

Database Services During Run2

WLCG Collaboration Workshop

Okinawa April 11th, 2015

Luca Canali, CERN IT-DB

Outline

- Status, evolution and readiness for RUN 2 of Oracle database services
- DB on Demand service
- Scale-out Databases on HADOOP



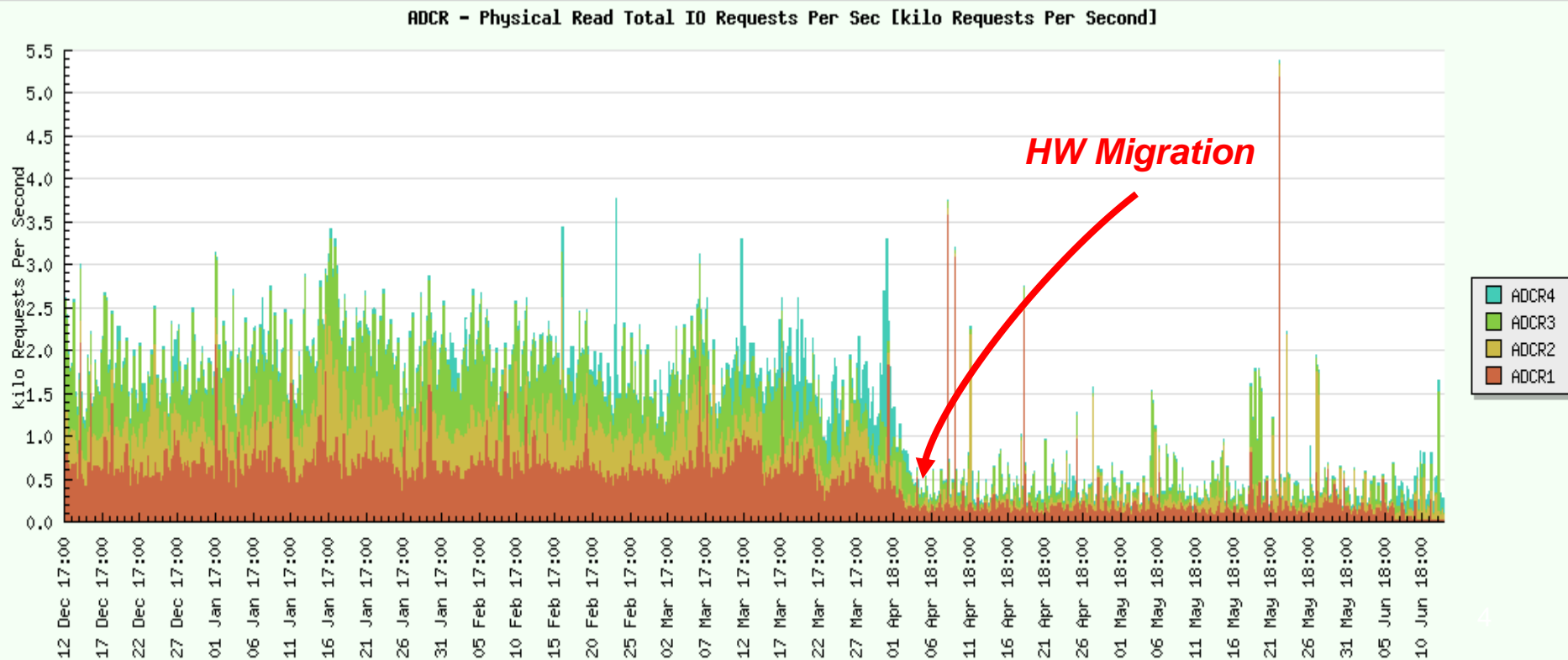
Oracle Database Services – Evolution in LS1

- Software upgrades
 - **Upgraded Oracle** version for all DBs
 - Moved to RHEL6 and **puppet**
- Hardware upgrades
 - **New generation** of servers and storage
 - Production now in the BARN (critical power)
 - **Disaster-recovery** from Safehost to **Wigner**
- SW upgrades have combined with HW move
 - Has allowed to reduce downtime and risk

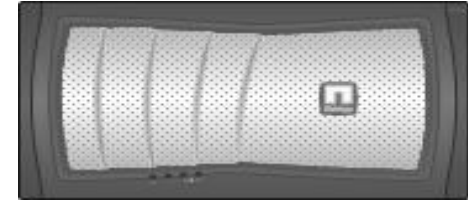
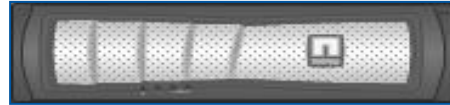
New Servers – More Capacity

