

Frequently asked questions

Please check also the official LHC Grid Fest Media Pack [☞](#).

- **Why Grid for LHC ?** (answer provided by Ian)
 - ◆ Computing funding locally in major regions and countries
 - ◆ Efficient analysis of data everywhere
- **What is the LHC Grid Service ?** (answer provided by Ian)
 - ◆ The LHC Grid Service is a worldwide collaboration between:
 - ◇ 4 LHC experiments
 - ◇ About 140 computer centres that contribute resources
 - ◇ International Grid projects providing software and services
- **What is a Tier-0, Tier-1, Tier-2 ?** (answer provided by Ian)
 - ◆ Tier-0, Tier-1s and Tier-2s are defined by their role in processing data (answer provided by Ian)
 - ◇ Tier-0 is the center acquiring data responsible for data recording, initial data reconstruction and data distribution
 - ◇ Tier-1 (11 centres) provide permanent storage, they are in charge of data reprocessing and data analysis
 - ◇ Tier-2 (about 130 centres) are in charge of simulation and end-user analysis
- **How did the Grid evolve ?**
 - ◆ First computing Grids (distributed computing over WAN) came already back in 1986 and 1988 when the Condor project started at the University of Wisconsin in U.S.A.
 - ◆ First ideas about Data Grids came with the CERN Monarc project back in 1994-1995. The project presented a partially decentralized model for data where event data were replicated at five regional centres and data transfer could take place either via network or movable media.
 - ◆ First proposal for "The Grid" was made by the group lead by Ian Foster and Carl Kesselman in 1999.
 - ◆ The European Data Grid project commissioned by the European Union was launched at the beginning of 2001 together with other partner projects such as DataTAG. It was followed by continuation projects such as EGEE1 and EGEE2. In parallel, in U.S.A. other projects were launched: GriPhyN, iVDGL, PPDG, Grid3 and today's OSG.
 - ◆ LCG was launched in 2002 to provide a computing infrastructure for the LHC experiments. It evolved in LCG 1 and LCG 2.
 - ◆ Today WLCG embraces other international Grid projects such as EGEE, NorduGrid (in the Northern countries) and OSG.
- **What do you see as the future of the Grid ?**
 - ◆ It is hard to say how the Grid as we use it today in the research field will evolve. We are assisting now in a commercialization of Grid-like distributed services that allow business to grow quickly, even without big initial resources and investments. That's utility computing, the paradigm of cloud computing and the so-called data centers. I personally believe that there's a huge need to share: experience, resources, instruments. We are eager to know and learn within some limits. The technology is just working toward this direction and objective.
- **The Grid is ready for LHC. When can a generic person use it from home ?**
 - ◆ If we think back about the Web, it was in 1989 that Tim Berners-Lee invented the Web. Only at the end of 1990 the first Web page appeared and this new technology was available to scientists at CERN. It took about 10 years for the Web, as we know it today, to really be available to the general public. For the Grid the process seems to be a bit faster, since already today we have Grid-based services made available to the business world. That's how very well-known utilities such as Facebook or YouTube were born. We are already using the Grid, we just do not realize it yet!
- **What can the Grid do for me ?**
 - ◆ The Grid is a Web with computing and storage power. It enables the sharing of data of all types, while providing an enormous computing power to process them. In the future, through

the Grid it will be possible to foresee catastrophes, to correlate several fields of sciences and allow for a better understanding of currently incurable illness, such as cancer or the Parkinson disease. It will be possible to better understand our environment and compare what we currently know about the animals with the environmental changes. Furthermore, the Grid is currently used for studies on the genome to prevent and act on diseases.

- **Can I use the Grid now if I want to ?**
 - ◆ Demonstration testbeds are available to allow new research communities and students to get familiar with the Grid. An example of this is GILDA [GILDA](#). You can connect to their Web site to make the exciting experience of using the Grid!
- **What is "Grid operations" ?**
 - ◆ Grid Operations monitors and maintains the data storage and analysis infrastructure used by the High Energy Physics community of LHC.
- **How can the Grid ensure fair share of resources among its users ?**
 - ◆ A new community of researcher that wants to join WLCG/EGEE normally comes with a set of resources at given sites. They make available their resources to share them with other research communities and in exchange to use other available resources. Through the EGEE infrastructure, a negotiation with the EGEE sites takes place to ensure that the proper share of resources will be granted to the new research community. Grid services are instrumented to guarantee a fair share of resource usage among the different communities.

WLCG numbers

I have extracted the following numbers from the presentation that Ian will give during the LHC Grid Fest. Other "official" numbers are available in the USB that was prepared for the press. If you run Window, you can find a copy here: G:\Users\laacook\Public. The official LHC Grid Fest Media Pack with official numbers and answers to questions is available here [here](#).

