PROCEEDINGS OF SCIENCE

ATLAS Distributed Computing Operations in the First Two Years of Data Taking

I. Ueda[1] for the ATLAS collaboration

The University of Tokyo, International Center for Elementary Particle Physics
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
E-mail: i.ueda@cern.ch

The ATLAS experiment has had two years of steady data taking in 2010 and 2011. Data are calibrated, reconstructed, distributed and analysed at over 100 different sites using the World-wide LHC Computing Grid. Following the experience in 2010, the data distribution policies were revised to address scalability issues due to the increase in luminosity and trigger rate in 2011. The structure in the ATLAS computing model has also been revised to optimise the usage of the resources, according to effective transfer rates between sites and site availability. Some new infrastructures were introduced for the software installation at the sites and for database access to reduce the bottlenecks in the data processing. Issues in the end-user analysis were studied and automated control system of the analysis queues based on functional tests has been introduced. The monitoring and accounting tools have been developed and provide views of the ATLAS activities by categories. In this talk, we will report on the operational experience and evolution in the ATLAS Distributed Computing and on the system performance during the first two years of operation.

The International Symposium on Grids and Clouds (ISGC) 2012
Academia Sinica, Taipei, Taiwan
February 26 – March 2, 2012

[1] Speaker

PoS(ISGC 2012)013

1. Introduction

The Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN) has been delivering stable beams colliding at 7 TeV since the first collisions at the end of March in 2010. ATLAS [1, 2], one of the general-purpose experiments at the LHC, has been taking data with a good efficiency, accumulating about 4 PB of data from the detector over the two years.

The ATLAS distributed computing system [3, 4, 5] consists of three classes of "Regional Centres", named as Tier-0, Tier-1 and Tier-2 following the WLCG MoU [6], with different roles and requirements for the pledged resources. The Tier-0 centre is located at CERN and associated with the CERN Analysis Facility. There are 10 Tier-1 and 38 Tier-2 centres over the world hosting ATLAS activities. A "Regional Centre" can be a federation of multiple sites and the number of Tier-2 sites becomes about 70 by counting site-by-site. There are other "ALTAS Grid Centres" without pledged resources, which can be named as "Tier-3 centres", but their roles in the ATLAS distributed computing activities are very much similar to those of the Tier-2 centres and therefore grouped in the description of the Tier-2 in this paper. The number of sites in total is about 130 including all the Tier-0, Tier-1, Tier-2 and other ATLAS Grid centres. In 2011, the pledged resources at the Tier-0 and the CERN Analysis Facility were 75000 HEP-SPEC06 for CPU, 7000 Tbytes for disk and 12200 Tbytes for tape and at the Tier-1 centres in total were 250208 HEP-SPEC06, 26869 Tbytes and 31959 Tbytes respectively. At the Tier-2 centres 281228 HEP-SPEC06 for CPU and 34203 Tbytes for disk were pledged in total and no requirement for tape.

The raw data acquired with the ATLAS detector (RAW) are recorded at a nominal rate of 200 Hz with average event size of 1.6 MB and stored into tape immediately at the Tier-0. The data files are registered to the data management system and sent out promptly from the Tier-0 to the 10 Tier-1 centres via the network, while the files are still on disk in the tape buffer at the Tier-0, and stored on tape at the Tier-1 centres. By having a replica of each file on tape at a Tier-1 in addition to the original at the Tier-0, a long-term protection against a possible data loss is ensured. After the calibration of the data is performed at the CERN Analysis Facility, the first-pass processing of the RAW data with event reconstruction is carried out at the Tier-0 [7, 8], producing Event Summary Data (ESD), Analysis Object Data (AOD) and other types of derived data (dESD, dAOD, NTUP, etc.) that are also distributed over the Grid to Tier-1 and Tier-2 centres for further analysis. The average event size of ESD and AOD are approximately 1 MB and 100 kB, respectively.

