Journal of Physics: Conference Series **219** (2010) 072046 doi:10.1088/1742-6596/219/7/072046

Readiness of the ATLAS Spanish Federated Tier-2 for the Physics Analysis of the early collision events at the LHC

E Oliver[1], J Nadal[2], J Pardo[3], G Amorós[1], C Borrego[2], M Campos[2], L Del Cano[3], J Del Peso[3], X Espinal[4], F Fassi[1], A Fernández[1], P Fernández[3], S González[1], M Kaci[1], A Lamas[1], L March[3], L Muñoz[3], A Pacheco[2], J Salt[1], J Sánchez[1], M Villaplana[1], R Vives[1].

[1] Instituto de Física Corpuscular (IFIC) (centro mixto CSIC – Univ. Valencia), E-46071 Valencia (Spain)
[2] Institut de Fisica d'Altes Energies (IFAE) Facultat de Ciències UAB, E-08193 Bellaterra (Barcelona, Spain)
[3] Universidad Autónoma de Madrid (UAM) Dpto. de Física Teórica, 28049 Madrid (Spain)
[4] Port d'Informació Científica (PIC) Campus UAB Edifici D E-08193 Bellaterra (Barcelona, Spain)

E-mail: elena.oliver@ific.uv.es

Abstract. In this contribution an evaluation of the readiness parameters for the Spanish ATLAS Federated Tier-2 is presented, regarding the ATLAS data taking which is expected to start by the end of the year 2009. Special attention will be paid to the Physics Analysis from different points of view: Data Management, Simulated events Production and Distributed Analysis Tests. Several use cases of Distributed Analysis in GRID infrastructures and local interactive analysis in non-Grid farms, are provided, in order to evaluate the interoperability between both environments, and to compare the different performances. The prototypes for local computing infrastructures for data analysis are described. Moreover, information about a local analysis facilities, called Tier-3, is given.

1. Introduction

The Large Hadron Collider (LHC) is the new accelerator that replaces the Large Electron-Positron collider (LEP) at the European Centre for Nuclear Research (CERN) in Geneva, Switzerland. In normal working conditions, it accelerates two proton beams of opposite directions to an energy of 7 TeV each, and leads them colliding at four different points corresponding to the localization of the four experiments: ATLAS, CMS, ALICE, and LHCb.

Around 15 Peta-bytes a year of data will be produced. 5000 scientists from 500 research institutes and universities are expected to access and analyze these data [1]. To allow this, a distributed computing model has been developed in order to achieve these challenges: the LHC Computing Grid (LCG). The costs of maintaining and upgrading the resources and services are easily handled in this model, sharing responsibilities between institutes and organizations.

The LCG computing model is built as a hierarchical structure in tiers. For ATLAS, the Tier-0 facility is located at CERN. It archives and distributes the primary RAW data taken from ATLAS, and, after a first processing, transfers the produced ESD (Event Summary Data) data to the 10 Tier-1s sites. There, these data are reprocessed with better calibration and AOD (Analysis Object Data) are produced. Tier-1s archive the ESD data and copies of the AOD data and transfer copies of the AOD data to the Tier-2 facilities. The Tier-2s resources are distributed over more than 30 sites in the world. They mainly provide the simulated events production data for the experiment, and are also deeply involved in the data analysis activities. Each Tier-1 manages few Tier-2s sites that are associated to it, defining then the so-called Cloud. In the Tier-3, which are local-site facilities for data analysis, the AOD data are processed to produce the so-called Derived Physics Data (DPD), a kind of n-tuples, to be directly analysed in ROOT [2]. The Tier-3 facilities own also a storage capacity for end-users analysis.

