EGI CF14 Helsinki - Posters & Demos

Report of Contributions

Contribution ID: **2** Type: **Demonstration**

Data Accessibility Reproducibility and Trustworthiness

Description of content and intended audience</br>- the outcome you expect to achieve.

The present demo aims to demonstrate the achievability of the following vision:

- A scientist can search on a wide plethora of resources for a specific term, subject, author or publisher (those already integrated in the CHAIN-REDS Knowledge Base by using the project Semantic Search Engine).
- He/she can discover new knowledge by linking the retrieved results via a semantic enrichment.
- He/she can retrieve the article and associated raw data of interest and either replicate the previous experiment or perform a new study with those data.
- He/she can seamlessly run applications on HPC machines, Grids and Clouds compatible with those retrieved data;
- The cloud-tenant of a real or virtual organisation can seamlessly and easily manage Cloud resources pledged by providers owning/operating infrastructures based on different middleware stacks.
- The new produced data (and publication) can be stored again using the same standards and being assigned to a specific PID so the cycle can be initiated again.

The previous items above will be addressed by using the CHAIN-REDS tools already integrated in the project website and by accessing current repositories already hosted by initiatives such as Zenodo or DataCite..

The standards on which this demo is based on are OAI-PMH, Dublin Core, SPARQL and on the use of PID.

Relevant URL (if any)

http://science-gateway.chain-project.eu
http://www.chain-project.eu/knowledge-base
http://www.chain-project.eu/linked-data
http://www.chain-project.eu

Printable summary: this is the only </br> section of the abstract that will
be published in the Book of Abstracts.

This demonstration is presented on behalf of the CHAIN-REDS project (www.chain-project.eu) and aims at demonstrating Data Accessibility Reproducibility and Trustworthiness. By using metadata standards already implemented in the project Knowledge Base (http://www.chain-project.eu/knowledge-base) and the Semantic Search Engine (http://www.chain-project.eu/linked-data), any user can

find repositories including a term and lately retrieve the raw data for performing a new calculation. The latter, is executed on distributed computing infrastructures, including Grid, local clusters and Clouds, using OCCI and SAGA as standard interfaces and the CHAIN-REDS Science Gateway (http://science-gateway.chain-project.eu) as virtual research environment. Last, the obtained results (raw data and publication) are able to be stored again in a way that they are searchable again.

In this sense, this demo is a step forward in the previous interoperability demo already performed by CHAIN-REDS at the EGI Technical Forum in Madrid during September 2013. The integration of data capabilities into the current CHAIN-REDS tools is a key goal for the project.

Primary authors: RODRIGUEZ-PASCUAL, Manuel (CIEMAT); MAYO-GARCIA, Rafael (CIEMAT); BARBERA, Roberto (University of Catania and INFN)

Co-authors: Mr RUBIO-MONTERO, Antonio J (CIEMAT); Ms CARRUBA, Carla (INFN); KANELLOPOULOS, Christos (GRNET); SCARDACI, Diego (INFN); RUGGIERI, Federico (INFN - Roma Tre); Dr LA ROCCA, Giuseppe (INFN); Ms INSERRA, Giuseppina (INFN); PRNJAT, Ognjen; Mrs RICCERI, Rita (INFN)

Presenters: RODRIGUEZ-PASCUAL, Manuel (CIEMAT); MAYO-GARCIA, Rafael (CIEMAT); BARBERA, Roberto (University of Catania and INFN)

Contribution ID: **3** Type: **Poster**

Non commutative simulations on GRID

Description of content and intended audience</br>- the outcome you expect to achieve.

The fuzzy disk is a discretization of the algebra of functions on the
two-dimensional
disk using finite matrices which preserves the action of the rotation
group. We define a
phi^4 scalar field theory on it and analyze numerically three different
limits for the rank of the matrix going to infinity. The numerical simulations reveal three different phases:
uniform and disordered phases already present in the commutative scalar
field theory and a non-uniform ordered phase as non commutative effects. Since this
simulations have an high grade of parallelism the use of the GRID allow us to process all considered different configurations of the
model (approx 10000 cases) potentially completely in parallel. All the submissions can be achieved using a parametric job together a suitable
wrapper script which allows to use multiple parameters.

Printable summary: this is the only </br> section of the abstract that will
be published in the Book of Abstracts.

