The ATLAS Tier 1 Computing Centre at TRIUMF

by Rèda Tafirout & Isabel Trigger

Deep below the vineyards and farmlands of the French-Swiss border, members of the ATLAS collaboration are racing to complete the assembly of the largest and most complex particle detector ever built. ATLAS is a colossal endeavour. Forty-six metres long, twenty-five metres in diameter, and weighing about 7,000 tonnes, ATLAS is the bigger of the two general-purpose detectors at the Large Hadron Collider (LHC) at the CERN laboratory near Geneva. Its final assembly, in a cavern a hundred metres underground, is like putting together a giant jigsaw puzzle. About 1,500 physicists from more than 150 universities and laboratories in 35 countries are working on the project. Liquid argon hadron calorimeters, comprising hundreds of tonnes of precisely machined copper plates and sandwiching delicate readout electronics, which were designed and assembled at TRIUMF, are set to take their place in the puzzle.

ATLAS will deal with huge energies and extremely high data rates. The LHC will collide bunches of about 100 billion protons, 40 million times every second. Each proton will have a kinetic energy of 7 TeV, the energy that an electron would have after moving through a potential difference of 7 trillion volts. Each beam will store 362 Megajoules of energy, enough to melt about 400 kg of copper. Although ATLAS is big, many of its components must be aligned within tens of microns, literally a hair's breadth. Thousands of sensors will track the movements of everything from huge metallic masses to tiny silicon wafers, and their output will be recorded in databases.

Forty million times a second, twenty or more individual pairs of protons will collide and interact. Only a small fraction of the interactions will be "interesting" ones in which most of the kinetic energy of the colliding protons is converted into new, much more massive particles. These could be Z and W gauge bosons or top and bottom quarks, particles which have been produced at other accelerators, but which will be made more copiously at the LHC. More excitingly, some collisions could release enough energy to produce the elusive and much sought-after Higgs boson (the key to why matter has mass), or perhaps other new particles predicted by models which invoke supersymmetry or gravity propagating in additional dimensions.

To avoid multiple occupancy, where ionization charges from several particles in this high-rate environment hit the same wire or calorimeter cell at the same time and send overlapping signals, ATLAS has a very fine granularity, with millions of electronics channels. It is impossible, however, to analyze millions of channels 40 million times a second, so a complex trigger system looks for specific features, or "signatures," and throws away most of the uninteresting data. After all filtering, ATLAS will still produce about 320 Megabytes of raw data every second (27 Terabytes per day). As a comparison, if a year's data were stored on double-density DVDs, even without cases, the stack of DVDs would be as tall as the CN Tower.

This is where the LHC Tier 1 centres enter the picture. It is not practical to create a giant computing centre at CERN to store all the data and run all the analysis jobs. Instead, data will be distributed to ten Tier 1 centres around the world, using the Worldwide LHC Computing Grid (WLCG).

A Grid connects computers in different places, running different hardware, different software, different security systems, and using different methods of storage access without requiring individual users to have accounts on all the systems involved. This connectivity is achieved by running so-called "Middleware" (which, despite its intriguing name, has nothing at all to do with corsets and cummerbunds) on all of the individual nodes that make up the Grid. Middleware provides a convenient way to distribute large tasks to geographically distributed, heterogeneous systems.

As part of the world's most advanced network and computing Grid, TRIUMF is hosting Canada's Tier 1 computing centre for the ATLAS experiment. Funds for the centre in the amount of $10.5M ($8.0M in capital, $2.5M in operating) have been secured by a consortium of Canadian universities from the Canada Foundation for Innovation (CFI) through its Exceptional Opportunities Fund program. An additional $4.0M grant has been secured from the BC Knowledge Development Fund. There will also be a very important contribution from the computing industry in the form of discounts on hardware purchases, equivalent to several million dollars.