The Port d'Informació Científica (PIC) in Barcelona, is the ATLAS Tier-1 which defines, with its two associated Tier-2s, the PIC-Cloud (or Iberian Cloud). These two Tier-2s are the Portuguese and the Spanish ones that consist respectively of federations of institutions: LIP-Coimbra and LIP-Lisbon for Portugal, and, IFIC-Valencia, IFAE-Barcelona and UAM-Madrid for Spain. A new additional Portuguese centre is expected to join the federated Portuguese Tier-2 during spring 2009 (LIP-LNEC/FCCN) [3].

In section 2, the federated ATLAS Spanish Tier-2 (T2-ES) is presented. A summary of its contribution to the production of simulated events for the ATLAS experiment is presented in section 3 and its network performances in section 4. The distributed analysis issues are discussed in section 5, while performances of T2-ES is compared with other ATLAS Tier-2s in section 6.

2. The federated ATLAS Spanish Tier-2

T2-ES consists in a federation of three Spanish institutions: the Institut de Física d'Altes Energies de Barcelona (IFAE), the Universidad Autónoma de Madrid (UAM), and the Instituto de Física Corpuscular de Valencia (IFIC), which has the responsibility to coordinate the activities of the T2-ES federation.

2.1. Resources and features of T2-ES

The T2-ES infrastructure is expected to contribute around 5% to the whole ATLAS Computing effort. IFIC has to provide 50% of the total resources of T2-ES, while IFAE and UAM share equally the other 50%. Table 1 shows the distribution of the computing resources inside T2-ES (values at march 2009) as well as the share of these resources between the simulated events production and the data analysis activities. The increase of the computing resources at T2-ES, as expected following the pledges of ATLAS, is shown in Figure 1. As it could be seen the major effort to increase the resources is made during 2007-2008 where the resources grew more than 300%.

Table 1. Resources and features of each site.

	UAM-Madrid	IFAE-Barcelona	IFIC-Valencia
Contribution to T2-ES	25%	25 %	50%
CPU (kSi2k)	276	201	438
Disk Capacity (TB)	147	104	198
Share Simulation/Analysis	60% / 40%	50% / 50%	60% / 40%
Storage System (SE)	dCache	dCache/disk+SRM posix	Lustre+StoRM

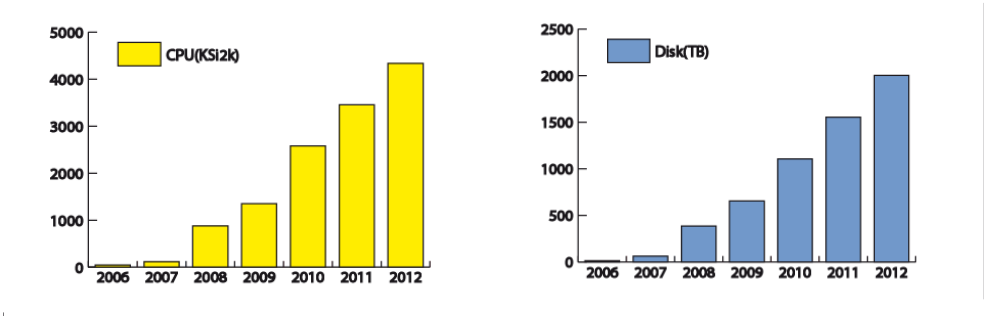

Figure 1. Pledges for CPU (left) and Disk (right) at T2-ES up to 2012.

2.2. Availability and Reliability

The Site Availability Monitoring (SAM) [4] defines the availability and the reliability metrics for the ATLAS sites as follows:

$$availability = \frac{Uptime}{Total\ time - Time\ Status\ was\ unknow}$$

$$reliability = \frac{Uptime}{Total\ time - Scheduled\ Downtime - Time\ Status\ was\ unknow}$$

The SAM runs a range of critical tests at regular intervals of time throughout the day over the sites. These sites are considered to be available and/or reliable when these tests complete successfully. The Management Board considers that reporting both of these metrics is necessary for ATLAS.

Applying these metrics to T2-ES sites, Figure 2 shows that T2-ES has been working with good levels of availability and reliability during the last year.

