2003), most of the projects that are developing software have not reached the maturity level of other communities yet and have difficulty to identify and adopt standards. Validating the compliance with standards often requires the design of custom test suites and a constant attention to any proposed change. Interoperability amongst middleware and application software developed in order to be used on Grid and other distributed environments is usually a very complex issue.

Plenary / 141

European Grid Initiative Design Study (EGI_DS)

KRANZLMUELLER, Dieter ¹

¹ GUP, JKU Linz

Corresponding Author: kranzlmueLLer@gup.jku.at

To successfully complete the set objectives of the EGI vision (<http://www.eu-egi.org/vision.pdf>), EGI_DS has committed to coordinate and support the following actions:

- 1) Consolidate the requirements for an EGI organisation from the NGIs and other important stakeholders such as application communities, infrastructure operators, related projects, NRENs.
- 2) Define the functional role of the EGI organisation with respect to the NGIs at the start of this organisation, and plan for the evolution of its functions as it matures.
- 3) Establish a legal, financial and organisational basis (supported by the member states) for the EGI organisation to undertake these functions.
- 4) Establish the EGI organisation and manage the transition from the existing project-based grid support to a sustainable production service.
- 5) Ensure that all stakeholders within the member states, international standards bodies, research grid services in other countries are aware of the EGI and have relationships.

Demonstrations / 142

Bioinformatics portal on Grid: the GPSA - Grid Protein Sequence Analysis - case.

Mr. MICHON, Alexis ¹; Mrs. ELOTO, Christelle ¹; Dr. COMBET, Christophe ¹; Dr. BLANCHET, Christophe ¹

¹ CNRS IBCP

Corresponding Author: christophe.blanchet@ibcp.fr

Bioinformatics analysis of data produced by high-throughput biology, for instance genome projects, is one of the major challenges for the coming years. Two of the requirements for this analysis are access to up-to-date databanks (of sequences, patterns, 3D structures, etc.) and access to relevant algorithms (for sequence similarity, multiple alignment, pattern scanning, etc.). Since 1998, we have been developing the Web server NPS@ (Network Protein Sequence Analysis), that provides the biologist with many of the most common resources for protein sequence analysis, integrated into a common workflow. We have adapted this portal to the EGEE Grid. The bioinformatics grid portal GPS@ ("Grid Protein Sequence Analysis") simplifies and automates the EGEE grid job submission and data management mechanisms using XML descriptions of available bioinformatics resources: algorithms and databanks.

Demonstrations / 145

EnginFrame Genius Grid Portal and VOMS Proxy creation.

Dr. VENUTI, Nicola ¹; Mr. FALZONE, Alberto ¹; Dr. ARDIZZONE, Valeria ²; Prof. ROBERTO, Barbera ³; Dr. SCARDACI, Diego ²

¹ *NICE srl*

² *INFN Catania*

³ *INFN Catania AND University of Catania*

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

With Enginframe user can interact with files on the UI, can submit and monitor jobs to Grid and manage data and job output inside the VO belongs to. The Web portal eliminates any problems and needs about particular Operating System running on the client, the user can interact with the Grid from everywhere and with everything:

a Java compliant web browser is only required. The purpose of VOMS ProxyInit service is to create the user's grid proxy file, with or without VOMS Extensions, using its personal certificate and private key from his usb pen drive attached to the local workstation, and finally, by means of an encrypted and mutually authenticated connection, transfer just the created proxy file to the server where is installed Enginframe with GENIUS on top. Globus Security Infrastructure provides communication integrity between elements of a computational Grid and support for single sign-on for users and it was implemented in the VOMS ProxyInit service to client-server communication

Workflow and Parallelism / 147

Non-parametric parallel GRID Harris-affine detector

BELLAVIA, fabio ¹; VALENTI, cesare ¹; TEGOLO, domenico ¹

¹ *Univ. Palermo*

Corresponding Author: fbellavia@math.unipa.it

Because most-wanted of image analysis applications as three-dimensional reconstruction, mosaicing, object recognition and classification can rely on feature detection methodology as a primary stage, it can be used to satisfy many requests of these items and more in general in the field of computer vision. Feature descriptors can be applied to identify similar regions on different images; it is clear that some characteristics have to be owned of good descriptors.

