ATLAS Tier-3 within IFIC-Valencia analysis facility

M. Villaplana, S. González de la Hoz, A. Fernández, J. Salt, A. Lamas, F. Fassi, M. Kaci, E. Oliver, J. Sánchez, V. Sánchez-Martinez for the ATLAS Collaboration
IFIC (CSIC/UV) Edificio Institutos de Investigación, 22085, E-46071 Valencia, Spain

Miguel.Villaplana@ific.uv.es, Santiago.Gonzalez@ific.uv.es, Alvaro.Fernandez@ific.uv.es, Jose.Salt@ific.uv.es

Abstract. The ATLAS Tier-3 at IFIC-Valencia is attached to a Tier-2 that has 50% of the Spanish Federated Tier-2 resources. In its design, the Tier-3 includes a GRID-aware part that shares some of the features of IFIC Tier-2 such as using Lustre as a file system. ATLAS users, 70% of IFIC users, also have the possibility of analysing data with a PROOF farm and storing them locally. In this contribution we discuss the design of the analysis facility as well as the monitoring tools we use to control and improve its performance. We also comment on how the recent changes in the ATLAS computing GRID model affect IFIC. Finally, how this complex system can coexist with the other scientific applications running at IFIC (non-ATLAS users) is presented.

1. ATLAS computing model.

ATLAS is one of the experiments that study the 7 TeV pp collisions provided by the LHC (Large Hadron Collider) [1]. The ATLAS Computing Model [2] allows the scientific community to store, access and analyse the enormous quantity of data produced at the ATLAS experiment. It has exported more than 25k TB of data to Tier-1s since January 2010 and the integrated luminosity recorded at the end of 2011 was 5.25 fb⁻¹. Faced with the difficulty that entails analysing such amount of data in a single place or even with the combined effort of several computing centres, the LHC experiments chose a model with a high degree of decentralization and the possibility of sharing resources, a model based on GRID technologies.

The WLCG (Worldwide LHC Computing GRID) [3] gives support to the computing models of the LHC experiments. It groups the different types of computing centers of the ATLAS collaboration in a tiered hierarchy that ranges from Tier-0 at CERN, to the 10 Tier-1 centers and the nearly 80 Tier-2 centers distributed world wide, and down to the end-user private analysis facility (Tier-3).

The Spanish Tier-1 is located at PIC (Port d'Informació Científica) at Barcelona. In Spain, there is a federated Tier-2 [4] made up of IFIC (Instituto de Física Corpuscular de Valencia), IFAE (Instituto de Física de Altas Energías de Barcelona) and UAM (Universidad Autónoma de Madrid). IFIC represents 50% of the ATLAS Spanish resources and it has the responsibility to coordinate the activities of the Spanish Tier-2 federation.

Typically, a physics analysis has two parts. In the first stage, physicists run their analysis on a huge number of collision events. These events are often stored in several datasets that are spread at different sites. At this step, the distributed computing and data management tools, based on GRID technologies, are used exhaustively. The outputs of this first step are the so-called D3PD that are ROOT [5] ntuples...
containing only the information relevant for a given analysis. In the second stage, the physicists analyze the D3PDs interactively in order to get the final plots, to refine the analysis, etc.

During the first year of operation the described model has been found insufficient. Some of the tiers were not fulfilling the required level of utilization, especially Tier-2 like the one at IFIC. This situation was mainly due to two issues. On one hand, Tier-2 usability was limited due to the strict dependence on the Tier-1 in order to run jobs. As a consequence, when the Tier-1 was in scheduled downtime its associated Tier-2 were affected. On the other hand, job input files, as well as IFIC job outputs had to be transferred to associated Tier-1. For example, if some data produced at IFIC was needed in a Tier-2 at another cloud, dataset had to be transferred via source and destination associated Tier-1.

