Oracle Services for Physics

Maria Girone

6th openlab Major Review,
January 2009

Introduction

Oracle Services for Physics Key Technologies

- RAC/ASM for availability
- Streams for data distribution
- Data Guard for data protection

This talk will cover last 4 months updates on

- Streams
- Monitoring
- Data Guard (including tests on 11gR1 Active Data Guard)

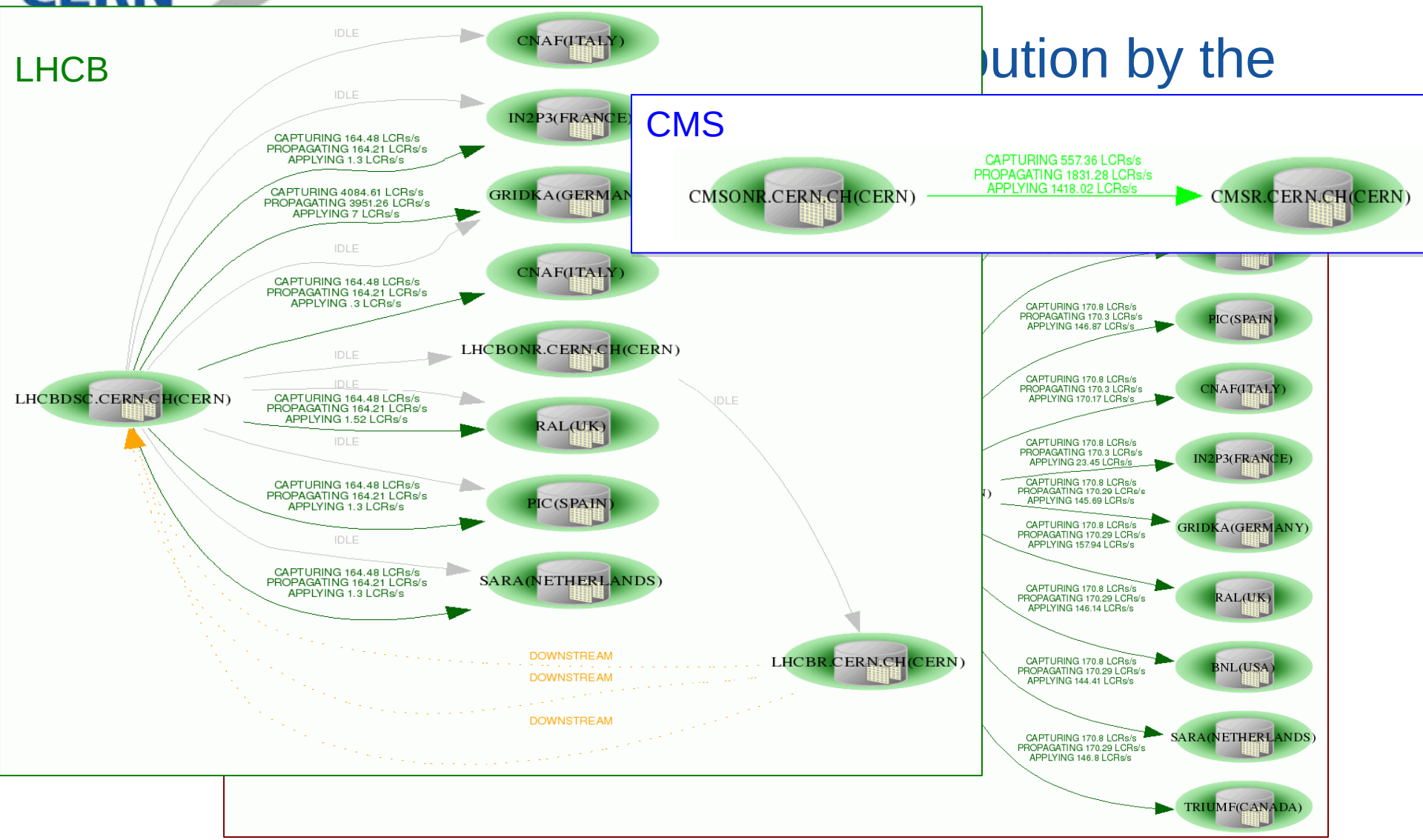
Future work on

- New version testing
  - ASM, ACFS, Active Data Guard, Streams

RAC/ASM

- Building block architecture for the Distributed Database Services at CERN and Tier1 sites
  - Key to providing the reliability, scalability, flexibility and required service level
- ~25 clusters in production at Tier0, ~20 at Tier1 sites
- Rolling upgrade capabilities essential for service continuity
- Expansion to this level of users / applications / data would have been impossible within resource constraints using individual disk servers

Streams


LHCB
CMS
ution by the
CNAF(ITALY)
IN2P3(FRANCE)
GRIDKA(GERMAN
CNAF(ITALY)
LHCBONR.CERN.CH(CERN)
RAL(UK)
PIC(SPAIN)
SARA(NETHERLANDS)
LHCBDSC.CERN.CH(CERN)
LHCBR.CERN.CH(CERN)
IDLE
CAPTURING 164.48 LCRs/s
PROPAGATING 164.21 LCRs/s
