ATLAS Distributed Computing on the Grid

Dario Barberis

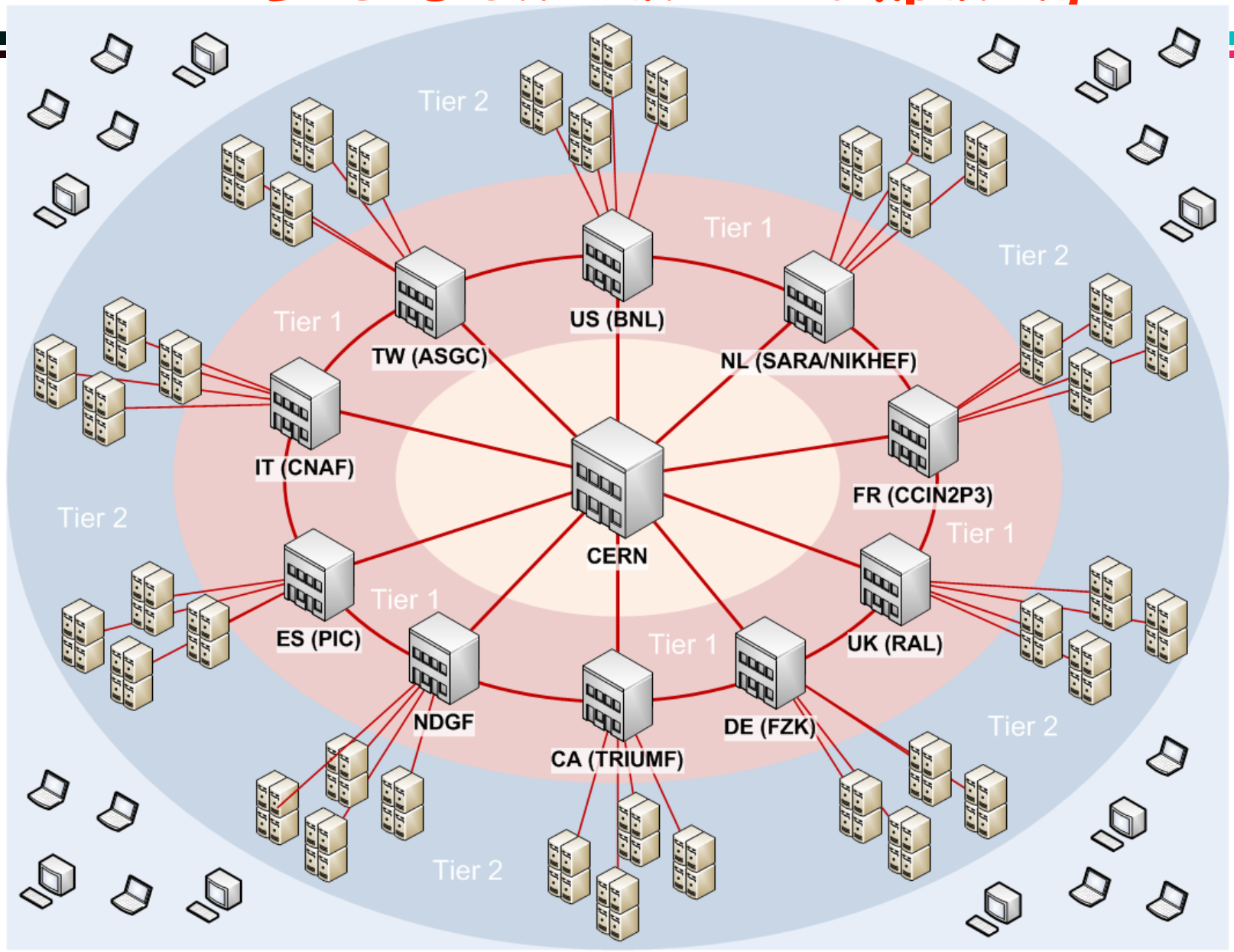
Genoa University/INFN & CERN

Overview

- Summary of the ATLAS Computing Model
- Some of the key technologies:
  - Data management and distribution with DQ2
  - Workload management with Panda
  - Conditions Databases with Frontier
- Some of the key activities:
  - Grid Data Processing
  - Distributed User Analysis
  - User Support
  - Tier-3s (on Friday)