Each Tier 1 centre will be coupled to a set of Tier 2 centres, which are based primarily at participating universities and which will produce the simulated datasets required to understand ATLAS data. The simulated data will be uploaded to the Tier 1. User analysis will also be done at the Tier 2 centres. In Canada, by Rèda Tafirout & Isabel Trigger

the two federated Tier 2 centres, one in the East and one in the West, will be "virtual" facilities pulling together resources located at participating universities. Canada's Tier 2 centres will share equipment and resources with other research clients of the high-performance computing consortia collectively known as Compute Canada. Funding for these shared resources is secured through CFI's National Platform Fund program.

Managing the Centre

In early 2006, the TRIUMF centre moved from being a Tier 1 prototype to a fully functional system providing production-quality service for ATLAS on a continuous, around-the-clock basis.

The main missions of the TRIUMF Tier 1 centre are to receive, store and perform the final-pass reconstruction of real data from the Tier 0 centre at CERN, and to receive and store simulated data from the Tier 2 centres. Freshly reconstructed data from the ATLAS detector at CERN are copied to TRIUMF, and the reconstruction may be re-done at the Tier 1 if new calibrations need to be applied. If the load on the Tier 0 is too heavy, some of the first-pass reconstruction may also be done at TRIUMF. In either case, final reconstruction of the Analysis Object Data, which physicists ultimately analyze, is done at the Tier 1 centres. Meanwhile, sets of simulated data produced at the Canadian Tier 2 sites are also sent to TRIUMF, where they are catalogued and copied to sites around the world.

To accomplish all this, the Tier 1 runs various Grid services which provide access to the Centre's resources. A Compute Element distributes tasks to hundreds of CPUs. The Storage Resource Manager allocates Terabytes (and, in the future, Petabytes) of disk and tape storage. A File Transfer Service controls data movement channels between LCG sites. An LCG File Catalog contains information about the locations of millions of files, and the metadata needed to organize them and prevent unnecessary duplications and file transfers. An ATLAS Virtual Organization service handles datasets and file subscription, determining who orders what to be copied where and when. The Tier 1 also provides a critical database service using an Oracle Real Application Cluster, which was deployed as part of the LCG Distributed Database Deployment project. The database will hold ATLAS conditions data, including readings from all the thousands of temperature sensors, alignment monitors and gas purity gauges, which determine the calibration and alignment constants, as well as event summary data suitable for quickly selecting events of interest in a particular analysis.

Lead system administrator Denice Deatrich and networking administrator Chris Payne examine the dCache servers.

The challenge is therefore to build, maintain and operate a large-scale data-intensive analysis facility for 24/7 operations. Compliance with the WLCG Memorandum of Understanding, which was signed by TRIUMF in October of 2006, allows for a maximum integrated downtime of just one week per calendar year. Carefully planned interruptions for system maintenance, improvements and expansion will be necessary, but the rest of the time, all Grid services must be available and reliable. The system is therefore designed to be highly fault tolerant and scalable for smooth growth so that each year more computing resources may be deployed to the experiment.

The ATLAS Tier 1 Computing Centre at TRIUMF

by Rèda Tafirout & Isabel Trigger

To test the LHC experiments' computing models, several Service Challenge phases have been planned. The Service Challenges are meant to test the robustness of Grid Middleware services, storage access and networking at several sites to ensure readiness for LHC start-up in 2007. In early 2006, a test of file transfers between CERN and several Tier 1 sites, including TRIUMF, achieved an aggregate rate of one gigabyte per second throughput — a major success. Imagine copying a typical six-gigabyte movie from Switzerland to sites around the world in just six seconds and sustaining that transfer rate for days on end! (Note that this is the aggregate rate for transfers to all sites; the rate to TRIUMF alone is about one tenth of that, so if we wanted the whole movie in our example at TRIUMF, it would take about a minute to get here.)

