

# Common Accounting System for Monitoring the ATLAS Distributed Computing Resources

**E Karavakis<sup>1</sup>, J Andreeva<sup>1</sup>, S Campana<sup>1</sup>, S Gayazov<sup>2</sup>, S Jezequel<sup>3</sup>, P Saiz<sup>1</sup>, L Sargsyan<sup>4</sup>, J Schovancova<sup>5</sup> and I Ueda<sup>6</sup> for the ATLAS Collaboration**

<sup>1</sup>CERN, European Organization for Nuclear Research, Switzerland

<sup>2</sup>Budker Institute of Nuclear Physics, Russia

<sup>3</sup>Centre National de la Recherche Scientifique, France

<sup>4</sup>A I Alikhanyan National Scientific Laboratory, Armenia

<sup>5</sup>Brookhaven National Laboratory, USA

<sup>6</sup>University of Tokyo, Japan

Email: [Edward.Karavakis@cern.ch](mailto:Edward.Karavakis@cern.ch)

**Abstract.** This paper covers in detail a variety of accounting tools used to monitor the utilisation of the available computational and storage resources within the ATLAS Distributed Computing during the first three years of Large Hadron Collider data taking. The Experiment Dashboard provides a set of common accounting tools that combine monitoring information originating from many different information sources; either generic or ATLAS specific. This set of tools provides quality and scalable solutions that are flexible enough to support the constantly evolving requirements of the ATLAS user community.

## 1. Introduction

The ATLAS experiment at the Large Hadron Collider (LHC) [1] has been collecting collision data since November 2009. The ATLAS data are distributed, processed and analysed at more than 130 grid and cloud sites across the world within the Worldwide LHC Computing Grid (WLCG) [2]. The total throughput of transfers is more than 5 GB/s and data occupies more than 120 PB on disk and tape storage. At any given time, there are more than 100,000 concurrent jobs running and more than a million jobs are submitted on a daily basis by more than 220 users.

The large scale activity of ATLAS Distributed Computing (ADC) increases the level of complexity of the system and, thus, increases the probability of failures or inefficiencies in the involved components. Effective monitoring provides a comprehensive way to identify and address any issues with the infrastructure. It is also a key factor in the effective utilisation of the system.

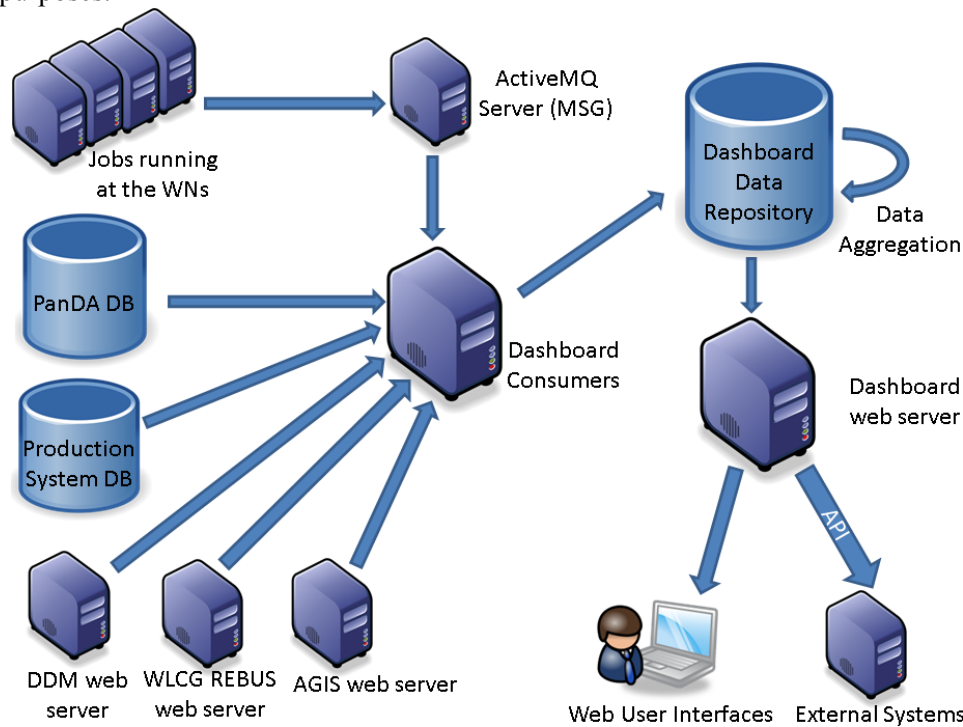
A significant effort has been invested over the last three years within the Experiment Dashboard [3] project to ensure effective and flexible monitoring. The Experiment Dashboard system [4] provides generic solutions that cover all areas of ADC activities [5], such as data distribution and data processing over a large number of sites, and these solutions are extensively used by different categories of ATLAS users ranging from daily operations to resource management.

