The ATLAS Distributed Computing: the challenges of the future

Hiroshi Sakamoto[1] for the ATLAS Collaboration

International Center for Elementary Particle Physics, the University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan
E-mail: `sakamoto@icepp.s.u-tokyo.ac.jp`

The ATLAS experiment has collected more than 25 fb-1 of data since LHC has started its operation in 2010. Tens of petabytes of collision events and Monte-Carlo simulations are stored over more than 150 computing centers all over the world. The data processing is performed on grid sites providing more than 100.000 computing cores and orchestrated by the ATLAS in-house developed job and data management services. The discovery of the Higgs-like boson in 2012 would not have been possible without the excellent performance of the ATLAS Distributed Computing. The future ATLAS experiment operation with increased LHC beam energy and luminosity foreseen for 2014 imposes a significant increase in computing demands that ATLAS Distributed Computing needs to satisfy. Therefore, a development of the new data-processing, storage and data-distribution systems has been started to efficiently use the computing resources exploiting current and future technologies of distributed computing.

1

    Speaker

1. Introduction

The biggest news in the particle physics field in 2012 must be the discovery of the Higgs(-like) [2] particle announced in July by ATLAS[1] and CMS[2]. For example, Fig. 1 shows the recent result on the search for new particles decaying into 4 leptons (exactly speaking, electrons and muons) [3]. A clear peak, which is thought to be originated from the newly discovered particle, is found around the 125GeV/c^2 region. It should be emphasized that the peak is composed of a very limited number of events of order few tens. These events are selected by scanning a very

Fig.1. Four lepton invariant mass distribution based on analyses of 2011-2012 data taken by ATLAS [1].

huge set of raw data taken in these 2 years. We recorded 3.3 billion events in 2011 and 8.4 billion events in 2012. It should also be noted that, in the analysis, simulated events with the
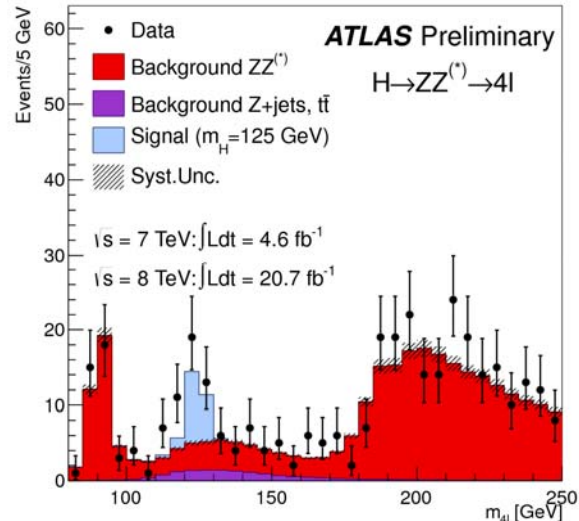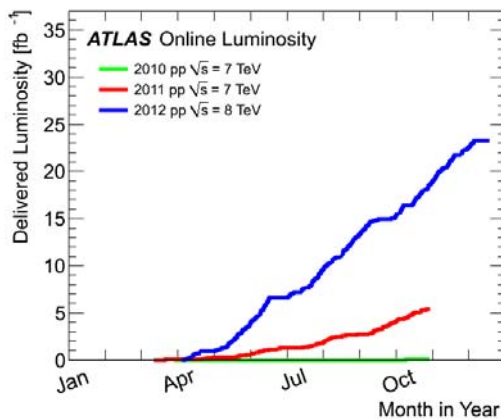

Fig.2. Progress of the integrated luminosityies measured at the ATLAS experiment. The blue line indicates the progress in 2012 and it is 5 times more than that in 2011 (the red line) and 500 times than in 2010 (the green line).
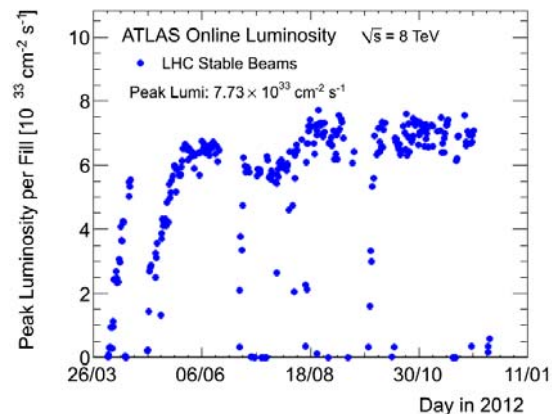

Fig.3. Instantaneous luminosity observed at ATLAS in 2012 runs. The peak luminosity of 7.73×10^{33}/cm^2/s was achieved in this year.