Data movements among the Tiers are a fundamental component of the ATLAS computing model. They require high-speed dedicated networks connecting the various Tier layers. In November of 2006, a five-gigabit/s CERN TRIUMF link was commissioned in accordance with a Memorandum of Understanding among HEPNET, CANARIE and TRIUMF to establish the proper Tier 1 connectivity as part of the LHC Optical Private Network. Two dedicated one-gigabit/s lightpaths were established in 2006 to connect TRIUMF to the Canadian Tier 2 resources at the University of Victoria and the University of Toronto, and more recently a dedicated link was established between TRIUMF and McGill University. Connections to Simon Fraser University and to the University of Alberta currently use the one-gigabit/s WestGrid link but will have dedicated links in the future.

The data centre is housed in the ISAC-II building and has required important infrastructure work to bring in power, cooling and networking. The available floor space is less than a hundred square metres, so the cluster design has been carefully optimized for the space limitations. The current estimates for power consumption are 175 kW for the computing nodes (CPU), 75 kW for the disk storage system and 25 kW for the remaining components (tape library and drives, network switches, Grid services and RAC system). Think about the air conditioning you would need in a typical two-bedroom apartment illuminated by three thousand 100-Watt light bulbs. The Tier 1 air conditioning/cooling system will use about 130 kW of power at full capacity. Its design is critical for the centre, and various options have been explored. A solution using liquid-cooled heat exchangers was selected due to the very high heat density.

All personnel for the system administration and operational aspects of the Tier 1 centre were hired by the end of 2006, including a lead system administrator, and experts in Grid computing, storage, databases and networking. While each staff member has a specific area of expertise, all must be knowledgeable about other operational aspects of the Tier 1 centre to provide backup when the expert is away. Some of the Tier 1 personnel are required to travel frequently to CERN to represent the TRIUMF centre at meetings and workshops.