This paper covers a common accounting system used to monitor the utilisation of the available computational and storage resources of ATLAS. This system provides quality and scalable solutions that are flexible enough to support the constantly evolving requirements of the ATLAS user community.



## 2. Architecture

The monitoring system is loosely coupled to the information sources adding flexibility to the system. The architecture (figure 1) is shared between ATLAS and CMS [6][7] even though they are using different job submission systems and execution back-ends. For the ATLAS job processing activity [8], jobs submitted via Ganga [9] are instrumented to report any job status updates from the Worker Nodes (WNs) to an Apache ActiveMQ Messaging System for the Grid (MSG) [10] server. Ganga is an easy-to-use front-end for Grid job definition and management. A Python agent, the Dashboard Consumer, collects the job status updates and stores them in the Dashboard Data Repository which is implemented in Oracle. Two more Dashboard consumers collect information from the databases of the ATLAS Production System and the PanDA Workload Management System [11] and then import it into the Dashboard Data Repository [12]. There are approximately 10 to 11 million raw records inserted per day for the real-time job status updates, which correspond to 1.1 million job records per day. On top of the real-time data, scheduled database procedures further aggregate them to approximately 65 thousand records for the hourly, daily and 5 minute snapshot statistics for accounting purposes.



**Figure 1.** Architecture of the Experiment Dashboard accounting solutions for ATLAS.

For the accounting information of ATLAS data transfers, the system collects information from ATLAS Distributed Data Management (DDM) [13] that provides daily summary dumps in the form of comma separated value (CSV) files. These daily summary dumps contain approximately 1.6 million rows per day. A Dashboard Consumer performs an initial aggregation on the data before they are inserted into the Dashboard Data Repository. Scheduled database procedures perform the final aggregation of the data to approximately 75 thousand records per day.

