Computing Activities at the Spanish Tier-1 and Tier-2s for the ATLAS experiment in the LHC Run 3 period and towards High Luminosity (HL-LHC)

Santiago González de la Hoz[1,*], *Vanesa* Acin[3,4], *Esther* Accion[3,4], *Carles* Acosta-Silva[3,4], *Javier* Aparisi[1], *Pablo* Collado Soto[2], *Jose* del Peso[2], *Álvaro* Fernández Casani[1], *Jose* Flix[4,5], *Carlos* García Montoro[1], *Gonzalo* Merino[3,4], *Andreu* Pacheco Pages[3,4], *Elena* Planas[3,4], *Javier* Sánchez[1], *Jose* Salt[1], and *Miguel* Villaplana Perez[1] on behalf of the ATLAS Computing Activity

[1]Institut de Física Corpuscular (IFIC), University of Valencia and CSIC, Valencia, Spain
[2]Departamento de Física Teórica y CIAFF, Universidad Autónoma de Madrid, Spain
[3]Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, Bellatera (Barcelona), Spain
[4]Port d'Informació Científica (PIC), Campus UAB, Bellaterra (Barcelona), Spain
[5]Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (Ciemat), Madrid, Spain

ATL-SOFT-PROC-2023-011
23 August 2023

Abstract. The ATLAS Spanish Tier-1 and Tier-2s have more than 18 years of experience in the deployment and development of LHC computing components and their successful operation. The sites are actively participating in, and in some cases coordinating, R&D computing activities in the LHC Run 3 and developing the computing models needed in the HL-LHC period. In this contribution, we present details on the integration of some components, such as HPC computing resources to execute ATLAS simulation workflows; the development of new techniques to improve efficiency in a cost-effective way; and improvements in Data Organization, Management and Access through storage consolidations, the use of data caches, and improving experiment data catalogues, through contributions such as Event Index. The design and deployment of novel analysis facilities using GPUs together with CPUs and techniques like Machine Learning are also presented. ATLAS Tier-1 and Tier-2 sites in Spain, are, and will be, contributing to significant R&D in computing and evaluating different models for improving performance of computing and data storage capacity in the LHC High Luminosity era.

1 Introduction

In the original computing model of the ATLAS experiment [1] at the LHC at CERN, Tier-1 centers had associated Tier-2 centers, which were close-by in terms of network connectivity, and they formed a cloud, which in ATLAS means a regional setup of one Tier-1 and its associated Tier-2s in a certain geographical area. All data flowed to and from Tier-2s via their associated Tier-1. However, an increasing number of Tier-2 sites gained exceptional global

*e-mail: santiago.gonzalez@ific.uv.es
Copyright 2023 CERN for the benefit of the ATLAS Collaboration. CC-BY-4.0 license.

network connectivity and the ability to directly exchange data among themselves. As a result of these advancements, the ATLAS computing model underwent a transformation to a Mesh model, which includes Nucleus sites and Satellite sites [2]. Tier-2s with a significant storage and excellent network connections are designated as "Nucleus" sites, passing job production on to smaller Tier-2s (Satellites) in any cloud, exchanging data directly. Currently, all the Spanish ATLAS Tier-2 sites are of the Nucleus type.

2 Spanish ATLAS Tier-1 and Tier-2s

The Spanish cloud (ES) [3] inside ATLAS has one Tier-1 at PIC in Barcelona providing 4% of Tier-1s data processing of CERN's LHC experiments ATLAS, CMS and LHCb. Also, it consists of one federated Spanish Tier-2 (ES-ATLAS-T2) at IFIC in Valencia, at IFAE (co-located in PIC) in Barcelona, and at UAM in Madrid with 60%, 25% and 15% of the resources respectively. ES-ATLAS-T2 represents 3% of the total Tier-2s resources and together with the Tier-1, they are integrated in the World Wide LHC Computing Grid (WLCG) project [4] and strictly follow the ATLAS computing model.

In Table 1 the Spanish resources in April 2023, together with their availability and reliability, are shown. The ES is at the top of availability and reliability ranks.

Table 1. Resources and performance of the Spanish cloud facilities for the ATLAS experiment.