---

[2] In the press release on March 14th, 2013, CERN announced that the discovered particle is more likely to be a Higgs particle. http://home.web.cern.ch/about/updates/2013/03/new-results-indicate-new-particle-higgs-boson

same or better statistics were generated for signal and background channels as shown in the figure, by fully utilizing computing resources distributed worldwide. This figure is a good indication explaining that the ATLAS distributed computing has been playing a very important
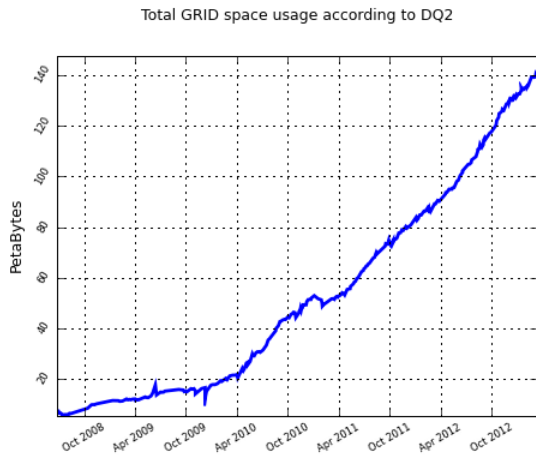

Fig.4. Progress of the total volume of data files managed by the DQ2 data management system since October 2008. It reached to 140PB at the end of 2012.
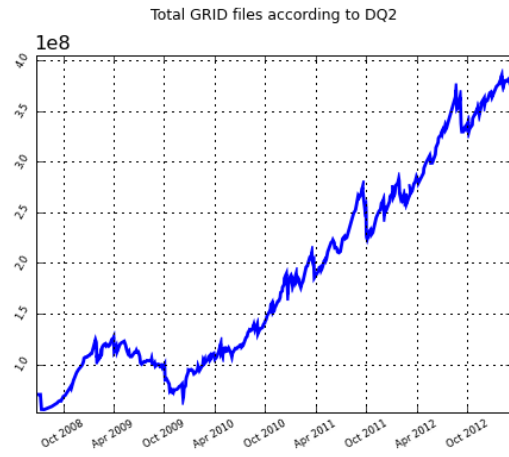

Fig.5. Progress of the number of files managed by the DQ2 data management system since October 2008. The number of files exceeded 350 million files

role for the discovery.

The LHC accelerator delivered 23fb^{-1} integrated luminosity in 2012 at the collision energy of 8TeV (Fig.2.). The highest instantaneous luminosity of $7.73 \times 10^{33}/cm^2/s$ was achieved in 2012, as shown in Fig. 3.

The ATLAS experiment recorded 3.43PB of raw data in 2012. It contains 10.79×10^9 events distributed on 6.11×10^6 files. The average data acquisition rate was around 450 Hz which was much higher than the design value of 200 Hz. This required more computing resources and more efficient use of them than originally expected.

2. ATLAS Distributed Computing in 2012

In the ATLAS Distributed Computing (ADC) system, we have one Tier 0 at CERN, 10 Tier1 centers and more than 130 Tier2/Tier3 sites around the world. The raw data are recorded and processed first at Tier0 and then the raw data and derived data are exported to Tier1s and then Tier2s. Monte Carlo production is performed both at Tier1s and Tier2s and the output of the simulation production is stored at Tier1s. In recent production, computing resources on the grid are used more flexibly and efficiently, by relaxing the function assignment to Tier1s and Tier2s.

2.1 Achievements

The ATLAS Distributed Data Management (DDM) system called DQ2 is now managing 140PB of replicas consisting of various types of data including real data and simulation data (Fig.4). The total number of files managed is reaching 4 billion (Fig.5).

These files have been transferred across the world throughout the year. The average transfer rate is 5GB/s and the overall success rate of these transfers is around 92.8% (Fig.6).