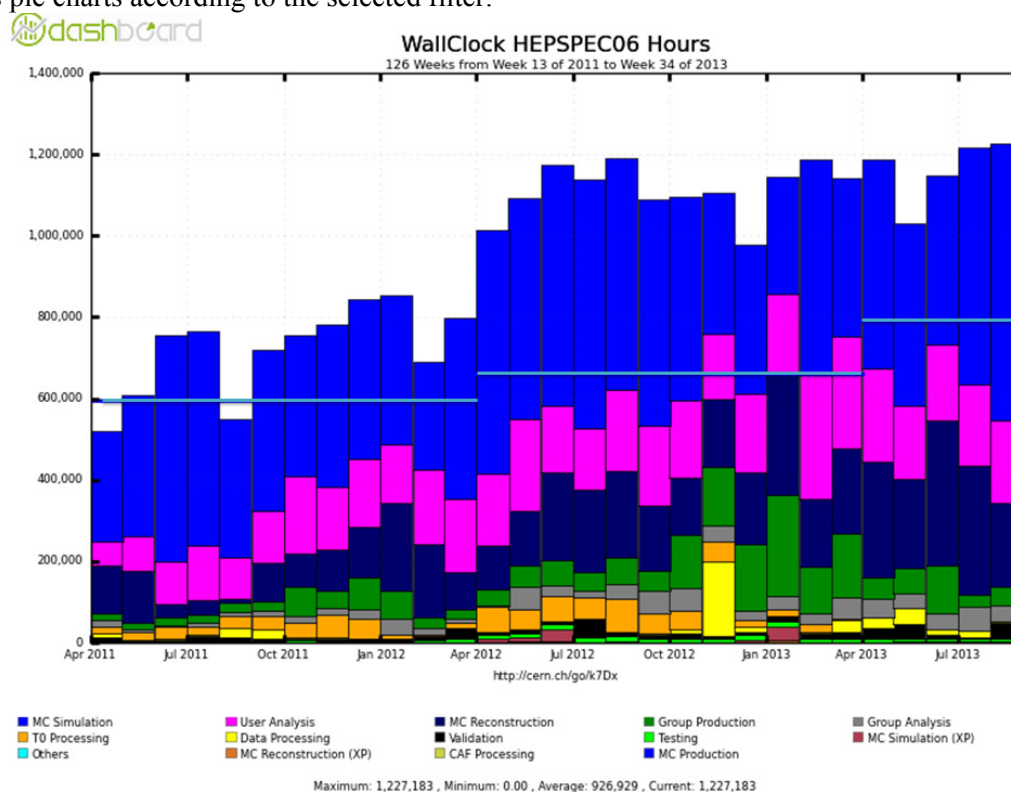
Information regarding pledged CPU / HEPSPEC06 and Storage resources is taken from the WLCG REBUS system [14] and information regarding site topology is taken from the ATLAS Grid Information System (AGIS) [15].

All the collected information is then presented to the user via the Web User Interfaces and the Dashboard web APIs. The same data repository is used by multiple Dashboard applications within the ATLAS Virtual Organisation (VO). Each application is focused on a particular use case and targets a

specific category of users. The Web User Interfaces are decoupled from the data storage technology, thus, it makes no difference if the data back-end is implemented in Oracle, MySQL, any NoSQL solution [16] or even a simple file on the user's hard disk, as long as the information that is being passed to the Web User Interface is in the correct format. The Web User Interfaces for the accounting of the job processing activity are shared between ATLAS and CMS. Although the accounting for the data transfers is quite specific to ATLAS, the Web User Interface is based on the common job accounting interface. The information stored in the Dashboard Data Repository is also consumed by other applications, either internal Dashboard or external third party, in machine-readable format through a set of well-defined APIs.

### 3. Experiment Dashboard Historical Views

The workload accounting is visualised in the Historical Views dashboard. The tool provides views focused on accounting information about completed, submitted, pending and running jobs, CPU consumption and efficiency, processed data, successes/failures, error statistics, activities, and resource utilisation. It provides a set of powerful filtering options: by sites, by physics groups requesting a workload to be processed, by input data types and data projects, by ATLAS cloud (region of sites), by activity and by pledged and unpledged resource type. The filters allow the user to define the time interval and time binning for the plots, and group the results by any of the filters mentioned above. Each of the views contains plots of instantaneous and cumulative value evolution over time, and total values as pie charts according to the selected filter.



**Figure 2.** Utilisation of the ATLAS grid resources between April 2011 and August 2013. The wall-clock time (HEPSPC06 hours) spent on all ATLAS sites over the yearly pledged resources (horizontal light blue lines) is shown.

