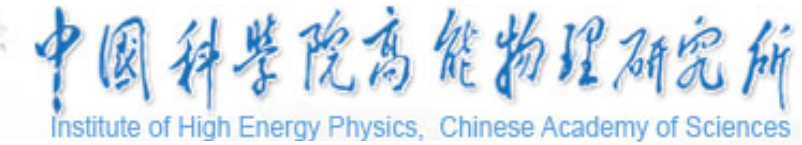
Data Management in ATLAS

Angelos Molfetas on behalf of the ATLAS DQ2 team

ATLAS DDM COLLABORATION

A.Molfetas (CERN), F.Barreiro (CERN), A.Tykhonov (Jožef Stefan Institute), V.Garonne (CERN), S.Campana (CERN), M.Lassnig (CERN), M.Barisits (Vienna University of Technology), D.Zang (Institute of high energy physics, Chinese Academy of Sciences), C.Serfon (LMU Munich), P.Calfayan (LMU Munich), D.Oleynik (Joint Institute for Nuclear Research), D.Kekelidze (Joint Institute for Nuclear Research) , A.Petrosyan (Joint Institute for Nuclear Research), S.Jezequel (IN2P3), I.Ueda (University of Tokyo), Gancho Dimitrov (Deutsches Elektronen-Synchrotron), Florbela Tique Aires Viegas (CERN)



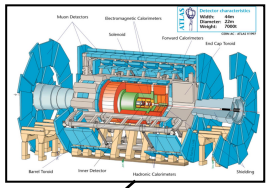


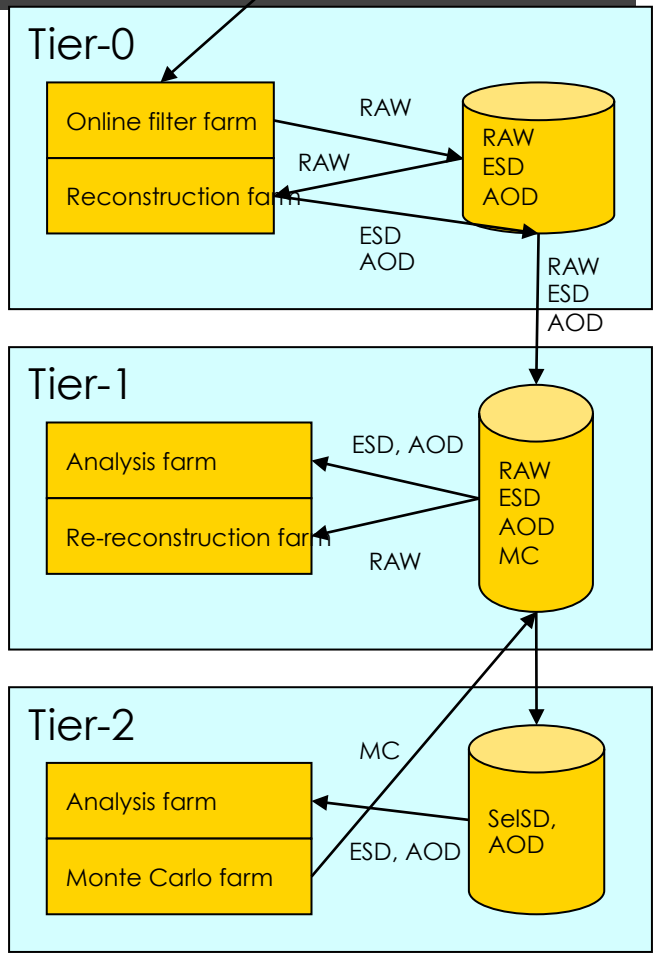
Introduction

- Presentation intended for a general audience
- Current issues & trends
- Covers some of the issues we are facing in ATLAS Distributed Data Management (DDM)
- ATLAS grid:
  - Over 800 end points
  - Petabytes of data managed on the grid
  - System responsible for this is DQ2 middleware