Here we present a first implementation in the GRID framework of a set of simulations in the field of the non commutative geometry.Such calculus (typically Monte Carlo like simulations) are usually done on local resources such as local clusters, our goal here is to open the way to the GRID paradigm to such models or even to the “fuzzy” community.
We will study then a quantized phi^4 scalar field theory approximating field with
NxN matrices. We are interested in particular to the phase transitions of the theory as we change the parameters of the action. The
quantity of interest are susceptibility and specific heat, as well as other order parameters,
which we will describe below.
Is worthwhile compare the present method with the simulations of such theories on the lattice. In general the simulations of scalar theories on fuzzy spaces are slower than in their lattice counterparts since the fuzzy models due to the self-interaction term phi^4 are intrinsically non local and for higher power of the self-interacting the number of operations to calculate each Monte Carlo step deltaS grows even faster. However, we can expect some advantages with the simulations of other symmetric field theories.
The implementation of the GRID paradigm in this field of theoretical research could push the complexity of the simulations (aka the
matrix rank) in such a way to obtain a much better result of even new results in the same amount of time spent in the current local simulations.

Primary author: Dr SPISSO, Bernardino (Federico II Napoli and INFN)

Presenter: Dr SPISSO, Bernardino (Federico II Napoli and INFN)

Contribution ID: **5** Type: **Demonstration**

Data transfer between storages with different protocols by Data Avenue

Description of content and intended audience</br>- the outcome you expect to achieve.

During the Data Avenue demonstration visitors can see how the Data Avenue interface looks like. The usage will be demonstrated by transferring files between storage resources that use different protocols and by uploading and downloading data from different storage resources using the Data Avenue service.

The intended audience is anyone who has ever had to transfer data for an e-science application and had difficulties with the data transfer, for example because the data was moved between different DCIs or storage resources that use different protocols.

The expected outcome of the demonstration is that visitors will understand the reasons why Data Avenue would be an essential tool for everyone who is dealing with data transfer between storage resources that are accessible with different tools through different protocols.

Preferred Day if any (Demos - Mon, Tue, Wed)

Tuesday

Relevant URL (if any)

https://data-avenue.eu/

Printable summary: this is the only </br> section of the abstract that will
be published in the Book of Abstracts.

Some of the data that scientists use for their e-science applications might be stored on different storage resources, making it difficult for them to browse, access or update it. Different protocols are accessible with different tools, some with ergonomic user interface, but some with only command-line interfaces. Data Avenue was created to ease the life of scientists by the widespread support of protocols, enabling the easy data moving between various DCIs (such as grid, cloud, cluster, supercomputers) storage resources by various transfer protocols (HTTP, HTTPS, SFTP, GSIFTP, SRM).

The Data Avenue operation is similar to a file commander tool: using Data Avenue, you can up- and download your data to storage services for scientific computation. Additionally, you can copy, move and delete files as well as you can create and copy folders. With Data Avenue, there is no need to use a local environment as temporary file storage during transfer.

The demonstration is intended to show the usage of the Data Avenue and highlight the advantages of the service.

Primary author: VARGA, Kitti (MTA SZTAKI)

Co-authors: KACSUK, Peter (MTA SZTAKI); FARKAS, Zoltan (MTA SZTAKI)

Presenters: VARGA, Kitti (MTA SZTAKI); FARKAS, Zoltan (MTA SZTAKI)

Contribution ID: **6** Type: **Poster**

SAGrid-2.0

Description of content and intended audience</br>- the outcome you expect to achieve.

This contribution describes new services provided by SAGrid which are now in beta. These have been developed in collaboration with several roleplayers in the distributed computing universe, essentially supported by the CHAIN-REDS and ei4Africa projects. These consist of essential core services, essential for infrastructure interoperability, as well as security services (CSIRT and CA) and provision for identity federation.

There are also four new services:
1) Executable Infrastructure: using tools like Ansible and puppet, a small team of experts can "code" the services, providing version control and verification of site configuration.
2) Self-service application porting : The Jenkins Continuous Integration server allows users or application support teams to test their applications against pre-defined test to ensure that they will run on the WN.
3) Automated application delivery : Once applications pass all tests, they are staged to the CVMFS repository
4) Easy access to heterogeneous services via a science gateway concept, based on SAGA and Liferay.

The intended audience is grid application developers, national operations teams and peer infrastructure liaison

Relevant URL (if any)

http://www.sagrid.ac.za
http://www.chain-project.eu

Printable summary: this is the only </br> section of the abstract that will
be published in the Book of Abstracts.

The South African National Grid (SAGrid) is a federation of universities, national laboratories and research groups which operates a distributed computing infrastructure. Starting in 2009 and initially based on gLite-3.2 middleware, SAGrid allowed better participation to WLCG and other large distributed e-Infrastructures. However, a long tail of individual researchers in South Africa and the rest of the region were excluded or poorly-served. With the evolution of grid and maturity of cloud middleware, the potential of the infrastructure has increased. With the support of the CHAIN-REDS project, a signature of a Resource Infrastructure Provider MoU with EGI.eu has also improved interoperation, collaboration and reliability of the infrastructure.
In this contribution, we report on how we further improve the usability and sustainability of the infrastructure by adopting new methodologies and technologies to support technical and user communities. While still offering a "grid" infrastructure, SAGrid is now also offers user- and developer-friendly interfaces to new services, while the adoption of executable infrastructure improves reliability and scalability of operations.

