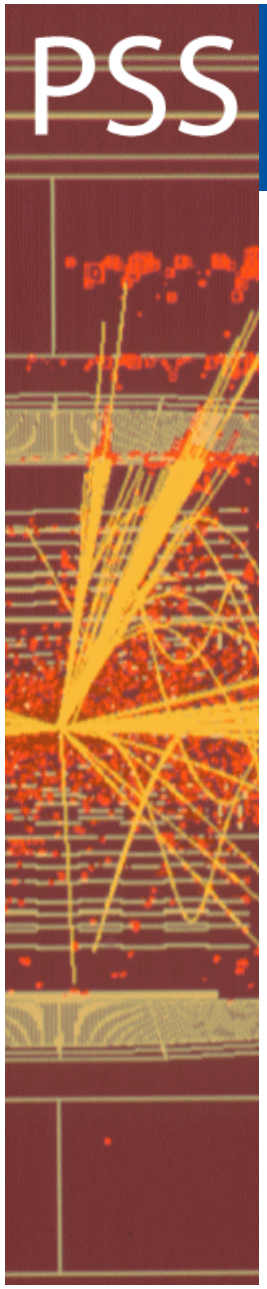


PSS Physics Services Support

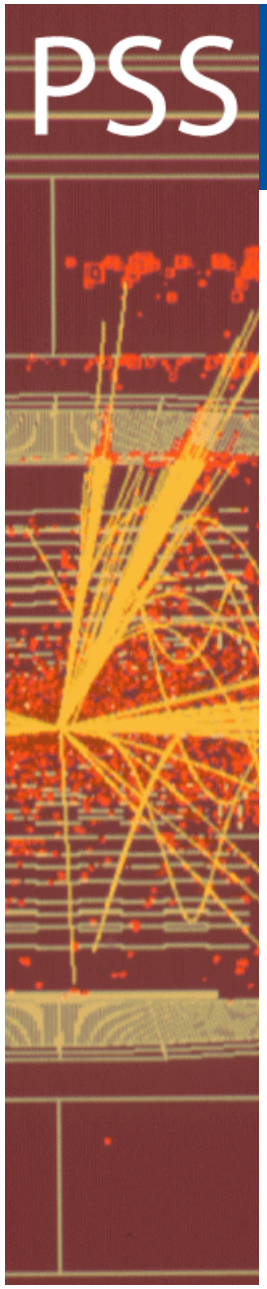
CERN IT Department

High Availability Databases based on Oracle 10g RAC on Linux

WLCG Tier2 Tutorials, CERN, June 2006

Luca Canali, CERN IT

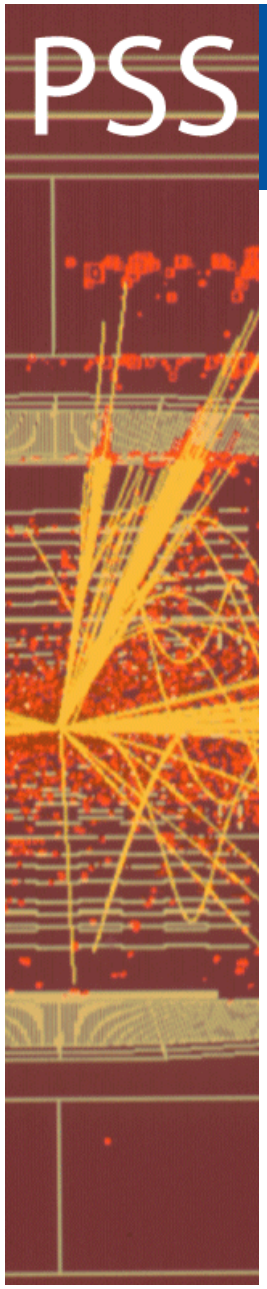
CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it



Outline

- Goals
- Architecture of an HA DB Service
- Deployment at the CERN Physics Database Service
- Focus on: what you need to do to build an HA DB with Oracle 10g RAC