One of the most appropriate question in match methodology regards which detector has to be used to characterize the region of interest. They can be classified as global or local, some of them are characterized by global information and other are modelled on local values. In this contribution a detector, based on local information, will be used. It rejects the same model

of the parametric Harris Affine detector with the peculiarity to have no parameters, and to be focussed on the each sections of the image.

Posters / 148

Enhancement and Breast Hierarchical Segmentation on Digital Mammograms

Mr. ALIU, Azir¹; Prof. KON-POPOVSKA, Margita²

¹ *SEEU*

² *professor*

Corresponding Author: azir.aliu@seeu.edu.mk

Our algorithm enhances the results of digital mammogram processing. For image enhancements and appearance improvement, noise or error elimination, or to highlight certain image features, the algorithm uses density measures based on a normalized breast representation, method of image equalization and Kuvahara filter. This first phase is designed to have a very high sensitivity; the large number of false positives is acceptable, since they will be removed in the second phase. In the second phase moments are used for image description and as its intensity distribution shape indication. This phase automatically generates the boundary values and segments the mammograms hierarchically. Using the grid improved both the image processing and mammogram segmentation. We hope that grid infrastructure can be used clinically for early detection of subtle signs of breast cancer, including calcifications and speculated masses.

Workflow and Parallelism / 153

The gLite Workload Management System

GIACOMINI, Francesco¹; CECCHI, Marco¹; GHISELLI, Antonia¹; GUARISE, Andrea¹; MARASCHINI, Alessandro²; MONFORTE, Salvatore¹; PERINI, Laura¹; SGARAVATTO, Massimo¹; GIANELLE, Alessio¹

¹ *Istituto Nazionale di Fisica Nucleare (INFN)*

² *Elsag Datamat*

Corresponding Author: francesco.giacomini@cern.ch

The WMS accepts requests concerning the execution of a computation, whose description is expressed in a flexible language, based on Condor ClassAds, as a set of key - value pairs. The WMS is then responsible to translate the logical description to concrete operations, in order to bring the execution of a job to a successful end. Several types of jobs are supported: simple, intra-cluster MPI, interactive, collections, parametrics, simple workflows in the form of DAGs, with on-going development of a generic workflow engine. Additional benefits concern sandbox management, with support for multiple transfer protocols, data-driven match-making, the availability of multiple mechanisms of error prevention and management, isolation from the heterogeneity of the infrastructure, the capability to implement optimizations based on non-local information, such as bulk submission and match-making.

The progress of the job is safely recorded into the complementary Logging and Bookkeeping Service.

Life Sciences / 154

GRID distribution supporting chaotic map clustering on large mixed microarray data sets

TULIPANO, Angelica ¹; Prof. ANGELINI, Leonardo ¹; DE SARIO, Giulia ²; DONVITO, Giacinto ¹; Prof. MAGGI, Giorgio ¹; Dr. MARANGI, Carmela ³; Dr. GISEL, Andreas ²

¹ INFN Bari

² Istituto di Tecnologie Biomediche, CNR

³ Istituto per le Applicazioni del Calcolo, CNR

Corresponding Author: andreas.gisel@ba.itb.cnr.it

To find correlation between genes within different experiments, clustering is a good and challenging analysis method for data sets of such size and complexity. We have chosen an unsupervised hierarchical clustering algorithm based on the cooperative behaviour of an inhomogeneous lattice of coupled chaotic maps, the Chaotic Map Clustering.

Analyzing data sets of 587 samples we were able to retrieve stable groups of genes. Using the biological knowledge of the gene ontology, we could show, applying a Fisher exact test, that each of the clusters have a set of over-represented functionalities and in most of the cases also clearly different functionalities from cluster to cluster.

In order to evaluate the vast number of clusters found by this process we use a cluster validation method based on resampling subsets of the data under investigation are constructed randomly, and the cluster algorithm is applied to each subset.

Measures of sensitivity and of positive predictive value are pro

Demonstrations / 155

Solving Data Transfer Level Grid Interoperability among SRM, SRB and OGSA-DAI data resources

Dr. LOVAS, Robert ¹; Mr. SIPOS, Gergely ¹; Mr. KISS, Tamas ²; Dr. CEVAT, Sener ³

¹ MTA SZTAKI

² Univ. of Westminster

³ Middle East Technical University

Corresponding Author: rlovas@sztaki.hu

The input files of a node (job or service call) of a P-GRADE workflow can come from file systems, like SRM or SRB, or from database management systems via OGSA-DAI and the results can also be fed into any of these solutions. Both the file systems and the databases can be located in different production grids, and the jobs of the workflow can also be mapped to different grids. These grids can be based on different grid middleware and may require different user certificates for authentication.