The solution to improve this situation was taken by ATLAS in the form of flattening the model from a tier to a mesh [6]. Now Tier-2 can directly exchange data with Tier-1 of different clouds, and even with other Tier-2 globally.

2. ATLAS Tier-3 at IFIC-Valencia

Tier-0 at CERN and Tiers 1 and 2 around the world are well defined within the computing model and have been thoroughly tested in the last years. Tier-3 centres, on the other hand, are institution-level non-ATLAS funded or controlled centres that participate presumably most frequently in support of the particular interests of local physicists (users at the local facility decide how these resources are used). Tier-3 facilities must provide the software tools to access data and to perform local analysis. These local resources can vary widely in size, they must be highly reliable and they must have low latency. Within the ATLAS model such sites are mostly used for interactive or batch analysis of D3PD data sets. It is up to the different institutions to propose possible Tier-3 configurations and software setups that match the requirements of their users’ analysis.

2.1. Design

At IFIC we have the computing scheme which has been adopted by ATLAS for a Tier-3 centre in its minimal version. IFIC Tier-3 [7] is attached to IFIC Tier-2. It currently has around 100 TB (60 TB under ATLAS Distributed Data Management (DDM tool) [8] control plus 40 TB under IFIC control).

An important feature of IFIC Tier-3 is that it uses the same storage system as the Tier-2. Its central component is the Lustre file system [9], a shared file system for clusters. The Lustre file system is available for Linux and provides a POSIX-compliant UNIX file system interface. This interface allows users to access the file system easily. Another important component of Lustre is the meta-directory Server (MDS), a catalogue that, in the case of IFIC, is the only shared resource between Tier-2 and Tier-3. In this case 3 disk servers are dedicated exclusively to Tier-3 to avoid overlap with Tier-2.

PROOF [10] enables interactive analysis of large sets of ROOT files in parallel on clusters of computers or many-core machines. Following a typical analysis workflow, a user first needs to find the input files to be analysed. These files can contain either simulated or real data and have a size around the TB. In the second phase, the user tests the analysis software locally before using the GRID to run on the whole data set. The output from the GRID phase typically contains all the processed information the user decided to be useful for the analysis. In the last phase, Tier-3 resources are used in order to obtain the final results.

The user can follow two different procedures in order to retrieve datasets generated in the distributed analysis phase:

- Request a subscription to the ATLAS DDM system to replicate the dataset in the LOCALGROUPDISK area. This disk space is allocated at the Tier-3 facility of the institute and it is connected to a storage element on the GRID through which the ATLAS DDM can perform the replication. This way the data can later be accessed locally. The ATLAS DDM manages the replication automatically.
- Use the ATLAS DDM client tools to download the dataset to the local disk space.
Three User Interfaces (ui04.ific.uv.es, ui05.ific.uv.es and ui06.ific.uv.es) and one CE are currently available at IFIC. Thanks to this the user can:

- Have the ATLAS software (production releases & DDM tools) installed automatically,
- Perform local checks to develop analysis code before submitting larger jobs via GRID,
- Search data in the GRID and copy them locally,
- Send jobs to the GRID,
- Use Lustre (/lustre/ific.uv.es/grid) as a local file system. Lustre being the high performance file system that is dedicated only to our SE.

GRID tools (Ganga, Panda and DDM clients) are installed in the AFS cell of IFIC. AFS is the common file system in the institute. Therefore, every desktop/laptop can use them.

IFIC uses StoRM as the SRM storage element. StoRM is the interface of IFIC storage services with the GRID and it provides information about the location of the data available at IFIC.

The computational facility must be highly reliable and must have low latency. These properties are particularly important in the last phase of the analysis due to the high frequency of the jobs that are running on the Tier-3 to obtain the final results. The CPU resources are organized in different architectures. This way the Tier-3 adapts better to the various needs the users might have:

- Some home-built user interfaces (UI) are used to perform interactive analysis on the final datasets produced in the distributed analysis phase.
- A local batch farm to provide additional computing power for analysis that need to run on local resources can be accessed via GRID tools.
- A PROOF farm for parallel processing of ROOT analysis jobs.