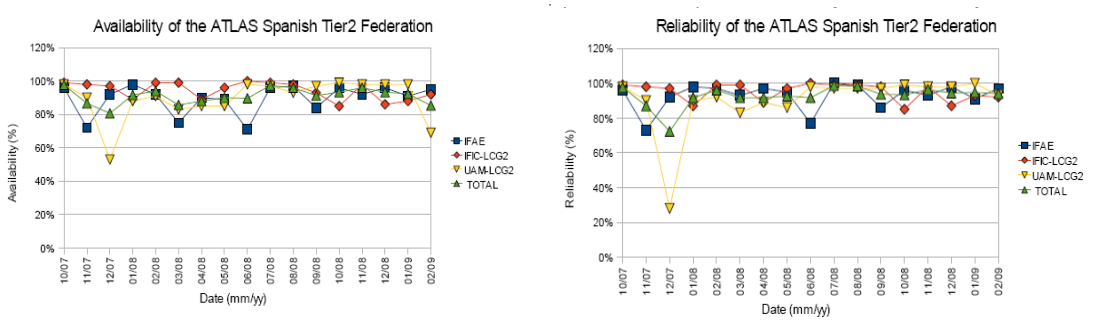

Figure 2. Availability (left) and Reliability (right) from October 2007 to February 2009 for T2-ES.

3. Simulated Events Production for ATLAS

The ATLAS Production System manages the official continuous production of simulated events using the Monte Carlo (MC) method [3]. In this system the physics groups define the jobs that have to be run on the GRID. These jobs are then placed in the central production database (DB) and the supervisor (Bamboo) drives this information to the production engine (Production And Distributed Analysis System: PANDA). Bamboo assigns the jobs to the clouds and updates the job status in DB.

PANDA submits the jobs with a python client interface where users define job characteristics like job settings, input and output files, etc… Job specifications go to the PANDA server via a secure HTTP connection, and the submission information returns to the client. The security of the connection is established by using a GRID certificate proxy for the authentication. A brokering module assigns work on the basis of job type, priority, input data and its locality, available CPU resources and other brokerage criteria. The jobs go to the site where the input files are located because jobs are not run until their input data arrives at the site.

An independent subsystem manages the delivery of pilot jobs to worker nodes via a number of scheduling systems (pilot job factories). A pilot launches on a worker node and contacts the dispatcher. Then, the pilot receives an available job appropriate to the site. If the job is not appropriate, the pilot could exit or pause and ask again later. The distributed data management (DDM) system stores the produced data on the adequate storage resources at different sites and registers them into the defined catalogs.

The simulated events production activity at T2-ES covering the period of time from January 2006 to February 2009 is shown in Figures 3 and 4 through the number of production jobs processed and their corresponding spent walltimes. The walltime is the time that the production job takes to run.

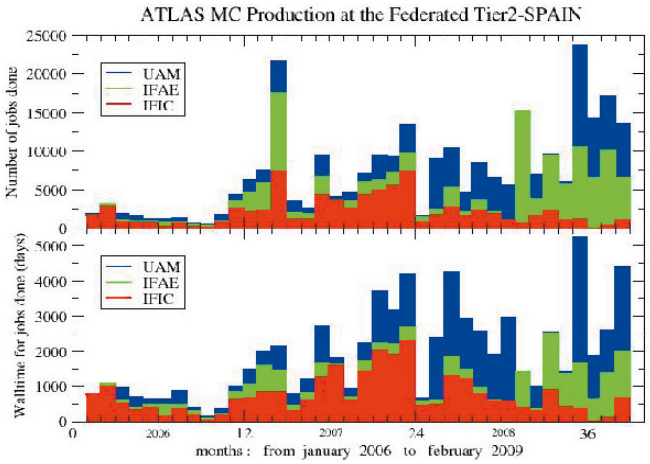

Figure 3. Distribution of the number of jobs run successfully and their corresponding walltimes spent at the sites of T2-ES (IFIC, IFAE and UAM) for the simulated events production during the period January 2006 to February 2009.

