
CERN

Eric Grancher
Dawid Wojcik

“Database On Demand” at CERN Empowering Users

Oracle Open World 2012

Agenda

- About **CERN**
- **Database as a Service** – Rationale
- Architecture
  - Virtualization Solution – Oracle VM
  - MySQL
- Empowering users
  - Instance administration
  - Backup & Restore
  - Monitoring
  - One button upgrades
- Summary

CERN

- European Organization for Nuclear Research
  - World's largest centre for scientific research, founded in 1954
  - Research: Seeking and finding answers to questions about the Universe
  - Technology, International collaboration, Education

Twenty Member States

Austria, Belgium, Bulgaria, Czech Republic, Denmark, Finland, France, Germany, Greece, Italy, Hungary, Netherlands, Norway, Poland, Portugal, Slovakia, Spain, Sweden, Switzerland, United Kingdom

Seven Observer States

European Commission, USA, Russian Federation, India, Japan, Turkey, UNESCO

Associate Member States

Israel, Serbia

Candidate State

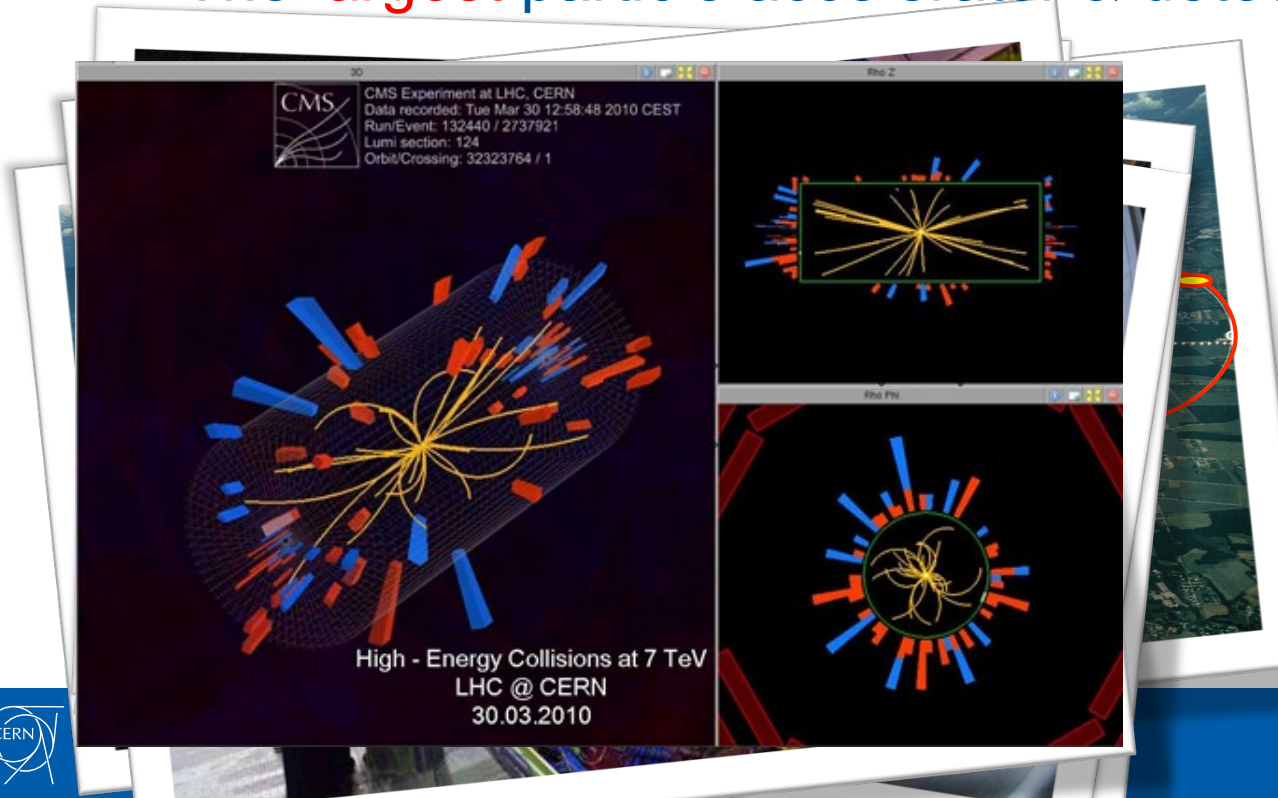
Romania

People

~2400 Staff, ~900 Students, post-docs and undergraduates, ~9000 Users, ~2000 Contractors

