CERN DB Services: Status, Activities, Announcements

Distributed Databases Operations Workshop

November 16th, 2010

Luca Canali, IT-DB

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

Database SERVICES

CERN

Outline

- Review of DB Service for Physics
  - Availability
  - Incidents
- Notable activities
  - Infrastructure activities, projects, planned changes of general interest
- Announcements
  - Upcoming changes on interest for experiments and Tier1s

Status of Services

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

Database SERVICES

CERN

Service numbers

- Infrastructure for Physics DB Services
  - ~115 quadcore machines
  - ~2500 disks on FC infrastructure
- 9 major production RAC databases.
- In addition:
  - Standby systems
  - Archive DBs
  - Integration systems and test systems
  - Systems for testing streams and 11.2

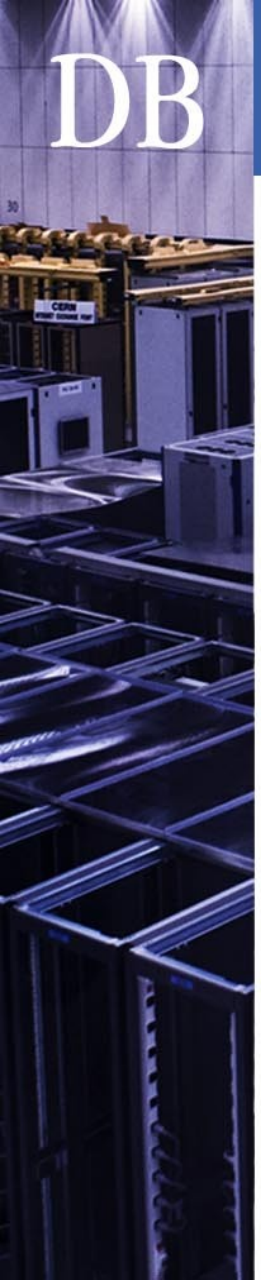
Services and Customers

CERN IT Department

- Offline DB Service of LHC experiments and WLCG
- Online DB Service
- Replication from online to offline
- Replication from offline to Tier1s
- Non-LHC
  - Biggest user in this category is COMPASS
  - and other smaller experiments

DBA Support

- 24x7 support for online and offline DBs
  - Formalized with a 'CERN piquet'
  - 8 DBAs on the piquet
    - Temporary reduced in Q2 and Q3:
    - 1 DBA in maternity leave, 1 post for CMS-founded DBA being recruited
  - Note, replication from offline to Tier1s
    - is 'best effort', no SMS alert (only email alert)
    - on-call DBA checks email 3 times per day



Service Availability

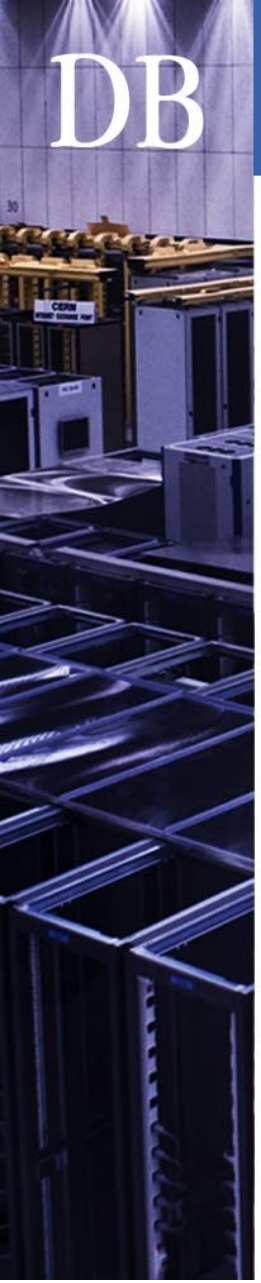
- Focus on providing stable DB services
  - Minimize changes to services and provide smooth running as much as possible
  - Changes grouped during technical stops
    - Security patches, reorg of tables
    - Major changes for end-of-the-year technical stop
- Service availability:
  - Note these are averages across all production services
  - Offline Service availability: 99.96%
  - Online Service availability: 99.62%

PSU non-rollingness

- Non-rollingness of April Patch
  - Security and recommended patch bundle for April 2010 (aka PSU 10.2.0.4.4)
  - Contains patches marked as rolling
  - Passed tests and integration
- Two issues show up when applied in production
  - Non rolling on clusters of 3 or more nodes with load
  - On DBs with cool workload
    - Symptoms: after ora-7445 and spikes of load appear
- Ora-7445
  - Reproduced on test and patch available from Oracle
  - Thanks to persistency team for help
- Non-rollingness
  - Reproduced at CERN, Oracle support does not have patch yet