Primary author: BECKER, Bruce (South African National Grid)

Co-author: Mr RIEKERT, Stefanus (UIniversity of the Free State)

Presenter: BECKER, Bruce (South African National Grid)

Contribution ID: **7** Type: **Demonstration**

Dynamic management of virtual infrastructures

Description of content and intended audience</br>- the outcome you expect to achieve.

This demo will demonstrate the usage of the Infrastructure Manager service using both the CLI tool and the Web interface, to deploy virtual infrastructures on cloud deployments.
It will show the Resource Application Description Language (RADL) language used to describe the requirements of users'customized virtual infrastructures. Then it will show how the user can create a virtual infrastructure using the CLI and the Web interfaces. It will also show the infrastructure can be elastically modified adding or removing VMs on runtime and how it will be reconfigured.
Finally the Elastic Cloud Computing Cluster (EC3) tool will be shown to demonstrate the usability of the IM tool in combination with an energy management system for High Performance Computing called CLUES to create elastic clusters that are self-managed based on the cluster workload.
The demo will show the description of how to configure and set-up a CE+WN EGI node from basic VMIs. Demonstration of this use case will take too long for an interactive demo session, but attendees will have a deeper understanding of the potential.

Relevant URL (if any)

http://www.grycap.upv.es/im
http://www.grycap.upv.es/clues
http://www.grycap.upv.es/VMRC

Printable summary: this is the only </br> section of the abstract that will
be published in the Book of Abstracts.

Cloud infrastructures are becoming an appropriate solution to address the computational needs of scientific applications. However, the use of public or on-premises Infrastructure as a Service (IaaS) clouds require users to have non-trivial system administration skills. Resource provisioning systems provide facilities to choose the most suitable Virtual Machine Images (VMI) and basic configuration of multiple instances and subnetworks. Other tasks such as the configuration of cluster services, computational frameworks or specific applications are not trivial on the cloud, and normally users have to manually select the VMI that best fits, including undesired additional services and software packages and even install additional software by their own. This demo will present a set of components that ease the access and the usability of IaaS clouds by automating the VMI selection, deployment, configuration, software installation, monitoring and update of Virtual Appliances. In addition it integrates a contextualization system to enable the installation and configuration of all the user required applications providing the user with a fully functional infrastructure. Moreover, the contextualization agent included in the framework supports horizontal (increase/decrease the number of resources) and vertical (increase/decrease resources within a running Virtual Machine) by properly reconfiguring the software installed, considering the configuration of the multiple resources running.

Primary author: CABALLER, Miguel (UPVLC)

Co-authors: Mr GARCÍA, Andrés (UPVLC); Mr DE ALFONSO, Carlos (UPVLC); Dr ROMERO, Eloy (UPVLC); Dr MOLTÓ, Germán (UPVLC); Dr BLANQUER, Ignacio (UPVLC)

Presenters: Dr BLANQUER, Ignacio (UPVLC); CABALLER, Miguel (UPVLC)

Contribution ID: **8** Type: **Poster**

The Grid Observatory 3.0 - Towards reproducible research and open collaborations using semantic technologies

Description of content and intended audience</br>- the outcome you expect to achieve.

–

Relevant URL (if any)

http://grid-observatory.org

Printable summary: this is the only </br> section of the abstract that will
be published in the Book of Abstracts.

The Grid Observatory 3.0 evolves the Grid Observatory (G.O.) and Green Computing Observatory (G.C.O.) along the Open Linked Data and reproducible research concepts.
The first objective is to make analysis easier and more productive, by addressing the technical heterogeneity of the data (EGI services logs), and the wide range of potential usage. Semantic web technologies address these by (i) creating an OWL ontology of the EGI software architecture, (ii) converting the traces from selected services of the grid into an ontology compatible RDF format and (iii) organizing them in SPARQL-enabled triple stores. These technologies expedite and make transparent the personalized integration of multiple, independent sources, required for analysing the behaviour of the EGI grid, as well as long-term sustainability of the GO and GCO repositories. Moreover, the scientist's activity can be exploited to refine the ontology in a collective knowledge building process
The second objective is to encourage reproducible science by providing ways to repeat in silico experiments based on the GO data and stored queries over data and processing algorithms. A catalogue of customizable queries will be provided to show examples of queries and processing over the published data. Cloud-based hosting and processing capabilities will be offered to scientists to store and share their processes and algorithms through a collaborative platform in order to encourage open collaborations.