The ATLAS Computing Model

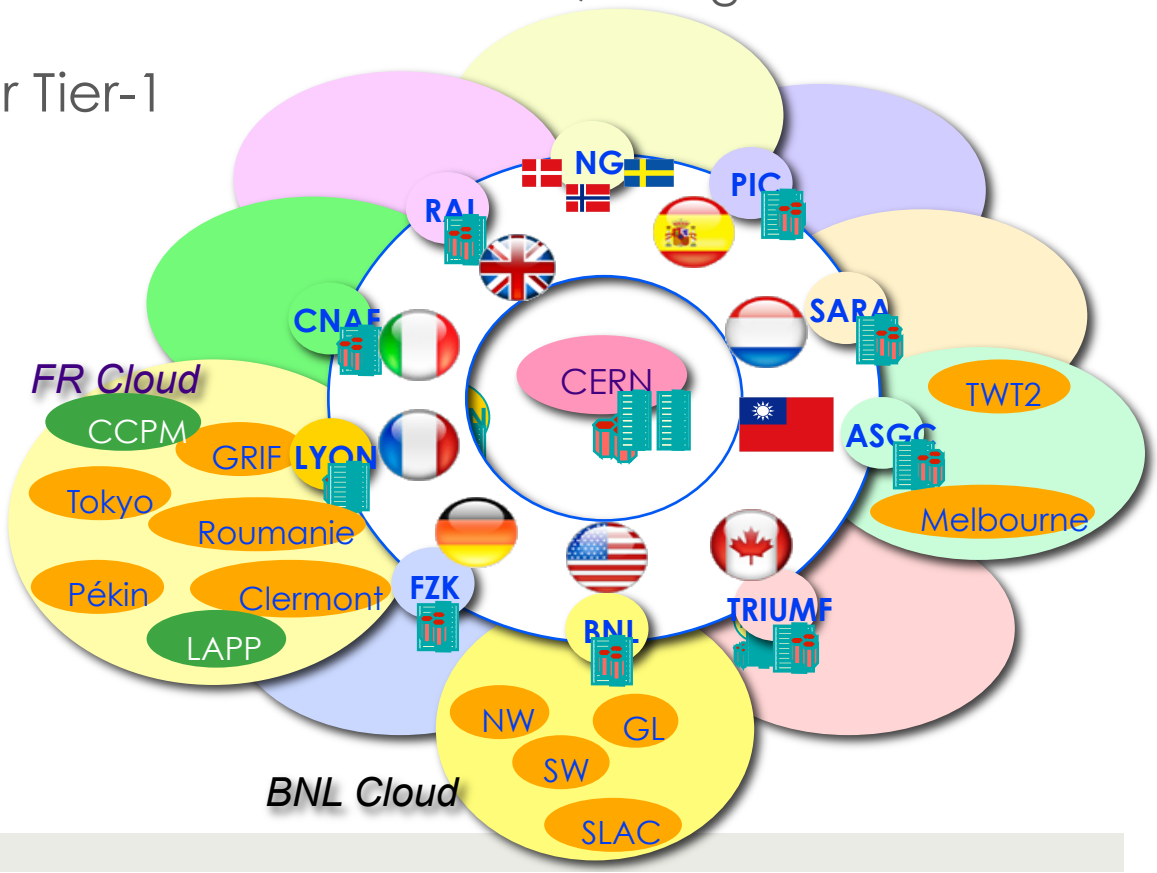
- Grid Sites are organised in Tiers
  - Tier-0
    - record RAW detector data
    - distributed data to Tier-1s
    - calibration and first-pass reconstruction
  - Tier-1s
    - permanent storage
    - capacity for reprocessing and bulk analysis
  - Tier-2s
    - Monte-Carlo simulation
    - user analysis





