

Grid Site Testing for ATLAS with HammerCloud

Johannes Elmsheuser¹, Friedrich Hönig¹, Federica Legger¹, Ramon Medrano LLamas¹, Francesco Giovanni Sciacca³, Daniel van der Ster²
on behalf of the ATLAS collaboration

¹ Ludwig-Maximilians-Universität München, Germany

² CERN, Switzerland

³ Universität Bern, Switzerland

E-mail: johannes.elmsheuser@physik.uni-muenchen.de

Abstract. With the exponential growth of LHC (Large Hadron Collider) data, distributed computing has become the established way to analyze collider data. The ATLAS grid infrastructure includes more than 130 sites worldwide, ranging from large national computing centers to smaller university clusters. HammerCloud was previously introduced with the goals of enabling VO- and site-administrators to run validation tests of the site and software infrastructure in an automated or on-demand manner. The HammerCloud infrastructure has been constantly improved to support the addition of new test work-flows. These new work-flows comprise e.g. tests of the ATLAS nightly build system, ATLAS MC production system, XRootD federation FAX and new site stress test work-flows. We report on the development, optimization and results of the various components in the HammerCloud framework.

1. Introduction

The LHC at CERN is colliding protons or heavy ions at unprecedented center-of-mass energies and these collisions are recorded by several experiments including the ATLAS experiment [1]. The large number of collision events and resulting data volumes require an analysis on distributed computing resources [2]. HammerCloud was previously introduced [3] as a solution to stress the Worldwide LHC Computing Grid sites and execute functional testing on these distributed resources.

This paper will first briefly describe the HammerCloud functionality and various test options used within the ATLAS experiment. These tests options will be described in more detail in the context of FAX (Federating ATLAS storage systems using XRootD) testing [4]. Some statistics of the HammerCloud usage will be given before the conclusions.

2. HammerCloud overview

HammerCloud is a grid testing service for the ATLAS, CMS and LHCb experiments. It is hosted by CERN IT and operated by the experiments grid operations teams. Figure 1 shows a schematic view of the HammerCloud architecture and its various components. HammerCloud is a Python application using GANGA [5] for test job submission and monitoring. It is using Django as a database access layer and presents the test results to the users and operators on web pages. GANGA is used as the backend job management tool since it provides with its plug-in architecture a convenient access to the interfaces of the ATLAS Grid middle-ware stack, namely



PanDA [6] for the workload management and DQ2 [7] for the distributed data management, and also provides the functionality to submit, clone, resubmit and monitor a series of identical or similar jobs.

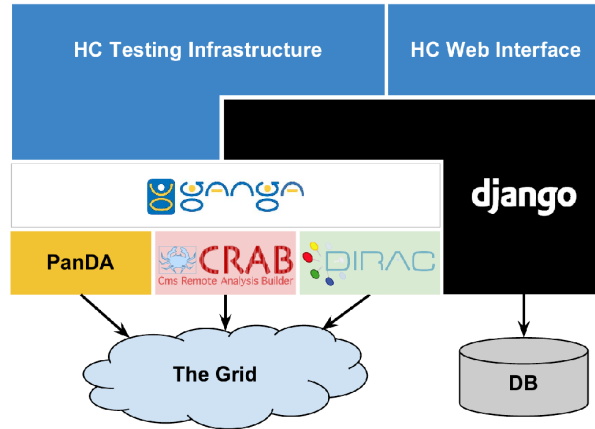


Figure 1. Schematic view of the HammerCloud (HC) architecture and its various components.

3. HammerCloud tests

The ATLAS experiment originally developed HammerCloud as a Grid stress testing system before the start of LHC data taking. Now, there are several extensively pursued use cases in addition:

- Stress testing scheduled on demand: these tests are used to measure experiment applications under different site and application configurations and evaluate their performances.
- Site functional tests: a continuous flow of jobs is sent to the PanDA system analysis and production system queues to validate their functionality. These so called “analysis functional tests” (AFT) and “production functional tests” (PFT) are used to automatically exclude and include sites into operations if they fail or subsequently succeed again a series of test jobs. Email notifications are sent out to shifters, site and cloud support teams. All the procedures have been optimized in the past months. The system works very robustly and reliably.
- PanDA pilot development test jobs: the PanDA system pilot code is the middle-ware component that executes the actual experiment payload job, such as the analysis or Monte Carlo simulation code on the site worker node and interacts with all the Grid related components. The pilot code needs to adapt to the latest Grid middle-ware code updates and experiment work-flow changes. This requires constant code validation which is done with a set of continuous HammerCloud functional tests that uses the latest pilot code version before it goes into large scale production.
- With the arrival of multiple core processors there are also special PanDA queues to use these resources in a dedicated way. These queues are tested with a constant but low rate of validation jobs similar to the previously described PFT. A queue exclusion/inclusion algorithm has been developed to white-list and black-list these special queues corresponding to the status of the regular site queues that are tested with the PFT.
- Cloud computing resources, Tier0 and higher level trigger resources are accessible through dedicated PanDA queues. The functionality of these resources are validated with the previously described AFT and PFT jobs.