The workflow level data resource integration allows the seamless interoperation of SRB catalogues, GridFTP file systems and EGEE storage elements (based on SRM) at the level of P-GRADE workflows. We will demonstrate that jobs of an urban traffic simulation workflow are running in different grids (US OSG, UK NGS, EGEE) and utilise data resources based on different technologies (SRB, SRM, GridFTP, local) from these different grids.

From research to production grids: interaction with the Grid'5000 initiative / 157

Expo : an experiment framework for dedicated platforms

Mr. VIDEAU, Brice¹; Dr. RICHARD, Olivier¹

¹ INRIA / LIG

Corresponding Author: brice.videau@imag.fr

Expo is used to analyze the performances of the file broadcasting tool Kastafior. Kastafior broadcasts a single file onto a given set of nodes. The aim of this experiment is to study Kastafior's performances across Grid'5000 when the file size and the number of nodes vary.

The script used to conduct the experiment is only 15 lines long. Expo interprets this script and issues reservation commands. When the described resources are obtained they are checked to verify that they suit the experiment.

Once used resources are determined, each measurement is launched by a command module. This module archives every outputs produced for future analysis. It also records the status of commands and thus monitors the proper unrolling of the experiment.

When all measurements are done, resources are freed and a complete report on the experiment can be stored on disk.

In the end the experimenter just has to analyze the results.

Computational Chemistry & Material Science / 158

The study of Cytochrome C oxidase on EGEE Grid

Dr. GERVASI, Osvaldo¹; Prof. FARANTOS, Stavros C.²; Dr. DASKALAKIS, Vangelis³; Dr. GIATROMANOLAKIS, Manos³

¹ *Dept. of Mathematics and Computer Science, University of Perugia*

² *Department of Chemistry, University of Crete, Iraklion 711 10 and Institute of Electronic Structure and Lasers, Foundation for Research and Technology-Hellas*

³ *Institute of Structure and Lasers, Foundation for Research and Technology-Hellas*

Corresponding Author: osvaldo@unipg.it

The current project involves Molecular Dynamics calculations on cytochrome c oxidase. CcO is the terminal enzyme of respiratory chains found in the inner mitochondrial membranes or in many bacteria and the last acceptor of electrons from oxidizing processes involving nutrient molecules.

The biophysical interest of this project stems on long standing problems which concern the assignment of difference spectra of isotopically substituted ferryl oxygen.

Data Management / 159

gLibrary/DRI: A grid-based platform to host multiple repositories for digital content

Dr. CALANDUCCI, Antonio ¹; Dr. RAMOS POLLÁN, Raúl ²; Dr. RUBIO DEL SOLAR, Manuel ²; Dr. GONZÁLEZ MARTÍN, Juan Manuel ³; Dr. TCACI, Dorin ³

¹ *INFN Catania*

² *CETA-CIEMAT*

³ *MAAT-G Knowledge*

Corresponding Author: tony.calanducci@ct.infn.it

A gLibrary/DRI repository is made of large digital content (as image files, video, etc) and metadata associated with it (annotations, descriptions, etc). In a typical scenario, new repository providers could use the built in mechanisms to store repository items (e.g. studies made of textual data and multiple medical images) in a combined GRID and federated RDBMS by simply describing the structure of their data in a set of XML files following the gLibrary/DRI specification. In a more elaborated scenario, repository providers can implement specific data management policies and use custom viewers for their specific data structures, still relying on the platform for navigation and management of their repository.