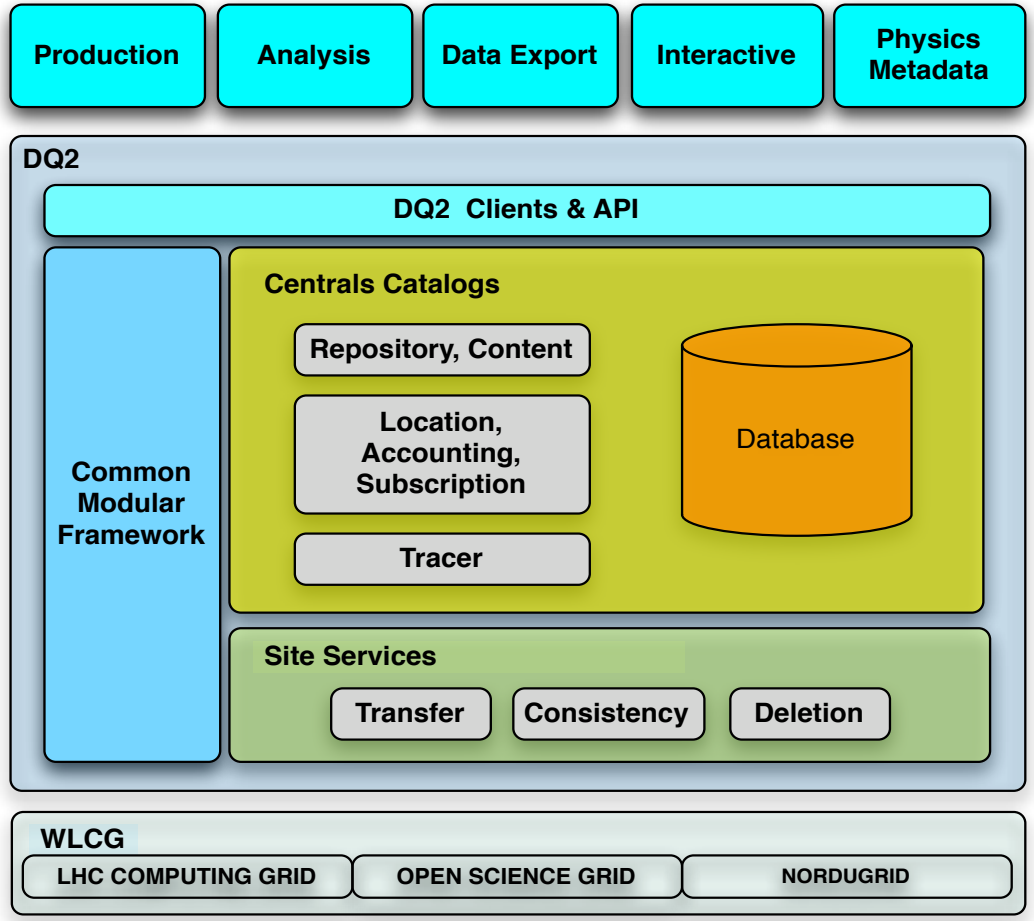
The ATLAS Computing Model

- Sites are also organised in clouds
  - not the "computer science" definition of clouds, though!
- Every cloud has a major Tier-1 and associated Tier-2s
- Mostly geographical and/or political
  - support
  - deployment
  - funding



DQ2 (Don Quijote 2)

- DQ2 enforces dataset
  - placement
  - replication
  - deletion
  - access
  - consistency
  - monitoring
  - accounting



Managing Heterogeneous resources

- Users need to be able to:
  - Download/Upload data from the grid
  - Transfer data between sites
  - User should not need to know about each storage system
- Many different mass storage systems are used - we need a simplified interface that hides the grid's heterogeneity.
  - Not trivial
  - In ATLAS this is done by DQ2 middleware and abstraction layers like SRM
- For example:
  - User downloads dataset by CLI: "dq2-get user.angelos.xxxxxxx"
  - No specific knowledge is required about castor, dcache, xrootd, etc.

Catalogs

- Maintain global state of data (central catalog of all datasets on the grid)
  - This has to scale
  - Central point of failure
- In ATLAS we have Local File Catalogs (LFC) which also have to be maintained.
- For example, uploading data to the grid:
  - Dq2-put –s files_location user.angelos.xxxxxxx
  - Has to handle different storage systems
  - Has to register files in central catalogs
  - Has to register files in LFC
- Not trivial. E.g. order of operations in dq2-put can create dark data