Acknowledgements: D. van der Ster, K. Bos, S. Campana , D. Front, A. Klimentov, M. Lamanna, D. Smith

ATLAS Distributed Computing



Computing Tasks per Tier

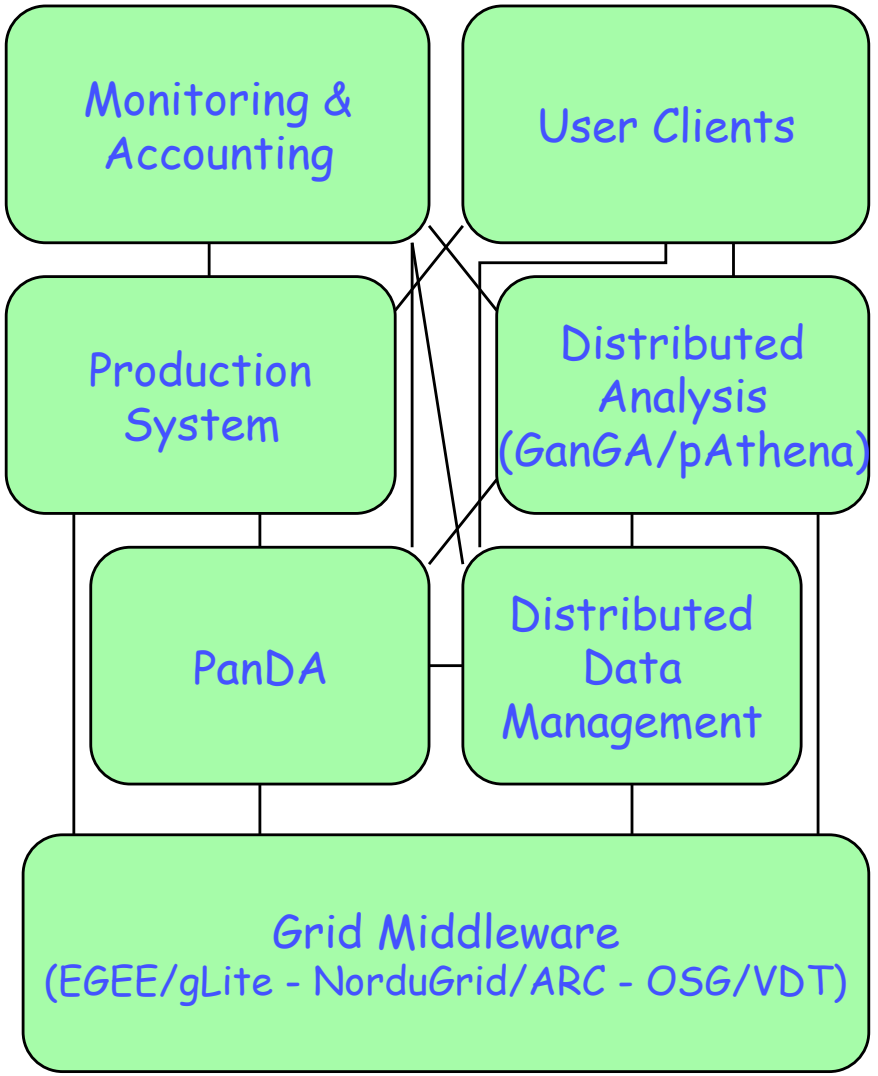
- Tier-0 (CERN)
  - RAW Detector Data Acquisition and archive to tape
  - Calibration and Alignment
  - First processing
  - Data distribution to Tier-1s
- Tier-1s (10 big computer centres)
  - One Tier-1 at the head of each *cloud*
  - Archive a share of the RAW Detector Data to tape (2nd copy)
  - Re-process those data when needed (new software, new calibration)
  - Archive simulated data to tape and reconstruct when needed
  - Bulk analysis jobs but also user analysis in some cases
  - Data distribution to Tier-2s
- Tier-2s (70 mid-size computer centres)
  - Many attached to a Tier-1 to form a cloud
  - Simulation production
  - User analysis
- Tier-3s (100 (?) home institutes, faculty facilities)
  - End user analysis
  - Non-pledged resources; not under central ATLAS control

