ATLAS operations in the GridKa T1/T2 Cloud

G Duckeck\textsuperscript{1}, T Harenberg\textsuperscript{2}, S Kalinin\textsuperscript{2}, G Kawamura\textsuperscript{3}, K Leffhalm\textsuperscript{4}, J Meyer\textsuperscript{5}, S Nderitu\textsuperscript{6}, A Olszewski\textsuperscript{7}, A Petzold\textsuperscript{8}, J Schultes\textsuperscript{2}, C Serfon\textsuperscript{1}, J E Sundermann\textsuperscript{9}, R Walker\textsuperscript{1}

\textsuperscript{1}LMU München
\textsuperscript{2}Universität Wuppertal
\textsuperscript{3}Universität Mainz
\textsuperscript{4}Desy-Zeuthen
\textsuperscript{5}Universität Göttingen
\textsuperscript{6}Universität Bonn
\textsuperscript{7}INP PAN Krakow
\textsuperscript{8}KIT Karlsruhe
\textsuperscript{9}Universität Freiburg

E-mail: gduckeck@lmu.de

\textbf{Abstract.} The ATLAS GridKa cloud consists of the GridKa Tier1 centre and 12 Tier2 sites from five countries associated to it. Over the last years a well defined and tested operation model evolved. Several core cloud services need to be operated and closely monitored: distributed data management, involving data replication, deletion and consistency checks; support for ATLAS production activities, which includes Monte Carlo simulation, reprocessing and pilot factory operation; continuous checks of data availability and performance for user analysis; software installation and database setup. Of crucial importance is good communication between sites, operations team and ATLAS as well as efficient cloud level monitoring tools. The paper gives an overview of the operations model and ATLAS services within the cloud.

1. Overview
The ATLAS GridKa cloud is one of the largest and most heterogeneous of the 10 ATLAS Tier-1 clouds. It consists of 13 Tier-1 and Tier-2 sites (Fig. 1) in 5 countries. In addition there are several Tier-3 sites with non-pledged resources which participate in cloud operations. The combined resources add up to about 50 kHS06 CPU (about 10k cores) and 5.5 PB disk storage. A further increase by about 30\% is planned for spring 2011.

Over the last years a well defined and tested operation model evolved. Several core cloud services need to be operated and closely monitored:

\begin{itemize}
\item Distributed data management
\item ATLAS production (MC simulation and real data reprocessing)
\item Data access for analysis jobs
\item Software installation and conditions database setup
\end{itemize}
Equally important are regular processing challenges to optimize site performance and data throughput and to exercise new analysis workflows which show up after LHC start, e.g. large-scale Root-based ntuple analyses.

1.1. Organisation
In order to coordinate the regular operations there are weekly operation meetings between the Tier-1 and service coordinators. These are complemented by monthly cloud video meetings including Tier-1/Tier-2 contacts and service coordinators. Wiki pages are maintained for documentation and information.

2. Data Management - DDM
The distribution and management of ATLAS data [1] is essential within the cloud and is one of the most challenging tasks. It includes

- Distribution of data from collisions, MC and reprocessing
- Organising space management (space tokens at sites etc)
- Data aggregation from MC production
- Replication of user data on request
- Data consistency checks and cleaning
- Data-recovery in case of data loss

Since the start of LHC both data volume and traffic increased substantially, see Fig. 2.

- Overall more than 4 PB data are stored in the cloud
- About 50% of this is LHC collisions data from 2010
- Peak rates exceeded 1 GB/s
- Replication of user data contributes significantly
- Data loss at sites due to disk server failures occurs regularly
- Automatic procedures are in place for quick recovery (if data available elsewhere on the Grid) [2]
Several tools have been developed to monitor and analyse in more detail the data traffic and access patterns at the sites, complementing existing ATLAS DDM tools.

