Operation of the ATLAS Distributed Computing

Fernando Barreiro Magino¹, David Cameron², Alessandro Di Girolamo³, Andrej Filipcic⁴, Ivan Glushkov^{1,}, Federica Legger⁵, Tadashi Maeno⁶ and Rodney Walker⁵ on behalf of the ATLAS Collaboration***

¹University of Texas at Arlington, USA

²University of Oslo, Norway

³European Laboratory for Particle Physics (CERN), Switzerland

⁴Joseph Stefan Institute, Slovenia

⁵Ludwig-Maximilians-University, Germany

⁶Brookhaven National Laboratory (BNL), USA

Abstract. ATLAS is one of the four experiments collecting data from the proton-proton collisions at the Large Hadron Collider. The offline processing and storage of the data is handled by a custom heterogenous distributed computing system. This paper summarizes some of the challenges and operations-driven solutions introduced in the system.

1 Introduction

The ATLAS distributed computing(ADC) [1], [2] was designed to meet the computational and storage needs of the ATLAS experiment. The experiment has produced over 370 PB of data to date and it is adding new data at a rate of 1.5 PB per week (See figure 1). The data is handled by a central data distribution management system based on the Rucio data management system [3]. It is processed in different workflows centrally and by end users using the Athena framework. The traditional unit of processing work in ATLAS is the job. Every job is generated by the Panda workflow management system [4] from requests and tasks defined by the task definition system, ProdSys2, and dispatched to one of the 180 computing resources, in over 40 countries, according to the specific job requirements and resource properties. The computing resources at ATLAS' disposal consist of WLCG GRID sites [5], commercial and private Clouds, High-Performance Computers (HPC) and volunteer computing resources [6]. ADC currently executes on average 1.1 M jobs per day (figure 2) on 347k (figure 3) computing slots on average, peaking at 960k slots for several hours. The global average transfer throughput of the system is 12 GB/s (figure 4).

2 Operations

The ADC operations is a shared effort between two dedicated teams - DDMOps (Distributed Data Management Operations) and DPA (Distributed Production and Analysis) and a set of dedicated shifters spread over the globe to ensure 24/7 coverage.

*e-mail: Ivan.Glushkov@cern.ch

**Copyright [2018] CERN for the benefit of the [ATLAS Collaboration]. Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license

Figure 1. The Amount of ATLAS data collected by June 2018 is 370 PB.

Figure 2. Number of completed jobs per day.

2.1 DDM Operations

A few of the most important aspects of the DDMOps team responsibilities include but are not limited to:

- Optimal storage utilization - data rebalancing, replication and deletion according to the current data processing needs.
- File transfers - ensure expected bandwidths are achieved.
- Data quality and consistency - tracing, deleting and recovering of corrupted data, local site and central monitoring and reporting consistency checks.

Figure 3. Running job slots contribution per resource type - GRID, HPC, Clouds

- Site storages - central configuration, performance tests.
- Monitoring and reporting.
- Developer communication - reporting and following up on Rucio feature requests.

2.2 Distributed Production and Analysis

The main task of the DPA team is to ensure the efficient utilization of the computing resources in accordance with the current needs of ATLAS. Within the experiment there are a set of computational activities (workflows). Each workflow has appointed production managers who are responsible for following up on the requests for the physics groups and for fulfilling them within the deadlines. The priorities in sharing ATLAS computational resources between the different workflows are decided by the executive bodies of the collaboration and implemented on computing level via the "global shares" [7]. DPA also takes care of system troubleshooting as well as information flow between different stakeholders - the central production via production managers and analysis users via the Distributed Analysis (DA) and the DAST (Distributed Analysis Support Team) shifters on one side and the infrastructure and services providers, on the other. Another focus of the DPA team is the continuous optimization and update of the monitoring, accounting and alarms - all of which are vital for precise diagnostics and troubleshooting of the system. DPA also takes part in further system development and deployment of new services and resources.

Figure 4. Average ATLAS data throughput in the last year.