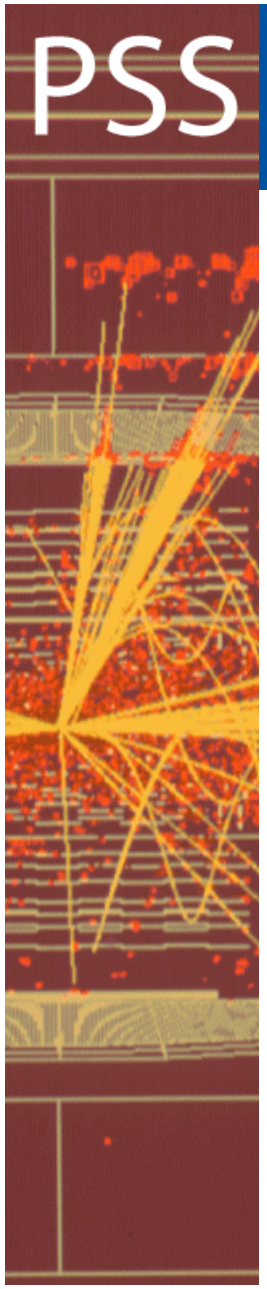
CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it



Goals

- Run database services to meet the requirements of the Physics experiments
  - Mission-critical: central repository for many LHC and grid applications
- Requirements
  - High Availability
  - High Performance and Scalability
  - Simplify implementation and administration
  - Provide a cost-effective solution

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



Architecture

CERN IT Department

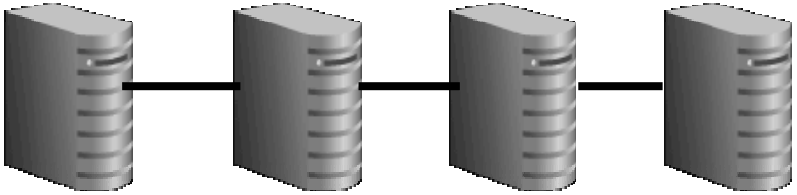
- The ‘big picture’: Database Clusters
  - An implementation of grid computing for the database tier for HA and load balancing
- HW
  - Many redundant server nodes
  - Network infrastructure
  - Cost-effective HW
- Software
  - Cluster-enabled database (Oracle RAC)
  - Cluster volume managers and filesystems