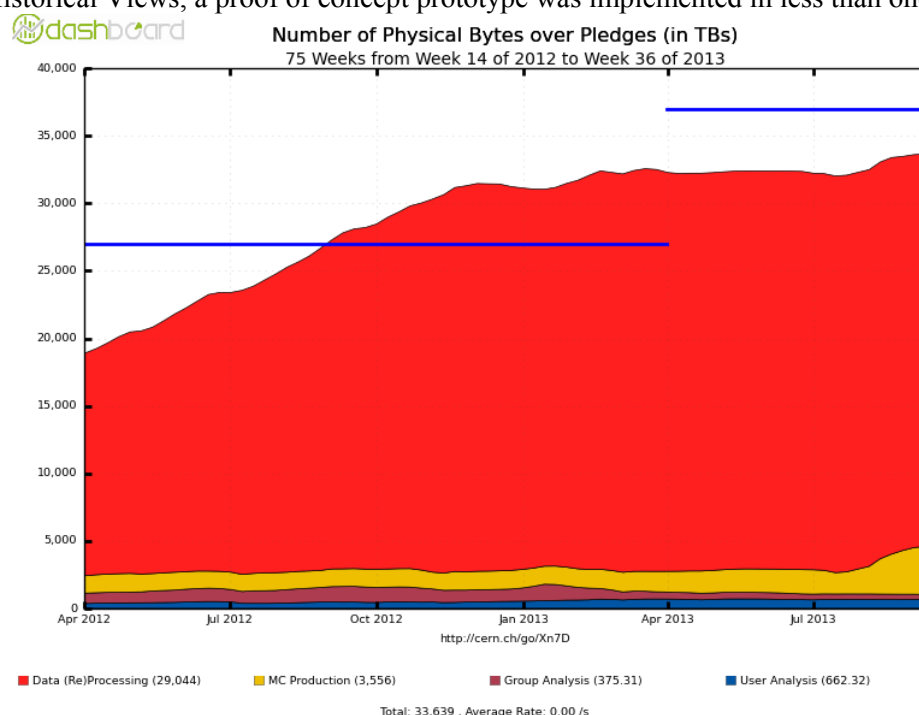
The tool has contributed to the success of the CERN scientific programme as it provides a useful source of information that can be data mined to analyse the performance of ATLAS grid resources, as well as a platform to interconnect information about the computing resources from different sources. An example of such an interconnection of the CPU resources information is shown in figure 2.

Examples of the use of the Experiment Dashboard Historical Views to illustrate the ATLAS job processing activity on the WLCG infrastructure are the inclusion of plots, taken from the tool, in ATLAS presentations at the CERN Higgs seminar and ICHEP2012, a major particle physics conference held in 2012 in Melbourne, Australia [17].

#### 4. Experiment Dashboard DDM Accounting

In order to prevent oversubscription of data to storage endpoints, ATLAS monitors how much space the site has pledged to the VO and how much space remains available at a site, tier or cloud (figure 3). The aim of all data replication policies is to make data available to the ATLAS physicists in a timely manner. Real-time data transfers are monitored with the DDM Dashboard [18].

In order to provide storage accounting reports to the ATLAS Collaboration, a new DDM accounting application to monitor the evolution of datasets, files and bytes over time was developed. The application provides advanced statistics either in historical or daily snapshot views and it offers wide flexibility to the ATLAS users. The development of the new application started in mid-May 2012 and thanks to the fact that it shares its implementation, to a large extent, with the Experiment Dashboard Historical Views, a proof of concept prototype was implemented in less than one month.



**Figure 3.** Utilisation of the pledged storage resources at CERN between April 2012 and August 2013 showing the physical bytes over the total site pledges (horizontal blue lines) in TBs.

#### 5. Conclusions

The ATLAS experiment has been collecting collision data since November 2009. The ATLAS data are distributed, processed and analysed at more than 130 grid and cloud sites across the world. Such a complex system must be continuously monitored. The common accounting tools provided by the Experiment Dashboard help to identify and address possible issues as well as mine long-term data trends, or monitor and properly utilise available computational resources.

There has been significant progress within the Experiment Dashboard in the development of common accounting applications for monitoring the job processing and data transfer activities from 2010 onwards. Furthermore, despite increased use of the grid, as measured by job submission and data transfer, the accounting tools proved to be stable and responsive. This work is very important, since it contributes to the overall success of the ATLAS offline computing effort. These applications were

developed in very close collaboration with the ATLAS computing team and as a result, they correspond well to the constantly evolving needs of the ATLAS high-energy physics community.

