THE ATLAS COMPUTING MODEL

RWL Jones, Dept. of Physics, Lancaster University, Lancaster LA1 4YB, UK

D Barberis, Dipartimento di Fisica, Universita di Genova, IT-16146 Genova, Italy

For the ATLAS Collaboration

Abstract

The ATLAS Computing Model is under continuous active development. Previous exercises [1,2,3] focussed on the Tier-0/Tier-1 interactions, with an emphasis on the resource implications and only a high-level view of the data and workflow. The work presented here considerably revises the resource implications, and attempts to describe in some detail the data and control flow from the High Level Trigger farms all the way through to the physics user. The model draws from the experience of previous and running experiments, but will be tested in the ATLAS Data Challenge 2 (DC2, described in other papers in this volume) and in the ATLAS Combined Test-beam exercises. An important part of the work is to devise the measurements and tests to be run during DC2.

THE PROBLEM

.

Introduction

ATLAS is a general-purpose particle physics experiment to be used by ~1700 scientists 6 continents. The events surviving the online triggers and filters will still correspond to several petabytes of raw data per year, which will require millions of SpecInt2k to process and analyse. Only a Grid can satisfy our requirements. ATLAS being a global collaboration, it is required to create production and analysis tools that can run tasks and manage and replicate data across multiple Grid deployments. The computing model is being tested in Data Challenges that are already using three major Grid test beds deployed worldwide.

The Structure of the Distributed System

The ATLAS model is not strictly hierarchical, but distinct roles remain. The initial processing and Raw data archival will be at the Tier-0 at CERN. ~10 Tier-1s around the world will provide robust data access and curation. Tier-2s will host user analysis, with a lower disk/CPU fraction.

Event Sizes and Data Volumes

Key to the data access and resource models are the event rates and sizes (see Table 1). The Raw data is the output from the online system. The Event Summary Data contains the detailed objects from offline reconstruction. The Analysis Object Data is a compact dataset for general physics analysis. TAG sets contain keyed summaries of the AOD sets and pointers to allow selection and rapid data access. Small part-reconstructed raw samples allow detector tuning.

	Rate (Hz)	**sec/year**	**Events/y**	**Size (MB)**	**Total (TB)**
Raw Data	**200**	**1.00E+07**	**2.00E+09**	**1.6**	**3200**
ESD	**200**	**1.00E+07**	**2.00E+09**	**0.5**	**1000**
General ESD	**180**	**1.00E+07**	**1.80E+09**	**0.5**	**900**
General AOD	**180**	**1.00E+07**	**1.80E+09**	**0.1**	**180**
General TAG	**180**	**1.00E+07**	**1.80E+09**	**0.001**	**1.8**
Calibration					**44**
MC Raw			**2.00E+08**	**2**	**200**
ESD Sim			**2.00E+08**	**0.5**	**50**
AOD Sim			**2.00E+08**	**0.1**	**10**
TAG sim			**2.00E+08**	**0.001**	**0**

Table 1: Data volumes for a single copy of the various formats.

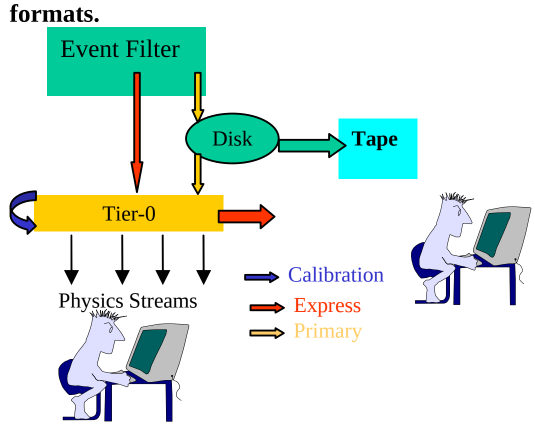

Figure 1: The ATLAS first-pass processing scheme.

The Production Model

The production model as shown in Figure 1 envisages three streams of data entering the T0: a primary physics stream of all events; a calibration stream to reduce the latency on the primary processing; and an express stream for new physics alerts and detector and algorithm tuning

After Tier-0 processing, the events are streamed to optimize access. The raw data is shared between the Tier-1 facilities for later processing each of N Tier-1 holds 2/N of the ESD on disk. It is intended to reprocess new data with new calibrations after 1-2 months. At the year-end, all of the cumulative Raw data will be reprocessed with improved algorithms and calibrations.

The Tier-1 cloud also allows scheduled physics group access to the full ESD sample, producing collections and Derived Physics Data (~100 full ESD reads per year). These are distributed to Tier-2 facilities for user analysis.

Small Raw and ESD samples are also copied to the Tier-2s to allow the user to develop algorithms and calibrations. The Tier-2s (and local group clusters) are the primary centres for chaotic user analysis. The Tier-2 also have a very important role in the production of the simulated data, which is then shipped to the Tier-1s for storage.

As well as the external user community (~600-700 users) there will be a significant number of ~100 CERN-based users and users wishing to make use of a super-Tier-2 facility. This does not carry the usual simulation load, but keeps larger than usual samples of data on disk and has occasional access to the data retained in the Castor mass store.

Deployment, Tests and Data Challenges

The ATLAS Computing Model is currently under test in a world-wide data challenge. This has three phases, all of which test the integration with the Grid deployments (the Challenge is exclusively Grid-based), bottlenecks and the effectiveness and accuracy of the information systems:

- The production phase, testing the production system, real-world network and middleware response and generating data for the later phases
- The Tier-0/Tier-1 phase, emulating the first-pass data processing and data transfer to the Tier-1 centers.
- The data analysis phase, testing the event model, data distribution and chaotic access patterns.

