

**DM** Data Management CERN IT Department


## Distributed Database Services

A Fundamental Component of WLCG Service

### Experience and Outlook

CHEP 2009, 23 March 2009

*Maria Girone, CERN - IT, Data Management Group*



CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it

**DM** Outline CERN IT Department

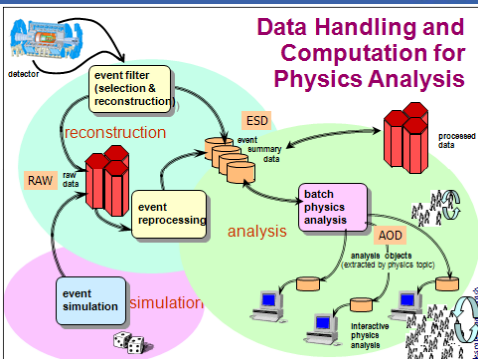
- Databases in the LHC Computing Grid
  - Key Technologies
    - Real Application Clusters (RAC)
    - Streams
    - Data Guard
- Lessons Learned in 2008
- Preparing for the 2009-2010 running

CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it

Maria Girone CERN Database Techniques and Experience 2

**DM** Databases and LHC CERN IT Department

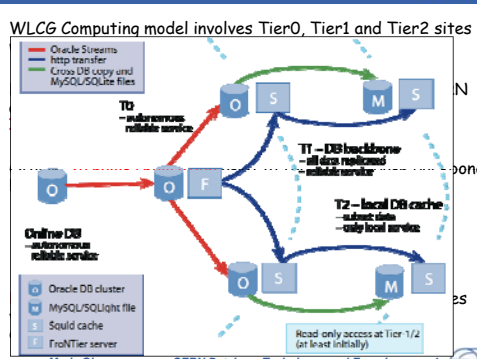
### Data Handling and Computation for Physics Analysis



CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it

**DM** Distributed Database Operations CERN IT Department

- WLCG Computing model involves Tier0, Tier1 and Tier2 sites



CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it

Maria Girone CERN Database Techniques and Experience 4

**DM DBA & Developer Community** CERN IT Department

- "Effective Oracle between DBAs and practices"
  - Regular tutorials, early 2005
- Consultancy to the
  - Performance tuning
  - Concurrency and str
    - Most issues of perf
  - Atlas Scalability tests for conditions data (see R. Walker's talk)
- Small DBA community: essential to collect and share knowhow among the (often limited) resources
  - Emphasis on homogeneity (e.g. RAC/Linux, SAN infrastructure)
    - Sharing policies and procedures
  - Distributed Database Operations regular meetings and workshops
    - now including discussions on CASTOR@Tier1

CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it

Maria Girone CERN Database Techniques and Experience 5

**DM Service Key Requirements** CERN IT Department

- Data Availability, Reliability and Scalability
  - Oracle Real Application Clusters (RAC) with Automatic Storage Management (ASM)
  - Hardware with redundancy for fault tolerance
- Data Distribution
  - Oracle Streams: for sharing information between databases
- Data Protection
  - Oracle RMAN on TSM for backups
  - Oracle Data Guard: for additional protection against failures (human errors, disaster recoveries,...)

CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it

Maria Girone CERN Database Techniques and Experience 6

**DM Data Management** CERN IT Department

**Key Technologies**

CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it

**DM Real Application Clusters** CERN IT Department

- Building-block architecture for the Distributed Database Services at CERN and Tier1 sites
  - Key to providing availability, reliability, scalability, consolidation and required service level
- ~25 clusters in production at Tier0, ~20 at Tier1 sites
  - Similar h/w set-up, SAN, same OS and Oracle versions
- Rolling upgrade capabilities essential for service continuity
  - Maintenance with no user-visible downtime
    - 0.04% services unavailability (2008) = 3.5 hours/year
    - 0.22% server unavailability (2008) = 19 hours/year
- Expansion to this level of users / applications / data would have been impossible within resource constraints using individual disk servers

CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it

Maria Girone CERN Database Techniques and Experience 8

### DM Oracle RAC/SAN Architecture CERN IT Department

**RAC: The Cluster Database**

Centralized Management Console  
 High Speed Switch or Interconnect  
 Clustered Database Servers  
 Hub or Switch Fabric  
 Mirrored Disk Subsystem

Network  
 Private Interconnect  
 Real Application Clusters  
 Storage Area Network

Dual-CPU quad-core 2950 DELL servers, 16GB memory, Intel 5400-series "Harpertown"; 2.33GHz clock  
 Dual power supplies, mirrored local disks, 4 NIC (2 private 2 public), redundancy of FC switches, dual HBAs, "RAID 1+0 like" with ASM.

