The ATLAS DDM accounting and Storage Usage Service

To cite this article: Fernando H Barreiro Megino et al 2010 J. Phys.: Conf. Ser. 219 072026

View the article online for updates and enhancements.

Related content
- Data management tools and operational procedures in ATLAS - Example of the German cloud
  Cédric Serfon
- Data Storage Accounting and Verification at LHC experiments
  C-H Huang, E Lanciotti, N Magini et al.
- Experience commissioning the ATLAS distributed data management system on top of the WLCG service
  S Campana and Atlas Distributed Computing
The ATLAS DDM Accounting and Storage Usage Service

Fernando H. Barreiro Megino, Vincent Garonne, Stephane Jezequel, Miguel de Oliveira Branco, Mario Lassnig

E-mail: fernando.harald.barreiro.megino@cern.ch, vincent.garonne@cern.ch, jezequel@lapp.in2p3.fr, miguel.branco@cern.ch, mario.lassnig@cern.ch

Abstract. The ATLAS Distributed Data Management system is the system developed and used by ATLAS for handling large amounts of data. It encompasses data bookkeeping, managing of largescale production transfers as well as endusers data access requests. The multi-petabyte ATLAS data volume already under management requires an accounting and monitoring service that collects different data usage informations in order to show and compare it from the experiment and application perspective. In this paper we will describe the design and implementation of the DDM Accounting and Storage Usage Service, built to meet the monitoring requirement.

1. Introduction

The ATLAS Distributed Data Management (DDM) (also known as Don Quijote 2 (DQ2) [1]), manages over 10 petabytes of data, corresponding to 120 million files and 1.3 million datasets (see figure 1) that are distributed over more than 500 Grid storage endpoints.

Given the complexity of the system, it is necessary to develop a monitoring tool, capable of showing the accounting and storage space information from different perspectives and, where possible, associate available metadata (for example information required by physics groups).

Figure 1. Plots showing the evolution of the amount of data managed by DDM between the 27 March and 25 April 2009.
As the information available in the DDM Central Catalogues can be incorrect, it is vital to compare the DDM registered data volume with the real data volume on the sites that can be obtained through the sites’ SRM interface.

2. System architecture
The system consists of two agents that collect the SRM and DDM information on a daily basis and store them in the Accounting Database. The system’s front-end is a web interface, capable of dynamically generating the plots required in each view (see figure 2).

![Diagram of system architecture]

Figure 2. Plots showing the evolution of the amount of data managed by DDM between the 27 March and 25 April 2009.

---

1 SRM Storage Resource Manager- is a common protocol for managing the different storage elements independently of their implementation technology (i.e. HPSS, CASTOR, dCache, DPM, StoRM).
2.1. DDM agent
The DDM agent is in charge of querying the DDM Central Catalogues daily to discover the files and datasets registered in each of the Grid sites. This information is then stored in the specific Accounting Database. With the information collected by this agent, the used storage space at each site can be calculated. However, given that the information in the DDM Central Catalogues is not always correct (reasons explained in section 4), it needs to be compared with the information available directly from the SRM.

2.2. SRM agent
The SRM information is queried by this agent, that obtains the used and total storage space through the sites’ SRM interface. The SRM Agent follows these steps:

(i) First, the agent updates the list of ATLAS associated sites by reading out the central information system.

(ii) The more than 500 Grid storage endpoints are then queried using the lcg-stmd command. To reduce running time, endpoints are queried in parallel.

(iii) Finally, the information is stored in the Accounting Database.

As well as populating the Accounting Database, the SRM agent is also used for other systems:

- the Service Level Status (SLS) monitoring pages [2]
- the PanDA production and workload management system [3]
- based on the SRM information from each site, the system also sends out emails to the ATLAS expert on call and cloud administrators, alerting them about sites that are nearly full. The conditions to send out an alarm are based on the maximum amount of data transfer foreseen to the site. In order to avoid repeated alarms about the same site, the agent saves the last alarms that have been sent out for the different sites.

2.3. Accounting Web Interface
The system front-end is a dynamic web interface, developed in Python using the Django web framework, and that uses Matplotlib libraries for generating the plots. To speed up the page generation and avoid unnecessary interactions with the Accounting Database, the

![Figure 3. DDM vs SRM plot for FZK-LCG2_MCDISK.](image)
Figure 4. Cumulative plot showing distribution of occupied storage space by physics groups in FZK-LCG2_MCDISK.

web interface checks if the requested plots already exist on the system before generating them again. Additionally, some plots require the CMS Graphtool extension for Matplotlib [http://t2.unl.edu/documentation/using-graphtool].

The general appearance of the user interface is shown in figure 2. All pages show a common structure, including a header with an extensible menu, where the desired view can be selected. The user can also select the period for the displayed information (last week, last month or all the information available in the Accounting Database).

The information presented in the different views is described next.