Computing resources (CPUs) have been also fully utilized throughout the year 2012. Around 100,000 jobs of centrally managed production and up to 50,000 user analysis jobs had been running concurrently (Fig. 7).

Number of jobs processed per day hits 400,000 for the central production and their success rate was 92.9% throughout the year 2012 (Fig. 8).
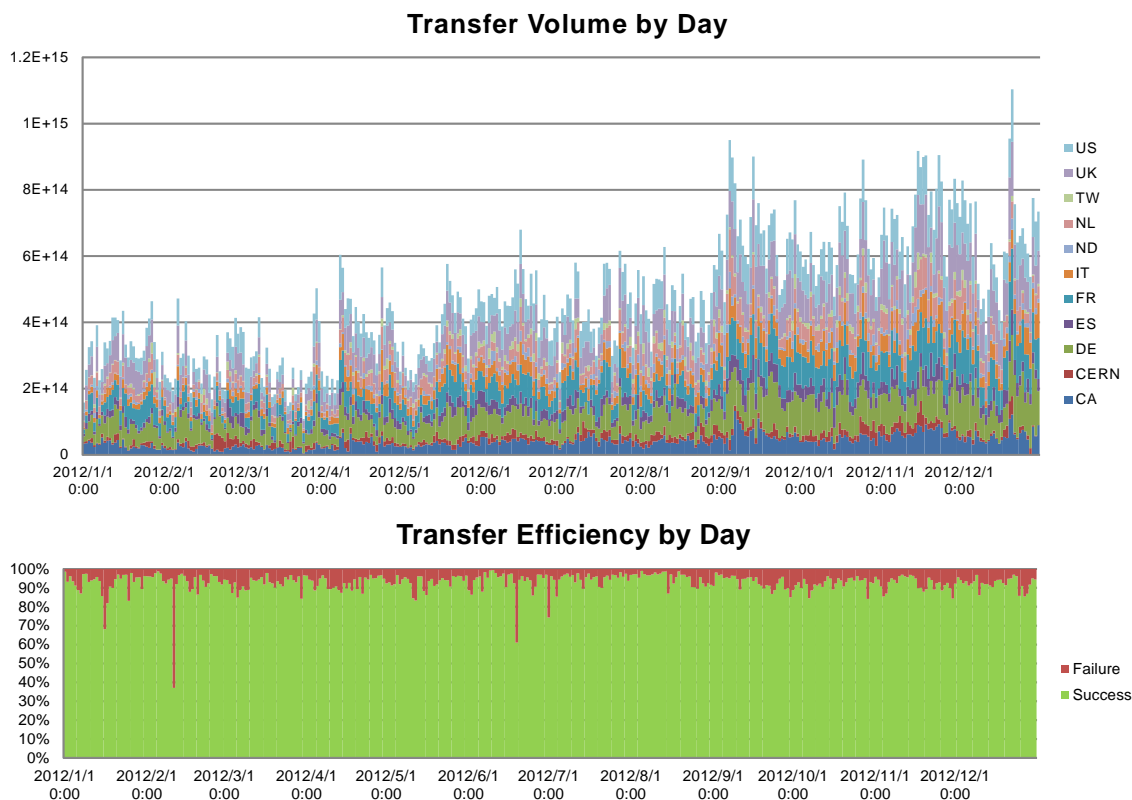