Primary authors: GERMAIN-RENAUD, Cecile (CNRS); NAUROY, Julien (CNRS); RAFES, Karima (Inria)

Presenter: NAUROY, Julien (CNRS)

Contribution ID: **9** Type: **Poster**

An OpenStack based IaaS installation for providing high quality services for scientific applications in the contest of EGI federated cloud task force

Description of content and intended audience</br>- the outcome you expect to achieve.

–

Printable summary: this is the only </br> section of the abstract that will
be published in the Book of Abstracts.

A highly customized deployment of the OpenStack platform, implemented within PRISMA, a project founded by the Italian Ministers of Instruction and Economic Development. This project aims at providing an open source cloud solution that can be easily used from both SME (Small and Medium Enterprises) and Local Public Administration. Within this activity, we have carried detailed tests, to provide optimal solutions in terms of back-end storage, High Availability configuration, Object Storage performance.
In particular we focus, also, on the provisioning of critical and CPU intensive services to support diverse scientific communities (Bioinformatics, Biomedicine, Astrophysics, etc) and different use-cases (Data Preservation, Seismic Risk, e-Government, etc) in the context of the EGI Federated Cloud Task force.

Primary authors: ITALIANO, Alessandro (INFN); MONACO, Alfonso (INFN-Bari); SANTERAMO, Bruno (INFN-Bari); TINELLI, Eufemia (UNIBA); Dr DONVITO, Giacinto (INFN); MINIELLO, Giorgia (UNIBA); MAGGI, Giorgio (INFN); ANTONACCI, Marica (INFN); NOTARANGELO, Pasquale (INFN-Bari); VALENTINI, Roberto (INFN-Bari); NICOTRI, Stefano (INFN); SPINOSO, Vincenzo (INFN)

Presenter: Dr DONVITO, Giacinto (INFN)

Contribution ID: **10** Type: **Demonstration**

Unity - one stop IAM & FIM

Description of content and intended audience</br>- the outcome you expect to achieve.

Intended audience: service providers, ICT project representatives, infrastructure providers.

Contents: Highly dependent on an available timeslot. The base idea is to present:
- basics of Unity features:
– users catalogue setup
– attributes
– groups
– members maintenance
– registration form definition
– registration process
- integration with upstream Identity Providers using various protocols
- features related to authentication management
- concrete use-case studies with their realisation

Expected outcome: establishing contact with prospective users of the system, collecting new requirements, initiating collaboration with new partners/projects.

Relevant URL (if any)

http://unity-idm.eu

Printable summary: this is the only </br> section of the abstract that will
be published in the Book of Abstracts.

The demonstration presents an emerging Unity system - a complete identity and (inter-)federation management solution.

Unity allows its users to enable authentication (or login) to their web services using various protocols, with different configurations for many relaying parties. The actual authentication can be performed using the built-in, feature-rich users database or can be delegated to one of supported upstream identity providers (IdPs). The information obtained from upstream IdPs can be flexibly translated and merged with the local database (if needed) and re-exported using other protocols.

While Unity is a new undertaking, being only one and a half year old, it already offers a very high level of service. Hierarchical groups, configurable credentials, automatically and manually assigned attributes with rich value syntax settings and attribute classes are only few examples of what is supported in the local users database. The users management is complemented with a flexible registration support. All of this can be seamlessly integrated with upstream and downstream services.

What is important for the Grid community, Unity contains a dedicated UNICORE plugin. Using this plugin, which extends a standard Unity's SAML IdP endpoint, UNICORE 7 services can be accessed by users who do not have certificates - in a fully web-alike mode. There are ongoing plans to implement other Grid-related plugins for Unity.

Primary author: BENEDYCZAK, Krzysztof (UWAR)

Presenter: BENEDYCZAK, Krzysztof (UWAR)

Contribution ID: **11** Type: **Poster**

ATLAS Event Index prototype for cataloguing large amounts of data

Description of content and intended audience</br>- the outcome you expect to achieve.

This project consists in the development and deployment of a catalogue of events for experiments with large amounts of data, such as those currently taking data with the LHC accelerator at CERN. The ultimate goal is to have available for our experiments, and make available to other users, a consistent set of software packages that allow fast and efficient searching of events of interest among the billions of events recorded in millions of files scattered in a world-wide distributed computing system.

The most innovative part of this project is the adaptation and application, for the first time, of the NoSQL technology for the cataloguing of data of a large experiment. ATLAS alone, but other experiments have similar numbers, accumulated since 2011, 2 billion real and 6 billion simulated events. This enormous amount of data (many TB) can be managed in Oracle database only by investing in hardware and operating personnel, and dividing the database into parts that for practical and financial reasons are located in different computing centres in different countries. With this work we want to explore the usage of NoSQL technologies for this kind of indexing, and evaluate aspects of performance and scalability that are extremely important to have a product that will be useful to the scientific community.

Printable summary: this is the only </br> section of the abstract that will
be published in the Book of Abstracts.