In the ATLAS Computing Model [4, 5], the main roles of the Tier-1 centres are to store the replicas of RAW data permanently, to store the reconstruction outputs on disk serving as repositories with faster access, and to perform the second and the further processing of RAW data hosted at the site (i.e. reprocessing). The major role of the Tier-2 centres is to serve as the main facility for end-user analysis and host input data for the analysis jobs on disk. Each Tier-1 centre has a group of Tier-2 centres associated to it. The data distribution over the Grid onto the Tier-1 and the Tier-2 centres are managed in an organised way with this association. The data to

be stored at a Tier-2 centre are delivered via its associated Tier-1 centre. The Tier-2 centres are also the main resources for producing simulated data.

2. Data distribution over the Grid

The data to be distributed on the Grid are registered to the ATLAS distributed data management system (DDM) [9]. Since the first collision at 7 TeV in 2010, ATLAS has registered more than 10 PB of data in DDM (figure 1). In the beginning, the data distribution was made following the ATLAS Computing Model. However, the model was not necessarily applicable to the situation of the first years. For example, many analysis jobs used ESD as input in 2010. Detector performance studies and physics analysis required information available only in ESD. While those studies were supposed to use the 'derived' ESD (dESD) and AOD respectively, their contents were not well tuned, nor were the analysis codes well adapted to the latter formats. The data distribution model was revised accordingly to the needs including creation of extra replicas of ESD. In 2011, ATLAS decided to decrease event filtering rate and take as much data as possible, broadening possibilities in physics studies, that led to a higher event recording rate up to 400 Hz. ATLAS also decided to put RAW data on disk for the "discovery mode", to provide prompt access to any candidate events of new particles. In order to keep the disk usage and the data export throughput within the available resources, ATLAS introduced (a) compression of RAW data that gives about factor 2 of reduction of data volume, and also (b) a limited lifetime of ESD with reduced number of replicas, that allows studies of detector performances during the period and making space to store the RAW data on disk. As figure 2 shows, the data export rate from the Tier-0 varied due to the changes in number of replicas and the composition of data on disk, while the volume on tape was following that of RAW data in figure 1.

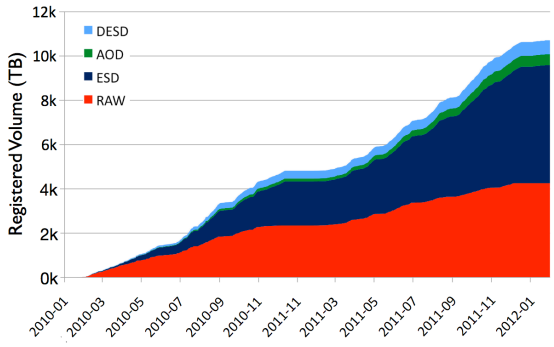

Figure 1. Cumulative data volume registered at the Tier-0 over the two years by the data type.

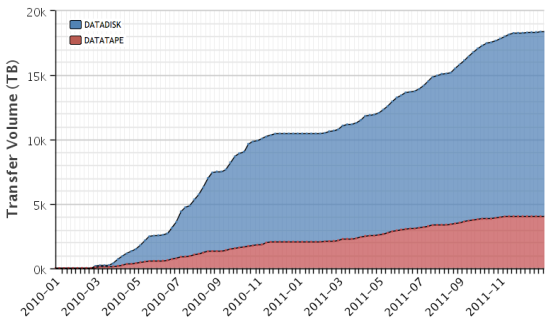

Figure 2. Cumulative data volume exported from the Tier-0 to the Tier-1 centres over the two years by the destination storage type.