Computing Model Principles

- RAW data master copy stored at CERN
- RAW data distributed over all Tier-1s
  - Tier-1 is responsible for preserving data on tape
  - And recall it for re-processing
- Cloud independence: all derived data available in each cloud
  - Generally, there should be a cloud with free CPUs
  - Generally, data should not have to move between clouds
- All data is pre-placed in each cloud
  - For controlled processing in Tier-1s
  - For user analysis in Tier-2s
    - (but evolution in progress right now)
- New data produced in a cloud should be archived there
  - Only Tier-1s are required to have tape archives
  - Also true for the Tier-0 (CERN)

ATLAS Grid Architecture

- ATLAS runs on 3 middleware suites:
  - gLite in most of Europe and several other countries
  - ARC in Scandinavia and a few other small European countries
  - VDT in the USA
- ATLAS Grid tools interface with the middleware and shield the users from it
  - They also add a lot of functionality that is ATLAS specific
- The ATLAS Grid architecture is based on few main components:
  - Distributed Data Management (DDM)
  - Distributed Production System (ProdSys/ PanDA)
  - Distributed Analysis (GanGA/pAthena)
  - Monitoring and Accounting
- DDM is the central link between all components
  - As data access is needed for any processing and analysis step!



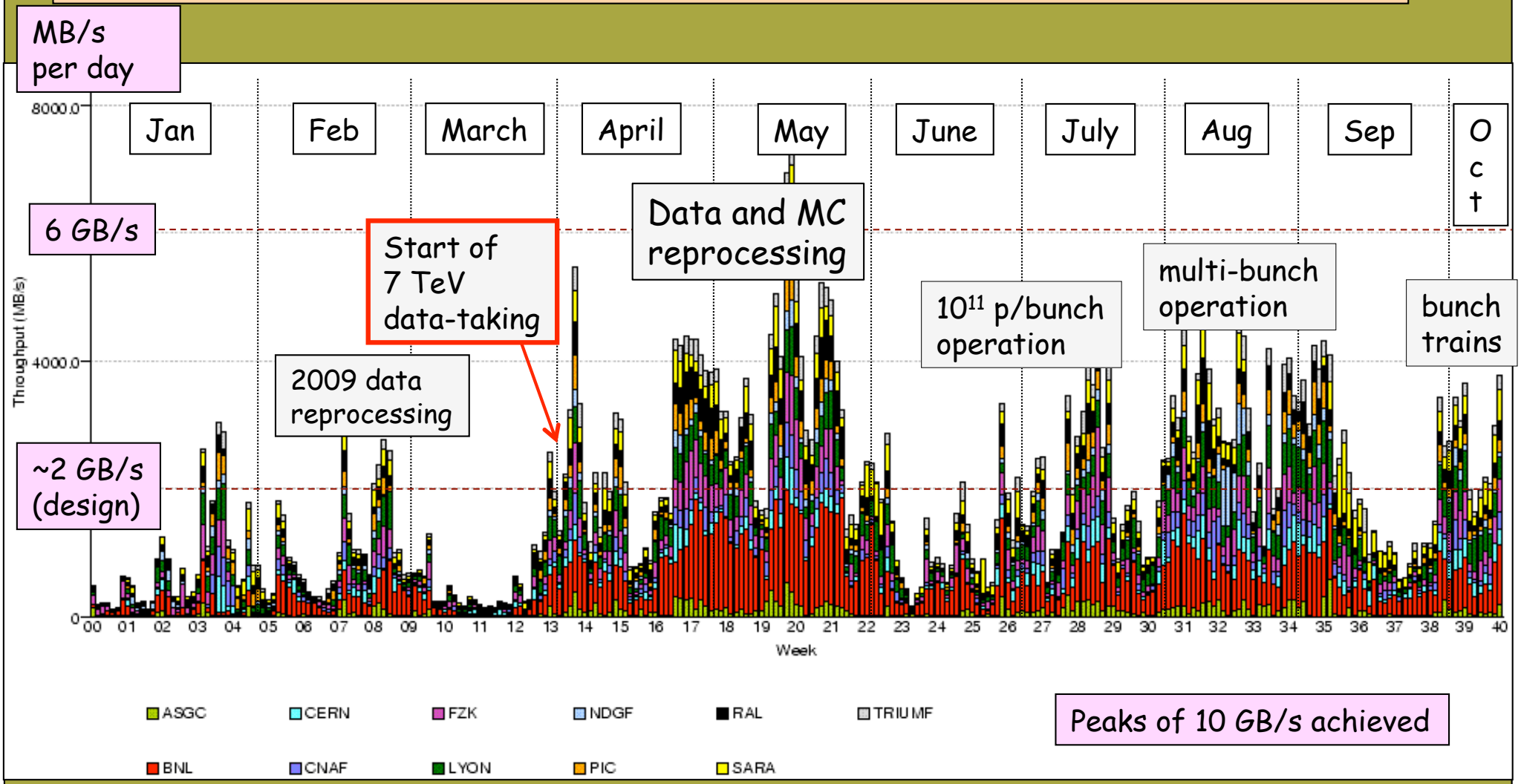
Distributed Data Mgmt: DDM/DQ2

- The Distributed Data Management (DDM) architecture is implemented in the current DQ2 tools

- The unit of storage and transfer is the dataset:
  - A dataset contains all files with statistically equivalent events
- DDM takes care of:
  - Distributing data produced by Tier-0 to Tier-1s and Tier-2s
  - Distributing simulated and reprocessed data produced by Tier-1/2s
  - Distributing user and group datasets as requested
  - Managing data movement generated by production activities
  - Cataloguing datasets (files, sizes, locations etc.)
  - Checking the consistency between the contents of ATLAS catalogues (LFC), local SRM databases and actual files on disk or tape
  - Providing usage information for each dataset replica
  - Deleting obsolete or unnecessary replicas of datasets from disk when unused
  - Providing end-users with client tools to operate on datasets (import/export/move etc)

Worldwide data distribution and analysis

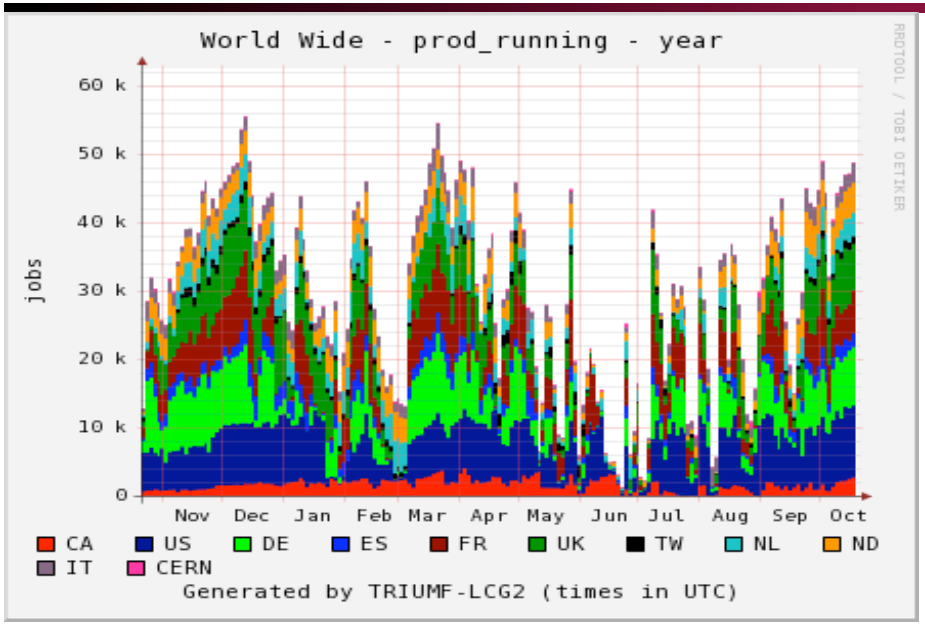
Total throughput of ATLAS data through the Grid: 1st January → mid-October