Note the trend of the increase of the number of jobs run at T2-ES over the years, due to the increases of the jobs in all ATLAS. A similar trend is observed for the measured walltimes. However, a little decrease in the contribution of T2-ES to ATLAS is observed for the year 2008, which coincide with the migration of the executor to PANDA and the introduction of the pilot jobs.

The monthly contributions of T2-ES to the whole production in WLCG sites (all sites contributing to ATLAS but excluding the USA and NorduGrid sites), in terms of jobs run and walltimes, for the simulated events production activity, are shown in the upper part of Figure 4, as well as the annual mean contribution ones. The global mean contribution of T2-ES to ATLAS is around 1.5 %. It is worth noting that there are more than 30 Tier-2 worldwide sites and 10 Tier-1 sites contributing to this effort of simulation.

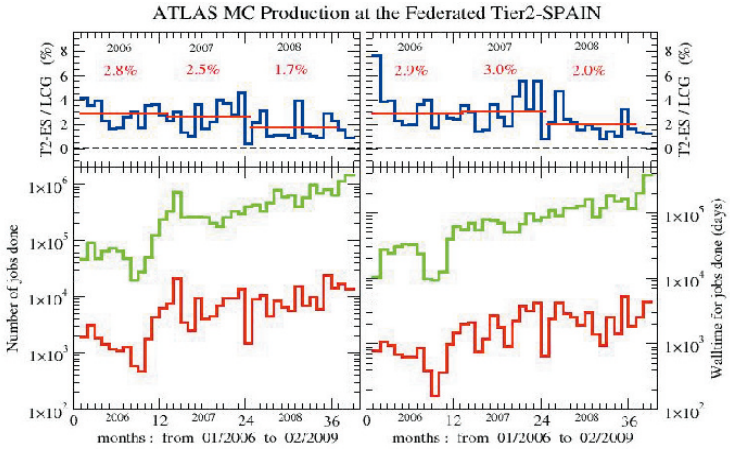

Figure 4. Contribution of T2-ES (red) in number of jobs and walltimes to the sum of the LCG (green) sites (all ATLAS sites but the USA and NorduGrid sites) during the period January 2006 to February 2009.

4. Network Performance and Data Efficiency

In T2-ES, the network connection is provided by the Spanish NREN RedIRIS at 10Gbps among the three sites. The ATLAS link requirement between Tier-1s and Tier-2s has to be 50 MBytes/s (400 Mbps) in a real data taking scenario.

4.1. Distributed Data Management exercises.

The Common Computing Readiness Challenge (CCRC'08) was a stress test designed to bring together all the four experiments of the LHC and to exercise the whole computing chain from data acquisition through to data analysis at the Tier-2 sites [5]. Figure 5 (up) represents the first exercise of data transfer on February 2008 from the Tier-1 site (PIC) to T2-ES [6]. The data transfers is done using the gridftp protocol. A maximum of 250 Mbps transfer rate was reached during these tests.