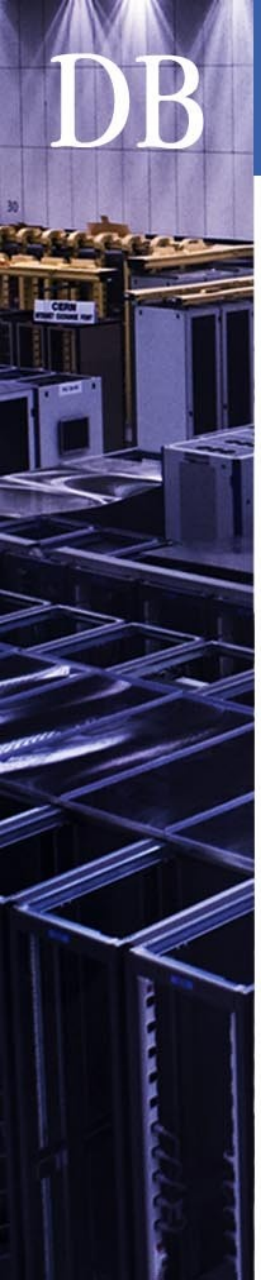
Power cut

- Two issues of unscheduled power cut at LHCB online pit
  - ~5 hours first occurrence (9/8)
  - ~2 hours for second occurrence (22/8)
- In first incident DB became corrupted
  - Storage corruption
  - Lost write caused by missing BBUs on storage after previous maintenance
  - Restore attempted from compressed backup, too time consuming
  - Finally switchover to standby performed
    - See also further comments on testing standby switchover in this presentation



Notable recurring issues

- Streams
  - Several incidents
  - Different parts of replication affected
  - Further discussions in Streams-related presentations
- High loads and node reboots
  - Sporadic but recurrent issues
  - Instabilities caused by load
  - Run-away queries
  - Large memory consumption makes machine swap and become unresponsive
  - Execution plan instabilities make for sudden spikes of load
  - Overall application-related. Addressed by DBAs together with developers

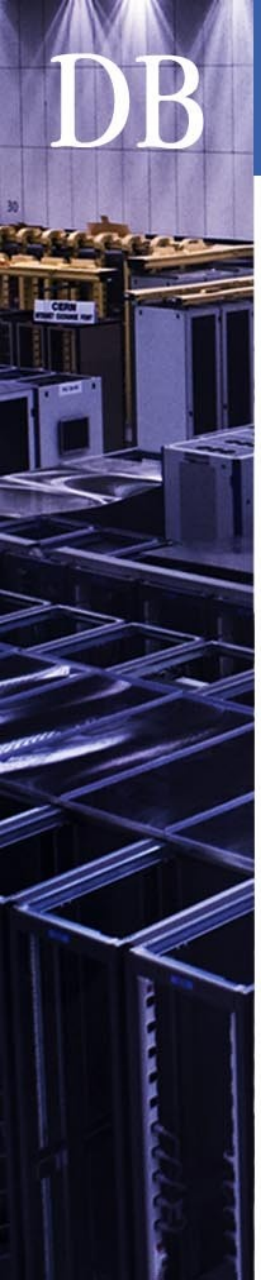


Activities and Projects

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

Database SERVICES

CERN



Service Evolution

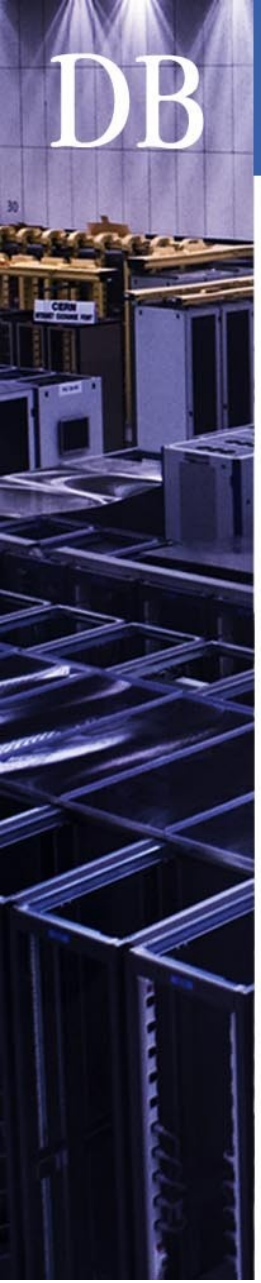
- Replaced ~40% of HW
  - New machines are dual quadcores
    - Old generation was based on single core Pentiums
  - New storage arrays use 2TB SATA disks
    - Replaced disks of 250GB
- New HW used for standby and integration DBs
  - New HW (RAC8+RAC9): 44 servers and 71 storage arrays (12 bay)
  - Old HW (RAC3+RAC4): 60 servers and 60 storage arrays (8 bay)