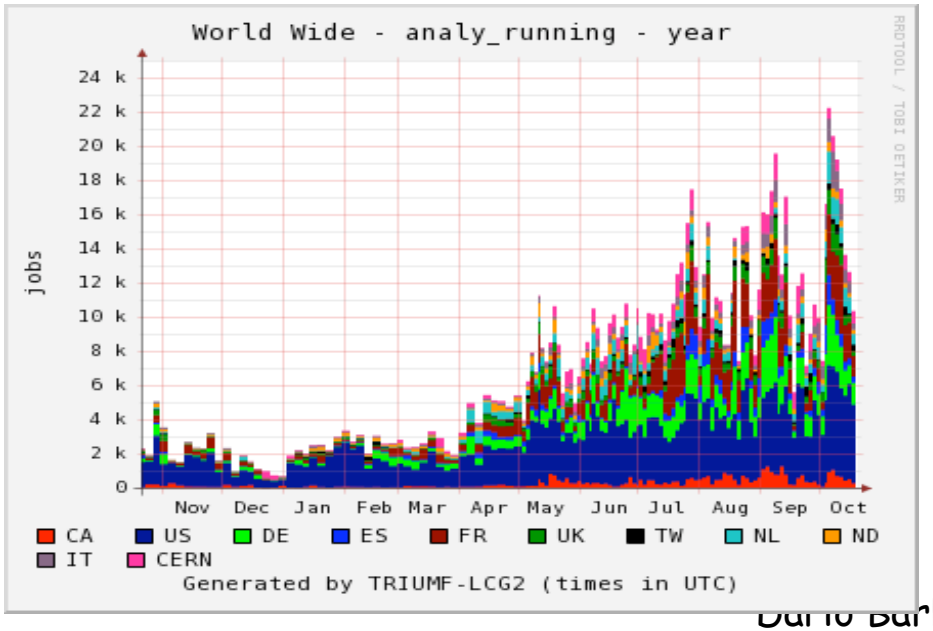


Grid-based analysis in Summer 2010: > 1000 different users; > 15M analysis jobs

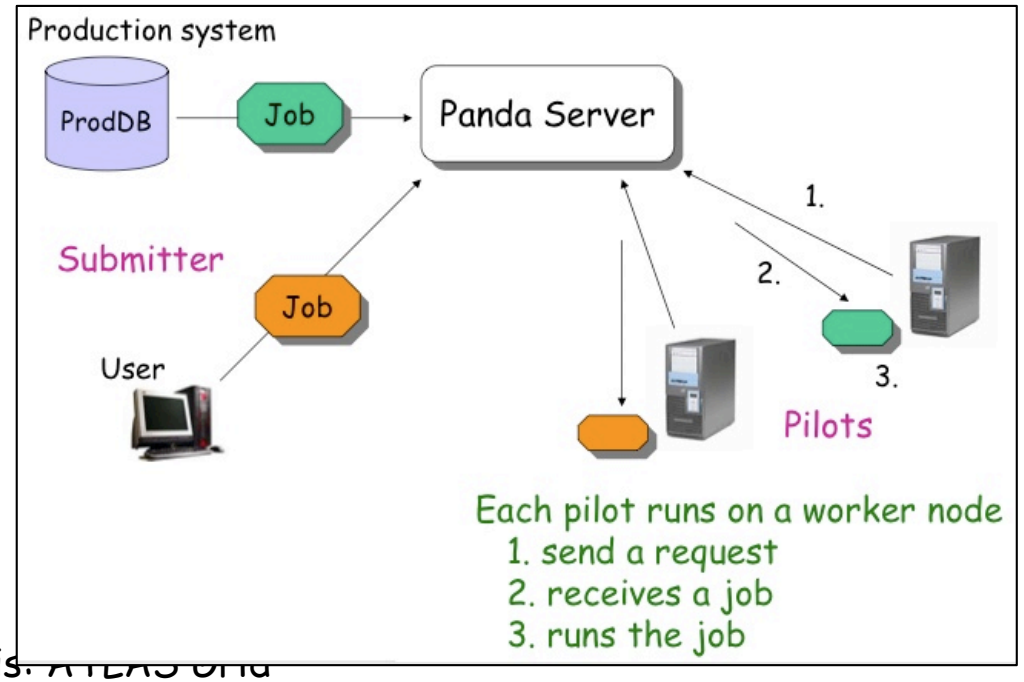
The excellent Grid performance has been crucial for fast release of physics results.
E.g.: ICHEP: the full data sample taken until Monday was shown at the conference on Friday