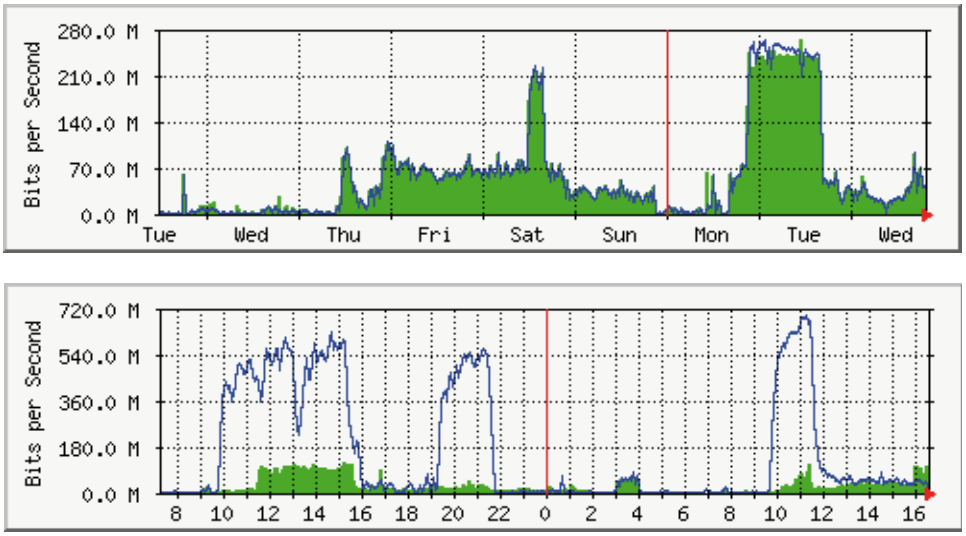

Figure 5.: transfer rates reached during the CCRC'08 for data transfer from PIC to T2-ES (up); and transfer rates between internal pool disks servers at IFIC (down).

On the other hand, IFIC site carried out a local file transfer test on March 4, 2008 [6] to look for potential bottlenecks in the local infrastructure for a real scenario. The transfer was done between internal pool disk servers. 720Mbps were reached improving the ATLAS requirement link of 400 Mbps between Tier-1 and Tier-2, as we can see in Figure 5 (down). It implies that IFIC has enough capacity to process the data transfer to its site.

4.2. The Data Transfer Efficiency

As would be expected for such a highly complex computing system as Grid, not all the data transfers between sites were done successfully. Some of them failed because of timeout connection or site problem. Figure 6 shows the efficiency measured for data transfers of datasets from PIC-Tier-1 to T2-ES, during the year 2008. Real improvement of data transfer is observed for the last months of 2008, where the efficiency reached around 90%.

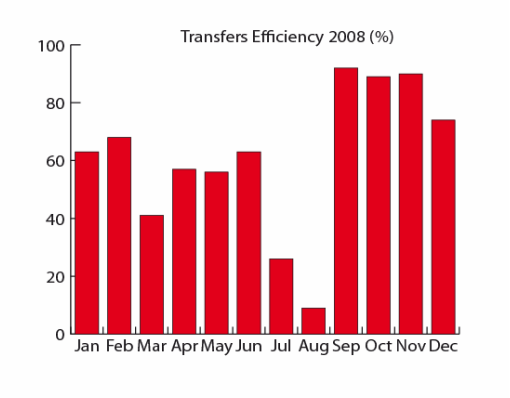

Figure 6. Efficiencies measured for the data transfer operation between the PIC-Tier-1 and T2-ES, during the year 2008 [7].

5. Analysis Status of Spanish Tier-2

The physicists community has to be able to access the data and run analysis jobs in the best conditions. Therefore, the monitoring of the analysis activities helps solving the user problems and to develop the Tier-3 (see section 5.3), which is adapted for the users' needs.

5.1. Efficiency of Simulation Jobs vs Analysis Jobs

The study of the status of the simulation jobs is very useful to value the operation of the Tier-2 (see the section 3). Nevertheless, this is not the case for the analysis jobs because the flow of the jobs is very variable and some errors are only caused by the users. However, the comparison between the efficiencies of the simulation jobs and those of the analysis jobs at T2-ES shows similar performance as can be seen in Figure 7. These efficiencies concern the activities during the year 2008. Outstanding that, in December, both efficiencies reached closely the 100%.

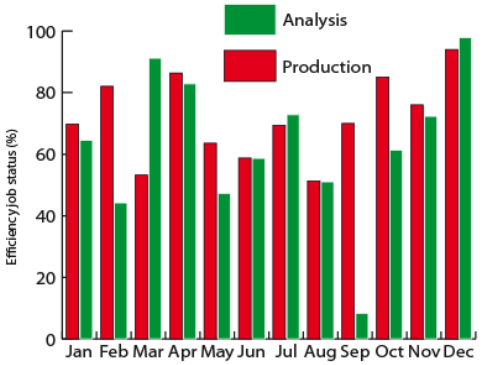

Figure 7. Comparison between the efficiencies measured for the simulation jobs production and the analysis jobs, during the year 2008 [7].