In revising the data distribution model, the concepts of 'primary' and 'secondary' replicas were introduced. The 'primary' replicas are the base replicas guaranteed to be available on disk as the minimal baseline of the model. The 'secondary' are extra replicas that are created for popular data using the remaining available disk space. The strategy was first to distribute the minimal replicas as 'primary' and some extra as 'secondary' that are foreseen to be used, adding

more 'secondary' replicas following the usage and needs, and removing unused 'secondary' replicas to ensure enough free space for further prompt replication, especially for new data or popular datasets. A system to measure data set popularity was established that recorded the number of accesses per dataset and per file from different activities, and another system for auto-cleaning that selects dataset replicas to be deleted based on the popularity accounting was developed for that purpose [8, 10].

The revised data distribution system is composed of pre-defined distribution based on the model, dynamic data placement based on the usage, and on-demand replication. The pre-defined data distribution creates 'primary' replicas at Tier-1 sites for redundancy and at Tier-2 sites for end-user analysis, as well as 'secondary' replicas to give larger opportunity for analysis. The replicas at the Tier-1 sites are either exported from the Tier-0 or copied from the other Tier-1 sites. The replicas at Tier-2 are created from the replicas at the Tier-1 sites. The dynamic data placement is implemented in the distributed analysis system to increase 'secondary' replicas based on the usage of the data, in order to reduce the waiting time of analysis jobs [11]. The first implementation was made in mid 2010 and some tuning was made in beginning of 2011 after the experiences acquired with the system and the observations on the analysis job statistics. The on-demand replication covers special cases or specific requests with approval by the responsible people. As figure 3 shows, the major transfer activities for the Tier-1 sites are "Tier-0 export" and "pre-defined distribution" from the other sites, while the "dynamic data placement" plays a larger role for the Tier-2 sites. The peaks in May and November 2010 correspond to the data replication right after the reprocessing campaigns that produced large amounts of data in short periods.

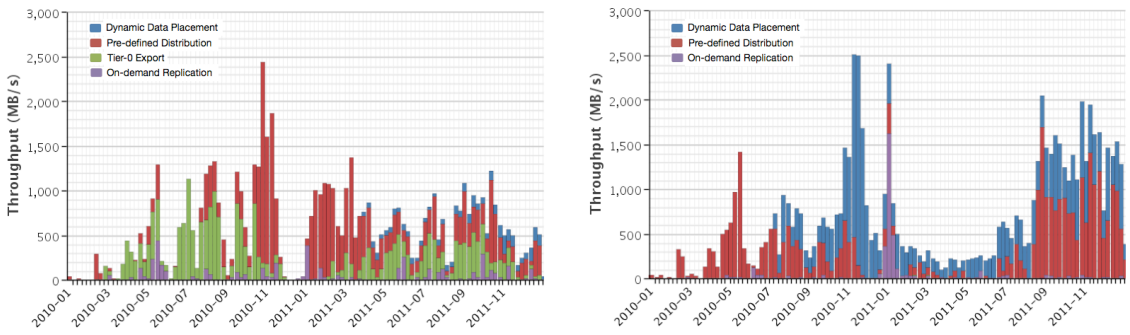

Figure 3. Data distribution to the Tier-1 sites (left) and to the Tier-2 sites (right) per activity. Only the selected activities to a specific space named ATLASDATADISK are plotted in order to see only the data distribution components. Data transfers such as needed for production activities are filtered out (see section 3).

3. Data processing over the Grid

Data processing for ATLAS can roughly be divided into two activities, central production and end-user analysis. Among the central production activities, Monte Carlo simulation has always been running even before the start of data taking and has proven that the ATLAS data processing system [11] is robust and performing well at the scale of the ATLAS distributed

computing. Reprocessing of detector data is another major central production activity, which has been carried out several times in 2010 and 2011, providing essential data for ATLAS to produce physics results. The end-user analysis activities on the grid have also been running before 2010, using the simulated data, but they started rising significantly since the start of data taking (figure 4). In between the central production and the end-user analysis activities, there are group analysis activities, where the detector performance study groups and the physics analysis groups produce common data for end-user analysis from the central production output such as ESD and AOD. In the beginning, the group analysis activities were run in a form of end-user analysis, submitted by the individuals who were responsible for producing the group data. The activities have features very much similar to those of central production and have been formalized as "group production" once their software became standardized.