Consolidation of Standby DBs

- New HW installations for standby DBs
  - Quadcore servers and high-capacity disks
    - This has increased resources on standby DBs
    - Provided good compromise cost/performance in case of switchover operation (i.e. standby becomes primary)
  - Installed in Safehost (outside CERN campus)
    - Reduce risk in case of disaster recovery
    - Used for stand by DBs when primary in CERN IT

Oracle Evolution

- Evaluation of 11.2 features. Notably:
  - Evaluation of Oracle replication evolution:
    - Streams 11g, Goldengate, Active Dataguard
  - Evolution of clusterware and RAC
  - Evolution of storage
    - ASM, ACFS, direct NFS
  - SQL plan management
    - for plan stability
  - Advanced compression
- Work in collaboration with Oracle (Openlab)

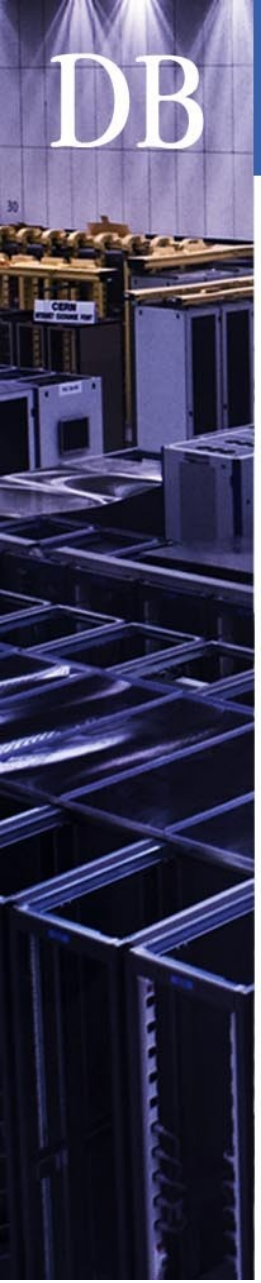


10.2.0.5 upgrade

- Evaluation of possible upgrade scenarios
  - 11.2.0.2, vs 10,2.0.5, vs staying 10.2.0.4
  - 11g has several new features
    - Although extensive testing is needed
    - 11.2.0.2 patch set came out in September and with several changes from 11.2.0.1
  - 10.2.0.4 will go out of patch support in April 2011
  - 10.2.0.5 supported till 2013
    - 10.2.0.x requires extended support contract from end July 2011
  - Decision to upgrade to 10.2.0.5 (following successful validation)
  - See also talk on application testing

Activities on Backup

- Backups to tape using 10gbps
  - have been successfully tested
    - Speed up to 250 MBPS per 'RMAN channel'
- First phase of production implementation
  - Destination TSM at 10gbps
  - Source multiple RAC nodes at 1gbps
    - Typically 3 nodes
  - In progress, to be completed by end technical stop
- Other activities
  - Moving backup management to a unified tool inside the group
  - Test recoveries also covered



Activities on Monitoring

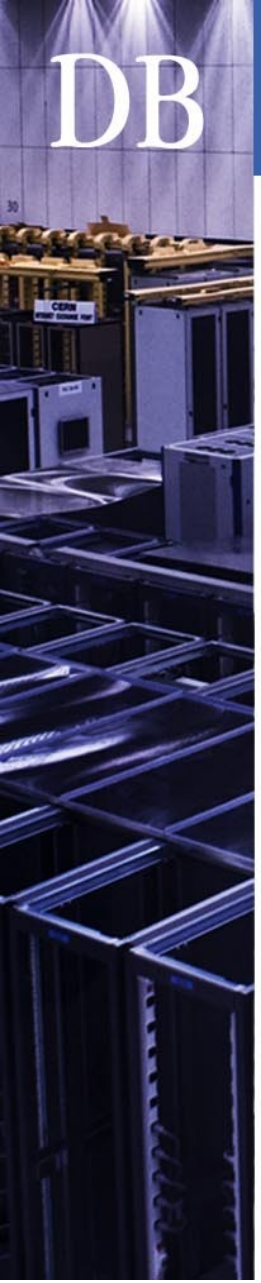
- Improvements to streams monitoring
  - Added Tier1 weekly reports
  - Maintenance and improvements to streammon
    - DML activity per schema, PGA memory usage
- OEM 11g
  - Currently deployed at CERN
  - When stabilized at CERN will be deployed to 3D OEM (schedule to be defined)
- Internal activities on monitoring
  - We are unifying monitoring infrastructure across DB group