Site	CPU (kHEP-SPEC06)	Disk (PB)	TAPE (PB)	Availability (2022)	Reliability (2022)
PIC-Tier-1	57.20	6.20	15.8	100%	100%
IFIC-Valencia	58.60	3.41	-	99.8%	99%
IFAE-Barcelona	17.50	1.60	-	100%	100%
UAM-Madrid	16.57	1.08	-	98%	97%

3 Grid and HPC ATLAS Tier-1 and Tier-2 jobs

The Tier-1 and the ES-ATLAS-T2 have been built and are maintained to operate as ATLAS Grid computing sites. They can execute both data analysis and Monte Carlo event generation for the ATLAS experiment. As "Nucleus" we are stable enough to guarantee data processing and analysis. The performance in the last year at the Spanish cloud in terms of Grid jobs is given in Figure 1. Figure 1 shows that more than 10 million jobs (analysis and production) have been completed by the Spanish cloud in the last year of the Run 3. And more than 15k slots of running jobs have been used on average.

In the Spanish case, the main HPC center in Spain, the Barcelona Supercomputing Center (BSC [5]), included LHC computing in its strategic projects list. BSC's HPC Marenostrum 4 was integrated as a shared resource by the three WLCG centers that provide IT resources to ATLAS, located in Madrid (UAM), Valencia (IFIC), and Barcelona (PIC) [6]. An instance of the ARC-CE [7] has been implemented in each of the three centers. Marenostrum 4 only accepts the SSH protocol for job submission and data transfers and we are running only simulation jobs where the workflow was validated using singularity containers pre-placed at Marenostrum 4's GPFS file sytem. The jobs sent to the ARC-CEs of the LHC resource exploitation centers are transformed into local jobs within Marenostrum 4 by submitting them to the center's Slurm queuing system (see Figure 2). Once the jobs finish, the ARC-CE is in charge of copying the results into the ATLAS computer system, recording them as if the job had been executed on the Grid. It is important to note that the jobs are sent using a