New DB servers have more memory

- From **48GB** -> to **128GB / 256GB**
- Beneficial for performance of transactional workloads
- Also increased CPU capacity (N# cores from 8 to 16)



Storage Evolution: Increased Capacity and Performance



	NetApp FAS3240 (OLD)	NetApp FAS8060 (NEW)
NVRAM	1.0 GB	8.0 GB
System memory	8GB	64GB
CPU	1 x 64-bit 4-core 2.33 Ghz	2 x 64-bit 8-core 2.10 Ghz
SSD layer (maximum)	512GB	8TB
Aggregate size	180TB	400TB
OS controller	Data ONTAP® 7-mode	Data ONTAP® C-mode*

Notable Software Changes

- **Oracle version** evolution
 - Production upgraded to Oracle **11.2.0.4**
 - Proven to be a **stable** version, smooth upgrades
 - 12.1.0.2 is the latest available version
 - Currently upgrading 12.1.0.1 DBs to this release
 - Oracle has announced the next release, **12.2**, for 2016
- **PVSS upgrades**
 - Upgraded all WINCC/PVSS schemas to latest version 8.9 CERN 1.2

Notable New Oracle DB Services

QPSR

- Quench Protection System
- WinCC/PVSS data, sustained rate **~150 K rows/second**
- **1M rows/second peak** achieved during stress testing
- Performance of the new HW instrumental to achieve this

SCADAR

- Consolidated WinCC/PVSS archive repository
- Will store **~50K rows/second** (may increase in the future)
- the data retention varies depending on the application (from a few days to 5 years)

Replication Evolution

- Technology evolution:
 - Conditions replication to Atlas Tier 1 sites now uses **Golden Gate**
 - Replication from online to offline DBs
 - **Active Data Guard**
 - Some remaining use of streams set to phase out
- See also
 - **“Evolution of Database Replication Technologies for WLCG”**, CHEP Monday 13/4 at 14:15

Some Important Trends for Transactional Databases

- Server **capacity** is growing fast also for commodity HW
 - Large increase in available **memory**
 - **SSD** storage becoming more affordable
 - Servers have increasing **CPU** power (typically more cores)
- Opportunities
 - **Consolidation**: reduce HW and management **costs**
 - Review of application **architectures**
 - More and more DB workloads fit into server memory
 - Opportunity to **reduce complexity** in favour of simpler architectures

CERN Oracle Databases have run successfully in RUN1

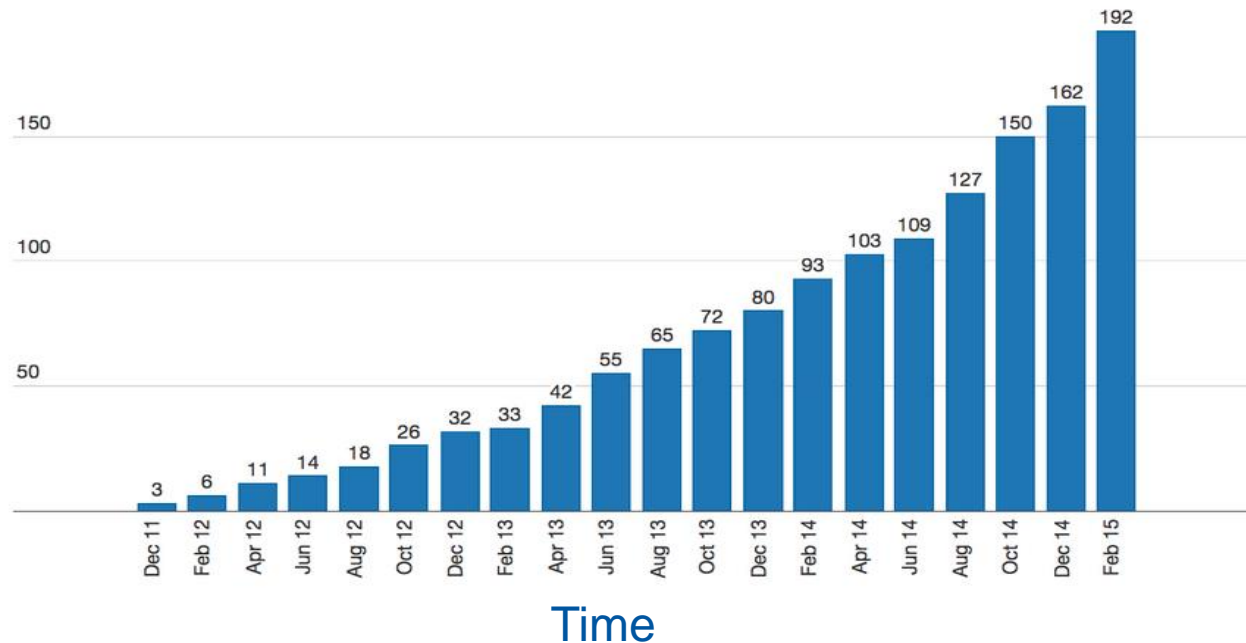
- ~100 Oracle databases, most of them RAC
- ~500 TB of data files for production DBs in total
- Used for physics, IT, accelerators, administration
- Oracle is feature-rich
 - Enterprise class DB
 - Solutions for high availability, disaster recovery, backup, monitoring, replication, proven for concurrent transactional processing
 - Oracle DBAs available for consultancy