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

Database Clusters

- Two different high-end DB architectures

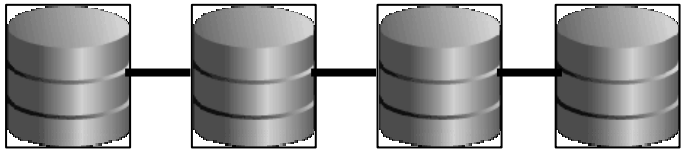
RAC



SMP, Scale UP

Grid-like, Scale OUT

ASM



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

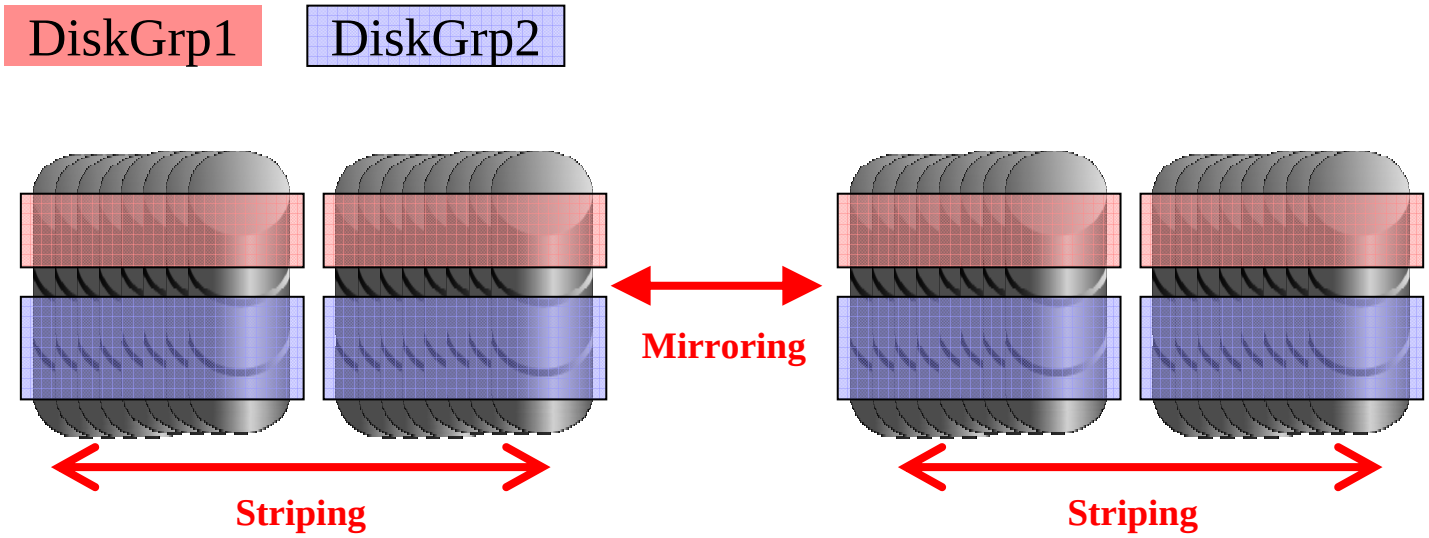
Oracle 10g RAC

- Oracle 10g RAC
  - A database engine that can scale DB workload across many cluster nodes
  - An HA and scalability solution
  - Applications tested on Oracle single node can be deployed on Oracle RAC
- Technology
  - Shared-everything clustering solution
  - Complex cache and distributed locking algorithms (cache fusion)

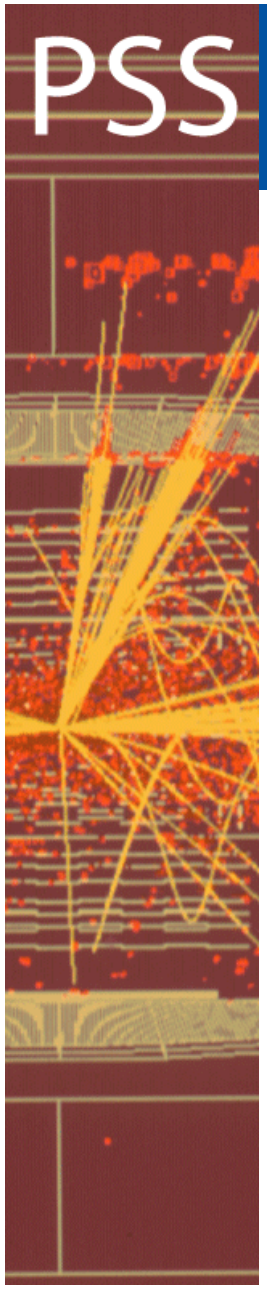
CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Oracle ASM

- ASM is a volume manager and cluster filesystem specialised for Oracle DB files
- Implements S.A.M.E. (stripe and mirror everything)
  - Similar to RAID 1 + 0: performance and HA
- Online storage reconfigurations (ex: in case of disk failure)
- Example of storage allocation with ASM:



