Grid Data Management

Storage and Data Management Technologies

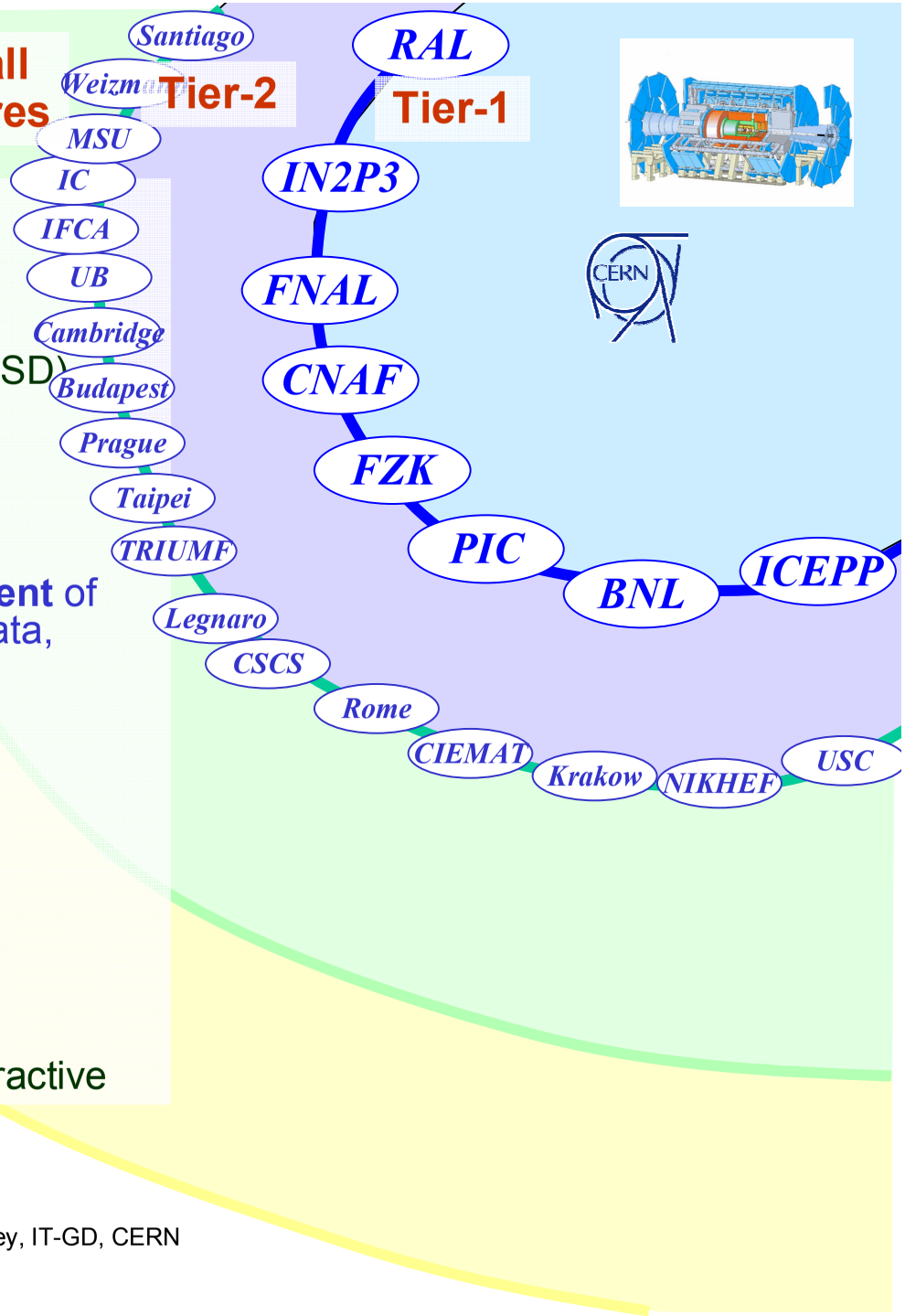
James Casey, IT-GD, CERN

DTI Mission to CERN in Distributed IT Applications

29th June 2004

Overview

- What is the data management problem for the LCG project?
- What do we have now?
  - Grid File Catalogs
  - The Storage Resource Manager
- What don't we have now?
- Summary