As example, we present a repository based on mammograms, composed of both a repository and a viewer application, to manage patient's mammograms and diagnostics. This includes both the patient's data (stored as metadata) and the mammography digital content (large images stored in SE)

Posters / 160

European Space Agency astronomy projects using Grid technology : XMM-Newton and Integral

Mr. ALVAREZ, Ruben ¹; Mr. ARVISET, Christophe ¹; Mr. GABRIEL, Carlos ¹; Mr. IBARRA, Aitor ¹; Mr. BELANGER, Guillaume ¹; Mr. TAPIADOR, Daniel ¹

¹ *European Space Agency*

Corresponding Author: vuerli@oats.inaf.it

With the introduction of Grid technologies in the European Space Astronomy Center (ESAC) and once the astronomers are getting used to them, new possibilities both for result enhancement and for collaboration with other Astronomy institutes have started to be explored. Here we present some examples of such usage, showing the current status and also the immediate future development: a) The Remote Interface for Science Analysis (RISA), which makes it possible to run SAS, the XMM-Newton data reduction software, through fully configurable web service workflows, enabling observers to access and analyse data making use of all of the existing SAS functionalities. The workflows run primarily but not exclusively on the ESAC Grid, directly connected to the XMM-Newton Science Archive.

b) The production of mosaics on the Grid from XMM-Newton data. c) The gain in processing time when processing Integral data on the Grid are the three examples presented of current usage of Grid at ESA.

Posters / 161

The Grid Application Platform: Development and Applications

Mr. LEE, Hung-Chun ¹; Mr. WEI-LONG, Ueng ¹; Mr. SHI-CHUNG, Chiu ¹; Mr. ERIC, Yen ¹; Dr. SIMON, Lin ¹

¹ ASGC

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

GAP is developed with the following aspects. It is easy to use for not only the end-users but also the grid application developers. GAP provides higher level of Java API which maps the problem domain model to programming domain model. GAP is easy to evolve for adapting new IT technologies as well, and the accommodation is transparent to both developers and users. The GAP abstracts the difference of grid middleware with an unified interface and could be extended to adapt new middleware. The GAP is light-weight in terms of the deployment effort and the system overhead. Its goal is to provide problem domain models for grid application and prevent developers from reinventing the wheels.

Demonstrations / 163

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

MANSET, David ¹; Dr. HUBER, Martin ²

¹ MAAT GKnowledge

² SIEMENS AG

Corresponding Author: david.manset@cern.ch

The HeC prototype is the result of 2 years of active R'nD, which has matured inside a private grid infrastructure. Amongst the 1st contributions, a security prototype was delivered as well as innovative domain specific client applications. Through a user-friendly single sign-on, clinicians access resources independently of their geographical location and connectivity. It allows them to enter, from within their hospital, the large grid spread over Europe to store anonymously patient records and further manipulate these. Medical images are processed, stored in the grid and referenced within the integrated case database. The system enables clinicians to look for similar patients and further process corresponding images, e.g. to extract 3D and 4D models of the heart, useful for better making decisions over particular cases. In cardiology, a 4D-mesh representing the right ventricle is computed and stored in the grid, from which various clinically relevant parameters can be further derived.

Data Management / 164

New results on a comparative evaluation of software providing access to different relational databases interfaced to the grid

Dr. DONVITO, Giacinto¹; Prof. MAGGI, Giorgio²; ATUL, Jain³; BARBERA, Roberto⁴; CAROTA, Luciana⁵; FIORE, Sandro⁶; ALOISIO, Giovanni⁶; CAFARO, Massimo⁶; VADACCA, Salvatore⁶; NEGRO, Alessandro⁶; TAFFONI, Giuliano⁷; VUERLI, Claudio⁷; BARISANI, Andrea⁷; PASIAN, Fabio⁷; MANNA, F.⁷; GISEL, Andreas⁸; CALANDRUCCHI, Tony⁵; AIFTIMIEI, Cristina⁵

¹ INFN-Bari

² INFN-Politecnico

³ INFN+Politecnico Bari

⁴ INFN+Università Catania

⁵ INFN

⁶ Università Lecce

⁷ INAF

⁸ CNR-ITB

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

A way to access widespread databases within a computational grid environment, through a set of secure, interoperable and efficient data grid services is very common in eScience projects.

Use cases in the bioinformatics and astrophysical communities span from very simple queries up to really stressing ones.

A stress tests has been set up exploiting the EGEE grid infrastructure by submitting jobs and monitoring them by means of the Job Submission Tool(webcms.ba.infn.it/cmssoftware/index.html/index.php/Main/JobSubmissionTool). In this configuration it is possible to easily reach the server software and hardware limits and to test the software in a real environment, taking into account the different latency between the server and all the clients, to put in evidence the efficiency of each software tool in delivering the output.