## References

- [1] Evans L and Bryant P, “*LHC Machine*”, 2008, *JINST* **3** S08001 doi:10.1088/1748-0221/3/08/S08001
- [2] Bird I, “*Computing for the Large Hadron Collider*”, 2011, *Annual Review of Nuclear and Particle Science*, **61**:99–118 doi:10.1146/annurev-nucl-102010-130059
- [3] Andreeva J et al, “*Experiment Dashboard - a generic, scalable solution for monitoring of the LHC computing activities, distributed sites and services*”, 2012 *J. Phys.: Conf. Ser.* **396** 032093 doi:10.1088/1742-6596/396/3/032093
- [4] Andreeva J et al, “*Experiment Dashboard for monitoring computing activities of the LHC virtual organizations*”, 2010, *J. Grid Comput.* **8** 323-339 doi:10.1007/s10723-010-9148-x
- [5] Schovancova J et al, “*ATLAS Distributed Computing Monitoring tools after full 2 years of LHC data taking*”, 2012 *J. Phys.: Conf. Ser.* **396** 032095 doi:10.1088/1742-6596/396/3/032095
- [6] Karavakis E et al, “*CMS Dashboard Task Monitoring: a user-centric monitoring view*”, 2010, *J. Phys.: Conf. Ser.* **219** 072038 doi:10.1088/1742-6596/219/7/072038
- [7] Karavakis E et al, “*CMS dashboard for monitoring of the user analysis activities*”, 2012, AIP Conference Proceedings, **1504**, 983-986 doi:10.1063/1.4771861
- [8] Karavakis E et al, “*User-centric monitoring of the analysis and production activities within the ATLAS and CMS Virtual Organisations using the Experiment Dashboard system*”, in proceedings of EGI Community Forum 2012 / EMI Second Technical Conference, 2012, PoS(EGICF12-EMITC2)110
- [9] Elmsheuser J et al, “*Distributed analysis in ATLAS using GANGA*”, 2010, *J. Phys.: Conf. Ser.* **219** 072002 doi:10.1088/1742-6596/219/7/072002
- [10] Cons L and Paladin M, “*The WLCG Messaging Service and its Future*”, 2012 *J. Phys.: Conf. Ser.* **396** 032084 doi:10.1088/1742-6596/396/3/032084
- [11] Maeno T et al, “*Overview of ATLAS PanDA Workload Management*”, 2011 *J. Phys.: Conf. Ser.* **331** 072024 doi:10.1088/1742-6596/331/7/072024
- [12] Andreeva J et al, “*ATLAS job monitoring in the Dashboard Framework*”, 2012 *J. Phys.: Conf. Ser.* **396** 032094 doi:10.1088/1742-6596/396/3/032094
- [13] Branco M et al, “*Managing ATLAS data on a petabyte-scale with DQ2*”, 2008 *J. Phys.: Conf. Ser.* **119** 062017 doi:10.1088/1742-6596/119/6/062017
- [14] WLCG REsource, Balance, USage (REBUS), <http://wlcg-rebus.cern.ch> retrieved 2013-10-21
- [15] A Anisenkov A et al, “*AGIS: The ATLAS Grid Information System*”, 2012 *J. Phys.: Conf. Ser.* **396** 032006 doi:10.1088/1742-6596/396/3/032006
- [16] Karavakis E et al, “*Processing of the WLCG monitoring data using NoSQL*”, to appear in Proceedings of the 20th International Conference on Computing in High Energy and Nuclear Physics 2013
- [17] EGI Document 745-v8, “*D6.9 - Annual Report on the HUC Tools and Services*”, 2013, EGI Document Server, <http://documents.egi.eu/document/745> retrieved 2013-10-21
- [18] Andreeva J et al, “*Designing and developing portable large-scale JavaScript web applications within the Experiment Dashboard framework*”, 2012 *J. Phys.: Conf. Ser.* **396** 052069 doi:10.1088/1742-6596/396/5/052069