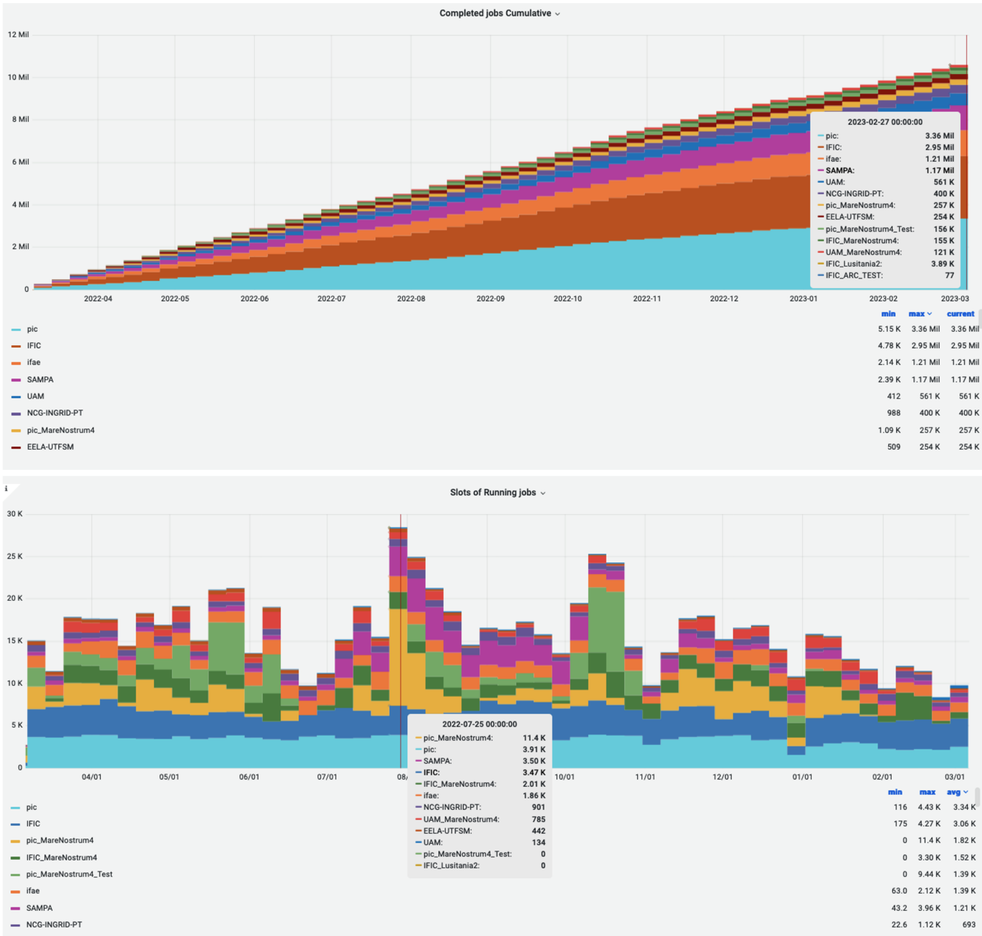

Figure 1. Top: More than 10 million jobs (analysis and production) completed at the ATLAS Spanish sites since the beginning of the Run 3 until April 2023. Bottom: More than 15k slots of running jobs on average used in the Spanish Tier-1 and Tier-2s in 2022.

Singularity image with CentOS7 and that while they are running, they do not make any network connection with the outside world. Unlike other HPC centers, the Marenostrum 4 compute nodes do not have outbound connectivity.

Since 2020 we have an agreement between the Spanish Ministry of Science and the BSC to have around 30 million hours approved at Marenostrum 4 every year by ATLAS through Spanish gateways, which corresponds to 50% of the simulation jobs assigned to Spain. During the beginning of the Run 3 and until March 2023, the Marenostrum 4 HPC provided 30% of the total contribution to ATLAS computing by the Spanish cloud (see Figure 3). Among all the different types of computing, the Marenostrum 4 HPC contributed only to the simulation effort.

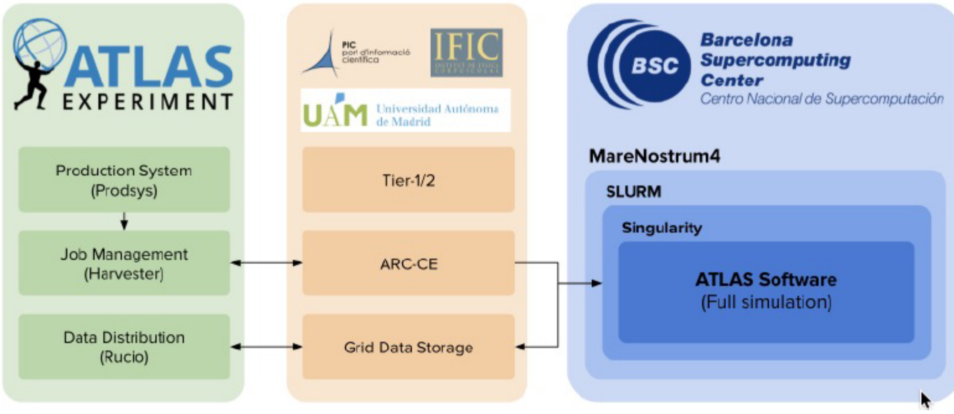

Figure 2. Flow diagram between the ATLAS production system and the execution of jobs on the Marenostrum 4 supercomputer.

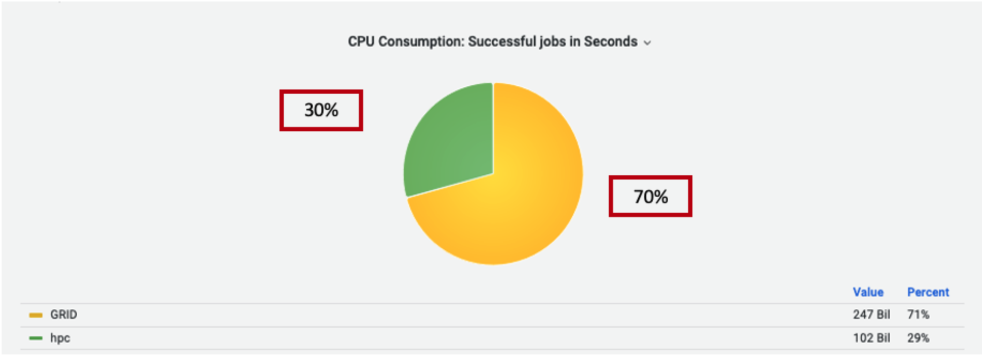

Figure 3. Proportion of HS06(s) provided by Grid resources (yellow) and the Marenostrum 4 HPC (green) of the total contribution to ATLAS computing by the Spanish cloud.