3 System Optimizations

In order to achieve as high a resource utilization efficiency as possible, a constant system optimization is ongoing. The areas with highest efficiency gain are failure rate minimization, network, disk and tape storage utilization optimization as well as local batch system performance optimization. Two of the latest optimizations implemented will be discussed below.

3.1 I/O Intensity Cut-off

The I/O intensity of a job is defined as the input data size normalized to its wall time. The I/O intensity of a task is set as the average I/O intensity measured from the first 10 finished jobs ("scouts"). Based on this value, and the location of the input data, the job brokerage decides on which site the next set of jobs generated should be dispatched. If the set value is below the cut-off value, the jobs are dispatched to the selected sites. Otherwise, the jobs are run where the input data is. The threshold is chosen to optimize resource usage whilst reducing the amount of file transfer required. In the future, this binary cut-off will be replaced by a weight in the standard Panda brokerage. The default I/O intensity cut-off is 200 kB/s. An increase of the cut-off value increases the number of running slots of high I/O intensity jobs, effecting the overall number of running slots (figure 5) at the expense of increased network utilization (See figure 6). Adjusting the I/O cut-offs properly is a key factor for system performance. Setting values too high results in transfer delays which in turn leads to low slot occupancy.

3.2 Panda Retrial Module

The Panda Retrial Module is a mechanism that can significantly reduce the lost CPU time from jobs which fail due to known software problems. It allows DPA to override the retrying

Figure 5. Effect of variation of the I/O cut-off value to the number of running job slots. I/O cut-off is defining when jobs within a task with a given I/O estimated value would be run - locally or on a remote site.

mechanism based on error exit message. The default action for a failed job is to retry it a number of times, where the number is either manually set by the task submitter at definition time or it is hard coded per workflow in the system. In case it is known that a certain error will appear in a certain percentage of the jobs when running the task, with the retry module one can set an action only for these cases based on the error message. The actions currently supported are:

- Do not retry.
- Increase memory requirement on the next retry
- Limit the number of retries to a new value.
- Increase CPU time on next retry.

A Web UI (figure 7) simplifies the introduction of new retry rules significantly.

4 Conclusion

The distributed computing system of the ATLAS experiment at the LHC is necessarily complex due to the diverse workload requirements and heterogeneous resources available. New workloads must be accommodated at ever higher scales, whilst the underlying resources change beyond our control. All this demands strong operational and shift teams, and a continuous optimization procedure. We have described two recent examples and expect further such incremental improvements to be needed as we approach Run 3 and beyond.

Figure 6. Effect of I/O intensity variations to the production input transfer throughput.

Figure 7. Web UI snapshot of Panda Retrieval Module Web UI.

References

[1] J. Elmsheuser, A. Di Girolamo (on behalf of ATLAS Collaboration), *Overview of the ATLAS distributed computing system*, in *Proceedings of this conference (CHEP 2018)* (2019), <https://cds.cern.ch/record/2625159>

[2] ATLAS Collaboration, JINST **3**, S08003 (2008)

[3] M. Barisits, T. Beermann, V. Garonne, T. Javurek, M. Lassnig, C. Serfon, Journal of Physics: Conference Series **1085**, 032030 (2018)

[4] F H Barreiro Megino and K De and A Klimentov and T Maeno and P Nilsson and D Oleynik and S Padolski and S Panitkin and T Wenaus and, Journal of Physics: Conference Series **898**, 052002 (2017)

[5] *Worldwide LHC Computing GRID*, <http://wlcg-public.web.cern.ch/about>, accessed: 2018-09-21

[6] D. Cameron, W. Wu, A. Bogdanchikov, R.M. Bianchi (on behalf of ATLAS Collaboration), *Advances in ATLAS@Home towards a major ATLAS computing resource*, in

Proceedings of this conference (CHEP 2018) (2019), <https://cds.cern.ch/record/2626282>

- [7] F.H. Barreiro Megino (on behalf of ATLAS Collaboration), *ATLAS Global Shares Implementation in the PanDA Workload Management System*, in *Proceedings of this conference (CHEP 2018)* (2019), <https://cds.cern.ch/record/2626911>