Life Sciences / 165

Molecular dynamics refinement and rescoring in WISDOM virtual screenings

RASTELLI, Giulio¹; Dr. DEGLIESPOSTI, Gianluca¹

¹ Università di Modena e Reggio Emilia

Corresponding Author: giulio.rastelli@unimore.it

After the docking screening of compounds contained in the ZINC database into the crystal structure, the docking results have been refined using molecular dynamics (MD) in order to validate and optimize the ligand orientation into the binding site of the target. Subsequently, the candidates have been rescored using more accurate scoring functions based on molecular mechanics Poisson Boltzman Surface Analysis (MM-PBSA) and molecular mechanics Generalized Born Surface Analysis (MM-GBSA) approaches. Such procedure was designed and validated on aldose reductase [1] and it is fully automated and able to prepare input files, efficiently refine the structures with MD, and rescore the compounds before the final selection of the best hits.

[1] Ferrari A. Degliesposti G. Sgobba M. Rastelli G. *Bioorganic & Medicinal Chemistry* 15 (2007) 7865-7877

Life Sciences / 166

BioinfoGRID: Bioinformatics Grid applications for life science

Dr. MILANESI, Luciano ¹

¹ *National Research Council - Institute of Biomedical Technologies*

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

The BioinfoGRID adopt high-level user interfaces, common to all the different BioinfoGRID applications, in order to exploit the Grid services provided within European Grid Infrastructures using a more user-friendly approach.

One of the activities within the project was to develop a Bioinformatics Portal, to simplify the services request and the jobs submission to the Grid, including the automation of Workflows in order to dynamically establish complex genetics analysis.

The project supports studies on applications for distributed systems for Microarray technology, for Gene expression studies, for Gene Data Mining, analysis of cDNA data, Phylogenetic analysis, Protein functional analysis and system biology simulations in GRID.

Grid Access / 167

Deploying e-Infrastructures for new User Communities in DORII project

Dr. NORBERT, Meyer ¹

¹ *PSNC*

Corresponding Author: meyer@man.poznan.pl

The DORII project aims to deploy e-Infrastructure for new scientific communities, where the ICT technology is still not present at the appropriate level. The DORII is focusing on the following selected scientific areas: earthquake community, with various sensor networks, environmental science community, experimental science community, with synchrotron and free electron lasers. Working closely with end-users, DORII will build solution upon the success of past and ongoing projects in such areas as remote instrumentation (GRIDCC, RINGrid), interactivity (int.eu.grid), software frameworks for application developers (g-Eclipse) and advanced networking technologies (GN2) with EGEE based middleware.

From research to production grids: interaction with the Grid'5000 initiative / 168

Panel Discussion

From research to production grids: interaction with the Grid'5000 initiative / 169

All-in-one graphical tool for grid middleware management

Dr. AMAR, Abdelkader ¹; Dr. CARON, Eddy ¹; Mr. LOUREIRO, David ¹

¹ *ENS-Lyon / INRIA / CNRS / UCBL*

From research to production grids: interaction with the Grid'5000 initiative / 170

Introduction

Plenary / 171

Gold Sponsor: SHARE

SOLOMONIDES, Tony ¹

¹ *University of West England*

Plenary / 172

Welcome from EGEE Project Director

Dr. JONES, Bob ¹

¹ *CERN*

Plenary / 173

Welcome from Local Organisation

Technical Plenary / 174

Computer Science Grids and research in Grid Computing

Dr. CAPPELLO, Franck ¹

¹ *INRIA*

Technical Plenary / 175

Application Porting in EGEE

SIPOS, Gergely ¹

¹ *Mr.*

Technical Plenary / 176

Challenges and success of the HEP GRID

MALEK, Fairouz ¹

¹ *LPSC Laboratoire de Physique Subatomique et de Cosmologie (LPSC)*

Posters / 177

ArchaeoGRID, a Laboratory for the Past on e-Infrastructures

PELPER, G ¹; PELPER, Pier Giovanni ²; CECCHINI, R ³; POLITI, A ⁴

¹ CSDC, Florence, Italy

² Dept. of Physics, University of Florence / INFN

³ INFN

⁴ CNR

Corresponding Author: pelfer@fi.infn.it

A primary goal of ArchaeoGRID as simulation engine is the development of simulation of integrated human-in-natural-systems models, which are treated as complex hypotheses, tested against the archaeological record and used for reconstructing the ancient societies history integrated with the Earth history.

