



CLOUD SERVICES IN THE GRID FRAMEWORK

CSIC-CNRS MEETING

14 DEC 2011, MADRID



VISION

CONTEXT

- CLOUD technology mature for business (and for many research applications)
Amazon, IBM, Google, Apple... + integration
- GRID technology mature (and complex) for LARGE scientific collaborations:
best example: LHC Worldwide Grid
- VERY DIFFERENT TARGETS AND (DIS)ADVANTAGES

MOTIVATION

- MORE EFFICIENT USE OF HPC RESOURCES**
including final users/developers/sys admin
- INITIAL AND RUNNING COST OF e-INFRASTRUCTURE**
Sustainability/**Business Model**

IDEAS

- Offer CLOUD services over GRID enabled infrastructure
- Enhance GRID access exploiting CLOUD experience
- FEDERATE CLOUDS exploiting GRID technology (AAA)

AN INFRASTRUCTURE IN OPERATION

Ibergrid: 132 Millions of CPU Hours Dedicated to Support

Ibergrid resources are being massively employed by the experimental particle physicist working on the data analysis of the LHC.

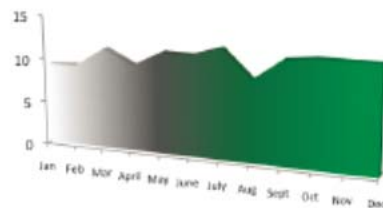
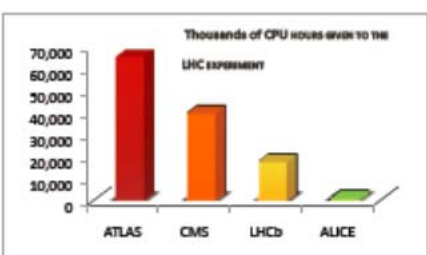
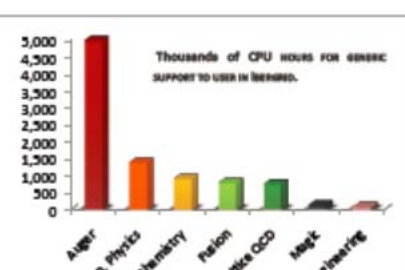
Over the year 2011 the Spanish Tier-1 PIC, and the Tier-2 centers located at Valencia, Santander, Madrid and Santiago de Compostela have jointly contributed with more than 115 millions of normalized CPU hours to the analysis of data for the LHC.

In 2012 we will see these numbers increasing because the LHC is going to run with more luminosity, implying a data production about 50 times bigger.

Along 2011 we have seen the consolidation of the usage of the Ibergrid infrastructure by user communities others than LHC.


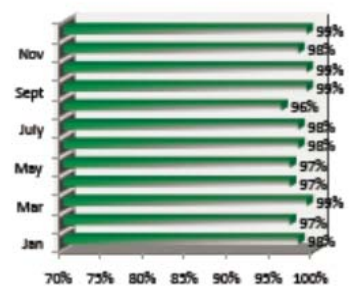
The analysis of data coming from the Pierre Auger Observatory of Cosmic Rays has led in the infrastructure as one of the major users.

Areas like Fusion, Computational Chemistry or Lattice QCD have consolidated a usage of about the million hours this year. The support to generic user communities has also been quite successful. In terms of number of users, over 54 currently, which altogether have employed about 1.5 millions of CPU hours.


High Availability and Reliability

IBERGRID IS ONE OF THE FIVE MORE RELIABLE PARTNERS WITHIN THE EUROPEAN GRID INFRASTRUCTURE, PROVIDING LEVELS OF ACCESSIBILITY OVER 96%, WELL BEYOND REQUIREMENTS (70%).

DURING 2011 THE EUROPEAN GRID INFRASTRUCTURE HAS PROVIDED TO SCIENTISTS OVER 1400 MILLIONS OF CPU HOURS WITH AN AVERAGE AVAILABILITY OF 93%.

THIS IS THE EQUIVALENT TO A DISTRIBUTED COMPUTING CENTER WITH ABOUT 160,000 CORES WORKING 24/7



WORK STARTING...