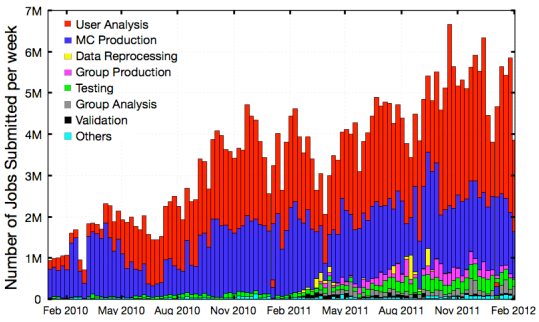

Figure 4. Data processing activities during the two years. End-user analysis started rising significantly since April 2010, whereas MC production has been running rather constantly. The group activities can be identified in the plot only after March 2011 since they were not defined in the monitoring system and classified as 'others' until then. Increase of group production is observed since June 2011.

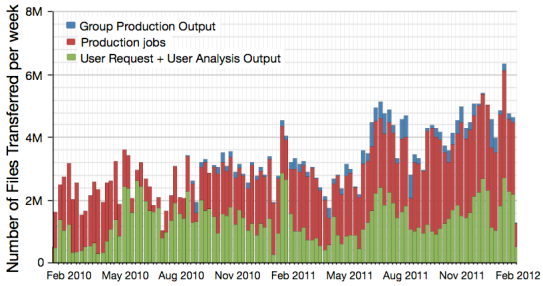

Figure 5. Data transfer activities related to data processing. Production jobs trigger transfers of input and output files, while analysis jobs are sent to the sites hosting input data and no transfers are triggered by default. However, users can request replication of the input data they are going to use, and can also specify the destination of the output. The two activities are comparable in terms of number of files.

Data processing activities also trigger data transfers. The central production jobs including the group production jobs run at Tier-2 sites require transfers of input data from their associated Tier-1 centres, unless they are already available at the site. The output files are sent back to the associated Tier-1 and aggregated there. Group production jobs are defined with destination group spaces allocated at various sites, and the output files are automatically transferred to the destination sites. End-user analysis jobs are brokered to the sites hosting the input data, but in case they have difficulties to run the jobs at those sites, the users can request replication of the input data to some other sites where they have no problem to run the jobs. Users can also submit analysis jobs with specifying a destination site for the output, which results in automatic data transfers as a set of requests in the on-demand replication system. All these transfer activities triggered by data processing can be identified in the data transfer monitoring, and as figure 5 shows, the transfers triggered by users including the output transfers to the destination sites are comparable in number of files to the transfers of input and output files for official and group production.

3.1 ATLAS software and database

The ATLAS software and the detector information are necessary to run ATLAS jobs. Since it would not be very efficient to send them to the site together with each of the jobs, the initial solution was to install the software and small file-based database at each site, and a larger database system at each Tier-1 site. The software installation is carried out by special jobs submitted to the site that download software releases from the central repository, install them on a local shared file system at the site, and validate the installation. The detector information is put into a database at each Tier-1 site synchronized with Oracle Streams from the database at the Tier-0. Some of the detector information such as detector geometry parameters needed for simulation jobs and detector conditions parameters for end-user analysis are put into file-based database and distributed to every site with the DDM. Thus, the initial model was to run at Tier-1 sites certain types of jobs, especially reprocessing jobs, which require the information not in the file-based database, and simulation and end-user jobs at mostly at Tier-2 sites although they can also be run at Tier-1. With this model, some bottlenecks were observed when many jobs accessed the shared file system or the file-based database simultaneously.