- Athena Nightly Build System: stable releases of the ATLAS experiment offline software Athena are distributed through the CVMFS file system to all ATLAS Grid sites. Different nightly builds of this software are distributed with the same procedure and are tested and validated with several low rate test jobs at several Grid sites. Different Linux operating system versions and storage element types are checked for incompatibilities. Better general overview monitoring for shifters is planned for these types of jobs.
- ROOT I/O and WAN tests: To ensure that changes within ROOT versions, or changes and modifications of the analysis input file layout are tested and verified on all different storage system technology type, test jobs are sent out at a low rate to various different Grid sites. The performance of different features in ROOT like e.g. TTreeCache for local and remote input file reading are tested.
- Federated XRootD (FAX): short and low rate test jobs are sent out to FAX enabled sites to test and validate the FAX infrastructure at the site and the various global FAX components. More details will be described in the next section.

All these various tests are defined in so called templates which define the application to be used, the input datasets to be processed, the job submission and processing rate and the sites to be tested. The test results about jobs successes and failures together with links to further monitoring and debugging information are displayed on the HammerCloud web page. Figure 2 shows an example of a HammerCloud test web page.

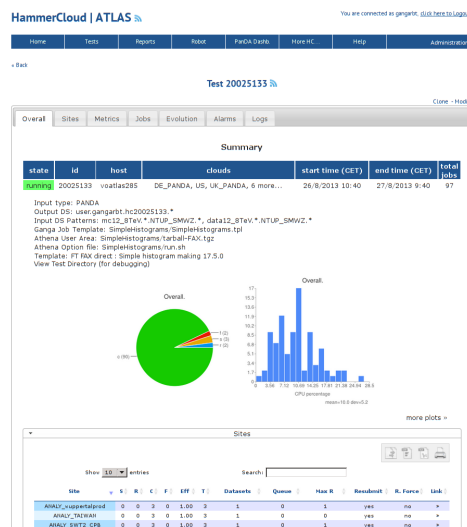


Figure 2. Example of an individual HammerCloud test web page. Here a functional test of the FAX system is shown.

4. HammerCloud FAX results

FAX (Federating ATLAS storage systems using XrootD) brings Tier1, Tier2 and Tier3 storage resources together into a common namespace, accessible from anywhere, thus relaxing the traditional requirement of data-CPU locality. Client software tools like ROOT or xrscp interact with FAX behind the scenes to reach data regardless of its location. Improved network bandwidth, reduced latency and data structure aware caching mechanism like TTreeCache makes this possible.

HammerCloud has been used to setup, test and validate the configurations of the FAX system especially in PanDA and the PanDA pilot. Since different storage technologies and

preferred input data access modes are used at every site, each site needs an individual setup for FAX testing. In these tests the default PanDA queue configurations are overwritten on a job by job basis using a PanDA pilot switch. These individual configurations are planned to be stored permanently in the ATLAS Grid information system AGIS for easy access in the future. To measure the performance of data throughput of the FAX system two typical ATLAS user analyses have been used to process regular ROOT ntuples, so called D3PDs, of detector and simulation data. Different access patterns in the FAX system have been tested: the input data access mode using direct I/O in ROOT, and the mode using a copy of the input data to the local worker node directory with xrscp. These modes have been tested with access to data at the local storage element or with remote access to the CERN storage. The FAX redirector to detect if data is available at the local site or only remotely is set up in a hierarchical manner. There is a redirector at the local site, in a so called cloud which combines a number of geographical close sites e.g. a single country and then the global redirector. The data access and discovery using the local and cloud redirector have been tested.

The tests have been performed in three steps. First the FAX configuration of the individual sites has been tested with single jobs submitted from HammerCloud to verify the correct setup. Then larger scale tests with several hundred jobs per site have been performed to test all the different data access modes described above. Finally three different functional tests with a low job submission rate have been setup to continuously monitor the FAX infrastructure functionality against site glitches or PanDA pilot software changes.

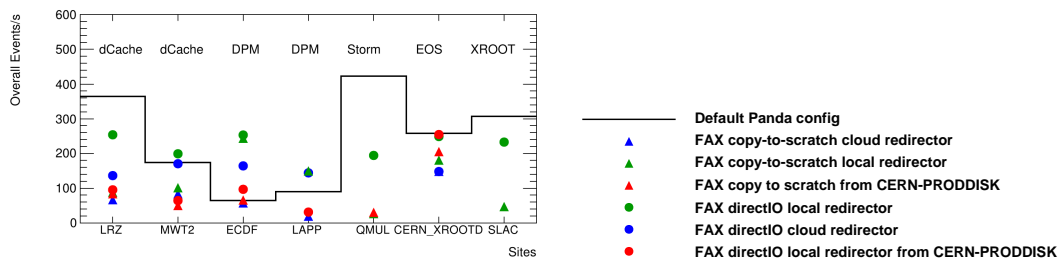


Figure 3. The distribution of several HammerCloud FAX infrastructure stress tests results. Different storage element types have been tested and event rates for several example sites are shown. The access modes have been varied throughout the different tests and are indicated with different markers as shown in the legend on the right.