Maintaining Consistency

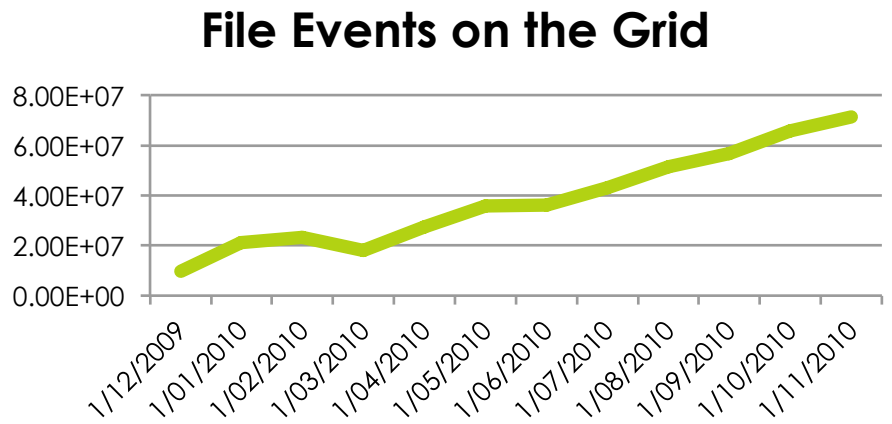
- Consistency service for identifying data corruption on the grid
- Have to maintain awareness of changing datasets on the grid. For example, if we replicate dataset user.angelos.xxxxx to site A, B, and C, and then this dataset changes, the changes have to propagated
- At the ATLAS scale we need to enforce concept of dataset immutability

Replication policy

- Replication largely driven by the ATLAS Computing Model
- Datasets are marked as:
  - Primary – mandated by the Computing Model
  - Secondary - in excess of the Computing Model
- Secondary replicas reduced by popularity
- Determining popularity of datasets
  - Collecting traces
  - Aggregating traces
- Problems with the current approach – dynamic approaches

Scalability

- At the grid level, scalability is a primary concern
- New technologies
- Seven fold increase of file events over the year
- Disk I/O is the bottle neck
- Parkinson's Law



Trends

- Moving towards meta data driven model, rather than hierarchical container -> dataset -> file
- Increased emphasis on searching by meta data
- Simplification of services, consolidation (e.g. consolidation of LFCs)
- Optimisation by simulation
- Move to open protocols

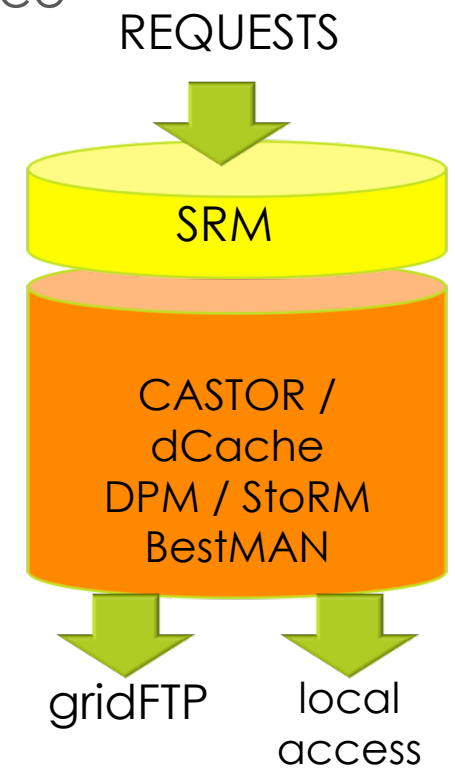
Summary

- Major Issues:
  - Scalability
  - Consistency
  - Replication policy
  - Heterogeneity
- Trends
  - Addressing scalability
  - Metadata
  - Simplification of services
  - Simulation

Backup slides

SRM and Space Tokens

- Storage systems implement a common interface
  - Storage Resource Manager (SRM)
    - gridftp as common transfer protocol
    - storage specific access protocols
  - Space Tokens
    - partitioning of storage resources according to activities
- Each ATLAS site is identified by a site name and according space token
  - DESY-ZN_PRODDISK



- `'srm': 'token:ATLASPRODDISK:srm://lcg-se0.ifh.de:8443/srm/managerv2?SFN=/pnfs/ifh.de/data/atlas/atlasproddisk/'`