4 ATLAS Event Index

The Event Index project [8, 9], aims to provide a catalog of real and simulated data of all events in all processing stages, which is needed to meet multiple use cases and search criteria. Billions of events (PetaBytes) have been indexed so far since 2015. Some of the use cases are: event picking, duplicated event checking, overlap, trigger check, event skimming, and trigger counter.

The latest developments are aimed st optimizing storage and operational resources in order to accommodate the larger amount of data produced by ATLAS in Run 3 (35 billion new real events each year). This is expected to increase in the future with a prediction of 100 billion events per year at the HL-LHC. At IFIC we have improved the data collection system [10], and the new storage schemas using HBase/Apache Phoenix for the final data backend. This new system has been in operation since last Spring 2022 and performing excellently with promising results.

5 Data Management

In the recent LHC running periods (Run 2 and Run 3) around 8 PB of data have been stored in the Spanish facilities as shown in Figure 4.

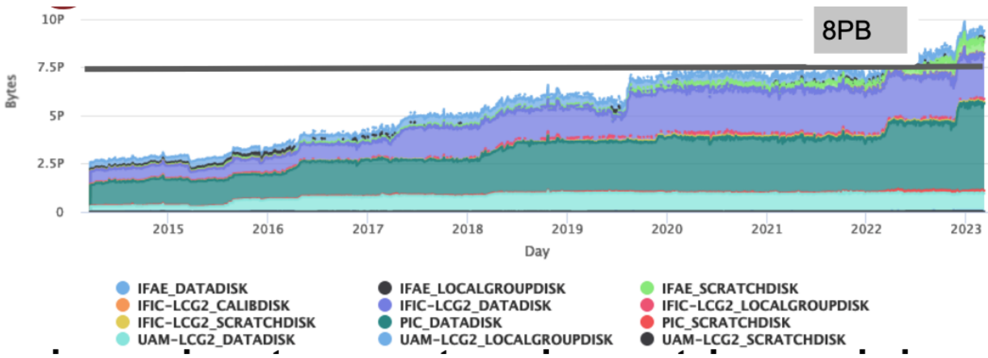

Figure 4. Disk usage in Run 2 and Run 3 in Spanish sites.

To enhance Data organization, Management, and Access (DOMA) in the Spanish cloud, a transition is underway towards a network-centric model known as the Datalake model [11]. This model aims to streamline operations by reducing the number of facilities responsible for storage services, minimizing data replication, and eliminating the requirement for CPUs and storage to be physically co-located. Achieving this objective entails delivering content over a high-speed (minimum 100 Gbps) and reliable WAN, as well as implementing caching mechanisms. By ensuring a robust and efficient network infrastructure, the feasibility of establishing a Datalake is significantly enhanced.

6 Analysis Facilities

In the original Grid computing model, Analysis Facilities, known as Tier-3s, played a crucial role in data analysis. These sites were often located alongside Tier-1s and Tier-2s but dedicated to local groups' private use. Unlike the standardized resources of the WLCG Grid deployment, Tier-3s were not pledged and did not require standard resources.

The primary objective of Analysis Facilities is to assist physicists in reducing the time-to-insight, enabling iterative data exploration. By leveraging these facilities, physicists can significantly accelerate the data analysis process, allowing for the efficient handling of larger volumes of data without proportional increases in time requirements. This significant reduction in waiting time will greatly enhance the productivity and competitiveness of our physics communities.

The components of the Analysis Facilities to be implemented should incorporate several key features:

1. The facilities should enable local access to reduced data samples (e.g., DAOD_PHYSLITE) with minimal latency from the compute infrastructure.

2. Adequate processing resources, including CPUs, GPUs, and memory, should be available, prioritized, and capable of handling the required workload.

3. The infrastructure should be designed for efficiency, employing techniques like scheduling based on HTCondor [12]. It should also be elastic, allowing for dynamic expansion to HPC/Cloud resources during peak loads.

4. The facilities should provide enhanced support for commonly used software tools such as the ROOT and the Python ecosystem, ensuring seamless integration and optimal performance.

5. They should actively to encourage and facilitate the use of shared repositories for analysis routines and workflows to foster collaboration and code reuse.

6. A knowledgeable data/code manager should play a crucial role as a liaison, facilitating access to the technology and assisting users in effectively utilizing the facilities.