Figure 3 shows some results of the HammerCloud infrastructure stress tests. The tests have been performed at more than 30 FAX enabled sites. The individual tests were executed while at the same time the regular operations of the sites continued. The test loads have been rather moderate with 20-50 parallel running jobs and lasted 48-72 hours each. The shown event rates are averaged over the test periods and have an error of about 10-20% each. Not all sites have been available for all tests due to downtimes or fully occupied resources. The plot shows the event rate for one of the described user analyses at different sites which are representative for each of the storage element technologies. Overall there is a large variation on reading speed across the sites using the default configured PanDA input access mode, which is direct I/O for dCache, Storm and EOS sites and copy-to-scratch on the worker node for DPM sites. For the FAX configuration the local redirector access with direct I/O performs the best at most sites. For DPM sites a much better performance is achieved using FAX compared to the default copy-to-scratch mode. As expected accesses to the cloud redirector or to the remote storage at CERN are slower.

The importance of TTreeCache usage in an user analysis is shown in Figure 4 (left).

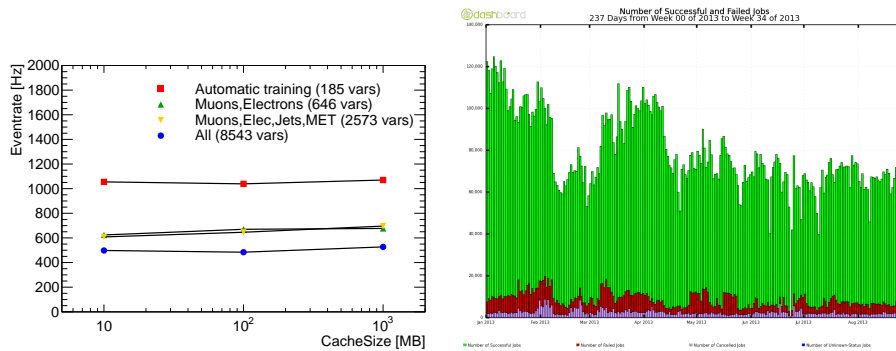


Figure 4. Left: eventrate of an example analysis reading with different ROOT TTreeCache settings. Right: Number of ATLAS jobs submitted with HammerCloud per days for 2013. Green indicates successful, red failed and magenta cancelled jobs.

TTreeCache acts as a file cache, registering automatically the baskets from the branches being processed when in the learning phase. This cache speeds-up considerably the performance, in particular when the TTree is accessed remotely via a high latency network. The user analysis described above has been tested in a HammerCloud test using different cache sizes and number of variables for training. The automatic training without previously defined variables shows the best performance independent of the cache size used.

The next steps in the FAX testing using HammerCloud are large scale stress tests over longer periods to determine the system behavior under high load. Furthermore sites using FAX will be included into the automatic PanDA queue exclusion mechanism similar to the existing system for analysis and production queues.

5. HammerCloud jobs

Overall HammerCloud is executing on average 50k functional test jobs per day. Periods with more jobs per day are due to additional running stress tests. Figure 4 (right) shows the number of ATLAS jobs submitted with HammerCloud per days for 2013.

6. Future plans and conclusions

HammerCloud is very successfully used in many aspects of ATLAS day-to-day Grid computing: functional testing with site black-listing, stress testing, and new Grid component validation as demonstrated with the FAX system setup. With the upgrade of several ATLAS Grid components before the 2015 LHC data taking there is further integration and validation work ahead for HammerCloud.

References

- [1] ATLAS Collaboration, JINST **3** (2008) S08003.
- [2] G. Duckeck *et al.* [ATLAS Collaboration], “ATLAS computing: Technical design report,” ATLAS-TDR-017; CERN-LHCC-2005-022, <https://cds.cern.ch/record/837738>.
- [3] D. C. van der Ster, J. Elmsheuser, M. Ubeda Garcia and M. Paladin, J. Phys. Conf. Ser. **331** (2011) 072036.
- [4] R. Gardner *et al.*, “Data Federation Strategies for ATLAS Using XRootD”. these proceedings.
- [5] J.T. Moscicki *et al.*, “Ganga: a tool for computational-task management and easy access to Grid resources”, Computer Physics Communications, Volume 180, Issue 11, (2009).
- [6] T. Maeno [ATLAS Collaboration], J. Phys. Conf. Ser. **119** (2008) 062036.
- [7] M. Branco *et al.* [ATLAS Collaboration], J. Phys. Conf. Ser. **119** (2008) 062017.