The Event Index project consists in the development and deployment of a catalogue of events for experiments with large amounts of data, such as those currently taking data like the ATLAS detector with the LHC accelerator at CERN. A database with the references to the files including each event in every stage of processing is necessary in order to later retrieve the selected events from data storage systems, and to be used as a reference index for final users, or for automated tools.

In this poster we present the architecture and the current implementation for the different parts involved in the project, including the data collection and upload to the central Hadoop server, and the designed infrastructure to access the stored information.

Data to be stored in the EventIndex are produced worldwide by all production jobs that run on Tier-0 or the Grid. For every permanent output file a snippet of information, containing the file unique identifier and for each event the relevant attributes is sent. In our first prototype we are using messaging technologies like Stomp protocol and ActiveMQ broker to convey the information. The estimated rate (in May 2013, during the LHC shutdown) is about ~20 Hz of file records containing ~3.4 kHz of event records, summing up a rate of 300GB of event information per day. During data-taking periods these numbers will be doubled; as both the data-taking rate and the Grid processing power are expected to increase by perhaps a factor two by 2015.

Primary author: FERNANDEZ, Alvaro (CSIC)

Co-authors: FAVARETO, Andrea (Università di Genova and INFN, Genova, Italy); BARBERIS, Dario (Università di Genova and INFN, Genova, Italy); MALON, David (Argonne National Laboratory, Argonne, IL, United States); GALLAS, Elizabeth (Oxford University, Oxford, United Kingdom); PROKOSHIN, Fedor (Universidad Técnica Federico Santa María, Valparaíso, Chile); SANTIAGO, GONZALEZ (CSIC); CRANSHAW, Jack (Argonne National Laboratory, Argonne, IL, United States); SÁNCHEZ, Javier (CSIC, Valencia, Spain); SALT, Jose (IFIC (Instituto de Física Corpuscular)); HŘIVNÁČ, Julius (LAL, Université Paris-Sud and CNRS/IN2P3, Orsay, France); NOWAK, Marcin (Brookhaven National Laboratory, Upton, NY, United States); BRIONGOS, Pablo (CSIC); YUAN, Ruijun Justine (LAL, Université Paris-Sud and CNRS/IN2P3, Orsay, France)

Presenter: FERNANDEZ, Alvaro (CSIC)

Contribution ID: **12** Type: **Poster**

Efficient Management of an OpenStack Cloud Infrastructure Through Multiobjective Programming

Description of content and intended audience</br>- the outcome you expect to achieve.

Description of contents: Our experiments analyse the influence of the ACPI module, the boot costs of a node configured with the different ACPI sleep status, virtual machine migration costs, comparisons between simulated and real cloud infrastructures and comparisons among the different alternatives explained in this poster.

Audience: System administrators, developers and data center directors

Outcome: The aim of this poster is to present our work, show our interest for energy and cloud technology, and meet new contacts.

Printable summary: this is the only </br> section of the abstract that will
be published in the Book of Abstracts.

Cloud Computing as IAAS (Infrastructure as a Service) has gained a rapid acceptance in the IT world producing implantations of clouds inside data centers. Focused on large cloud infrastructures, hundreds of machines are working together to offer users a base to execute their virtual machines. However, these cloud infrastructures consume large amounts of electrical energy, contributing to high carbon footprints to the environment. In addition, users expect the best reliability and availability of the cloud services when their business are depending on those services. Therefore, we need a Green Cloud Computing solution which does not only minimize energy consumption, but also achieves the efficient processing and utilization of cloud computing infrastructures.

In this poster, we work with a real cloud architecture based on the OpenStack platform to achieve a solution through real infrastructures instead of simulators. Our experiments analyse the influence of the Advanced Configuration and Power Interface (hereinafter ACPI) module, the boot costs of a node configured with the different ACPI sleep status, virtual machine migration costs and comparisons among the different alternatives explained in this poster (without threshold, with one threshold and with two thresholds). The result is the development of a set of daemons which minimizes the energy consumption, and also achieves the efficient processing and utilization of a cloud computing infrastructure.

Primary author: DIAZ, Miguel Angel (CETA-CIEMAT)

Co-authors: Dr PAZ, Abel (CETA-CIEMAT); Dr VEGA, Miguel Angel (University of Extremadura)

Presenter: DIAZ, Miguel Angel (CETA-CIEMAT)

Contribution ID: **13** Type: **Poster**

Introducing rOCCI-server and the rOCCI Framework

Description of content and intended audience</br>- the outcome you expect to achieve.

rOCCI is a new product developed inside EGI to facilitate interoperability within the Federated Cloud Platform. It is currently being deployed across EGI Federated Cloud sites, and users will be using it more and more frequently, be it directly through their OCCI-enabled clients, or indirectly through brokers or submission portals. This poster aims at informing FedCloud users about the new component in the Federated Cloud Platform.