DB on Demand

- Self-service for provisioning, management, backup
 - The number of deployed instances is growing
 - **MySQL** (85%), PostgreSQL (10%) and Oracle (5%)

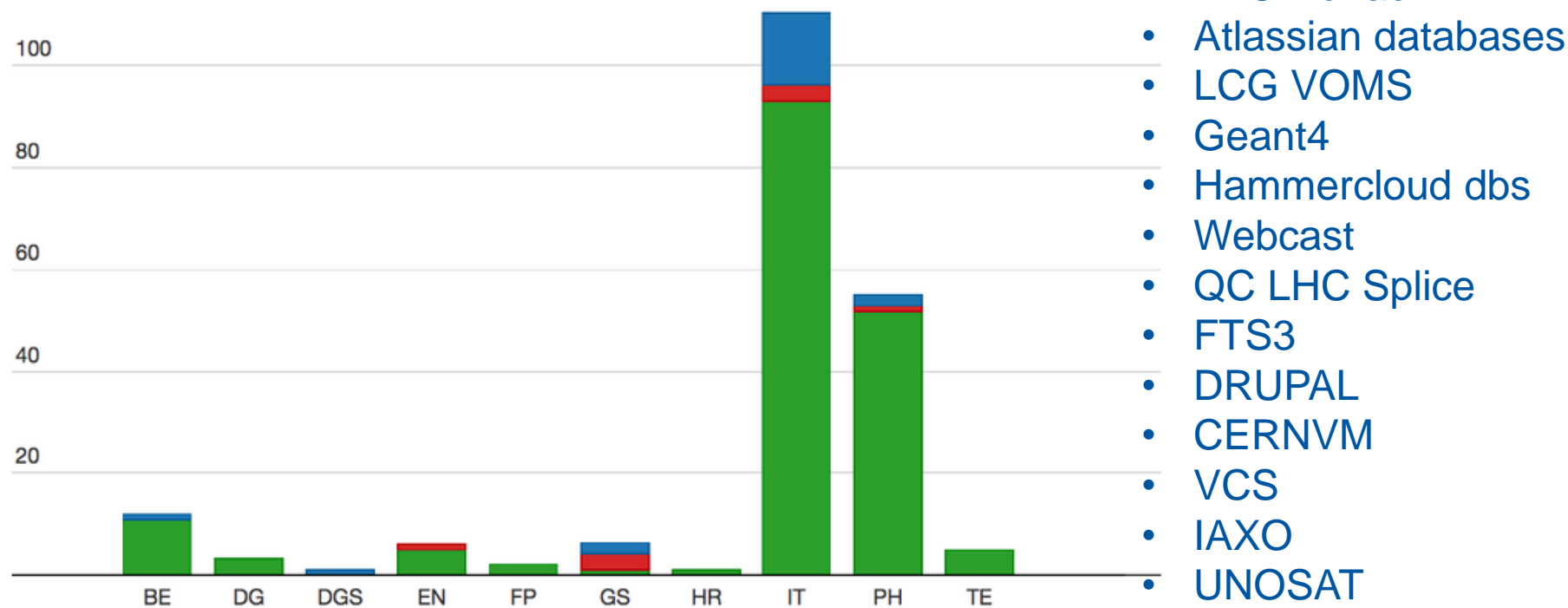
Number
of deployed
DBoD
Instances



Additional Stats for DBoD

DBMS's per Department

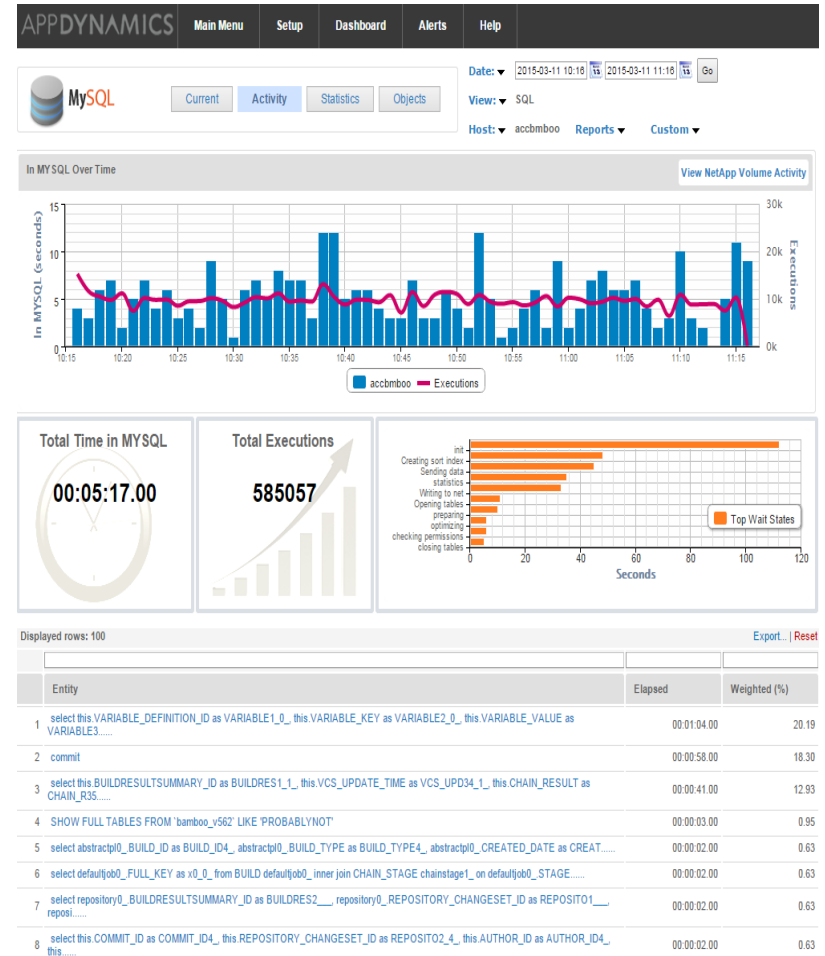
■ MySQL ■ ORACLE ■ PostgreSQL



- Openstack
- Puppetdb (MySQL)
- LHCb-dirac
- Atlassian databases
- LCG VOMS
- Geant4
- Hammercloud dbs
- Webcast
- QC LHC Splice
- FTS3
- DRUPAL
- CERNVM
- VCS
- IAXO
- UNOSAT
- ...

DBoD new monitoring tool

- Appdynamics
- Allows DBoD users to troubleshoot performance problems with a GUI interface



DBoD Evolution

- Completed:
 - Migration to **new HW** (servers and storage)
 - Migration to **CERN Agile Infrastructure**
 - **High Availability** cluster solution based on Oracle
- Plans for 2015:
 - **Automation**: instance creation, storage, puppet...
 - Reduce waiting time to obtain your instance
 - **Cloning**
 - Backups verification
 - Test schema/DBMS upgrades
 - **Replication**: MySQL & PostgreSQL
 - Business continuity and data protection
 - **Upgrades**: Oracle 12.1, MySQL 5.6.x, PostgreSQL 9.3.x
- See also: CERN **poster “Database on Demand”** by Ruben Gaspar Aparicio and Ignacio Coterillo