In our study cases archaeological and non archaeological data are multivariate geospatial and temporal data. Grid technology has been developed for general sharing of computational resources but has not been designed for the specialty of geospatial data. In order to make Grid technology applicable to geospatial data, it needs integrate the technologies for the geospatial data with the Grid technology. Grid service-oriented geospatial standards, compliant to Grid framework, are developed for giving to the researchers the possibility to build up their models, to execute them and to have back the desired geospatial products.

Posters / 178

Be Elegant on the Grid

PUGLIESE, Roberto ¹; CURRRI ¹; DEL CANO ¹; DEL LINZ ¹; OLAJUYIGBE, E.; SAMBRI, C.; TAFFONI, G.

¹ ELETTRA

Corresponding Author: pugliese@elettra.trieste.it

ELETTRA is now building a new light source FERMI@Elettra which is a single-pass FEL user-facility covering the wavelength range from 100 nm (12 eV) to 10 nm (124 eV). The advent of femtosecond lasers has revolutionized many areas of science from solid state physics to biology. This new research frontier of ultra-fast VUV and X-ray science drives the development of a novel source for the generation of femtosecond pulses.

ELETTRA is a large data producer. As a partner of EGEE ELETTRA is representing the new community of light sources. In this work we describe the use case of initiating this new community to the eInfrastructure and in particular the case of Elegant, an application program used to design the new light source FERMI@Elettra.

Technical Plenary / 179

VO-level Application Support in EGEE

Dr. SCIABA, Andrea ¹

¹ CERN

Corresponding Author: andrea.sciaba@cern.ch

Plenary - Summary Presentations / 180

Computational Chemistry & Material Science Summary

Plenary - Summary Presentations / 181

Earth Science Summary

Plenary - Summary Presentations / 182

Finance & Multimedia Summary

Corresponding Author: s.holsinger@metaware.it

Plenary - Summary Presentations / 183

Fusion Summary

Plenary - Summary Presentations / 184

Astronomy & Astrophysics Summary

Plenary - Summary Presentations / 185

Interaction with the Grid'5000 initiative Summary

Plenary - Summary Presentations / 186

Life Sciences Summary

Plenary - Summary Presentations / 187

Monitoring, Accounting & Support Summary

Plenary - Summary Presentations and Closing talk / 188

Data Management Summary

Dr. MENDEZ LORENZO, Patricia ¹; Dr. PAGANO, Pasquale ²

¹ CERN IT/GD

² CNR-ISTI

Corresponding Author: patricia.mendez@cern.ch

Plenary - Summary Presentations and Closing talk / 189

Grid Access Summary

Plenary - Summary Presentations and Closing talk / 190

Interoperability and Resource Utilisation Summary

Plenary - Summary Presentations and Closing talk / 191

Workflow and Parallelism Summary

Plenary - Summary Presentations and Closing talk / 192

Demo/Poster Session Summary and Awards

Plenary - Summary Presentations and Closing talk / 193

User Forum Closing Talk

Demonstrations / 194

Fusion Demonstration: Ion Kinetic Transport and Stellarator Optimization

Dr. CASTEJON, Francisco¹; Mr. VOZNESENSKIY, Vladimir²

¹ CIEMAT

² Inst. of Information Systems, RRC "Kurchatov Inst."

Corresponding Author: francisco.castejon@ciemat.es

Two representative fusion applications have been chosen to show the power of grid computing for this community. The first is the grid-based stellarator optimization, run in the RDIG VO; the second is Ion Kinetic Transport, which is running in the fusion VO.

Both applications have been successfully ported from a supercomputer to the EGEE infrastructure. The porting of the first implied the use of mathematical methods that allow asynchronous completion of parallel jobs without loss of efficiency, implying a shift from iterative (gradient-based) methods to results-based, spool-driven (genetic, stochastic interpolation) ones, together with tools for application-specific job and data management.

The second application was tailored to run in this kind of architecture. In its first version, all the jobs are serial and the huge amount of information needed for every job is sent to the catalogue. In its second phase, an iterative method is developed.

Workflow and Parallelism / 195

Workflow and Parallelism Overview

An overview of the of the workflow and parallelism session with information about how these themes fit into the User Forum programme and broad goals of the session.