Workload Management: PanDA



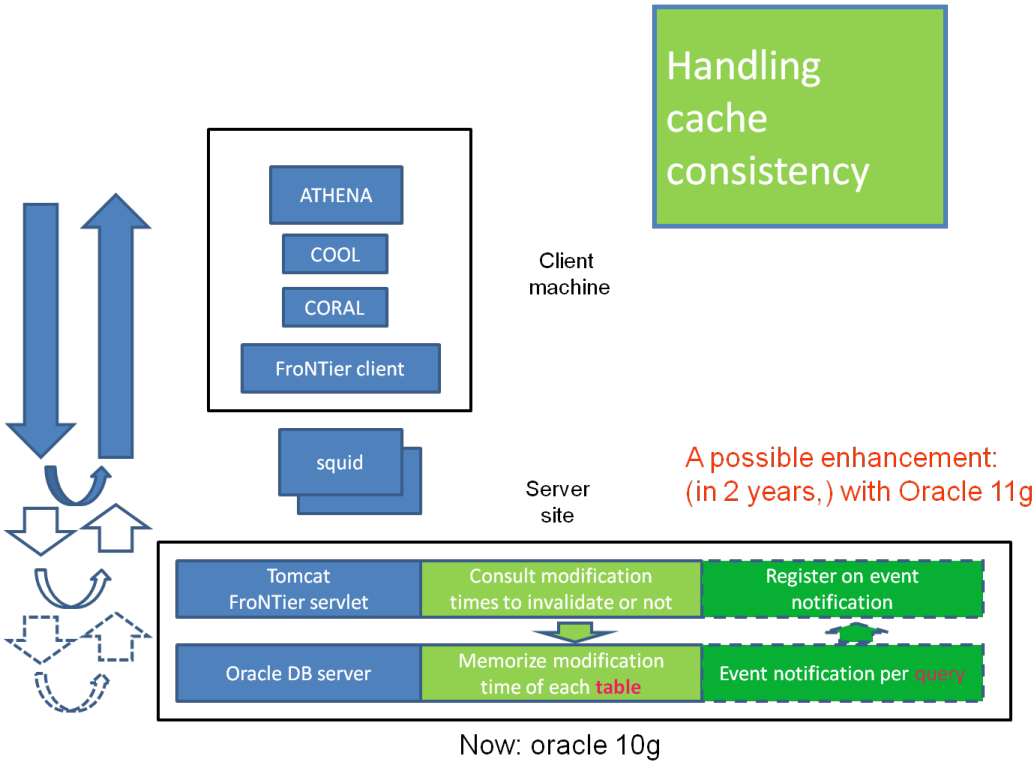


- PanDA is used to run all MC and reprocessing, and most of the user analysis worldwide
- PanDA@CERN deployed >1 year ago and is running successfully.
- The service was well prepared thanks to pre-exercises such as STEP'09