With these tools being operational, physicists at IFIC have access to the ATLAS software (Athena) and all analysis tools (Ganga, Panda, DDM, ROOT) from their desktop or laptop. Therefore, IFIC ATLAS users can perform analysis on simulated or real data and they can store their results locally.

**Figure 1.** Example of data transfer monitoring using the Cacti tool link.
Scalability tests point at the disk server interface as a possible bottleneck. The interface is closely monitored and the usage is safely below the limit. In the future, channel bonding or a better disk server interface will solve the problem.

2.2. Monitoring

Tools used for monitoring are the same for the entire infrastructure, including Tier-2 and Tier-3 resources, and in general all the scientific computing at IFIC.

To monitor data transfers we are using Cacti [11] as a tool for checking links to data servers (figure 1). We can check Tier-3 servers, which are isolated from others, enhancing performance and not disturbing Tier-2 transfers. Disks in these servers are also grouped in a particular Lustre pool, simplifying the management.

For node monitoring, including computing and storage, we are using Ganglia [12] as can be seen in figure 2.

2.3. Effect of the changes in the ATLAS computing model on IFIC-Valencia Tier-3

The changes in the ATLAS computing model not only affect Tier-2 operations in the infrastructure, Tier-3 operations are indirectly affected as well. In order to satisfy the new Tier-2 quality requirements, the availability and the connectivity of the site need to be constantly evaluated. This permanent monitoring of the infrastructure meant an increase from the original 8h per working day and the inclusion of weekends.
The infrastructure had to be updated in order to fulfill the new requirements, specially the availability and the transfer requirements imposed by the project. Even though Tier-3 infrastructure is out of the experiment pledges, as it is collocated with our Tier-2 and it shares some computing and storage resources, it also benefits from these updates.

On the other hand, the most demanding parts of the infrastructure are as isolated as possible to prevent interferences between them. For example, disk pools are different for the different projects (figure 3).

![Figure 3. Disk pool distribution at IFIC](image)

2.4. Coexistence with other scientific applications running at IFIC-Valencia

This Tier-3 facility is part of the e-science environment of IFIC consisting of two infrastructures targeting scientific and technological applications: the ATLAS Tier-2 in Spain and the GRID-CSIC [13]. The Spanish ATLAS Tier-2 project provides distributed computing and storage resources to generate Monte Carlo events and to analyse real data collected by the experiment. On the other hand, the GRID CSIC provide also distributed computing and storage resources for different scientific applications and with emphasis in multidisciplinary projects [14].

As it has been described, the ATLAS Tier-3 at IFIC is coupled to the Spanish ATLAS Tier-2 and it uses the achievements of the computing and the storage management system to analyse the ATLAS data. Taking into account the definition of e-Science (see ref [15]), this infrastructure holds all the requirements to be considered as a part of the whole IFIC infrastructure.

At IFIC there are more scientific applications running in GRID mode and/or using the resources of the GRID-CSIC infrastructure as it is shown in figure 4: neutrino physics, medical physics (in particular, hadron therapy), medical imaging, lattice QCD calculations, nuclear physics, etc. (see ref. [16]). The more frequent usage from the non-ATLAS physicists is the computing power obtained in a distributed system and the intensive computing of batch jobs but there is room for interactive analysis of derived data coming from the application of algorithms to big datasets. The next step would be to export the ATLAS Tier-3 experience to the other physics groups.
All these research activities, including Tier-3, are being coordinated within the context of group e-Science GRID by sectorial meetings of the various projects and collaborations.