The Analysis Facilities being deployed at the Spanish sites typically comprise dedicated storage resources with capacities on the order of several hundred terabytes. For processing, CPU resources are utilized interactively or through a batch system, predominantly HTCondor. Software delivery is primarily facilitated via CVMFS [13], with a growing presence of container-based solutions. Although GPU resources are available, they are not always exclusively dedicated to the ATLAS teams. The network infrastructure consists of a LAN with multiple 10 Gbps capacity, supporting high-speed data throughput between the facility's processing and storage nodes. Additionally, there is a wide-area network (WAN) connectivity of 100 Gbps, ensuring seamless access to the WLCG dedicated network for data lake operations.

At the PIC in Barcelona, a deployment of Jupyter Notebook instances [14] have been implemented. These instances can be created through a dedicated portal, providing the flexibility to select different resource profiles (CPU, memory, GPU) for resource allocation at the PIC farm. On the other hand, at IFIC in Valencia, a GPU infrastructure named ARTEMISA [15] is actively utilized for computational tasks and to take profit of these advantages for analysis.

7 Conclusions and Perspectives

The ATLAS Spanish Tier-1 and Tier-2s have more than 18 years of experience in the deployment of LHC computing components and their successful operation. The sites are actively participating in the evolution of the computing models through the integration and update of ingredients/tools provided by ATLAS. The HL-LHC provides unprecedented opportunities for particle physics, yet its implementation poses technical and logistical challenges. Therefore an essential plan for upgrading computing infrastructure and optimizing data analysis methods is necessary for fully realizing the HL-LHC's potential and overcoming its hurdles.

This work was partially supported by MICINN in Spain under grants PID2019-110942RB-C22, PID2019-104301RB-C21 and PID2019-104301RB-C22, and by the GenT program of Generalitat Valenciana under grant CIDEGENT/2019/027. The authors thankfully acknowledge the computer resources at Marenostrum, and the technical support provided by BSC (FI-2022-1-0002, FI-2022-1-0037, FI-2022-1-0007, FI-2022-2-0001, FI-2022-2-0008, FI.2022-2-006, FI-2022-3-0001, FI-2022-3-0032, FI-2022-3-002, FI-2023-1-0001,FI-2023-1-0017, and FI-2023-1-0004).

References

[1] ATLAS Collaboration, *ATLAS Experiment at the CERN Large Hadron Collider, JINST* **3**, S08003 (2008).

[2] J. Elmsheuser et al., *Overview of the ATLAS distributed computing system, EPJ Web Conf.* **214**, (2019).

[3] S. Gonzalez de la Hoz et al., *Computing activities at the Spanish Tier-1 and Tier-2s for the ATLAS experiment towards the LHC Run 3 period, EPJ Web Conf.* **245**, (2020).

[4] Worldwide LHC Computing Grid project: http://wlcg.web.cern.ch

[5] Spanish Supercomputing Network: http://www.res.es/en

[6] C. Acosta-Silva et al., *Exploitation of the MareNostrum 4 HPC using ARC-CE, EPJ Web Conf.* **251**, (2021).

[7] M. Ellert et al., *Advanced Resource Connector middleware for lightweight computational Grids, Future Generation Computer System* **23**, 219-240. doi:10.1016/j.future.2006.05.008 (2007).

[8] E. Gallas et al., *Deployment and Operation of the ATLAS Event Index for LHC Run 3,* these proceedings (2023).

[9] D. Barberis et al., *The ATLAS EventIndex: A BigData Catalogue for ALL ATLAS Experiment Events, Comput.Softw.Big Sci.* **7**, (2023).

[10] C. Garcia Montoro et al., *HBase/Phoenix-based Data Collection and Storage for the ATLAS EventIndex,* these proceedings (2023).

[11] I. Bird et al., *Architecture and prototype of a WLCG data lake for HL-LHC, EPJ Web Conf.* **214**, (2019).

[12] D. Thain, et al., *Distributed Computing in Practice: The Condor Experience Concurrency and computation: practice and experience*, **17(2-4)** 323-356. DOI: https://doi.org/10.1002/cpe.938 (2004).

[13] J. Blomer et al., *CernVM-FS: delivering scientific software to globally distributed computing resources,* Proceedings of the first international workshop on Network-aware data management.

[14] Jupyter: https://jupyter.org

[15] ARTEMISA infrastructure: http://artemisa.ific.uv.es