3. Views of the system
The system offers different plots and structural overviews in order to facilitate the work of the DDM operations team and the different cloud and site administrators. The following subsections will describe the content in each view.

3.1. Site view
The Site view shows the plots related to a particular spacetoken at a site. This view presents the DDM versus SRM plot, which compares the used space according to DDM accounting and to SRM (see figure 3). This plot is interesting for estimating the consistence of the information in the Central Catalogues. In order to monitor the free space, the plot also includes the total storage space available for that spacetoken. The curve for the total space is skipped in cases where the total value is over ten times the used values, so that the used curves are displayed in more detail.

Other plots in this view show the DDM evolution of the number of files and datasets registered for the particular spacetoken.

At the bottom of this view there are three cumulative plots that show the break down by physics group of the occupied space (figure 4), the number of files and the number of datasets registered for the site.
3.2. Cloud view
In this view, a summary table of a cloud is presented, including downsized plots for all the sites belonging to the cloud. Each plot links to the respective Site view page.

3.3. Tier1 summaries by spacetoken
For operational issues, it is very practical to have a summary view of the most important spacetokens (ATLASDATADISK, ATLASDATATAPE, ATLASMCDISK and ATLASMCTAPE) for all Tier1 sites. In this view, the corresponding DDM versus SRM plots are grouped together and the information is also presented in cumulative plots to enable the comparison of the amount of data registered at each site (see figure 5).

3.4. Tier1 and Tier2 reports
Additionally, on a weekly basis, summary tables and comparative plots are generated for Tier1s and Tier2s. These reports allow, at a glance, an overview of the space status at the different sites. The reports for Tier1s are more complete and include experimental summary plots (e.g. figure 5).

3.5. Predictions
Using the data from the previous seven days, the system tries to predict the behavior of the storage occupancy for the next week and shows the information on a simple Grid plot summary by cloud. The objective of the plot is to try to alert cloud administrators about which sites are becoming full.

3.6. Grid view
Under the Grid view, the user can see the data volume that is registered in DDM over the whole Grid, as well as the corresponding number of files and number of datasets (see figure 1). The information is also broken down by physics group in a cumulative plot.

4. Discussion
The system has been developed in close collaboration between the ATLAS DDM development and operations team. Since its first deployment it has helped to spot different types of problem:
• Dark data: Dark data is data not known by the Central Catalogues. It can be due to bad deletions, where a file has been deleted from the Central Catalogues, but not effectively removed from the Storage Element. As the Central Catalogues are not aware of this data, it will not be accessed and is a waste of storage space. Figure 7 shows sites with dark data:
  – The first image presents the typical symptoms of a site with dark data, as the plot has a constant shift of the red curve (SRM) over the blue curve (DDM).
  – In contrast, the second image shows a site where the site administrator has made the effort to delete dark data and thus made the two curves coincide by 16 October 2008.

• Multiple registration of files in overlapping datasets: Handling of overlapping datasets clearly complicates the system and in previous versions of DDM, it could happen that files belonging to different datasets were registered more than once. This lead to DDM reporting a higher data volume than SRM. The example in figure 8 shows an example of a site having this multiple registration problem and therefore the blue curve (DDM) has a constant shift with respect to the red curve (SRM).

• Publication problems: In some cases, sites are badly configured and do not respond correctly to the SRM commands. This problem has been corrected for most of the sites after dedicated effort from the DDM operations team. The example in figure 9 shows a past example, where the red curve with the SRM used space is always zero, although the site was publishing correctly the SRM total value (which appears in the legend).
5. Conclusions
The system is regularly used by the DDM operations team and many site and cloud administrators. In particular, the German and French cloud follow the DDM Accounting and Storage Usage Service to run regular consistency checks on their sites. Thanks to their constant work and the ongoing improvement of the DDM system, the DDM and SRM values are becoming more and more consistent.

The storage space data gathered for the accounting is not only used for monitoring, but is also used by PanDA for the job brokering and is being integrated into the DDM infrastructure to avoid pushing data to sites with full disks.

Acknowledgments
This work was supported by the Spanish Ministry of Science and Innovation (MICINN).
Figure 9. Publication problems.

References
[1] Branco M, Zaluska E, de Roure D, Salgado P, Garonne V, Lassnig M and Rocha R 2008 Managing Very-Large Distributed Datasets Proceedings of the OTM 2008 Confederated International Conferences. 775 - 92
[2] Di Girolamo A, Campana S, Santinelli R, Sciaba A and Kobliatz B 2009 The Service Level Status monitoring of the LHC Experiments Distributed Computing Proceedings of the 17th International Conference on Computing in High Energy and Nuclear Physics
[3] Nilsson P, Caballero J, De K, Maeno T, Potekhin M, Wenaus T 2008 The PanDA System in the ATLAS Experiment XII Advanced Computing and Analysis Techniques in Physics Research.