Evolutions came with CernVM-FS [12] and Frontier/Squid [13]. CernVM-FS is a network file system based on HTTP, with which files are downloaded and cached at the sites and on the worker nodes. The ATLAS software releases and the smaller file-based database are now installed on the server at CERN, and there is no more need to install them at the sites where CernVM-FS is used. This has removed the workload in software installation and the bottlenecks with the shared file systems. Frontier/Squid is a http-based system to access database with caching, avoiding a high load on the database and latency in accessing the database from remote sites. Introduction of the system has removed limits with the database access, allowing the jobs running at Tier-2 sites accessing the database at Tier-1 sites. With this, any type of jobs can now run at any Grid site that has these tools available.

4. Ensuring smooth activities

Running the activities over the components widespread around the world at many sites, and experiencing troubles frequently, constant flows of tests have been introduced to detect problems as soon as possible, notifying the responsible people, and in some cases, when necessary, to avoid the problematic components while waiting for fixes. The first such tests that were introduced were the data transfer functional tests to ensure a smooth data distribution. They are important especially for the Tier-1 sites where data replication from the Tier-0 is crucial while the Tier-0 export is not a constant flow. The tests were introduced after the early exercises of a large-scale data distribution in 2007, and the results are constantly monitored in order to avoid finding a problem only after starting a replication. There is also a system that collects site downtime and storage information periodically, and automatically stops data transfers from and to the sites that are in downtime and as well as to those sites with almost no free space.

Failures in data processing affect end-user jobs more seriously than production jobs because failed production jobs are re-submitted automatically by the system, whereas end-user analysis jobs are not resubmitted automatically. The reason is that production jobs in principle

are well validated and most failed jobs will finish successfully after resubmitting, whereas end-user jobs may fail due to the user codes. Since problems at sites are usually reported only after a large number of failures are observed, many failed user jobs could potentially occur before actions, such as closing the queues, are taken. Therefore, it is important to detect problems as soon as possible, and close the queues at problematic sites, and re-opening the queues only after having successful tests. In order to improve the end-user experiences in data processing on the grid, the analysis job functional tests together with automatic control of queue status have been introduced in 2010. There is also an on-going work to resubmit end-user jobs based on the category of the failures. When production jobs are failing, it is enough to close the queue manually to avoid submitting jobs to the site and to send test jobs once the problem is fixed. However, the manual interventions result in an increased load on the operations staff. After seeing successful automatic queue control with the analysis job functional tests, production job functional tests were introduced recently in 2012 in order to reduce the load of manual interventions by automatically sending test jobs and controlling the queue status based on their results. In addition to the queue control based on the functional tests, another system has also been introduced to automatically control the queues based on the downtime information.

5. Monitoring

Monitoring is a key to effective operations, and a significant effort has been invested to assure effective monitoring. The most basic monitoring tools are that for the ATLAS activities, made to understand the situation, such as number of successes and failures in data processing and data transfers, number of running and waiting jobs, throughput of data transfers. Based on the activity monitoring, a monitoring of site status has been built, so that one can see not only that the services provided by the sites are working, but also if the sites are used and are working properly in the ATLAS activities. The monitoring system also records declared downtime information of the sites that can be taken into account for site availability calculation.

Transfer statistics between the sites depend more on the network between the sites rather than the site status. The monitoring for the transfer statistics was built on top of the data transfer activity monitoring, but collecting throughput and data size rather than successes and failures. In order to monitor the statistics of the full mesh between all the sites, of the order of hundred times hundred, test transfers named as ‘sonar’ tests are injected to every direct link between the sites, whereas the real production transfers are routed via the well established links, i.e. between Tier-1 and associated Tier-2 sites and between Tier-1 sites. Monitoring of the bare network performance, such as throughput, latency and ‘traceroute’ results has also been started, using perfSONAR [14] with a couple of dedicated nodes at each site. In order to gain a good understanding of network by closely watching the monitoring results, the monitoring was started with a short list of sites, which is to be extended in future.