APPLYING 1.3 LCRs/s
CAPTURING 4084.61 LCRs/s
PROPAGATING 3951.26 LCRs/s
APPLYING 7 LCRs/s
CAPTURING 164.48 LCRs/s
PROPAGATING 164.21 LCRs/s
APPLYING .3 LCRs/s
CAPTURING 164.48 LCRs/s
PROPAGATING 164.21 LCRs/s
APPLYING 1.52 LCRs/s
CAPTURING 164.48 LCRs/s
PROPAGATING 164.21 LCRs/s
APPLYING 1.3 LCRs/s
CAPTURING 164.48 LCRs/s
PROPAGATING 164.21 LCRs/s
APPLYING 1.3 LCRs/s
DOWNSTREAM
DOWNSTREAM
DOWNSTREAM
CMSONR.CERN.CH(CERN)
CAPTURING 557.36 LCRs/s
PROPAGATING 1831.28 LCRs/s
APPLYING 1418.02 LCRs/s
CMSR.CERN.CH(CERN)
PIC(SPAIN)
CNAF(ITALY)
IN2P3(FRANCE)
GRIDKA(GERMANY)
RAL(UK)
BNL(USA)
SARA(NETHERLANDS)
TRIUMF(CANADA)
CAPTURING 170.8 LCRs/s
PROPAGATING 170.3 LCRs/s
APPLYING 146.87 LCRs/s
CAPTURING 170.8 LCRs/s
PROPAGATING 170.3 LCRs/s
APPLYING 170.17 LCRs/s
CAPTURING 170.8 LCRs/s
PROPAGATING 170.3 LCRs/s
APPLYING 23.45 LCRs/s
CAPTURING 170.8 LCRs/s
PROPAGATING 170.29 LCRs/s
APPLYING 145.69 LCRs/s
CAPTURING 170.8 LCRs/s
PROPAGATING 170.29 LCRs/s
APPLYING 157.94 LCRs/s
CAPTURING 170.8 LCRs/s
PROPAGATING 170.29 LCRs/s
APPLYING 146.14 LCRs/s
CAPTURING 170.8 LCRs/s
PROPAGATING 170.29 LCRs/s
APPLYING 144.41 LCRs/s
CAPTURING 170.8 LCRs/s
PROPAGATING 170.29 LCRs/s
APPLYING 146.8 LCRs/s