The analysis jobs have been sent using GANGA (a user-friendly job management tool, implemented in Python, and developed for ATLAS and LHCb.[8]).

5.2. Distributed Analysis Stress Test

A GANGA-based stress testing system called "HammerCloud" is being used currently to submit large number of analysis jobs to all the ATLAS facilities. These tests are executed in a regular basis in sites to spot potential problems at the sites.

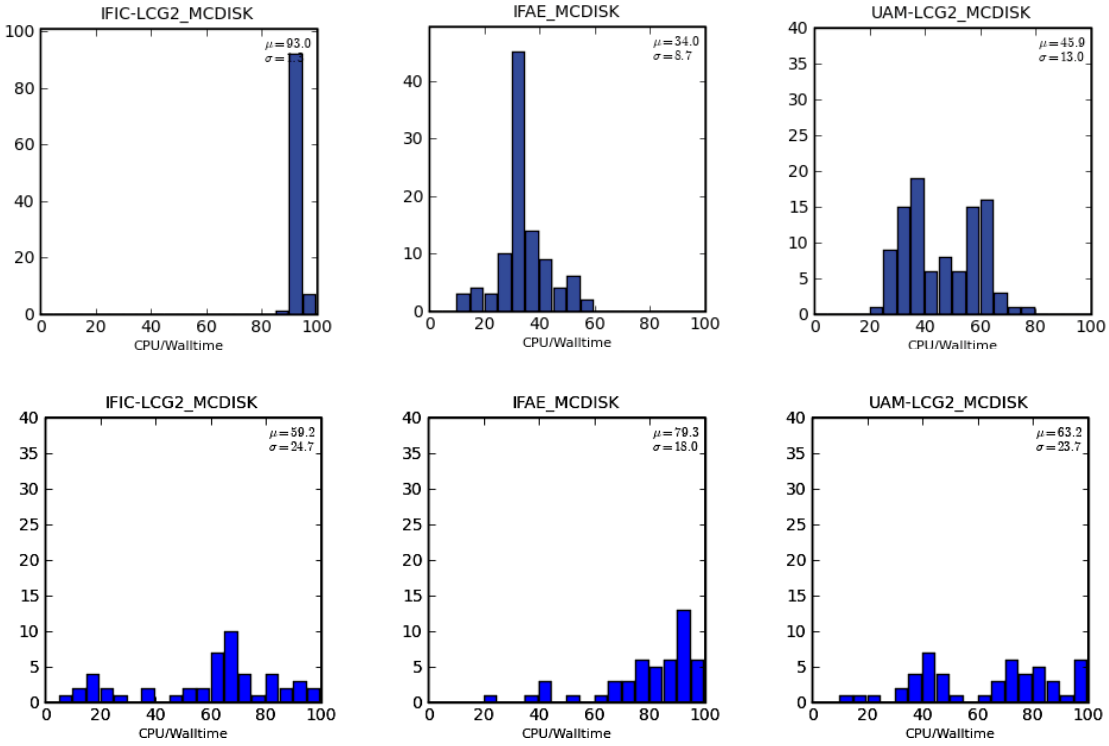

Figure 8. CPU/Walltime ratio without using file stager (up) and with file stager activated (down)

HammerCloud has been used to test the T2-ES sites by running 100 jobs per site. The efficiency of the analysis jobs obtained is around 98-100%. The overall CPU/Walltime ratio is shown in Figure 8. The upper part of the figure shows the result of the test when a file stager is not used. It consists in triggering copies in the background using LCG tools. The lower part of the figure shows similar results with the file stager activated. The performance is shown to be dependent on the used file system. Lustre (at IFIC) works better without using file stager while dCache (at IFAE and UAM) has better behaviour when the file stager is activated [3,9].