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



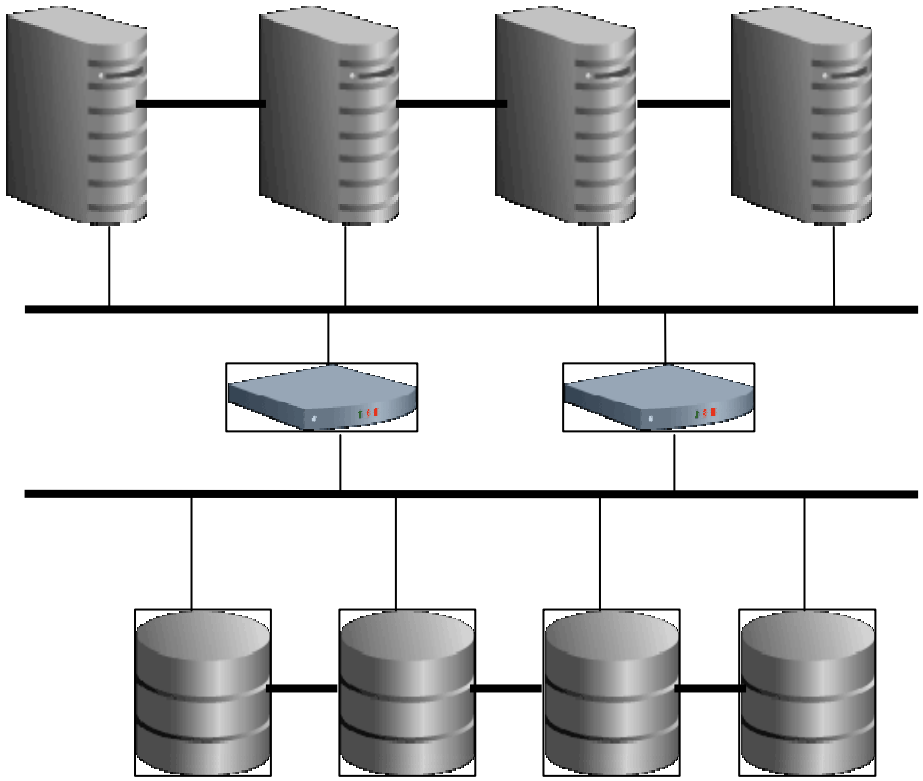
Network

- Public networks
  - Gigabit Ethernet for ‘SQL input-output’
- Cluster interconnects
  - Two gigabit Ethernet networks
  - Inter-node communication (cache transfer)
- Storage Area Network
  - Disk arrays are connected via SAN
  - Redundant Fiber Channel network (2Gbps)
    - Two SAN switches
    - Dual-ported HBAs

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

Growth on Demand

- Database clusters can grow to meet the experiments' demands.



DB Servers

SAN Switches

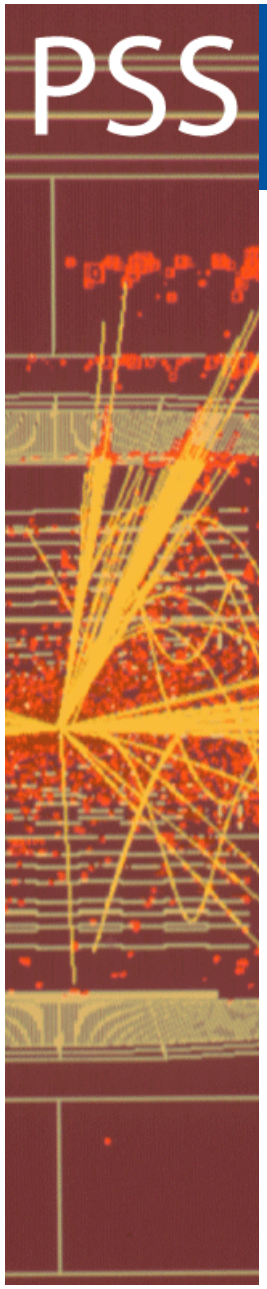
Storage Arrays

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

Scale Economies

- Homogeneous HW configuration
  - Clusters can be easily built and grown
  - A pool of servers, storage arrays and network devices are used as ‘standard’ building blocks
  - Hardware provisioning is simplified
- Software configuration
  - Same OS and database version on all nodes
    - EX: Red Hat Linux and Oracle 10g R2
  - Simplifies installation, administration and troubleshooting