Fig.6. Transfered volume and success rate of all file transfers of ATLAS per day in 2012.Annual record of transfers and success rates in 2012. The upper graph shows the total volume per day, and the lower graph shows the success rate per day. The average success rate throughout the year 2012 is 92.8%.
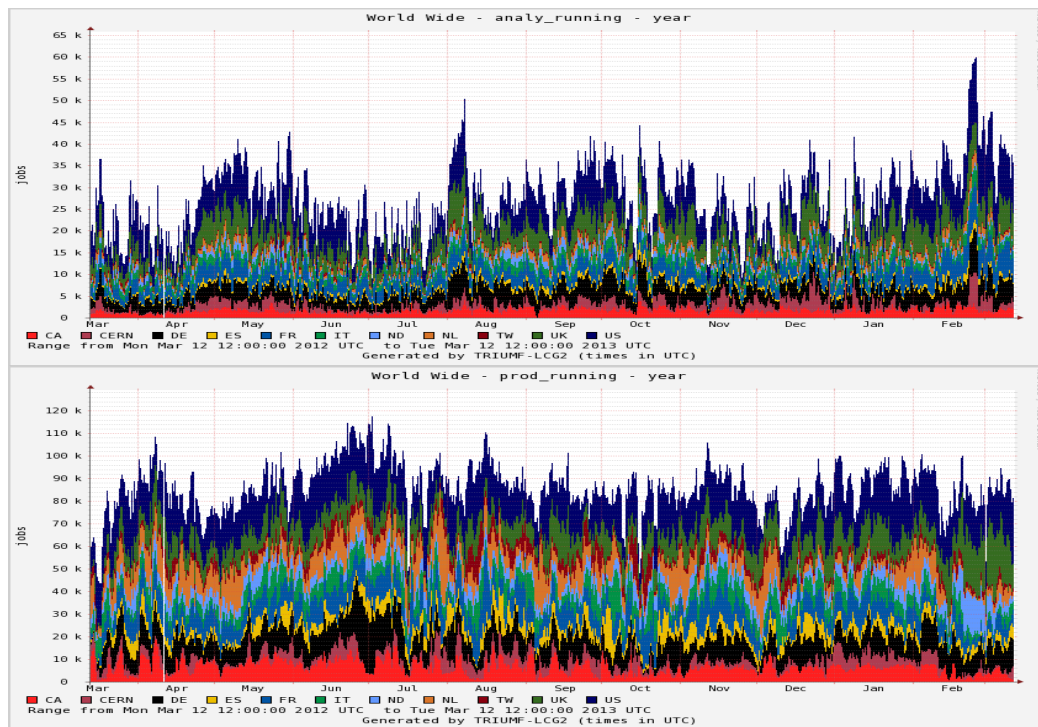
Fig.7. Number of concurrently running jobs from March 2012 to February 2013. The upper graph shows the number of user analysis jobs and the lower one shows the number of centrally controlled production jobs. The analysis jobs hit 50,000 jobs at the peak and the production jobs exceeded 100,000 jobs. In total, 150,000 jobs are running concurrently.

2.2 Operation Experiences

These operations have been monitored by on-site (at CERN) and off-site (remotely from home institutes) shifters on 24/7 basis. Incidents are reported by using several bug ticketing systems. One is GGUS (Global Grid User Support)[4], which handles site issues. In total, 1,707 tickets have been submitted in 2012 by the shift team, as shown in Fig. 9. Figure 10 shows the interval between the ticket submission and the fixing of the issue by the relevant site. Almost half of issues have been solved in one or two days. The pie chart in Fig. 11 shows categories of issues. Major issues are related to file transfers, file access or local storage systems.

Another ticketing system is the LCG Savannah[5], which treats issues on the central services or production tasks. In 2012, 762 tickets were issued on the ADC operation group. The number of Savannah tickets per month, sent to the ADC operation group is plotted in yellow bars in Fig. 12.

In order to achieve higher efficiency of operation, automated control is very desirable. Using the HammerCloud[6] test job submission scheme, the manipulations of production task queues have been automated since March 2012. By submitting several types of test jobs to a