- Panda load depends more on the number of resources (~70 sites), and less so with the amount of data



Conditions Databases

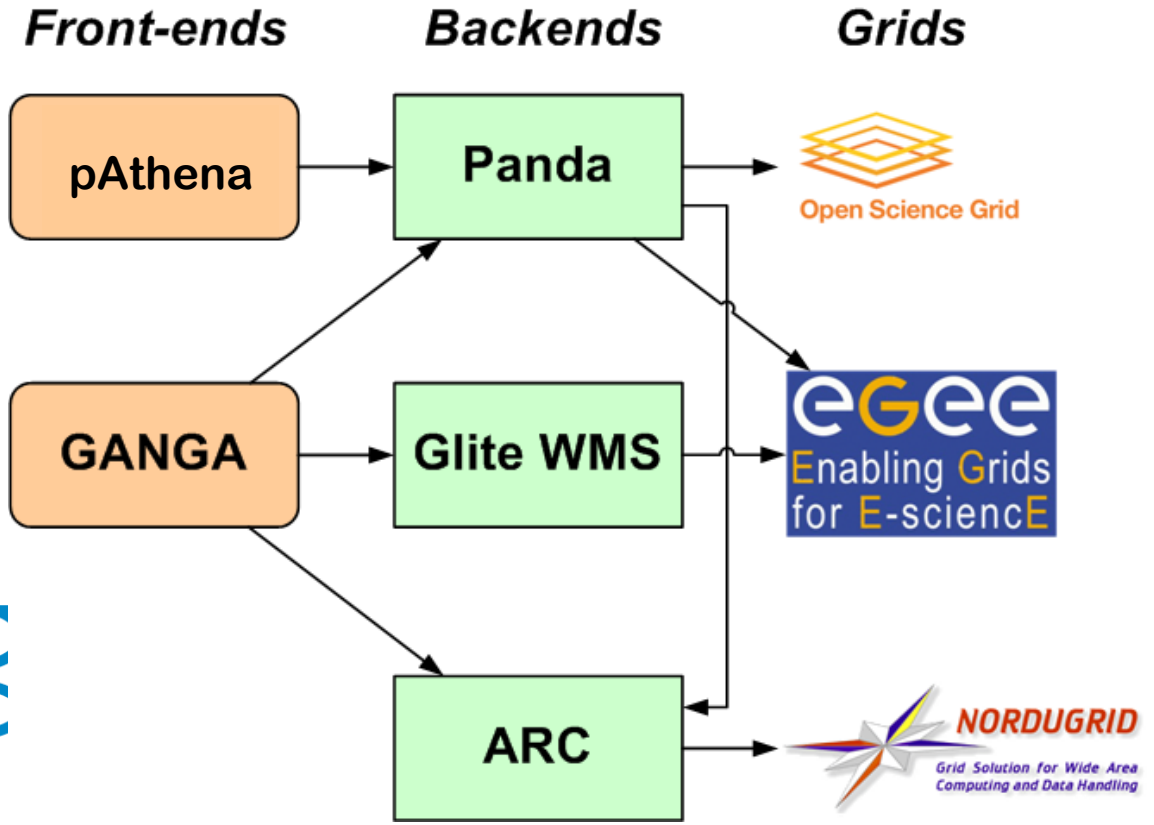
- Frontier deployed to enable distributed access to the conditions DB
- Working toward making it more transparent to the end users