5.3. IFIC Analysis Facility Tier-3

A Tier-3 site-located computing infrastructure is needed as Analysis Facility for the Spanish ATLAS end-user physicists. The use of the Analysis Facility Tier-3 could be faster than using directly the Grid for some kind of jobs. However, Tier-2-Tier-3 interaction is necessary in order to access AODs and DPDs with the Distributed Data Management tools (DDM-DQ2).

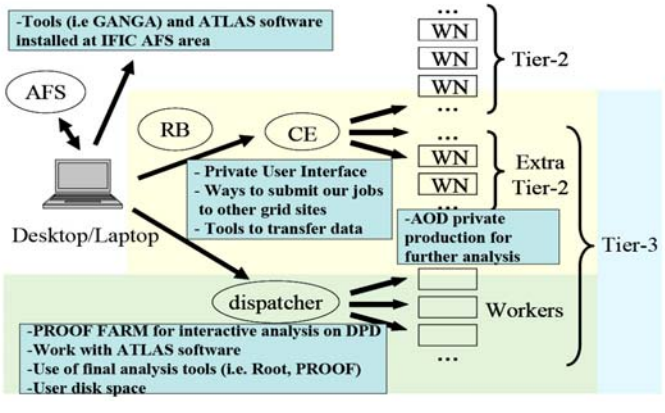

Figure 9. Tier-3 schema.

In a Tier-3 infrastructure [10] physicists from an institute or a university could perform physics analysis on site and have access to the different ATLAS simulation and analysis facilities: tools, data, etc. Actually, according to the ATLAS computing model, users should send analysis jobs to sites where data are available and extract relevant output as n-tuples or similar.

At IFIC the Tier-3 resources are being split into two parts as is shown in Figure 9:

- Some resources are being coupled to IFIC Tier-2 resources in a grid environment. These extra Tier-2 resources will be used preferentially by Tier-3 users. While resources are idle, then they can be used by the ATLAS community.
- A computer farm to perform interactive analysis outside the grid framework.

6. Several ATLAS Tier-2s behaviour

We evaluated four parameters: the reliability, the efficiency for data transfer, the efficiency for production jobs and the efficiency for analysis jobs. In order to evaluate the performance of T2-ES we have compared these parameters with other ATLAS Tier-2s. The reliability, obtained during January 2009 and the efficiencies for the period from November 2008 to February 2009 are shown in Figure 10.

The ATLAS Tier-2 sites used for the comparison with T2-ES are those presented in Table 2. The T2-ES behaviour does not differ appreciably from the others. The main differencies appears at the transfer efficiency and the analysis efficiency, however the last one should be taken carefully since it depens strongly on the specifications and requirements of the users that can be different for several

Tier-2s. The transfer efficiency has also a big dependence on the infrastructure at the site, it seems to favor distributed Tier-2s.

Table 2. : sites used for comparison with T2-ES

Tier-2 (Federation if so)	Number of sites
ScotGrid (UK)	3 sites
Romanian Federation (Romania)	2 sites
INFN ATLAS Federation (Italy)	4 sites
Polish Tier-2 Federation (Poland)	2 sites
ATLAS Federation, Munich (Germany)	2 sites
GRIF, Paris (France), Distributed	6 sites
FZU AS, Prague(Czech Republic)	1 site
Canada-West Federation(Canada)	3 sites
Austrian Tier-2 Federation (Austria)	1 site