Total Production Jobs by Day
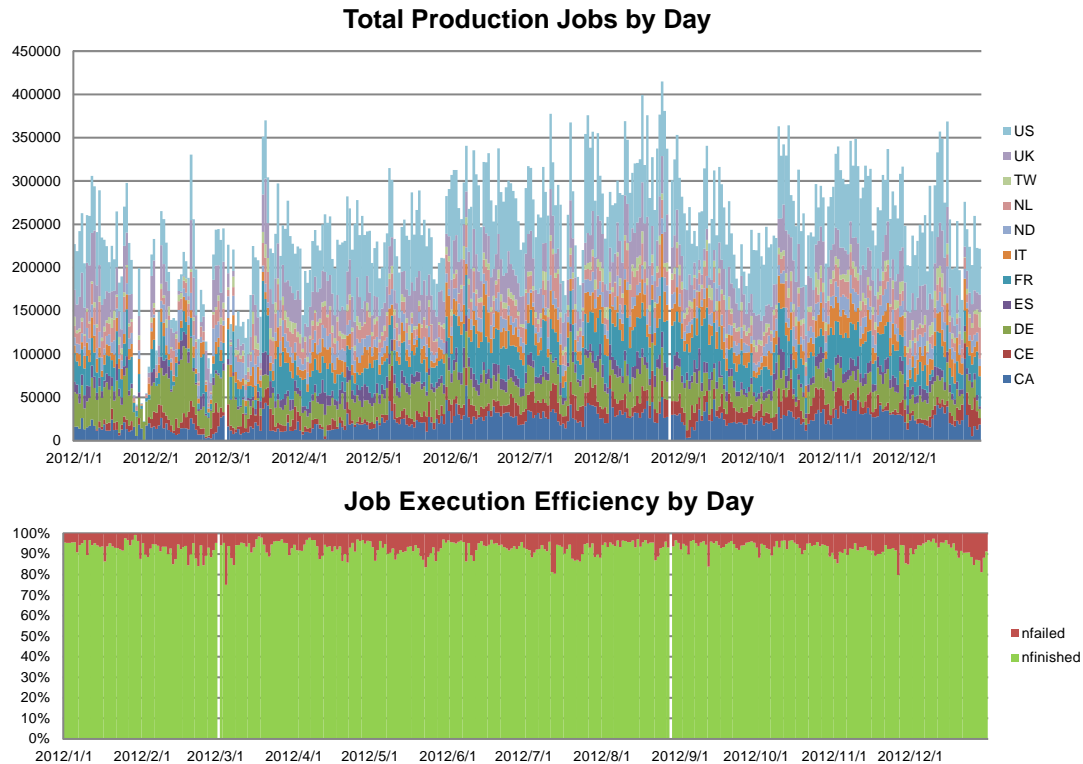

Job Execution Efficiency by Day

Fig.8. Number of production jobs executed per day in 2012, and their success rates. The total number of jobs exceeded 350,000 at the peak. The lower graph shows the success rate of the production jobs. The average success rate throughout the year 2012 is 92.9%.

production site, and by analyzing the job results, the production queues of the site can be disabled and enabled. Figure 12 shows the number of such exclusion incidents recorded in 2012, together with the number of GGUS and Savannah tickets already described above.

3. Long shutdown in 2013 and 2014

In years 2013 and 2014, the LHC accelerator is in a shutdown period to prepare for the operation at the increased collision energy of up to 14 TeV. In Run 2, starting from 2015, it is foreseen that not only the collision energy is raised, but also the luminosity is increased. Accordingly, the number of collisions per bunch crossing goes up to 27 in average, which means that the average event size grows accordingly. The expected recording rate will exceed 1 kHz, in order to keep as much as possible of the physics capability. These imply that the requirements to the distributed computing system will be much severer compared to those of previous years. Therefore, the system must have enough scalability to keep up with such heavy loads in Run 2. There are many upgrade programs running during the long shutdown.

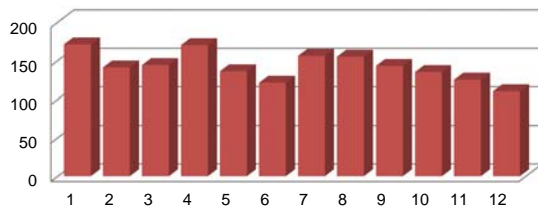**GGUS Team Tickets for ATLAS by Month in 2012**

Fig.9. Number of GGUS tickets submitted per month in 2012. In total 1,707 GGUS tickets had been submitted in 2012.

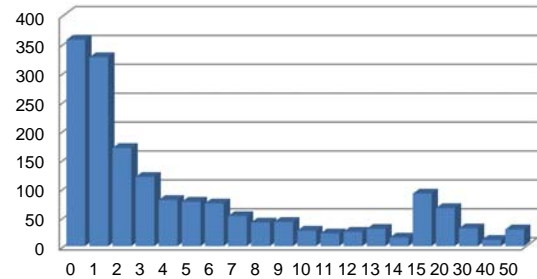**Number of days to close tickets**

Fig.10. Distribution of number of days to close a GGUS ticket. Around 800 tickets have been closed in two days.

3.1 New Distributed Data Management System

As already shown before, the existing data management system DQ2 has been demonstrating a good performance so far. However, there is still a room to improve. For example, it is an issue how quickly the replica information of a dataset can be retrieved from the database, as the total number of files which the system is managing is constantly increasing. The scalability of the system is our key concern.