Printable summary: this is the only </br> section of the abstract that will
be published in the Book of Abstracts.

The OCCI (Open Cloud Computing Interface) Standard by the Open Grid Forum is becoming more and more popular and, indeed, is one of the most widely adopted standards across the cloudscape. It focuses on integration, portability and interoperability, and also provides a high degree of extensibility. The rOCCI framework implements the OCCI standard and exposes essential client- and server-side functionality to simplify the use of OCCI by developers and –through providing its own command-line interface to OCCI –also to users. The rOCCI-server takes that simplification even further. Despite the above-mentioned high rate of adoption of OCCI, popular but OCCI-incompliant cloud management platforms still exist. The rOCCI server extends such services to provide OCCI capabilities by exposing its OCCI interface on the outside, and communicating internally with the cloud management platform through its arbitrary API by means of a product-specific backend. rOCCI is an EGI product currently being released into the EGI Cloud Federated Platform to bring OCCI capabilities to sites relying on the OpenNebula Cloud Manager. This poster introduces the basic principles of the rOCCI-server, and explains the overall architecture of cloud sites, which decide to employ it.

Primary authors: PARAK, Boris (CESNET); DVORAK, Frantisek (CESNET); SUSTR, Zdenek (CESNET)

Presenter: SUSTR, Zdenek (CESNET)

Contribution ID: **14** Type: **Poster**

CosmoHUB: database and web solution for access and distribution of cosmological catalogs

Description of content and intended audience</br>- the outcome you expect to achieve.

We will describe:
- The structure of our database and how we guarantee the privacy of the data.
- The web interface, statistics of access and relations with the database tables.
- The custom query service and the automatic execution of queries in the PIC computer farm
- The future development of CosmoHUB, with a forecast of the increase in number of catalogs, especially from the PAU Survey project data (coming soon!).

Intended audience:
- scientists working with data to be easily distributed and with strict privacy requirements,
- computing experts who want to find a solution for a database-web implementation that allow arbitrary queries and the download of big data volumes (several tenth of Gb per catalog).

We would like people to know the successful technical solution we found for such a case.
Astronomers interested in cosmological public catalogs we store will know that they can freely access them using CosmoHUB.

Relevant URL (if any)

http://cosmohub.pic.es

Printable summary: this is the only </br> section of the abstract that will
be published in the Book of Abstracts.

Projects like DES (Dark Energy Survey), PAU (Physics of the Accelerating Universe) Survey, MICE (Marenostrum Institut de Ciències de l'Espai) Simulations and Euclid are contributing to dark energy study mapping the large scale structure of the universe and producing catalogs of million of galaxies. From ICE-IEEC (Institut d'Estudis Espacials de Catalunya) and IFAE (Institut de Física d'Altes Energies) scientists working in these projects came the need of a centralized site to collect, access and distribute catalogs among the project group members and where to publish catalogs (or part of them) to the whole scientific community. At PIC (Port d'Informació Científica) we are taking advantage of the grid infrastructure to implement such a service. While catalogs in their original format are stored in the PIC disk storage system, we implemented a database fed with real and simulated cosmological catalogs with the relative access permissions, and a web portal called CosmoHUB to access the cosmological data of the database. Access to private data is restricted to groups members after administrators confirmation, while public data is of free access after registration. The available services in CosmoHUB are: download of entire or pre-built value-added catalogs, request of quick custom queries (limited in number of elements returned) for visual analysis, or of full custom batch queries, executed after web submission through a job automatically executed in the PIC computer farm.

Primary authors: Dr PISCIA, Davide (ICE-IEEC); Dr CARRETERO, Jorge (ICE-IEEC); Dr TONELLO, Nadia (PIC); TALLADA, Pau (PIC); SERRANO, Santiago (ICE-IEEC)

Co-authors: RODRÍGUEZ, José Antonio (ICE-IEEC); Dr FOSALBA, Pablo (ICE-IEEC); FIRPO, Roger (IFAE)

Presenter: FIRPO, Roger (IFAE)

Contribution ID: **15** Type: **Poster**

Grid- and Cloud-based Model for Evaluation of Simulated Annealing Method for Distributed Software Engineering

Description of content and intended audience</br>- the outcome you expect to achieve.

Many scientific grid and cloud computing applications helps to solve the problems of either the physical sciences or the life sciences. Only marginal part of all the applications is dedicated to other domains. Moreover IT and software engineering have a tiny (if any) piece even of this part. Our research focus is on the applications of the distributed infrastructures for the software engineering, including generative software development, modeling and optimization.
We do believe that such non-traditional application can be interesting to many participants especially to computer science and software engineering related ones. Our presentation can help to open more wide discussion on the place of Software engineering in distributed infrastructures context. We're expecting to find collaborators and potential partners.