Streams Update

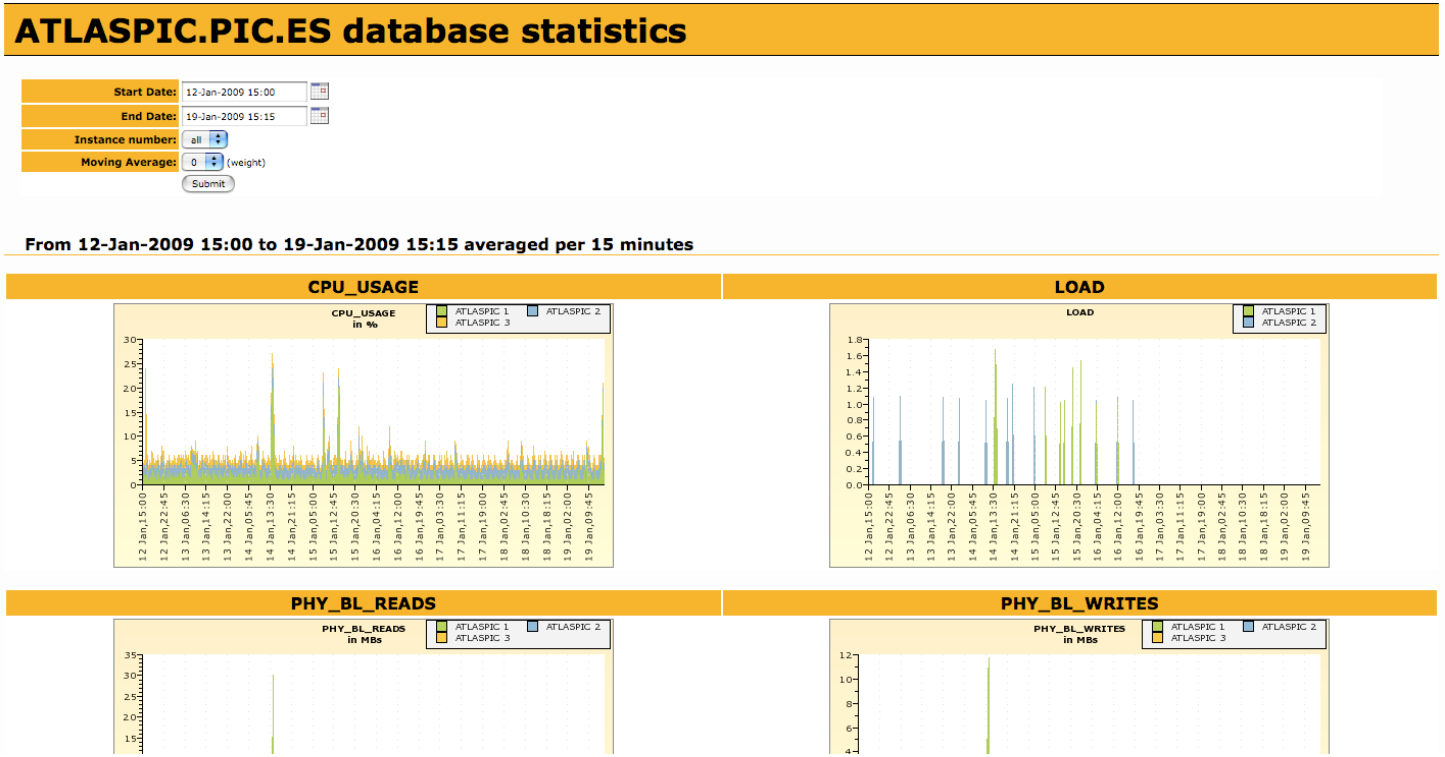
- Downstream cluster re-organization needed to increase space for spilled LCRs (from 2.6 GB to 10GB)
  - Larger time window for sites to be down without need of splitting them
  - New node allocated
    - 3 node cluster ➔ 4 node cluster
  - Downstream databases configured to run in different nodes
    - Before both databases shared 3 nodes
    - Now 2 nodes for each database

Streams Update (2)

