The Evolving role of Tier-2s in ATLAS with the new Computing and Data Distribution Model

S. González de la Hoz for the ATLAS Collaboration

IFIC (CSIC/UV) Edificio Institutos de Investigación, 22085, E-46071 Valencia, Spain

ATLAS Computing and Data Distribution Model

- Total Luminosity recorded at the end of 2011 was **5.25 fb^{-1}**.
- Data exported to the the tiered GRID hierarchy more than **25k TB** since January 2010.
- In the original Computing Model the role of **Tier-1, Tier-2** and **Tier-3** is:

Tier-1 (10 sites)
- Share raw and reconstructed data
- Reprocess data
- Distribute data to tier-2
- Host some grid services (LFC, FTS) for its cloud (Tier-1 and nTier-2s)

Tier-2 (~ 80 sites)
- Take care of analysis data on disk
- Provide computing power for analysis and simulation

Tier-3
- From Tier-2 like activity without pledge resources to local batch cluster

- Initially thought as a **hierarchical infrastructure:**
  - Tier-2 could only receive data through their associated Tier-1.
  - Task could run at the chosen Tier-1 and any of its associated Tier-2s.

- The **original model was a working starting point** but
1) Tier-2 activity was strongly linked to the associated Tier-1 reliability:
   - Logical File Catalog (LFC) downtime at Tier 1 -> No grid activity in associated Tier-2
   - Tier- 1 Storage downtime -> Stop production and data distribution to/from the Tier-2
2) Tier-2 had the computing resources to do reprocessing but limited due to required direct access to the Tier-1 database

- **Operational Improvements:**
  - The existing network provided good connectivity to many Tier-1s or Tier-2s -> possibility to make direct transfers.
  - Frontier/squid is a http-based system to access database from remote sites.
  - CernVM-FS is a network file system based on http. ATLAS software releases and the smaller file-based database are now installed on the server at CERN.

Evolution of Computing the ATLAS Model
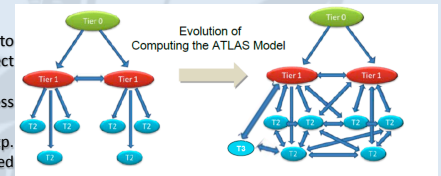
Data Distribution and Processing activities in the first years
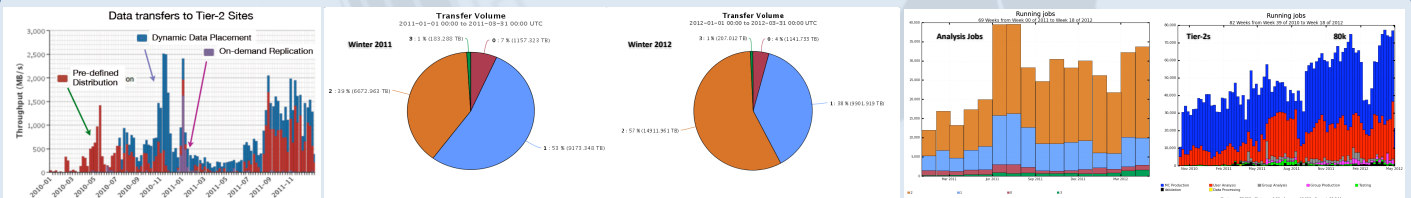
- New **data distribution policy and the new Dynamic data placement** algorithm were deployed in Sep11:
  - Pre-defined distribution explains the increase in Sep11. The goal was to preplace data in Tier-2 so that the first analysis jobs could run at Tier-2 instead of Tier-1.
  - To optimize the analysis and production activities.
  - It has significantly changed the data volume transferred to Tier-2 sites.
  - Data replicas are distributed at Tier-1s for redundancy and at Tier-2s for analysis.
  - Dynamic placement of data replicas at Tier-2s based on usage as well as an on-demand replication system.

- Tier-2s with good connectivity can get data from Tier-1s and other Tier-2s:
  - Data transfer volume in the Tier sites from winter 2011 (left) and winter 2012 (right). The replication for Tier-2s has significantly increased.
  - Multi-cloud production and direct inter-cloud transfer.
  - Tier-2 sites are getting now more datasets than Tier-1.
  - Extra replicas of popular data (used and reused very often) using remaining available disk space at Tier-2s.
  - There is more Monte Carlo simulation in Tier-2s than in Tier-1.

- Three main data processing activities: **Official productions, End-User analysis and Group activities.**
  - Official Monte Carlo simulation production has been running at Tier-1s and Tier-2 constantly since before the start of data taking together with the reprocessing of detector data.
  - End-user physics analysis on the Grid started rising since the start of data taking on March
  - Group activities started as "end-user analysis" of the group of physics analysis responsible of producing common data for end-user analysis.
  - A large part of the analysis and the MC production is done at Tier-2s.
  - The sudden increase of analysis activities in spring11 is due to summer conferences.

Network model: Availability and Connectivity

- **Test are done to check Tier-2 are well connected (T2D).**

  - At the moment, the limit is set to an overall transfer performance of at least 5MB/s to at least 10 of the 12 Tier-1s in ATLAS, for very large files (>1GB).
  - ATLAS monitors transfers among sites through the **File Transfer Service (FTS)**. ATLAS evaluates the connectivity from transfer speed of large files.
  - ATLAS can check for example how the transfers perform from and to a Tier-2 from every site, within the complete NxN matrix containing all the transfers.
  - ATLAS uses the **HammerCloud** framework to test its availability, by constantly submitting typical analysis jobs to every site.
  - Site availability for analysis jobs is defined based on the analysis panda queue status: (the amount of time queue status = online in the period) / (the whole period).
  - In terms of availability, every Tier-2 is analysed at the end of each month, and classified in one of **four categories :**
    - **alpha** – availability >90% if site is also a T2D (~ 30 sites)
    - **bravo** - availability >90% but site is not a T2D (~ 20 sites)
    - **charlie** - availability >80% (~5 sites)
    - **delta** - availability <80% (~10 sites)
  - The indicated numbers of sites correspond to the status of March 2012.
  - The site reliability is assumed to be correlated between two consecutive months:
    - Input data for analysis are preferentially distributed to "good" sites.
  - Sites in downtime do not get data

Conclusion and Prospect

- Computing and data distributed model continues to evolve and improve beyond the original data processing model.
  - ATLAS is monitoring all activities, sites, network, etc and running functional tests for that purpose.
  - LHCONE (LHC Open Network Environment) is going to provide a collection of access locations that are effectively entry points into the private LHC Tier-1/2/3 network in order to improve the connectivity and to complement the LHCOPN.

- Tier-2 activities are now less dependent to Tier-1➔Tier-2 participate to more critical activities in ATLAS.
  - Tier-2s are again receiving data immediately to have a higher contribution to analysis activity.
  - Their availability is reported every month, requiring a good connectivity for data transfer and production.
  - Central production jobs run at Tier-2s because input data are already available at the Tier-2 site. The brokering system (Bamboo) now take into account the replicas in Tier-2s.

GOBIERNO DE ESPAÑA MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

IFIC Instituto de Física CORPUSCULAR