This system is built upon many grid middleware components like replica catalogues and file transfer services. Among them, some component packages are going to retire. So it should be capable of adopting to the evolution of such middleware. It is important to simplify and reduce dependencies to such middleware packages.
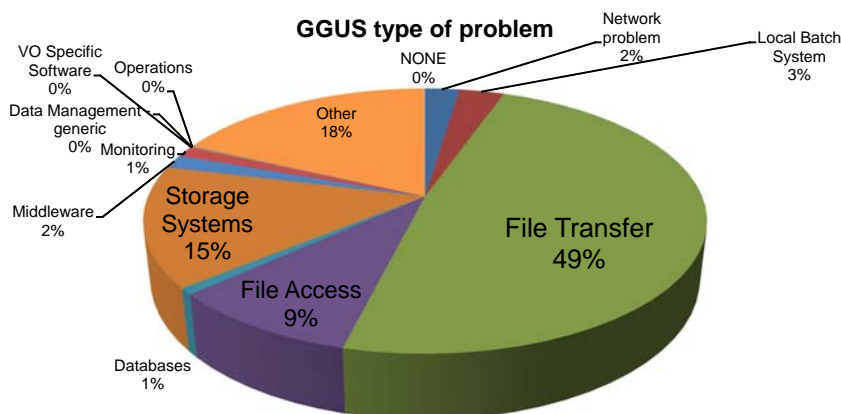

Fig.11. Problem types of GGUS tickets issued in 2012. The major fraction of the tickets are related to file transfer issues.
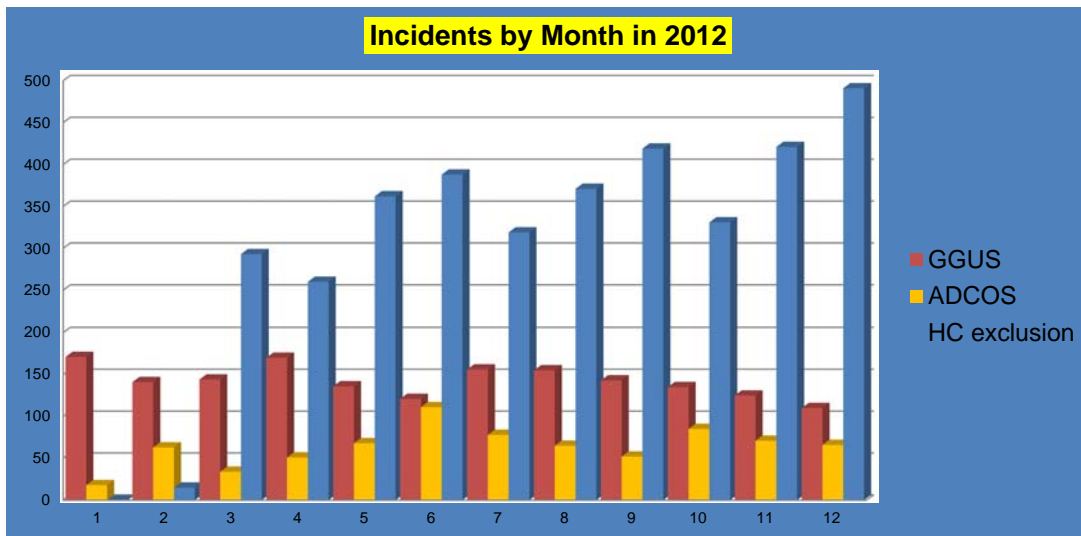
Fig.12. Number of incidents by month in 2012. The red bars indicate GGUS tickets and the yellow bars indicates the number of Savannah tickets issued to the ADC operation support group. The blue bars indicate the number of exclusion incidents of production queues per month. Actually, the automatic exclusion was introduced in Mach 2012.

Development work of a new distributed data management system, called Rucio [7], is now going on, reflecting the experiences learned in RUN1. Several new ideas are to be implemented in the system. By introducing a concept of 'account', more sophisticated resource allocation and monitoring becomes possible. A new mechanism to handle the metadata of files and their aggregations is also introduced in order to enable much faster database retrieval and smarter control of replications. Deployment of Rucio is foreseen in 2014 during the long shutdown.