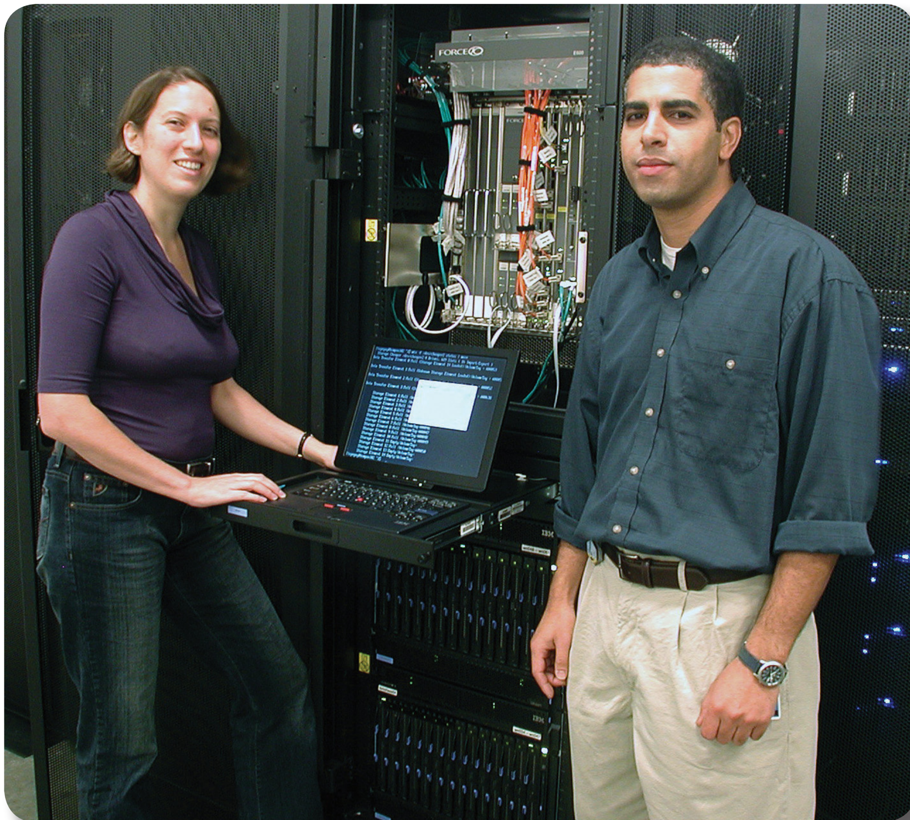
ATLAS-Canada Reaping the Benefits

When an individual Canadian physicist sits down at her computer to analyze ATLAS data, how does having a Tier 1 centre at TRIUMF benefit her? The beauty of Grid computing is that no matter where she is in the world, our ATLAS physicist can submit an analysis job to the LCG and tell it what data she wants to analyze. Grid tools then figure out where the data are stored and what nearest available processors are available. The tools can decide whether it is more efficient to copy the data to sites with free processors or to wait for resident processors. The highly refined and fully reconstructed Analysis Object Data (AOD) files, and the processors that will be available for user analysis jobs, will be at Tier 2 centres. It might therefore seem irrelevant to our physicist whether the nearest Tier 1 is in Vancouver or in Amsterdam. Indeed, as long as working on an analysis only requires access to the standard set of AOD files, it *should* be irrelevant on a day-to-day basis. On the other hand, the first few years of a particle physics experiment are usually eventful ones. The detector is new and has to be completely understood before subtle analyses can provide conclusive answers. Calibration and alignment data change rapidly. These are ingredients which feed into the final AOD production.

What then if our physicist is a calorimetry expert and his job is to calibrate the hadron calorimeters? He can't run on the AOD because his job requires access to the more detailed information only found on the Event Summary Data (ESD) stored at the Tier 1. Or perhaps he is performing a new analysis with a signal that requires special AOD files, which nobody else in ATLAS is using, or takes advantage of information which is only recorded at the ESD level.

by Rèda Tafirout & Isabel Trigger

In practice, there will be some additional resources at the TRIUMF Tier 1 over and above the Canadian share of LCG requirements, which will be dedicated to Canadian research activities. A Canadian "Virtual Organization" (VO) has been set up as a sub-group of the ATLAS VO in the LCG, so that priority queues can be set aside for Canadian users. This may be used, for example, to perform tasks like calorimeter calibration which require access to the ESD at the Tier 1.



Isabel Trigger and Rèda Tafirout with one of the control stations for the Tier 1.

Subtler advantages of having a Tier 1 centre in Canada come from the creation of a concentration of computing expertise. The 24/7/365 schedule of a Tier 1 requires that there be abundant local expertise in all fields necessary for operating the facility, as well as dedicated user-support personnel. Having these experts in one place ensures that they constantly exchange ideas and provide a pool of expertise that is a valuable resource for the Tier 2 centres.

Students at Canadian universities can apply for the appropriate Grid certificates and VO memberships. They can then plug in their laptops and run a few scripts to set up user interfaces to submit analysis jobs to the Grid. Jobs could now run anywhere on the LCG, but with priority access to Canadian processors, jobs will probably run faster if they run in Canada and copy the output back to Canadian storage elements. If students have difficulties with the software, they can ask for help from local experts. If they are in any doubt, they can consult Tier 2 personnel who, in turn, can refer them to Tier 1 personnel.

Having a Tier 1 centre gives Canada a higher profile within ATLAS and the LHC computing community and demonstrates our leadership in the field. TRIUMF will continue to attract top-quality post-doctoral fellows and computer scientists, who will transfer Grid computing expertise back into the larger community, where it will benefit Canadian high-tech industries. As computing becomes ever more critical to the field of subatomic physics, it is appropriate that Canada's national laboratory for subatomic physics should be home to a world-class computing facility.

The ATLAS Tier 1 centre at TRIUMF opens many doors: between the LCG and the heterogeneous Grid computing resources in Canada; between fundamental research and high-tech industry; between TRIUMF and our consortium of universities; between ATLAS and TRIUMF; between Canada and the world. Now is our chance to step through as many of those doors as possible, as TRIUMF and Canada join in the incredible voyage of discovery that is ATLAS.

Isabel Trigger is a TRIUMF Research Scientist with the ATLAS Group at TRIUMF.

Rèda Tafirout is a TRIUMF Research Scientist and the ATLAS Tier 1 Data Centre Manager.