Preferred Day if any (Demos - Mon, Tue, Wed)

May 20 (Tue)

Relevant URL (if any)

http://ik.su.lt/~vaigie/socosys/

Printable summary: this is the only </br> section of the abstract that will
be published in the Book of Abstracts.

We have proposed the method for optimal component set selection taking into account non-functional properties of components. This method have been used in the extended SoCoSys system and can be used as an extension of other proofs-as-programs methods and component-based software development systems. However the algorithm for the evaluation and improvement of the component set presents the class of algorithms. It is possible to change the possibility of global optimum detection of our optimization algorithm using different values of parameters α, β, T and g. In order to investigate the impact of the values of parameters we use massive computations using distributed computing infrastructures. In this presentation we introduce the model of using e-infrastructures in order to solve particular software engineering problem. We present our experience and the role and implementation-related issues working with ARC middleware and Windows Azure. This is one of examples of the applications from various domains covered by Lithuanian National Grid Initiative.

Primary author: Dr GIEDRIMAS, Vaidas (Siauliai University)

Co-author: Prof. SAKALAUSKAS, Leonidas (Vilnius University)

Presenter: Dr GIEDRIMAS, Vaidas (Siauliai University)

Contribution ID: **16** Type: **Poster**

Software compatibility check framework for grid computing elements

Description of content and intended audience</br>- the outcome you expect to achieve.

The target audience of this work are developers of software to be executed on the grid, system administrators and devops operators in charge of putting into production new software packages on the grid.

The goals of this work are basically:

- share our recent experience of our group related to the deployment of new software to be used on the gridand determing the compatibility between software and computing elements.
- receive feedback and suggestions about possible improvements for the proposed framework. There are several aspects of our proposed architecture that would benefit from some feedback from grid operators.

Relevant URL (if any)

http://bioinformaticslaboratory.nl/twiki/bin/view/EBioScience/CvmfsPublic

Preferred Day if any (Demos - Mon, Tue, Wed)

Wednesday

Printable summary: this is the only </br> section of the abstract that will
be published in the Book of Abstracts.

The proposed framework for software compatibility checking on grid computing elements (CEs) based on gLite middleware derives from the experience at the Academic Medical Center of the University of Amsterdam (AMC, NL) using the Virtual Organisation (VO) VLEMED on Dutch gLite resources of the European Grid Infrastructure.

The jobs submitted by VO VLEMED are based on a heterogeneous set of software for data analysis, which management on the CEs involves several challenges: library incompatibilities, lack of integration in package management systems, binary architecture and version conflicts, etc.

To improve the success rate of submitted jobs, we designed an implemented a framework to check software compatibility on grid CEs. This framework relies on predefined test jobs and their expected results.

CernVMFS, a read-only network file system optimised for software distribution, combined with the Environment Modules tools, overcomes many of the problems associated to the software distribution on the grid described above.

The grid job submission is based on Ganga, a Python-based front-end for job management and submission. Each software to be tested requires the definition of a Ganga grid job, including input data and scripts to validate the results from the CEs.

Finally, Nagios has been used to publish the compatibility results obtained by the framework. The cron daemon and several shell scripts provide automation features.

Primary author: FONT, Juan Luis (EGI.EU)

Co-authors: Dr LEE, Hurng-Chun (Radboud University); JAGHOORI, Mohammad Mahdi (Academic Medical Center, UvA); OLABARRIAGA, Silvia (Academic Medical Center of the University of Amsterdam)

Presenter: FONT, Juan Luis (EGI.EU)

Contribution ID: **17** Type: **Poster**

The EGI-DRIHM collaboration: how the european Grid and Cloud infrastructures can support Hydro-Meteorology Research

Description of content and intended audience</br>- the outcome you expect to achieve.

In the poster the architecture of the DRIHM science gateway (with the screenshot of the most important portlets), and the different components of the distributed computing infrastructure used in the project will be presented.
Furthermore, a box will present in details the considered HMR simulation chains, while another will present the "gridification" process of the HMR models.
As regards the outcome, the first goal is to present the experience and the results achieved by the DRIHM project so far. The second goal is to try to get more resources for the drihm.eu VO.

Relevant URL (if any)

http://www.drihm.eu/

Preferred Day if any (Demos - Mon, Tue, Wed)

Tue or Wed

Printable summary: this is the only </br> section of the abstract that will
be published in the Book of Abstracts.