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



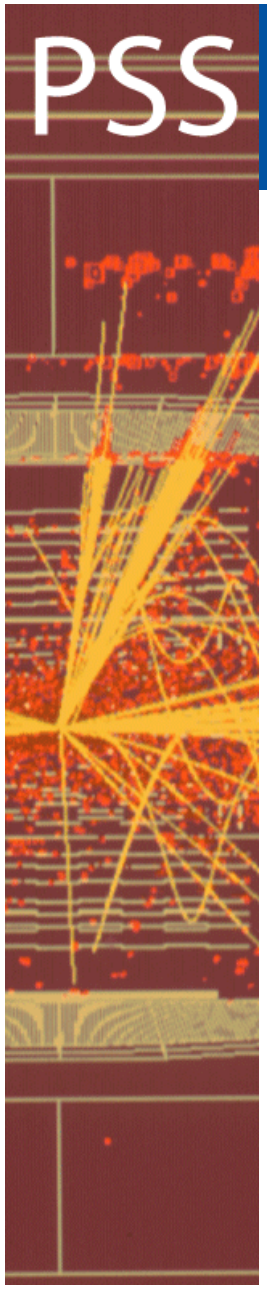
Backup and Recovery

CERN IT Department

- Technology:
  - RMAN (Oracle's primary solution for HA)
  - Media manager (ex: Tivoli)
- Backup to tape using RMAN
  - No need to stop the DB, 'hot backups'
  - Incremental policy: reduces the performance overhead
- Backup to disk with RMAN
  - Additional layer of protection, allows quicker recoveries

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

CERN



Recap of the infrastructure

CERN IT Department

Database HA requires redundant HW:

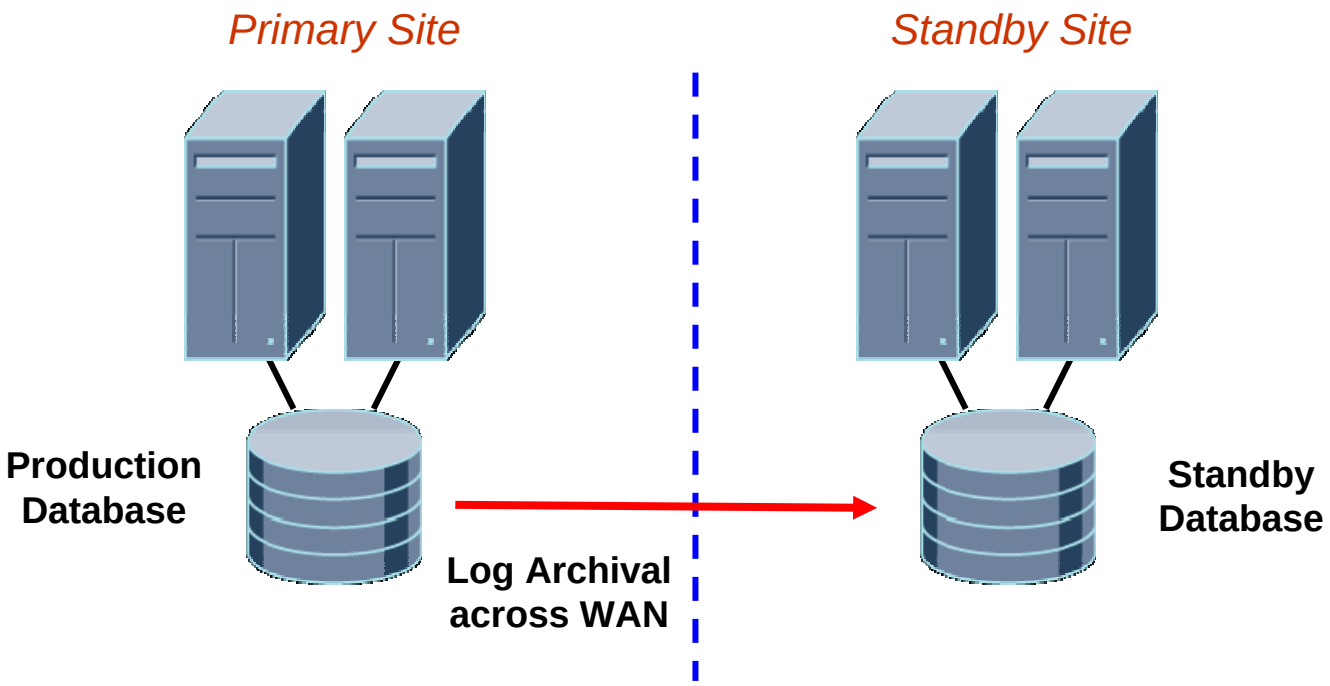
- DB servers
- Storage Arrays
- Ethernet networks (public and interconnect)
- Fiber Channel networks (SAN)
- Redundant power supplies and UPS