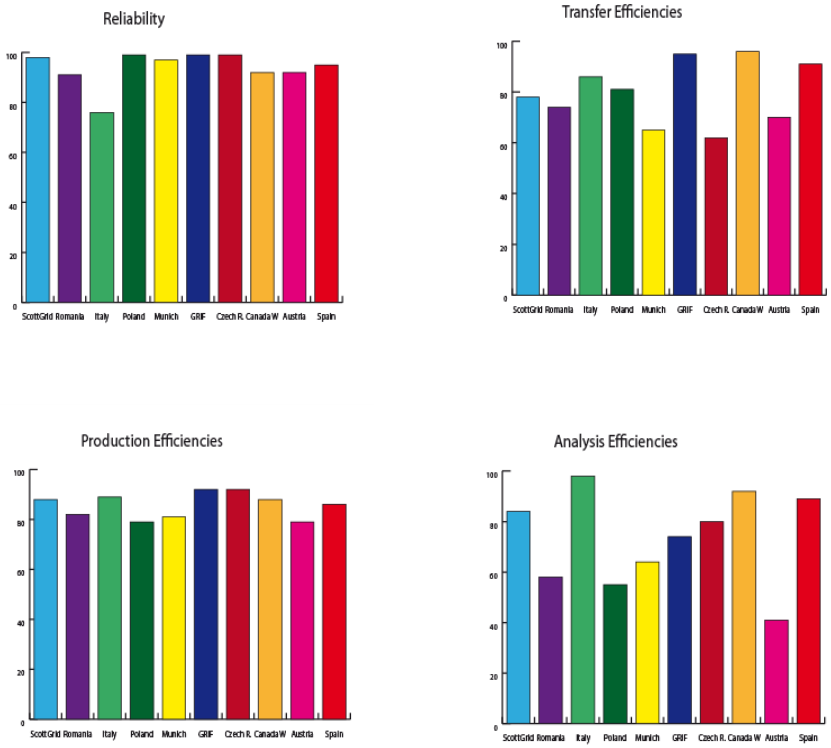

Figure 10. Reliability [4] and efficiencies [7] of T2-ES (last column) compared to the Tier-2s from Table 2 as the same order.

7. Conclusions

The different tests using simulated data have showed that the Spanish ATLAS Tier-2 is ready for data taking: A reliability of T2-ES greater than 90% over several months has been proved. There is a continuous production of ATLAS Simulated Events and a high Data Transfer rate between T2-ES and its associated Tier-1 (PIC) has been achieved. Efficiencies on Data Transfers, Simulated Events Production and Analysis Jobs are similar to other Tier-2s. Computing resources are expected to increase according to the ATLAS schedule.

8. References:

[1] March L *et al* 2008 Experience running a distributed Tier-2 in Spain for the ATLAS experiment *International Conference On Computing In High Energy And Nuclear Physics* (Victoria, Canada, 2-7 Sep 2007) (*J. Phys.: Conf. Ser.* 119 052026)
[2] ROOT webpage: http://root.cern.ch
[3] Espinal X *et al.* 2009. To be presented to the IBERGRID Conference, Valencia. Iberian ATLAS computing: Facing data taking.
[4] WLCG official webpage: http://lcg.web.cern.ch/LCG/
[5] Bird I, Renshall H and Shiers J May 1 2008 *Cern Computer Newsletter* Computing models pass first test of readiness.
[6] Fernández A *et al*. 2008 E-science intrastructure T2-T3 for high energy physics data analysis *Cracow Grid Workshop 2008* (Cracow, Poland, 13.15 October 2008) (*Proceedings of the Cracow Grid Workshop 2008* ISBN 978-83-61433-00-2.) pp 69-77
[7] ATLAS Dashboard: http://dashb-atlas-job.cern.ch/dashboard/request.py
[8] Ganga webpage: http://ganga.web.cern.ch/ganga/
[9] GangaRobot: http://gangarobot.cern.ch/
[10] Gonzalez de la Hoz S *et al.* 2008. *Eur.Phys.J.* C Analysis facility infrastructure (Tier-3) for ATLAS experiment. 54 Heidelberg 691-697

Acknowledgements

We are greatly indebted to the Spanish National Program of High Energy Physics for their support coming from project FPA2007-66708-C03-01.