6. Evolution of the ATLAS computing model

Based on the operational experiences and the monitored transfer throughput and reliability, it was decided to evolve the computing model to make flexible data transfer routing, enabling more efficient job distribution to the sites in data processing.

6.1 Evolution of the transfer routing model

In the original model of data transfer routing, it was assumed that efficient network existed only between Tier-1 sites and between associated Tier-1 and Tier-2 sites. Thus, the transfers between two Tier-2 sites that are not associated to the same Tier-1 were routed as T2A – T1A – T1B – T2B, where T2A and T2B are the two Tier-2 sites and T1A and T1B are their associated Tier-1 sites correspondingly (figure 6a). Direct transfers between Tier-2 sites were carried out only between 'close' sites that are associated to the same Tier-1 site. However, in reality, some Tier-2 sites have no problem in data transfers from or to other Tier-1 or Tier-2 sites not in association (fitures 6b and c). Based on the full mesh transfer tests from every site to every site as described in the previous section ('sonar' tests), an auto-routing algorithm was implemented in the transfer system, that decides which routing, either via Tier-1 sites or direct, is more efficient for a transfer of a certain file size between a pair of sites based on the measured statistics.

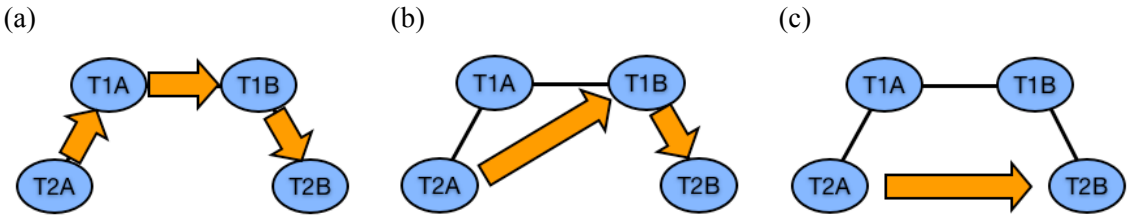

Figure 6. The evolution of the data transfer routing model. (a) The original model relied on the presumably efficient network between Tier-1 sites and between associated Tier-1 and Tier-2 sites. (b) Some Tier-2 sites are well connected to many other Tier-1 sites and can skip a transfer routing via their associate Tier-1. (c) The transfer efficiencies between some Tier-2 sites not associated to the same Tier-1 sites are good enough to make direct transfers without routing via Tier-1.

6.2 Evolution of the data processing model

With the help of the full-mesh monitoring, the Tier-2 sites that are well connected to most of the Tier-1 sites have been identified. This allows more flexible association in data processing than the original model. In the original model, each production task is assigned to a Tier-1 site where the input data is available and the jobs of the task are run at the Tier-1 and its associated Tier-2 sites. To run the jobs at Tier-2 sites, corresponding parts of the input data are sent to the Tier-2 sites, and the output data are aggregated at the Tier-1 site (figure 7a). A shortcoming of this model is that the sum of the CPU capacity at the Tier-1 and Tier-2 sites to produce data may not be in balance against the disk capacity at the Tier-1 to host the aggregated output. Following the network monitoring, the Tier-2 sites with good network connection are associated to multiple Tier-1 sites (figure 7b). As an extension, it is also possible to associate even a Tier-1 site with other Tier-1 sites so that it can contribute to the task assigned to the other Tier-1 (figure 7c).

This multiple association does not change the total CPU capacity that ATLAS can utilize, but increases the possible maximum CPU power that can be utilized to run jobs for a task

assigned to a Tier-1 site. As a result, high priority tasks can be completed more quickly. It is also expected that the CPU and the disk capacities are brought into balance in the end.