- Recommended patches for Streams applied on all production databases
- Automatic Split and Merge procedures now possible after the problem with dropping propagation was solved by Oracle
  - Merge procedure might cause the capture process to start in a old archived log file
- Streams queue tables maintenance
  - Dequeue IOT tables grow in size, affecting dequeue performance
  - Recommended to perform manually space management tasks (dynamic shrink) of the AQ objects on regular basis
  - New daily job being validated on our Streams test environment

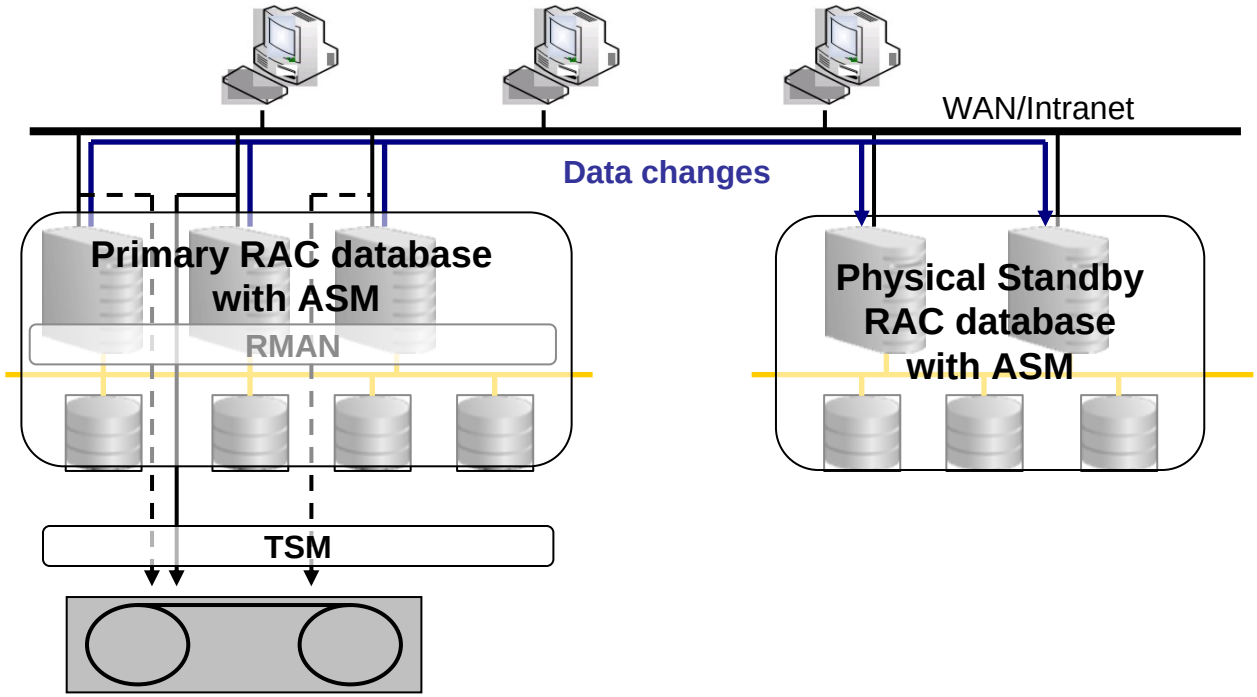
Monitoring Update

- Web conference on EM Streams Enhancements in the next version. New functionality meets our needs. Waiting for testing it.
- DB resource monitoring to Tier1 sites as extension of StreamMon
  - Evolution of database usage for certain metrics
    - Well appreciated by experiments and Tier1 sites
  - https://oms3d.cern.ch:1159/dstrmon/index



Data Guard

- Data Guard Physical Standby databases have been deployed for all the critical production systems
  - Another step towards Oracle MAA