3.2 New Production System

There is also a plan to upgrade our production system. The workload management system is based on a pilot job scheme. A pilot process resides on a worker node and takes care of its payloads on the node. When a payload becomes available, the pilot pulls a job from the waiting queue assigned to the site. Implementation of this scheme, called 'Panda' is very successful. However, a cleverer job dispatcher is required in order to improve CPU usage and to reduce job failures.

Design work of a new production system is now on-going. An idea of 'meta-task' is introduced. In general, production jobs process very large number of files, by changing input parameters and conditions. So, a series of jobs should be described in a simpler way, so that a user can submit his/her jobs with much fewer job description statements.

Jobs are distributed to job execution systems around the world according to the job parameters. The jobs are always monitored by the production system and once completed, the job outputs are sent to their proper destination. If a job fails, the information is returned to the

production system, and the job will be submitted again. If it fails repeatedly, the parameters of the task can be dynamically modified according to the failure pattern and resubmitted with much safer parameters.

The development of the new production system is tightly connected with that of the new distributed data management system explained above (Rucio). First, the new production system will be built with the existing data management system at the end of 2013, and then, after the deployment of the new DDM, this production system will switch to interface to the new DDM. It will take place in 2014.

3.3 LHCONE: Open Network Environment

In the distributed computing system, the wide area network (WAN) is playing an essential role. In recent years, traffics running on WAN are rapidly increasing because of evolving bandwidth-consuming applications like video streaming, and congestions have been observed frequently everywhere on WAN worldwide. So, it is very important to understand and control the usage of WAN. In order to share the WAN bandwidth more efficiently and stably with other scientific disciplines, LHCONE (LHC Open Network Environment) [8] is being deployed.

LHCONE is a community of layer 3 private network services for LHC experiments implemented on various national research and education networks (NRENs). These LHCONE services are mutually connected at several exchange points on the internet. This architecture allows us to design a better routing and to allocate necessary bandwidth with less congestion. The number of sites connected to the LHCONE services is now rapidly increasing.

Monitoring of the LHCONE network service is the key for its stable operation. Each LHCONE site has perfSONAR-PS[9] instances and periodic measurements of latency, bandwidth, routing and packet loss are done as shown in Fig. 13. This is a really powerful tool to diagnose the LHCONE network.
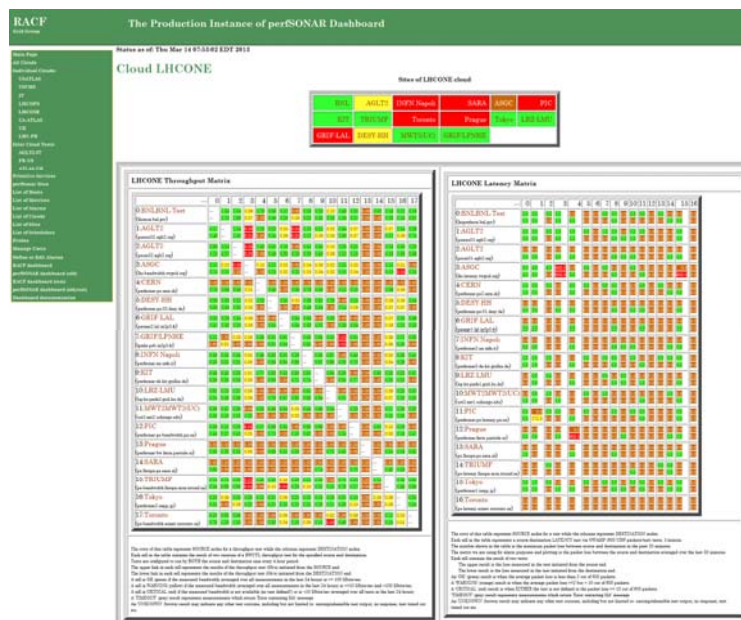

Fig.13. A snapshot of the LHCONE monitoring page. Network connectivities among sites are shown in colored boxes.

3.4 Monitoring and Automation

The ADC system is a quite complex system with very many features and very many control aspects. In order to keep the system healthy and stable, we need detailed monitoring of the system from various points of view. A lot of monitoring tools have been developed in these years to improve the operation. For example, links to 45 monitoring tools are displayed on the ADC monitoring portal page as shown in Fig. 14. ADC shifters are checking these pages so that any incident can be detected as quickly as possible.