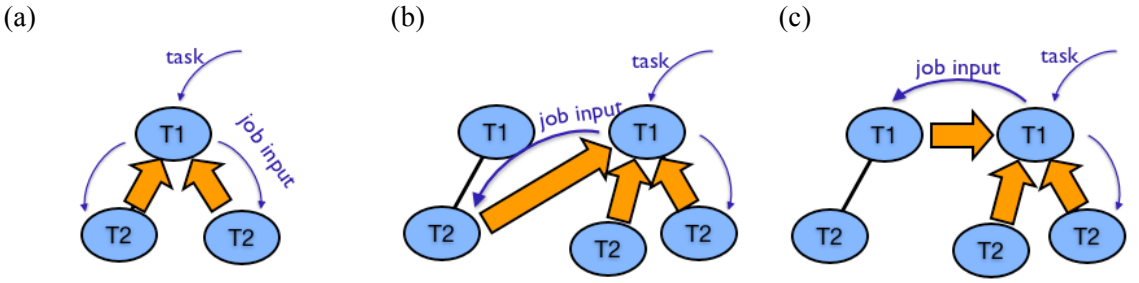

Figure 7. The evolution of the data processing mode. (a) In the original model, the jobs of a task assigned to a Tier-1 site, are run at the Tier-1 and its associated Tier-2 sites, with the input data sent to the Tier-2, and the output data aggregated at the Tier-1. (b) The Tier-2 sites well connected to another Tier-1 site can contribute to the tasks assigned to the Tier-1 as well as its associated Tier-1. (c) Even a Tier-1 site can contribute to the tasks assigned to another Tier-1.

7. Conclusions

The ATLAS distributed computing system has been running stably with the large amount of data for the activities such as data distribution from the Tier-0, production of simulated data and its distribution, group and end-user analysis jobs. The system has been evolving and improving without facing scalability issues, for example, data placement with various components to optimize the data distribution dynamically and automatically, group analysis integrated into the production system, monitoring of those activities, site status and network, constant flow of functional tests of data transfer, analysis and production to ensure smooth activities and automated actions against site instabilities, and the models for data transfers and data processing beyond the original ones. The result of this work is a better environment for physics studies of the collaboration, and we are looking forward to fruitful physics results.

References

[1] The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST 3 S08003

[2] The ATLAS Collaboration, *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, hep-ex/0901.0512

[3] The A TLAS Collaboration, *ATLAS computing : Technical Design Report*, CERN ATLAS-TDR-017, CERN-LHCC-2005-022

[4] D. Adams et al. on behalf of the ATLAS Collaboration, *THE ATLAS COMPUTING MODEL* CERN ATL-SOFT-2004-007, CERN-LHCC-2004-037/G-085

[5] R.W.L. Jones and D. Barberis, *The Evolution of the ATLAS Computing Model*, J. Phys.: Conf. Ser. 219 072037

[6] *Memorandum of Understanding for Collaboration in the Deployment and Exploitation of the Worldwide LHC Computing Grid*, http://cern.ch/LCG/mou.htm

[7] M. Elsing, L Goossens, A. Nairz and G. Negri, *The ATLAS Tier-0: Overview and operational experience*, J. Phys.: Conf. Ser. 219 072011

[8] I. Ueda for the ATLAS collaboration, *ATLAS Operations: Experience and Evolution in the Data Taking Era*, J. Phys.: Conf. Ser. 331 072034

[9] M. Branco et al., *Managing ATLAS data on a petabyte-scale with DQ2*, J. Phys.: Conf. Ser. 119 062017

[10] A. Molfetas et al., *Popularity Framework to Process Dataset Tracers and Its Application on Dynamic Replica Reduction in the ATLAS Experiment*, J. Phys.: Conf. Ser. 331 062018

[11] T. Maeno et al. for The ATLAS Collaboration, *Overview of ATLAS PanDA Workload Management*, J. Phys.: Conf. Ser. 331 072024

[12] *CernVM File System (CernVM-FS)* http://cernvm.cern.ch/portal/filesystem

[13] *Frontier/Squid* http://frontier.cern.ch/

[14] *perfSONAR* http://www.perfsonar.net/