2.5. Conclusions
IFIC analysis facility provides distributed computing and storage resources to different scientific applications. Even though its main e-Science user is the ATLAS community, the experience acquired is being exported to other scientific groups. The Tier-3 at IFIC has a GRID-aware part that shares some of the features of the Tier-2 and some private computing and storage resources. The Tier-3 allows ATLAS users to analyse the data coming from the experiment locally or via GRID and to store the results locally. Users can easily access data stored at IFIC as Lustre is a POSIX-compliant UNIX file system interface. The fact that users can test their programs locally before sending them to the GRID has contributed to improve the overall efficiency. The recent updates needed to fulfil the new requirements of the ATLAS computing model have benefited the whole infrastructure. The fact that the most demanding parts of the infrastructure are as isolated as possible has probed to be very useful because it avoids problems in a given area to spread and affect other parts of the infrastructure.

Acknowledgements
We acknowledge the support of MICINN, Spain (Plan Nacional de Física de Partículas FPA2010-21919-C03-01)

References
[1] ATLAS Collaboration, The ATLAS Experiment at the CERN Large Hadron Collider, JINST 3 (2008) S08003.
[2] R.W.L. Jones and D. Barberis, The Evolution of the ATLAS Computing Model, J. Phys.: Conf. Ser. 219 072037 doi:10.1088/1742-6596/219/7/072037
[3] http://lcg.web.cern.ch/lcg
[4] “Iberian ATLAS Cloud response during the first LHC collisions”, M Villaplana Perez et al; 2011 J. Phys.: Conf. Ser. 331 072068 doi:10.1088/1742-6596/331/7/072068
ATLAS Spanish Tier2 experiences during the STEP09 period“ S. Gonzalez de la Hoz et al. EGEE09 Conference. 21 September, 2009, Barcelona, Spain
[5] ROOT, http://root.cern.ch/drupal
[6] “The ATLAS computing model“, R Jones and D Barberis, 2008, J. Phys.: Conf. Ser. 119 072020, doi:10.1088/1742-6596/119/7/072020
[7] S. Gonzalez de la Hoz et al. “Analysis facility infrastructure (Tier-3) for ATLAS experiment”, Published in Eur.Phys.J.C54:691-697, 2008.
M. Villaplana et al. “First tests with Tier-3 facility for the ATLAS experiment at IFIC (Valencia)”, ISBN 978-84-9745-549-7, pags 212-220.
[8] “Managing ATLAS data on a petabyte-scale with DQ2”, M Branco et al, 2008, J. Phys.: Conf. Ser. 119 062017 doi:10.1088/1742-6596/119/6/062017
[9] Lustre Cluster File System, http://wiki.lustre.org
[10] “PROOF - The Parallel ROOT Facility”, http://root.cern.ch/twiki/bin/view/ROOT/PROOF
[11] CACTI, http://www.cacti.net
[12] http://ganglia.sourceforge.net
[13] http://www.grid.csic.es
[14] “Aplicaciones de Física utilizando la infraestructura de e-Ciencia del IFIC”
S. González de La Hoz, E.Oliver, J. Salt, E. Ros, M. Villaplana, C. Escobar, G. Amorós, A. Fernández, M. Kaci, A. Lamas, V. Sánchez, J Sánchez.
XXXIII Bienal de Física, Santander, 19-23 Sep 2011.
Published at: Resúmenes de las comunicaciones de la XXXIII Bienal de Física, A Vol. I, p. 20-21, ISBN: 978-84-86116-8, 2011-09- 19.
[15] http://www.egi.eu
[16] "Scientific Applications running at IFIC using the GRID Technologies within the e-Science framework"
G. Amoros, J. Ors, A. Fernandez, S. Gonzalez de la Hoz, M. Kaci, A. Lamas, L. March, E. Oliver, J. Salt, J. Sanchez, M. Villaplana, R. Vives y F.J. Albiol. The Third International Conference on Advanced Engineering Computing and Applications in Sciences (ADVCOMP 2009). 11-16 October 2009, Sliema, Malta. ISBN 978-0-7695-3829-7