Support Levels in RUN 2

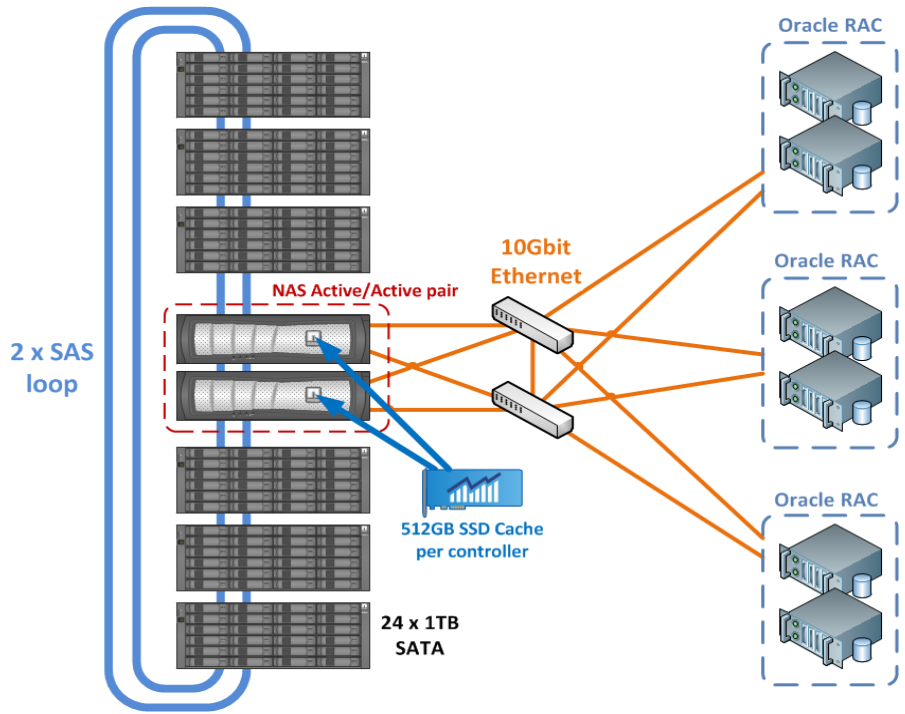
- Oracle Database Service support
 - 24/7 CERN piquet in Run 1
 - A review has been made with the experiments for Run2
 - Oracle piquet support will start mid May 2015
 - We will then re-evaluate in 2016 if needed
- DB-on-demand services have grown from being “nice to have” to being essential to both experiment and Grid services
 - New management interface and monitoring with AppDynamics
 - A best effort SLA will be provided for a number of critical databases, details to be announced later in 2015

Scale-out Databases

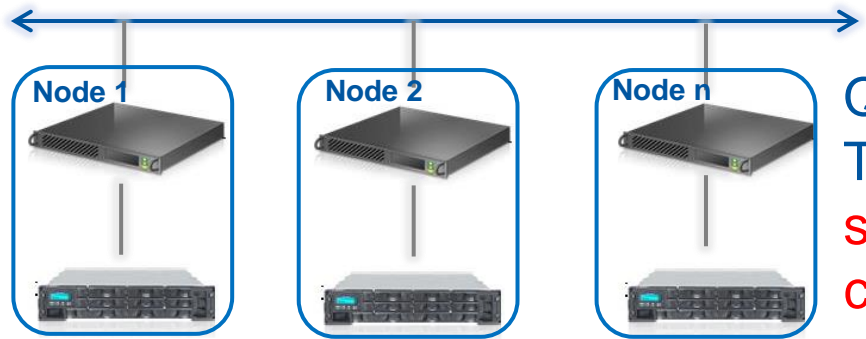
- Scale-out, shared-nothing architectures have appeared and are quickly growing in adoption
 - **Performance and low-cost**
 - Back-end: **HADOOP** or DB-specific
 - Engine on top of HADOOP cluster
- Useful for
 - Data warehousing, reporting, analytics workloads
 - Ex: Logging systems, controls dashboards, auditing, archives



Example of Oracle RAC deployed with shared storage



Interconnect



Queries run in parallel on the cluster nodes
The **shared nothing** architecture allows to **scale** for high capacity and throughput on **commodity** HW

Hadoop Ecosystem and Databases

- It's a full ecosystem, many components
 - **HDFS** provides the shared-nothing scalability
 - Specialized data formats for
 - **compression, partitioning** and **column-oriented** storage
 - Query engines
 - Declarative (**SQL**) Impala, Spark SQL, Hive
 - Imperative: Yarn/Map Reduce, **Spark**
 - Loading tools also important (sqoop, flume, ..)

Offloading Data into HADOOP

- We are building HADOOP-based systems for **offloading** resource-intensive queries
 - Controls data (PVSS) write once read-many
 - Accelerator log DB is now 200 TB + **90 TB/year**
 - **Offline database** for SCADA (EN controls)
- **Analytics**: Working with CMS on popularity
- Projects in collaboration IT-DB and IT-DSS

- See also **CHEP poster** “Scale out database for CERN use cases” Z. Baranowski et al.

Conclusions

- **Oracle** database services
 - **Proven** and stable infrastructure and support levels
 - Hardware refresh and software upgrades during LS1
- **Database on demand** services
 - Established platform and growing fast
 - Offer **MySQL**, PostgreSQL and Oracle services
- **HADOOP**-based databases
 - For data warehouse, reporting and analytics
 - First results very **promising**, more development in progress
- Many **thanks** to the **experiments**, application **DBAs** and **Tier 1** DBAs for their help and collaboration