Automated control of the ADC system is also a desired feature which we are keen to introduce. The automation needs sophisticated decision algorithms based on the detailed observation of the situation, comparison with a large list of recorded failure patterns, and selection from wide variety of templates of actions to be taken. Such automation machineries are gradually being introduced. Since March 2012, the production queues of ATLAS have been turned on and off automatically, as explained above, and, since March 2013, file transfer endpoints have been disabled and enabled automatically using the results of transfer tests.

4. Future Prospects: RUN 2 since 2015

During the long shutdown, many improvements and consolidations are applied to both the accelerator and the ATLAS detector. The LHC accelerator will restart operation in 2015 at the much higher collision energy of 13 TeV with higher luminosity. The expected data acquisition
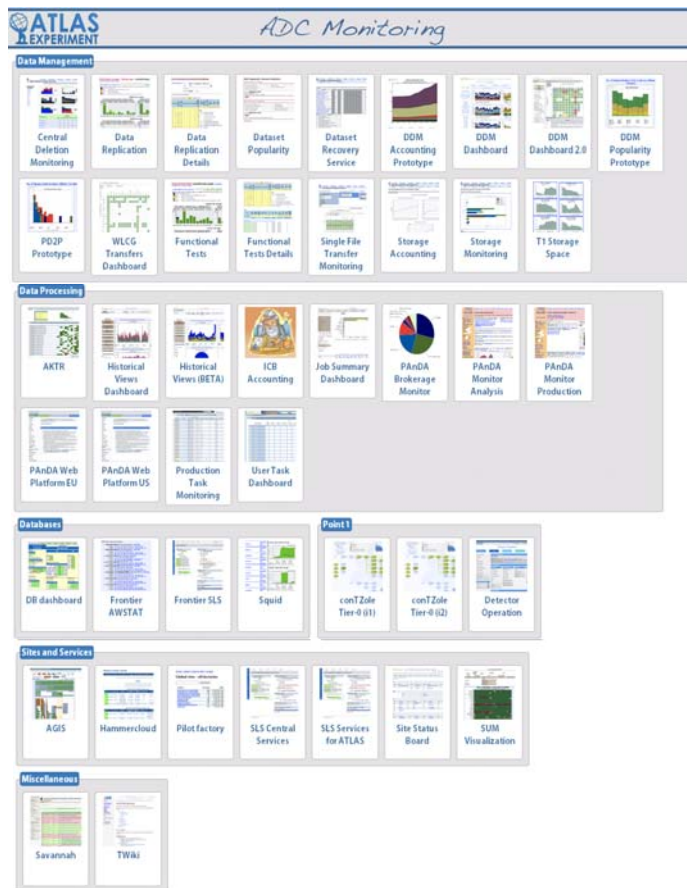

Fig.14. A snapshot of the ADC Monitor portal. Each box indicates an entry to a monitoring tool. There are 45 such tools shown on this page.

http://pos.sissa.it

rate will exceed 1 kHz and experimental data of more than ever will be collected. Even for ADC, its operation is very challenging, with an unprecedented scale of the infrastructure. The further investigation of the Higgs particle, and searches for the physics beyond the standard model, will be achieved only when the ADC system is fully exploited.

References

[1] The ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B 716(2012)1-29.

[2] The CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,* Physics Letters B 716(2012)30-61.

[3] The ATLAS Collaboration, *Measurements of the properties of the Higgs-like boson in the four lepton decay channel with the ATLAS detector using 25 fb−1 of proton-proton collision data*, ATLAS-CONF-2013-013, http://cds.cern.ch/record/1523699.

[4] *Global Grid User Support*, https://wiki.egi.eu/wiki/GGUS

[5] *The LCG Savannah,* https://savannah.cern.ch/

[6] *HammerCloud*, https://twiki.cern.ch/twiki/bin/view/Main/HammerCloud

[7] *The Rucio project,* http://rucio.cern.ch/.

[8] *LHCONE: LHC Open Network Environment*, http://lhcone.net/.

[9] *perfSONAR*, http://www.perfsonar.net/