Hydro-meteorological forecasts rely on the execution of complex, computationally intensive, simulation models. A full simulation - from rainfall to impact on urban areas - requires the execution of several models organized through workflows, whose accuracy is strongly dependent on an extensive set of configuration parameters.
The DRIHM project aims at setting the stage for a new way of doing Hydro-Meteorological Research (HMR). In particular, the goal is the
development of a science gateway that allows users, from scientists to environmental agencies and citizen scientists, to access and combine data and forecasting models using integrated services, user-friendly interfaces and resources from European - ad in perspective worldwide - infrastructures.
Within this framework the EGI-DRIHM collaboration focuses on the integration of the HMR model selected in the project on the Grid and Clouds resources made available within the drihm.eu virtual organization. In this contribution we present such experience.
The main focus is on the process of adapting HMR models for the deployment on an enabling Grid infrastructure that supports dynamic model chains.

Primary author: D'AGOSTINO, Daniele (Consiglio Nazionale delle Ricerche (CNR))

Co-authors: QUARATI, Alfonso (CNR-IMATI); CLEMATIS, Andrea (CNR-IMATI); GALIZIA, Antonella (CNR-IMATI); PARODI, Antonio (CIMA RESEARCH FOUNDATION); STRAUBE, Christian (LMU); DANOVARO, Emanuele (CNR-IMATI); SCHIFFERS, Michael (LMU); GENTSCHEN FELDE, Nils (LMU)

Presenter: D'AGOSTINO, Daniele (Consiglio Nazionale delle Ricerche (CNR))

Contribution ID: **18** Type: **Poster**

From DIRAC towards an Open Source Distributed Data Processing Solution

Description of content and intended audience</br>- the outcome you expect to achieve.

The DIRAC software framework allows easy extension of its already existing functionality for the needs of new user communities. A number of user communities in different domains already adopted the DIRAC software as the basis for their distributed computing systems.

The outcome expected is to raise awareness on new communities that might be interested on a tool like DIRAC to simplify their access to distributed computing.

Relevant URL (if any)

http://diracgrid.org

Printable summary: this is the only </br> section of the abstract that will
be published in the Book of Abstracts.

The DIRAC Distributed Computing Project (the "DIRAC Project") provides an open source software framework for building distributed computing systems. The DIRAC software framework is designed to create services for distributed computations in various environments such as grids, clouds and clusters. The DIRAC software was initially developed to meet the needs of the LHCb HighEnergy Physics experiment at CERN. It was first used for massive production of the LHCb modeling data and was later extended to support all the distributed computing operations, including Data Management and User Analysis. The DIRAC Project introduced several innovations focusing on the needs of large and distributed scientific collaborations.
Based on this framework, the DIRAC Project develops and maintains software systems supporting standard distributed computing tasks. The DIRAC Workload Management System provides all the necessary components to ensure efficient execution of user jobs using heterogeneous computing resources. It makes a special emphasis on fault tolerant operations in unstable distributed computing environments. Support is provided for large user communities with complex internal policies of computing resource usage for massive data processing. The DIRAC Data and Storage Management Systems provide tools for seamless access to various types of data storage, data cataloging and classification as well as for efficient and reliable massive data replication.

Primary authors: TSAREGORODTSEV, Andrei (CNRS); Dr GRACIANI DIAZ, Ricardo (University of Barcelona)

Co-authors: CASAJUS, Adria (CERN); Dr POSS, Stephane (AlpesLaser); MENDEZ, Victor (UAB)

Presenter: Dr GRACIANI DIAZ, Ricardo (University of Barcelona)

Contribution ID: **20** Type: **not specified**

Experience in building heterogeneous cloud/grid infrastructure in BITP data centre

Printable summary: this is the only </br> section of the abstract that will
be published in the Book of Abstracts.

In this poster will be presented current installation of cloud platform in BITP data centre. The main reason to use the mix of cloud and physical resources is a variety of hardware families presented in data centres and job tasks for submission that require different quantity and type of CPU, RAM and storage capacity. Cloud can intelligently reallocate resources and organise more efficient management of available resources. In BITP we tried to create dynamic bridge between Torque scheduler that serves as a "head" of physical cluster and Openstack Havana (with Neutron) interacting by Openstack API requests (idea was based on usage of project Dynamic Torque); idea of package distribution for data analysis was based on usage of CVMFS: local NGI_UA's repository and ALICE experiment's repository managed by CERN. For authentication was selected keystone-voms plugin for managing VOMS authentication of grid users and compatible with usage in Federated Cloud. As a result we received stable, load balanced cloud/grid system, able to serve as data centre with dynamically allocation properties of resources and available for job submition and data analysis either between two parts of clouds (using availability zones) or directly to resources of EGI Federated Cloud or organise data analysis for LHC.

Description of content and intended audience</br>- the outcome you expect to achieve.

v

Primary authors: SHADURA, Oksana (BITP); SVISTUNOV, Sergiy (Bogolyubov Institute for Theoretical Physics NAS Urkaine)

Presenter: SHADURA, Oksana (BITP)