Based on the detailed information in the dCache billing logs we can determine the file-access frequency of certain dataname patterns over time, distinguishing local (dcap) and remote (gridftp) access (Fig. 3). In addition, one can analyse remote data distribution: For each transfer one can extract the remote peer and then determine how much traffic goes to which sites, distinguishing incoming and outgoing connections. An example is given in Fig. 4, which shows the peer domain for gridftp transfers to or from the dCache storage at the Wuppertal Tier-2 site.

3. Monte Carlo Production

Monte Carlo production is a continuous service within the cloud. There are two Panda pilot submission instances running to provide redundancy and scaling. A large fraction is handled by the Tier-2 sites and a few Tier-3 sites also contribute. The submission instances require
supervision and good contact is needed to the site managers to aid with problem spotting and solving.

The DE cloud contributes well to overall ATLAS Monte Carlo production:

- About 12% in 2010
- Up to 16000 jobs running simultaneously in the cloud.
- In recent years the job failure rate could be continuously reduced and we reach now a walltime-efficiency of 94% (average in the cloud).

Contributions from Tier-1 and Tier-2 are well balanced and the share of analysis jobs increased substantially since LHC start as shown in Fig. 5.

4. Distributed Analysis

Since LHC start in spring 2010 group and user analysis jobs become much more prominent. Typical analysis jobs have much higher IO requirements than MC production; it is a challenge for sites to optimize overall throughput between compute cluster and storage systems and achieve good job efficiency.
4.1. I/O tests and optimization

A dedicated tool (‘dCacheLoadTest’) has been developed in the cloud in order to systematically assess the IO performance between worker node cluster and storage systems. Series of jobs on the worker nodes read randomly files from a large dataset. This way one can determine both the integral IO rate as function of jobs running in parallel (Fig. 6) as well as the IO rate of single storage pools which helps to identify bottlenecks and optimize the overall IO rate.

4.2. ATLAS HammerCloud Tests

The ATLAS HammerCloud system [3] is an invaluable tool to systematically assess and optimize the analysis performance of a site. We run regular tests of distributed analysis on all sites in the cloud to evaluate the performance and reliability of the sites and to compare options for data access, namely file-stagein to local disk versus direct IO via dcap/rfio to dCache/dpm storage systems. Figure 7 shows a recent example of a test, important goals are low job failure rate and decent CPU-time/Wall-time ratio.
4.3. Large scale Root based ntuple analysis

Certain analysis use cases require multiple and fast iterations over large data samples. Parallel Grid/batch execution (e.g. splitting one 2-day job into 100 jobs of 30 min each) can be a competitive and widely available alternative to dedicated Proof clusters. However, this poses new challenges for the IO access patterns between worker-node and storage. Moreover, such an application needs a quick turnaround, less than one hour between job submission and finishing on the Grid would be desirable.

Based on a real analysis case a realistic benchmark test-suite has been derived (G. Brandt/J. Samson/W. Ehrenfeld, Desy) for standard batch system submission. With few modifications this was ported to ATLAS Grid sub-mission and tested on several sites. An example test is shown in Fig. 8. A 1.1 TB ntuple dataset consisting of some 2000 files was processed via the ATLAS GangaPanda system [4] at the Desy-ZN Tier-2 site. The task got automatically split into 80 sub-jobs. The typical processing time varied between 30 and 50 min (depending on CPU type). Since the site was not fully loaded at that time all jobs started within 10 mins and the whole dataset processing was finished within one hour.
5. Conclusions
The sites and operations in GridKa cloud reached a decent level of general performance and stability. The experience from running in 2010 showed that sites and operations can cope well with both the increased data volume and flow since LHC start as well as the much increased analysis activity.

References
[1] I. Ueda et al., “ATLAS Operations: Experience and Evolution in the Data Taking Era”, these proceedings.
[2] C. Serfon et al., “The consistency service of the ATLAS Distributed Data Management system”, these proceedings.
[3] D. Vanderster et al., “HammerCloud: A Stress Testing System for Distributed Analysis”, these proceedings.
[4] J. Elmsheuser et al., “Reinforcing User Data Analysis with Ganga in the LHC Era: Scalability, Monitoring and User-support Ganga”, these proceedings.