Activities on Architecture

- Evaluation of HW
  - In 2011 a large group of production machines goes out of warranty
  - HW renewal and occasion to profit from more recent HW for performance and capacity
  - Ideally time HW move with 11g upgrade
    - Upgrading on new HW with standby 'swing' is preferred way vs. local upgrades
  - Several technologies being evaluated
    - Use of SSD for caching
    - 10gbps Ethernet for interconnect
    - 8gbps Fiber Channel storage access
    - NAS at 10gbps with flash cache

Activities on Data Lifecycle

- Data life cycle for physics data
  - Databases are growing larger
  - Some data sets can be aged out
  - Activity launched in 2008
  - In 2010 more applications modified to use it
    - Data start to be moved to archive DBs
  - More work needed on the area
    - Joint work DB group and experiments/development
- See talk on the topic later on at this workshop



Activities on Security

- Internal application developed
  - To track access to DBs
  - Mining audit data
  - Allows to spot unusual access patterns
  - Can be source of info for defining white lists
- See also discussion later on in the agenda on firewalls and white lists

Activities for Online Setups

- ALICE, LHCB and CMS online
  - Installations of the DBs at the experiments' pits
  - HW is managed by experiments
- HW warranty expiring
  - Replacement under way
  - IT discussed with experiments on HW replacement
  - Goal of having similar HW at the pit as in IT to reduce maintenance effort and complexity
  - Deployment of new HW expected in Q1

Plans Concerning Standby Tests

CERN IT Department

- Pending activity are switchover tests
  - Activity will be discussed with experiments
  - Ideally during the technical stop after upgrade to 10.2.0.5
  - Activity needed to validate the disaster recovery infrastructure
  - Require downtime
    - ~1h to switch to standby and ~1h to switch back

Announcements

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

Database SERVICES

CERN

Announcements concerning Oracle 10.2 (upgrade)

CERN IT Department

- Upgrade to 10.2.0.5
  - CERN plans the upgrade during technical stop
  - Upgrade requires ~2-4 hours downtime
  - Schedule to be agreed with experiments early in December
    - Following successful testing
    - No show stopper found so far
  - Following successful upgrade at CERN we invite Tier1s to upgrade
    - Following agreement/best practice of keeping same config in Tier1s as at CERN
    - Schedule to be defined with experiments. Reasonably can be done in Q1

Announcements concerning Oracle 10.2 (support)

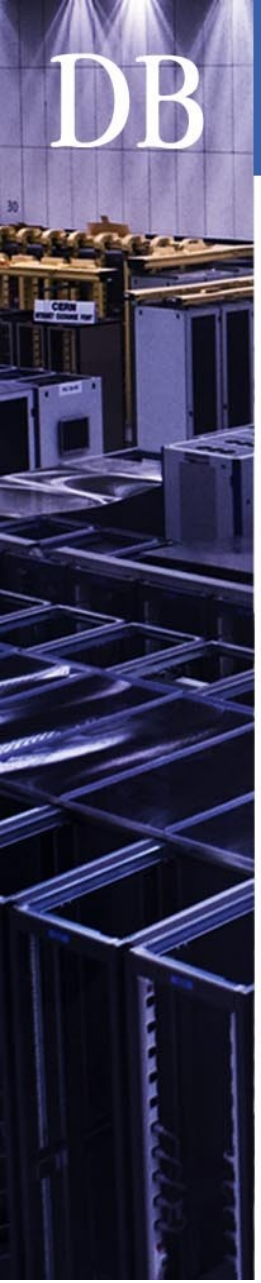
- Oracle 10.2 is in extended support since July 2010
  - From July 2011 extra support fees are required
- CERN will take care of paying the extra license fee for 1 year of extended support
  - For licenses acquired for Tier1s via CERN in 2006 and 2008.

Plans concerning Oracle 11.2

- We plan the upgrade to 11.2 in 2012
  - In Q1-Q2 during extended stop of LHC
  - Simultaneous move to new HW and RHEL5
  - Occasion to make some architectural changes
    - In particular possible changes in replication architecture
  - Dates and details to be defined
    - Depend on the LHC schedule too

Conclusions

- Focus on stability for DB services for Physics in 2010
  - Following several years of preparation
- Continuity on DB operations in 2011
  - Software evolution: upgrade to 10.2.0.5
  - Infrastructure activities on improving backups, archive, application testing, HW testing
  - Work on possible architecture changes and preparation for 11g upgrades in 2012



Acknowledgements

- This work was made possible by the collaboration of the Experiments, WLCG, Tier1 DBAs and to the Physics DBA team:
  - David, Dawid, Eva, Jacek, Kate, Luca, Marcin, Przemek, Sveto, Zbigniew

Q&A

CERN IT Department

Thank you for your attention!