Map of installed Squids

- Frontier reduces considerably the access time to DB data from remote sites
- It is particularly important for sites with low bandwidth and high latency towards Oracle servers

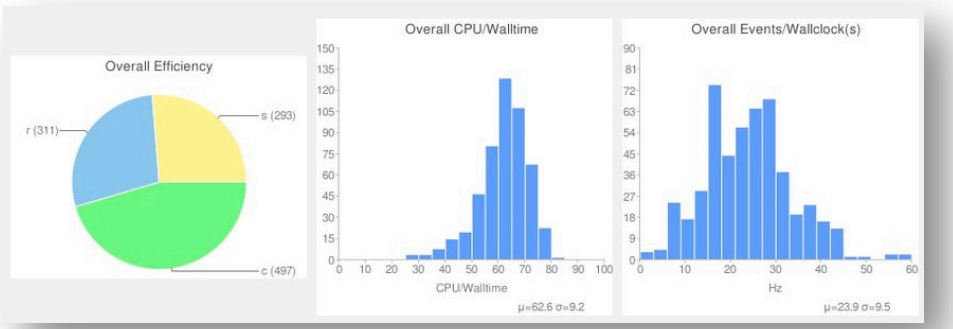
Distributed Analysis



- Basic model: data is pre-distributed to the sites, jobs are brokered to a site having the data
- Large dataset containers are distributed across clouds, so the front-ends do not restrict jobs to a cloud. i.e. DA jobs run anywhere in the world.

DA Functional and Stress Testing

- We pre-validate sites for distributed analysis with Functional and Stress tests:
  - GangaRobot is running a continuous stream of short user analysis jobs at all Grid sites
    - Results fed into SAM
    - Manual or automatic blacklisting
  - HammerCloud is used for on-demand stress tests spanning one or many sites
    - Used to commission new sites, tune the performance at existing sites, and to benchmark sites to make comparisons
- HammerCloud
  - Invested ~200k CPU-days of stress testing jobs since late 2008



Supporting a Thousand Users

- We have ~1000 active distributed analysis users
  - They should not need to be distributed computing experts – The Grid is a black box that should just work
  - Grid workflows are still being tuned – not everything is 100% naïve user-proof
  - Supporting the users to get real work done is critical (it will stay like this!)
- ATLAS introduced a team of expert user support shifters in fall 2008.
- DAST: Distributed Analysis Support Team
  - Class 2 (off-site) ATLAS shifts; week-long shifts in EU and NA time zones (Asia-Pacific shifters wanted…)
  - 1st and 2nd-level support: better incorporate new shifters and shares the load in times of high demand
  - DAST is a ~15 member team; each takes a shift every 4-8 weeks.
- Users discuss all problems on a single "DA Help" eGroup
  - Discussion about all grid tools, workflows, problems
  - Not just DA – also data management questions

Enabling the Tier-3s

- Enabling Tier-3 activity is the next essential step in ATLAS Computing
- Formed working groups in February 2010 to study:
  - Distributed storage (Lustre/Xrootd/GPFS)
  - DDM-Tier3 link
  - Tier-3 Support
  - PROOF
  - Software / Conditions Data
  - Virtualization
- WG's have wrapped up last Summer. Sites are starting
  to get connected.
- Tier-3 co-ordinators appointed:
  - Tier-3 co-ordinator: Andrej Filipčič (Ljubljana)
  - Technical co-ordinator: Doug Benjamin (Duke)
- Much more on Tier-3s Friday morning!

Summary and Outlook

- The ATLAS Distributed Computing infrastructure is working thanks to many efforts in preparation
- We are able to
  - process, distribute, and reprocess the data
  - analyse the data
  - provide support to our large community
- and we are tackling the next frontier: Tier-3s
- *As we get experience with reality we are looking at the evolution of the model and our implementations, e.g.*
  - Less-strict cloud model?
  - Better data distribution for analysis?