Data Guard Use Cases

- Limiting database downtime in the event of:
  - Multi-point hardware failures
  - Wide-range corruptions
  - Disaster
- Handling human errors
  - Possible if discovered and reported within configured redo apply lag (24 hours)
- Application change management (schema upgrades or data manipulation)
  - Standby database temporarily opened in a read-write mode with flashback logging enabled

Implementation Details

- Standby RAC sized for handling average load
  - half of hardware resources allocated on production
- Maximum performance mode
- Redo data transported asynchronously by the LGWR process
  - Standby redo logs
- No Data Guard Broker
- No fast-start-failover

Active Data Guard

- Experiments would like to use the standby DBs for reporting and monitoring
- Physical standby DB can be continuously opened for read-only access
- Extensive tests on 11gR1 performed
- Setup: two 2-node RACs with ASM, on RHEL 4 64bit and Oracle11.1.0.7
  - Primary and standby were installed in different locations of CC

Active Data Guard Functionality Tests

- Standby database configuration using active database duplication feature
  - RMAN “duplicate target from active database”
  - No problems spotted
  - Much faster than backup-based duplication
- Role transition tests
  - Smooth and easy
- Primary-Standby consistency
  - No issues detected (also long transactions)
- Very good stability
  - The configuration still running smoothly for almost 2 months

Active Data Guard Performance Tests

- The tests were mainly focused on measuring data propagation delay:
  - for different transactions' sizes
  - for different redo transport mechanisms
  - with Real-Time Apply enabled
  - with 1 or 2 standby nodes opened in read-only mode
- No performance issues detected so far

CERN openlab

FUTURE WORK

RAC and ASM

- New Release beta testing
  - The ASM and ACFS test plan has been accepted by Oracle beta program Coordinators
    - Functionality
    - Stability
    - Performance
- Evaluating possible new RAC storage options
  - iSCSI

Streams

- AMI "Atlas Metadata Interface" replication from Lyon (IN2P3) to CERN
  - Dataset selection application for ATLAS
  - AMI servers located at IN2P3
  - Request: Streams replication to CERN – ATLAS offline database
  - Currently being tested using CERN – ATLAS integration database
- Looking into the PVSS data replication for CMS (between online and offline databases)
  - Already working for ATLAS
- New version beta testing

Data Guard

- Data Guard on 10g
  - Configuration of Data Guard broker
  - Move backups to standby
- Active Data Guard
  - New version testing
    - Repeat tests performed on 11gR1
    - Backups to standby
  - Apply performance tests

Summary

Oracle Services for Physics Key Technologies

- RAC/ASM key DB services at Tier0 & Tier1s
- Streams for detector conditions: key for data (re-)processing
- Data Guard for data protection: critical databases

- Understand service implications & production deployment schedule for the next Oracle version

Projects, Contacts and Participants

- Oracle Streams and Data Replication Services
  - Single Point of Contact: E. Dafonte Perez (CERN) – G. Kerr (Oracle)
  - Participants: M. Girone (CERN) – P.McElroy (Oracle)
- Streams and RAC monitoring
  - Single Point of Contact: D. Wojcik (CERN) – G. Kerr (Oracle)
  - Participants: D. Wojcik (CERN)
- Oracle Enterprise Manager
  - Single Point of Contact: C. Lambert (CERN) – A. Bulloch (Oracle)
  - Participants: D. Wojcik, A. Wiecek (CERN) – G. Kerr (Oracle)
- Oracle Data Guard
  - Single Point of Contact: S. Kapusta (CERN) – G. Kerr (Oracle)
  - Participants: M. Girone, E. Grancher, S Kapusta (CERN)
- Database Virtualization
  - Single Point of Contact: A. Topurov (CERN) – G. Kerr (Oracle)
  - Participants: E. Grancher, C. Garcia-Fernadez (CERN)
- Highly available database services based on RAC/ASM
  - Single Point of Contact: D. Wojcik (CERN) – G. Kerr (Oracle)
  - Participants: M. Girone , J. Wojcieszuk, D. Wojcik (CERN)