- As of today 33 countries have signed the MoU
 - ◆ CERN (Tier 0) + 11 large Tier 1 sites
 - ◆ 130 Tier 2 sites in 60 "federations"
 - ◇ Other sites are expected to participate but without formal commitment
- Data volume is about **15 PetaBytes** of new data each year
 - ◆ This comes from the high rate * large number of channels * 4 experiments
- Compute power needed :
 - ◆ **100000** of today's fastest CPUs (about 2000KSI2000 each)
 - ◆ **45 PetaBytes** of disk storage
- Data acquisition rate per experiment:
 - ◆ ALICE = about 100MB/sec or 1.25GB/sec when colliding ions
 - ◆ ATLAS = about 320MB/sec
 - ◆ CMS = about 220MB/sec
 - ◆ LHCb = about 50MB/sec
- Dedicated **10Gbit/sec** network links between Tier-0 and Tier-1s
- The Grid concept really works. The CPU Usage in Early 2008 was:
 - ◆ 11% provided by CERN
 - ◆ 35% provided by Tier-1s
 - ◆ 54% provided by Tier-2s
- Number of jobs in WLCG:
 - ◆ About 2Million/month in Jan 2007
 - ◆ More than 10Million/month in May 2008
 - ◆ About 350K/day in May 2008
 - ◆ 1 job is about 8 hours use of a single processor
- Data transfer out of Tier-0
 - ◆ Full experiment rate needed is 650MB/sec
 - ◆ Desire capability to sustain twice that to allow for Tier 1 sites to shutdown and recover

- ◆ Demonstrated a sustained rate of more than 2GB/sec for a full day
- Tier 1s must store the data for at least the lifetime of the LHC = about 20 years
- EGEE is now in its 2nd phase with 91 partners in 32 countries.
 - ◆ The EGEE infrastructure spans on 240 sites, 43 countries. It offers about 45000CPUs, 12PetaBytes of data, more than 5000 users, more than 100 Virtual Organizations and sustains more than 100000 jobs/day.
 - ◆ Among the EGEE VOs: archeology, astronomy, astrophysics, civil protection, Computing Chemistry, earth sciences, finance, fusion, geophysics, high energy physics, life sciences, multimedia, material sciences.
- OSG is supported by the Department of Energy and the National Science Foundation
 - ◆ It has access to 45000Cores, 6PetaBytes of Disk Space, 15 Petabytes of tape
 - ◆ It offers more than 15000 CPU Days/day to physics (85%) and non-physics (15%) applications such as biology, climate and text mining.
 - ◆ It also offers for a 20% an opportunistic use of others resources.

A little story around the tools in demo

After introducing the audience to the RTM tool explaining the various elements on the screen, the presenter goes into the details of what it means to operate a Grid infrastructure. A few monitoring and accounting tools in use today are described to give a more in-depth feeling of what it means to operate a World-wide Grid infrastructure.

The RTM tool has highlighted the fact that users submit computational tasks to the Grid Service and get the results back. In order to achieve that, the sites collaborating to the Grid Service have to be fully functional and operational.

- The **GridMap** tool is used to have a quick overview of the sanity of the sites. The representation offered by GridMap conveys a lot of information in an intuitive manner, and is very practical for a quick identification of non-functional or degraded sites.
- The Grid service does not only offer computational power but also data storage and transfer services. The **GridView** tool allows an operator to verify that data transfers between Grid sites take place normally and efficiently. An in-depth description of this tool is provided by James Casey.
- Sites receive alarms about degraded performance of services or hardware failures or any other operational problems either through the Grid operators on shift or through monitoring tools installed at a site. The **Service Level Status (SLS)** is the tool used at CERN for this purpose to monitor Grid services and not only.
- Some times determining the cause of a problem implies a deep debugging of Grid services as well as the experiment's application. For instance, a Grid job can fail because of a non-functional Grid Service or because the application that such a job tries to execute has not been well written. The **Grid job dashboard** assists Grid operators and experiment scientists on shift in investigating possible failures. *Julia's comment : After looking in this application from the demo perspective, I think we rather skip it. The main reason is that inthe Dashboard we show success rate and people can notice that success rate is not always high (specially for analysis). We show RTM , so job processing is covered anyhow, we add the comment from Pablo about CPU used by 4 experiments in September, might be enough on this topic*
- Similarly to the Grid job dashboard, the **Grid Data Management dashboard** assists the physicists and the Grid operators during the investigation of transfer or data storage problems. *What we can show there: Example of the ATLAS Data Transfer Dashboard². In the 4 plots on the upper part of the page there are graphs showing troughput, amount of transferred data (in unites of GBytes and Files) done in one time bin, number of errors. In the table below one can see the status of group of sites (ATLAS cloud) which are 'attached' to a given Tier1. The status of the group is indicated by the colour. Pink colour indicates problems. Clicking on the name of the group, one gets the detailed status of transfer for all storage elements belonging to the sites of the group. Clicking on the problematic line, one gets detailed information about failures*

FAQ < LCG < TWiki

- It is important to check that the Grid is used efficiently by all its users. In fact, the WLCG infrastructure is not only used by High Energy Physics scientists to store, process and analyze their data, but also by other sciences. Therefore, it is important to be sure that the resources granted to given activities are indeed used and that the correct share is ensured among scientists. The **WLCG Accounting Portal** allows Grid operators and managers to check usage statistics and resources available.

-- FlaviaDonno - 01 Oct 2008

This topic: LCG > FAQ

Topic revision: r4 - 2008-10-02 - JuliaAndreeva



Copyright &© 2008-2022 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.
or Ideas, requests, problems regarding TWiki? use [Discourse](#) or [Send feedback](#)