Computing in Large High Energy Physics Collaborations

Sahal Yacoob[1]

School of Chemistry and Physics, University of Kwazulu-Natal, Durban

E-mail: Yacoob@ukzn.ac.za

Abstract. The ATLAS experiment at CERN is a large multi-purpose detector used to gather data from proton-proton and nucleon-nucleon collisions at the Large Hadron Collider. The Detector stands 5 stories high, and comprises greater than 150 Million readout channels, spread between a variety of detector technology in order to maximise the reach of the ATLAS scientific program The collaboration comprises over 3000 people. In order to fully characterise the detector output under different running conditions (like the energy and composition of the colliding proton bunches) which is necessary for precision measurement there is a large computing requirement to generate large samples for each possible configuration and gain insight to how the detector will behave. The smaller the uncertainties on the detector response the more precise our measurements will be. From a more physical side, ATLAS was designed to look for small signals over a large background, and requires large simulated samples in order to characterise these background physical processes (in a well calibrated detector), and reduce our uncertainties. Within this context I will outline the ATLAS distributed computing model, and the current resources available in South Africa.

1. Distributed Computing Grid

In addition to providing shared data storage and computing power, the grid computing infrastructure allows for shared expertise with trouble shooting system configuration and installation problems. ‘The Grid’ is envisioned as a global network of computers which function as one shared efficient computing resource. In practice this would be composed of local grid entities which interoperate. Grid infrastructure in South Africa (SAGrid) is independent of ATLAS -- some of these resources are shared as part of the ATLAS grid structure allowing for storage and analysis for ATLAS collaborators

2. Local Components

The first elements required for distributed computing to be successful is a local computing and networking infrastructure which can support the tasks envisioned. Local resources for distributed computing can be broken down into two major categories:

- Storage
  - the data storage system is referred to as a Storage Element (SE)
- Processing (via a local batch system)
  - local nodes accessed via the Computing Element (CE)

[1] Work done while at University of the Witwatersrand

The resources available then need to be published via an information system and paired with an authentication system in order to allow users with sufficient privileges to submit jobs to sites which host the required resources and data sets and have access to the final results.

2.1. Scale of Data Storage and Computational Tasks

Atlas has over 104 PB of data stored on the grid. Initially data (Simulated, and Real) are calibrated, reconstructed, distributed and analysed at over 100 different sites using the World-wide LHC Computing Grid and the tools produced by the ATLAS Distributed Computing project. ATLAS produces and stores information on detector status, luminosity, calibrations, alignments, and data processing conditions in relational databases, online and offline, which are made transparently available world-wide through an infrastructure consisting of distributed database replicas and web servers that exploit caching technologies[1].

3. The ATLAS Computing Model

ATLAS has a hierarchal model which is structured into tiers, as follows:

- Tier 0 (at CERN) is responsible for the archiving and distribution of the primary RAW data and provides prompt reconstruction of the calibration and express streams as well as first-pass processing of the primary event stream and automated calibration tasks.
- Tier 1 (10 worldwide) centres provide long term access and archiving of subsets of RAW data ability to reprocess local RAW data and provide access to up-to-date derived datasets for real and simulated data with low latency
- Tier 2 (approximately 80 worldwide) centres host datasets derived by individual physics groups and provide resources for analysis and simulation, there are large variations in available resources
- Tier 3 (hundreds) centres range from user desktops to centres which provide resources commensurate with the Tier 2 level. A major difference between a Tier 3 centre, and the higher tiers is that Tiers 1, and 2, have to maintain a level of availability in terms of resources and uptime and are depended upon.