-Collaborate with CLOUD Virtual Team in EGI



The screenshot shows the Ibergrid Wiki interface. At the top, there is a green header with the 'IBERGRID wiki' logo and a world map. To the right of the logo is a search bar with the text 'Buscar:' and two buttons labeled 'Títulos' and 'Texto'. Below the search bar is a navigation bar with buttons for 'RecentChanges', 'FindPage', and 'HelpContents'. The main content area is titled '1st Ibergrid Cloud Meeting' and contains the following text: 'Ibergrid is discussing the deployment of a federated cloud infrastructure composed by several resource centers of Ibergrid. On 13th-14th of December in Lisbon we will meet to discuss the current status and issues to deploy the infrastructure.' Below this text is a 'Tabla de Contenidos' (Table of Contents) with the following items: 1. 1st Ibergrid Cloud Meeting, 1. Agenda, 1. Wed 14 Dec, 2. Topics for discussion, 1. AuthN/AuthZ, 2. Virtual Images, 3. APIs, 4. EGI Virtualization Task Force, 5. Jobs vs Virtual Machines, 3. Resource Centers Experience, 1. IFCA HPC Cloud, 1. Issues, 2. CESGA Cloud. On the left side of the page, there is a sidebar with a yellow background containing a list of categories: Operations, User Support, and Software Provisioning, each with sub-items like FAQs, Technical Documents, Deliverables, Presentations, Project reports, Meetings, Agendas, Minutes, Publications, and Other.

OTHER INITIATIVES

- Start work with SME to offer HPC CLOUD Services to Companies
- Try to setup a (FP7 / CDTI / regional) project + involve IBM as “observer”

TOPICS

AUTHENTICATION/AUTHORIZATION

On Grids users are identified with a unique certificate issued by national level CAs. Users belong to different VOs and can be organized with attributes and groups. Authorization policies are usually defined using these VO groups and attributes. Products like ARGUS simplify the management of the policies by centralizing their definition at site level.

In the case of cloud, the authentication and authorization of users is independent on each cloud resource. Most systems use a certificate based authentication with CAs created ad-hoc at deployment time. Users cannot share their credential between different deployments and policies are usually defined at the user level (not at the group level as in the case of grid)

A federated cloud infrastructure may need a single-sign-on and a way to share authorization policies (similar to ARGUS or Shibboleth systems).

TOPICS

VIRTUAL IMAGES

On a federated cloud infrastructure, users should be able migrate their virtual machines from one resource provider to another. This requires that the container of the virtual machine is accepted in both providers and that the hypervisor is able to run the machine (Xen VMs may not run on KVM and viceversa). The Open Virtualization Format OVF initiative proposes a standard way to package and distribute virtual machines.

While some virtual machines can be run on a completely isolated environment that reduces the risk of security issues, for other virtual machines (WN) it may be necessary to certify, or at least to endorse the VM, so it can be trusted. See the HEPIX VWG Image transfer slides with info on this topic.

Repositories of Virtual Images (e.g. StratusLab MarketPlace) may be also of interest.

Another option for storing and indexing virtual images is VMRC (<http://www.grycap.upv.es/vmrc/>) developed at GRyCAP.

TOPICS

APIs and DATA ACCESS

There is a need for a standard API that allows interoperability between the different resource providers and to avoid vendor lock-in. Cloud Orchestrators normally provide their own proprietary API to manage the Virtual Machines. The Amazon EC2 API is also widely supported. There is also an initiative of OGF called OCCl (Open Cloud Computing Interface) that tries to provide an open standard for virtual machine management APIs.

The data area also presents similar issues. There are several proprietary specifications with Amazon S3 being implemented by most systems. The CDMI (Cloud Data Management Interface) also tries to provide an open standard for cloud data APIs.

EGI Virtualization Task Force

There is a virtualization task force created within EGI that tries to define a technical roadmap for the interoperable integration of virtualised resources from different resource providers to provide an integrated federated virtualised resources infrastructure for exploitation by EGI's user community.

COLLABORATION

INITIAL STEPS

The EGI-Virtualization task force has defined a set of scenarios that will be deployed in the following months. Some of the scenarios are rather trivial (first one is running a predefined virtual machine).

1) COLLABORATE IN AN INITIAL SETUP OF FEDERATED CLOUDS

CSIC offers technical personnel, infrastructure, and IBERGRID leadership
“Short” term project, with “long term” perspective within EGI.

2) ANALYZE A NON TRIVIAL USE CASE

Requirements for a simple but realistic application on BIODIVERSITY

Starting with a simple use case

-for example: DB implementation/access/exploitation

Covering an increasing number of related topics

-simple observation but geo-defined

-genetic sequences...

Explore possibilities at EU level (including LifeWatch)