The exercise also tests the human and organizational aspects of the model: the required production team size, roles and distribution; the role of local managers; the interactions with other users and other scheduling policies; and the role of mutual-coverage between sites in operations.

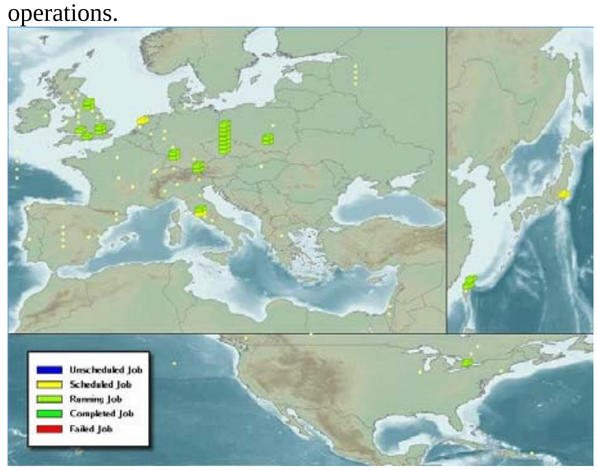

Figure 3: ATLAS Data Challenge 2 production uses facilities worldwide using the LHC Computing Grid deployment (pictured), Grid3 and NorduGrid.

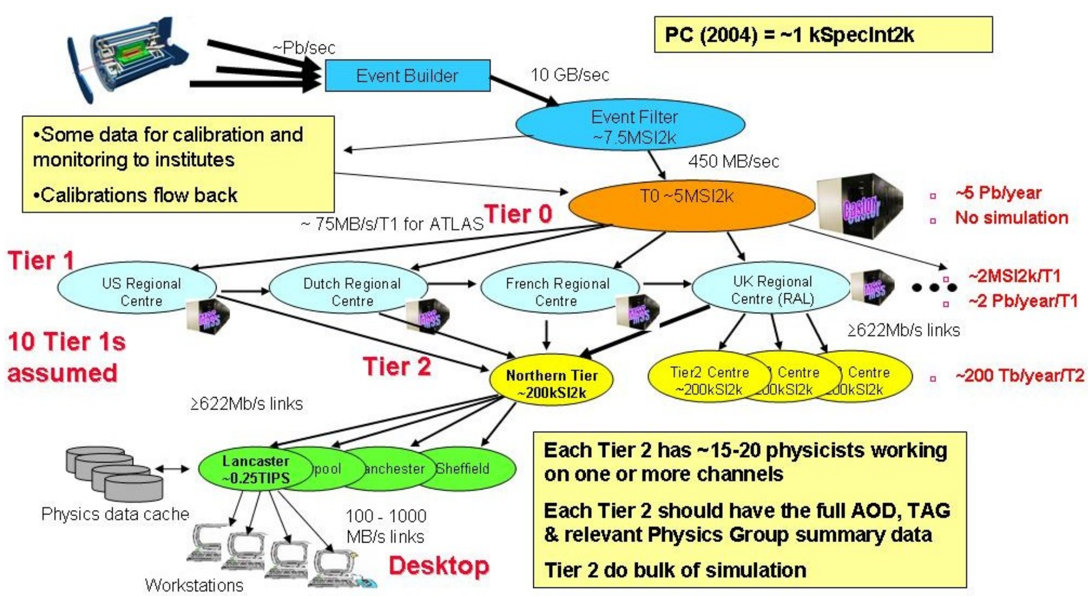

Figure 2: A schematic summarising the expected data flows and processing capacities in the ATLAS Computing model.

Network

The demands on the network between the Tier-0 and Tier-1s are well established at ~3Gbps (~75MBps raw rate for ATLAS), with similar capacities required between the Tier-1s. (Additional loads from traffic direct from the online system are being evaluated.) Future work will determine the network requirements between the Tier-2s and the other Tiers; initial work suggests 622Mbps would be sufficient, but 1 Gbps would be more effective. Dedicated tests of the network performance and data transfer tools are planned, with specific investigations of light path connectivity.

Conclusions and Plans

A document for external review describing the Computing Model will be produced by the end of the year, and the full ATLAS Computing TDR is due in Summer 2005. The model described above encapsulates the steady-state system for ATLAS computing. The overall resource requirements for the first year are shown in Table 2. At the start of data-taking, there will be particular pressures on the system, and more recourse will be had to the Raw and ESD data formats. Current work is focused on the best way to address this transient need within the resource limits set by the steady-state model, and without disrupting the final system.

	CERN	All T1	All T2	Total
Tape (PB)	**4.3**	**6.0**	**0.0**	**10.3**
Disk (PB)	**0.7**	**9.2**	**5.4**	**15.3**
CPU (MSI2k)	**5.6**	**16.6**	**5.7**	**27.9**

Table 2: The required overall capacity in the various Tiers for one year of data-taking. Approximately 10 Tier-1s and 30-40 Tier-2s are expected.

REFERENCES

[1] ATLAS Computing Technical Proposal, CERN LHCC/96-43 (1996).

[2] Report of the Steering Group of the LHC Computing Review. CERN/LHCC/2001-4

[3] Principles of cost sharing for the ATLAS Offline Computing Requirements, September 2002..