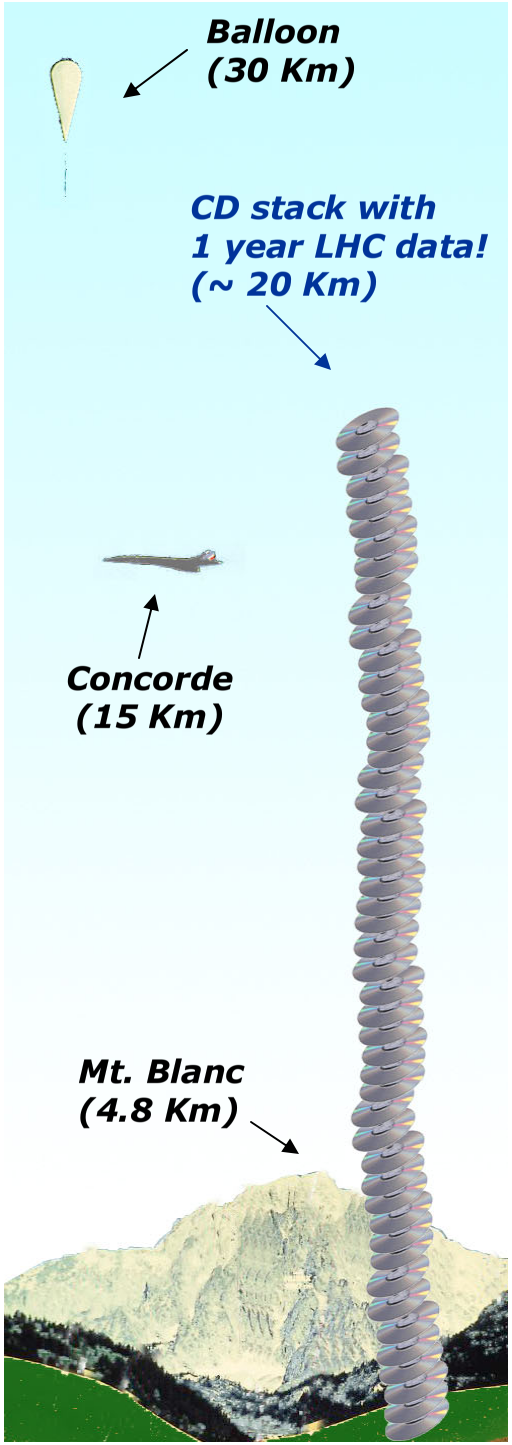


LHC Computing Model (simplified!!)

- Tier-0 – the accelerator centre
  - Filter → *raw data*
  - Reconstruction → *summary data*(ESD)
  - Record *raw data* and *ESD*
  - Distribute *raw* and *ESD* to Tier-1
- Tier-1 –
  - Permanent storage and **management** of *raw*, *ESD*, calibration data, meta-data, analysis data and databases → **grid-enabled data service**
  - Data-heavy analysis
  - Re-processing raw → ESD
- Tier-2 –
  - Well-managed disk storage – grid-enabled
  - Simulation
  - End-user analysis – batch and interactive
  - High performance parallel analysis (PROOF?)

How much data in one year?

- Storage Space
  - Data produced is ~15PB
  - Space provided at all tiers is ~80PB
- Network bandwidth
  - 10 – 40 Gb/s to all big centres
  - only for Tier-0 -> Tier-1 distribution
- Number of files
  - ~ 40 million files
  - assuming 2GB files



Problem definition in one line…

- "…to distribute, store and manage the high volume of data produced as the result of running the LHC experiment"
- Data comprises of
  - Raw data ~90%
  - Processed data ~10%
  - "relational" metadata ~1%
  - "middleware-specific" metadata ~ .001%
- Grid tools provide some of the solution…
  - We currently use tools from the European Data Grid project (and some home-brew tools to provide missing functionality)
  - Move to EGEE provided tools as they mature

File Catalogs

- Tracking ~40 million file per year is “easy” compared to storing the files!
  - “only” 80GB of replica catalog data per year
  - Metadata catalog could still pose problems (1% of 10PB)
- Use standard database solutions – See next presentation
  - Oracle Database + Application Server
  - Expose to the Grid as Web Services
- Availability and reliability more important problems than data size or data rate

Storage Challenges – 1/2

- Managing storage resources in an unreliable distributed large heterogeneous system
- Long lasting data intensive transactions
  - Can't afford to restart jobs
  - Can't afford to loose data, especially from experiments
- Type of failures
  - Storage system failures
    - Mass Storage System (MSS)
    - Disk system
  - Server failures
  - Network failures

Storage Challenges – 2/2

- Heterogeneity
  - Operating systems
  - MSS - HPSS, Enstore, Castor (see Olof's talk)
  - Disk systems – system attached, network attached, parallel
- Optimization issues
  - avoid extra file transfers - What to keep in each disk caches over time
  - How to maximize sharing for multiple users
  - Global optimization
  - Multi-Tier storage system optimization
- Management Issues
  - Need to manage more storage with less people
  - Storage is now distributed too

Storage Resource Manager - 1/2

- Storage Resource Manager (SRM)
  - Collaboration between LBNL, CERN, FNAL, RAL, Jefferson Lab
  - Becoming the GGF Grid Storage Management Working Group
  - http://sdm.lbl.gov/srm-wg/
- Provides a common interface to Grid Storage
  - Exposed as a Web Service
  - Negotiable transfer protocols (Gridftp, gsidcap, …)

Storage Resource Manager - 2/2

- MSS vendors (and users) provide their own implementation
  - CERN – Castor
  - Fermilab- ENSTORE, dCache SRM
  - LBNL – HPSS
  - JLab – JASMINE
  - RAL – ADS (in progress)
- No real “lightweight” solution for small Tier-2s
  - 1-10Tb of disk on multiple disk servers
  - No dedicated admin resources - requires very low maintenance overhead
- Lightweight disk pool solution under investigation

So is it all solved then?

- No ! We don't know how to (or haven't implemented yet)
  - Storing 150TB of relational metadata
  - Moving 40Gb/s of data around the world 24/7
    - Service Challenges being set up to test all parts of the system required for this. From disk and network to resilient replica management software
  - Hands-off management at Tier-2s of TB's of data
- Standards are good, but not enough
  - E.g. Still haven't got multiple interoperable SRM services – all slightly different
  - Have more SRM versions and specifications than implementations
    - SRM-v1, SRM-v2.1, SRM-Basic, SRM-Advanced
  - SOAP/XML over HTTP is not enough
    - But what is the future? OGSI, WS-RF ?

Summary

- Data storage and movement are two outstanding problems of Grid Data Management
  - Tier-0 to Tier-1 has very high data transfer rate
    - Need “self-managing” systems which require low intervention
  - With many distributed users and limited resources, optimization of data movement and access becomes important
- Current solutions becoming standardized in GGF – still open problems
  - Data optimization
  - Integration with other storage systems
- What are the future standards?
  - How can we influence them?