3.1. Institution / User Level resources

Assuming that local sites have sufficient network connectivity and hardware resources a middleware layer will be installed. ATLAS runs on 3 middleware suites:

- gLite in most of Europe and several other countries (including all A-P countries) primarily used in South Africa (choice made by SAGrid prior to ATLAS involvement).
- ARC in Scandinavia and a few other small European countries
- VDT in the USA

Once a middleware is enabled for a particular VO (Virtual Organisation), users from that VO will have access to the resources of site concerned

The next step is the installation of ATLAS specific Applications and Serviceware which interface with the middleware and shield the users from it. They also add a lot of functionality that is ATLAS specific

3.2. Atlas Specific Services and Tools

The ATLAS Grid architecture is based on few main components:

- An Information system which publishes information about available resources, (hardware and software), usage and available datasets upstream.
- Distributed data management (DDM) which controls movement and storage of datasets, the actual details of the storage method at any individual site is transparent to the user

- Frontier and Squid for caching remote information which is often accessed locally.
- A Distributed production and analysis job management system (PanDA) which is highly automated and includes online monitoring functionality alongside distributed production (ProdSys) and analysis (Ganga/pAthena/prun) interfaces.
- Monitoring and Accounting tools to avoid job submission to problematic sites, like Hammercloud, a Distributed Analysis testing system, that can excluded sites which do not pass the a constantly running series of test jobs.

A software package is available for Tier 3 centres which allows user access to local ATLAS tools for interactive and batch execution Tools are continually evolving

4. Local GRID development

SA-GRID the South African grid initiative, assists sites to get their clusters on the grid, and provides users with help optimising their analyses for running within the GRID framework. SA-GRID are also responsible for ensuring that sites are properly registered in the correct databases.

4.1. SA-GRID developnment

SA-GRID run two EPIKH schools in SA per year (one for site admins, one for users) and one ops team development course per year [2]. SAGrid supports many independent VO's such as: Biomed (life sciences), GILDA (training), E-NMR(protein structure, physical chemistry), ATLAS, ALICE, and GEANT4.

The ATLAS community is small (4 faculty, 3 post-docs, 5-10 students) and Grid infrastructure and resources are only sustainable (and worthwhile) if utilised by researchers across many fields.

4.2. Local ATLAS growth

Two universities (UJ and Wits) allowed the ATLAS VO as of the beginning of 2010, DDM end points and data transfer paths were configured successfully in March 2011 and the first successful grid analysis jobs ran in May 2011. Analysis queues at these sites passed the automatic test to be set online (Hammercloud) for all ATLAS user jobs in September 2011.

We have received lots of help and support from the IT cloud (our tier 1), and SAGrid, we currently exists as a set of tier 3 clusters, which means that we are not responsible for providing any services to ATLAS -- this should change and we aim to progress to providing a tier 2 level of services and resources in order to meaningfully contribute resources to the experiment.

5. Challenges Related to the Distributed Computing Model

Our infrastructure allows for large bandwidth between local sites, while we limited in our international bandwidth, in order to minimise the effect on operations each local site knows contacts other local sites first when searching for required data (datasets, or run conditions). For the same reason transferring RAW data to South Africa to be analysed and returned is not feasible -- but we should be able to generate and store simulated data, and derived data formats (like a tier 2) thus contributing meaningfully as part of the ATLAS computing community, and encouraging efficient local research.
We can benefit from local mirrors and servers to minimise reliance on international bandwidth when updating, or installing products.

6. Closing Statements

Our current resources are not being utilised to their fullest, since users are unfamiliar with them and weary of the lack of available resources when compared to GRID infrastructure available globally, but resource growth is tied to usage. As our user base grows we expect to encounter new challenges, and continue to grow as a resource. we will continue to push our sites to reach this goal and in so doing we will provide better services to local ATLAS collaborators encouraging them to use local resources, and hopefully encouraging the growth of the field. This, in turn, will build local capacity for all users not just ATLAS. Distributed computing along with advances in telecommunications offers developing countries the ability to contribute to research projects that been previously unavailable due to our limited resources. We must develop and use local resources to grow the local scientific community (including but not limited to ATLAS, and particle physics). Widespread usage will show, or possibly determine the value of distributed computing and encourage investment in local resources.

References

[1] D. Barberis ATL-COMSOFT-2011-033
[2] http://www.epikh.eu/