CERN IT Department  
 CH-1211 Geneva 23  
 Switzerland  
 www.cern.ch/it

### DM Oracle Streams Replication CERN IT Department

• Technology for sharing information between

LHCb  
 CMS

CH-1211 Geneva 23  
 Switzerland  
 www.cern.ch/it

Maria Girone CERN Database Techniques and Experience 10

### DM Data Guard CERN IT Department

- Limiting database downtime in the event of:
  - Multi-point hardware failures
  - Wide-range of corruptions
  - Disaster
  - H/W upgrades
  - Human errors
    - within configured redo apply lag (24 hours)
- Ad-hoc testing of major schema upgrades or data reorganization on the standby

Primary  
 Clients  
 Standby  
 Data changes

CERN IT Department  
 CH-1211 Geneva 23  
 Switzerland  
 www.cern.ch/it

### DM Data Management CERN IT Department

**2008 Experience and Lessons Learned**

CERN IT Department  
 CH-1211 Geneva 23  
 Switzerland  
 www.cern.ch/it

**DM** Service Priorities in 2008 CERN IT Department

- **Met requirements for LHC start-up**
  - H/W upgrade of Tier1s validated
- Took over more databases and services
  - Additional Oracle
- Standardized complexity
  - Oracle
- Enforce robust pre-production testing
  - Eliminate service problems at the production stage

CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it

Maria Gironé CERN Database Techniques and Experience 13

**DM** Service Enhancements in 2008 CERN IT Department

- **Data Guard** for critical databases
  - At CERN for physical standby: validated during CCRC08 and deployed in production prior to the LHC start-up
  - At CNAF for logical standby: for ATLAS LFC from CNAF to INFN-ROMA1 (see B. Martelli's talk)
- **Streams Replication**
  - Automatic Split & Merge procedures
  - Use of transportable tablespaces for site re-synchronization
- **Backup and Recovery**
  - On-tape backups performance improvements (from 30Mb/s to 70MB/s)
  - Automated test recoveries
  - LAN-free tape backup tests
  - On-disk image copies to speed-up recovery from physical and logical data corruptions

CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it

Maria Gironé CERN Database Techniques and Experience 14

**DM** Service Enhancements in 2008 (2) CERN IT Department

- **DB Monitoring**
  - Coherent tool for database (online, offline & standby) and streams monitoring/alerts integrated
  - Extended to display Tier1 status
- **Database Security**
  - Isolating the online DBs from external access (typically limited to experiments' networks)
  - Oracle Critical Security updates regularly validated and applied
- **Account Policies**
  - Enforcing use of reader and writer. Owner locked on production. Automatic tool for temporary unlocking

CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it

Maria Gironé CERN Database Techniques and Experience 15

**DM** Data Management CERN IT Department

Preparing for 2009-2010 running

CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it

**DM** 2009 Priorities CERN IT Department

- **H/W and O/S upgrade**
  - Study usage of systems resources: CPU, random IO (read), available storage space and correlate it with experiments' requests
  - Evaluation of RHEL5
- **DB Security**
  - Rule-based tool to detect malicious access to DB
- **Client connection to the DB**
  - Coral Server deployment (see A. Valassi's talk)
- **Data Life Cycle**
  - "old" data offline with possibility of accessing it on demand
    - (SAM, conditions, PVSS, etc)
  - multiple schema archiving, Oracle partitioning, compression
- **Regular Oracle Reviews** (discussed at WLCG workshop)
  - To follow-up service requests/bugs which affect WLCG Services

CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it Maria Girona CERN Database Techniques and Experience 17

**DM** Oracle 11g CERN IT Department

- Understand service implications & production deployment schedule for Oracle 11g version
  - Will be a coordinated effort with the Tier1 sites
- Tested 11.1 features
  - Streams
    - Performance improvements
    - Combined Capture & Apply
    - Source and target data Compare & Converge
  - Active Data Guard
    - Enables to continuously open a physical standby database for read-only access (critical to the online DBs)
  - Execution Plan Stability
  - Real Application Testing
  - ASM and cluster storage improvements
- Now involved in 11gR2 beta testing

CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it Maria Girona CERN Database Techniques and Experience 18

**DM** Summary CERN IT Department

- We are operating a complex world-wide distributed database infrastructure for LHC Computing Grid
  - RAC/ASM key DB services at Tier0 & Tier1s
  - Streams for detector conditions: key for data (re-)processing
  - Data Guard key for data protection: critical databases
- Large scale tests show that the experiments' requirements have been met
  - Testing and validation - hardware, DB versions, applications - proven key to smooth production
    - close cooperation between application developers and DBAs
- Monitoring of database & streams performance has been implemented
  - key to maintain and optimise any larger system
- Several successful data-challenges but data-taking starts soon..

CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it Maria Girona CERN Database Techniques and Experience 19

**DM** More Details CERN IT Department

- CERN Physics Databases wiki:
  - General advices, activities
  - Interventions, service procedures, policies
  - <http://cern.ch/phydb/wiki>
  - Support: [phydb.support@cern.ch](mailto:phydb.support@cern.ch)
- LCG 3D wiki
  - interventions, service procedures, performance pages
    - <http://cern.ch/phydb/lcg3d>

CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it Maria Girona CERN Database Techniques and Experience 20

**DM** Data Management CERN IT Department

## Backup Slides

CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it

**DM** Streams Procedures CERN IT Department

- Downstream capture to de-couple Tier 0 production databases from destination or network problems
  - redo log retention optimized to allow for a re-synchronisation window of 5-days

- If one site is down
  - LCRs are not removed from the queue
  - Capture process might be paused by flow control
- Split & (and then) Merge procedures
- Resynchronization when the site is up again
  - Transportable tablespaces

CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it

Maria Sironi *CERN Database Techniques and Experience* 22

**DM** Streams Throughput: Conditions CERN IT Department

- ATLAS Conditions data
  - 2 GB per day
  - 600 - 800 LCRs per second

CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it

**DM** ... and PVSS CERN IT Department

- ATLAS PVSS tests
  - 6 GB per day
  - 2500 - 3000 LCRs per second

CERN IT Department  
CH-1211 Geneva 23  
Switzerland  
www.cern.ch/it