- Other components:
  - Backup infrastructure
  - Monitoring
  - ‘Redundant’ sysadmins and DBAs

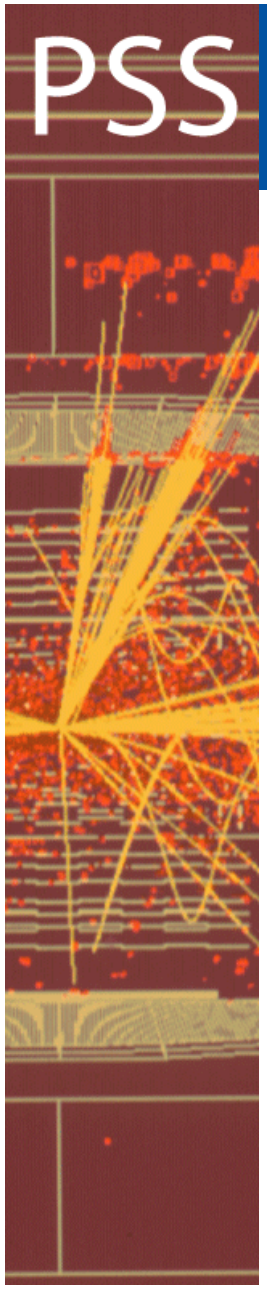
CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Disaster Recovery

- Disaster Recovery for HA:
  - With Oracle DataGuard a standby DB is kept current by shipping and applying redo logs



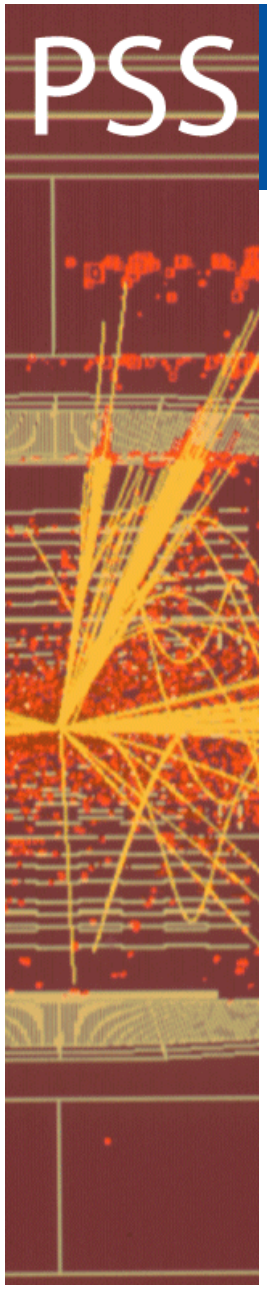
CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it



Distributed Databases

- HA can be achieved using distributed database technologies
- Examples (Oracle solutions):
  - Streams replication
    - DB changes are captured at source, propagated at destination and then applied
  - Logical standby databases over WAN
    - DB changes are replayed at destination from the redo logs of the source
  - Materialized views replication
    - DB tables are refreshed via DB links over WAN

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



Conclusions

- Physics Database Services run production Oracle 10g RAC services for Physics for HA and scalability
  - Currently 100 CPUs, 200TB of raw data
- Further links:
  - http://www.cern.ch/phydb/
  - https://twiki.cern.ch/twiki/bin/view/PSSGroup/HAandPerf

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