LHC

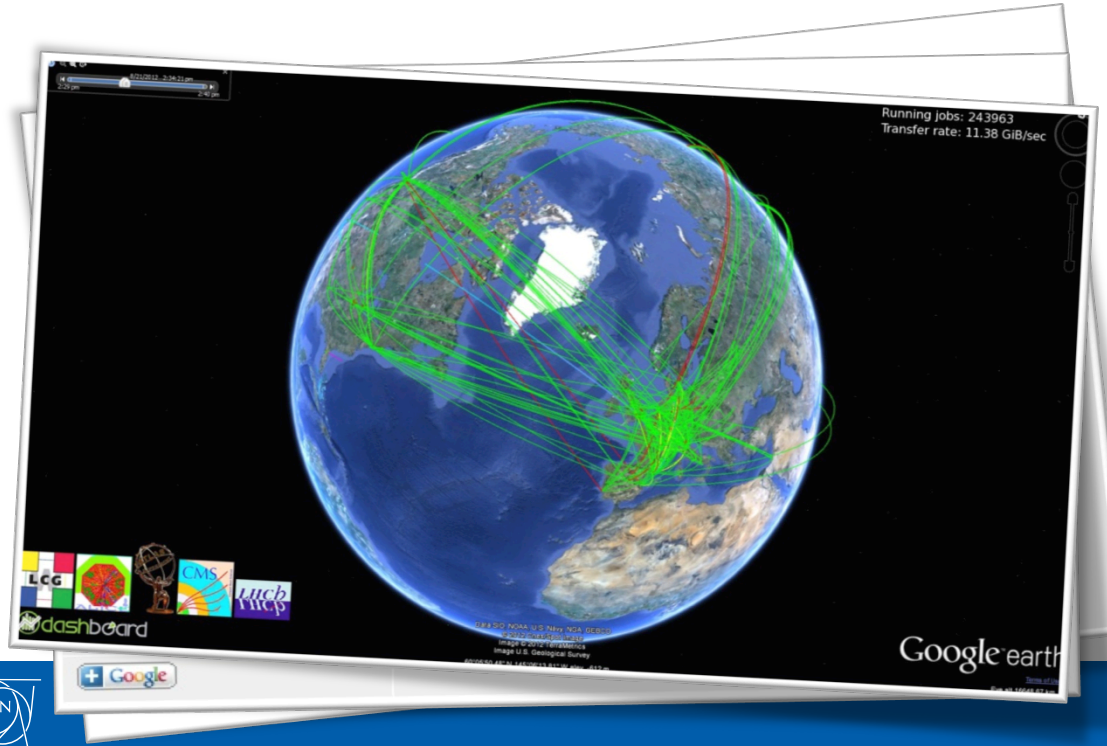
- The largest particle accelerator & detectors



17 miles (27km) long underground tunnel
Thousands of superconducting magnets
Coldest place in the Universe
-271.3 °C
but also...
...hottest place in the Universe :)
100k times hotter than the heart of the Sun
600 million collisions per second detected by about 150 millions of sensors

WLCG

- World's largest computing grid



More than 20 Petabytes
of data stored and analysed
every year

Over 68 000 physical CPUs
Over 305 000 logical CPUs

157 computer centres in 36
countries

More than 8000 physicists with
real-time access to LHC data

Databases at CERN

- CERN IT-DB Group manages Oracle databases used by LHC and its experiments, as well as financial and administrative ones
- >100 databases, most of them RAC
  - Mostly NAS storage plus some SAN with ASM
- >70 databases backed up to tapes
  - On average ~5.1 TB of redo daily, ~302 TB of datafiles in total
- MySQL on demand service now in production
- The biggest databases at CERN
  - LHC logging database ~145 TB, expected growth up to 70 TB / year
  - 13 production experiments' databases ~122 TB in total

Database as a Service – Rationale

- Empowering CERN IT and research community
  - Users can request and manage different database instances (currently MySQL and Oracle single instance)
  - Aimed at medium size and long-term projects
  - Users are provided with a self-service portal
    - Ease of administration
    - Integrated backup & recovery
    - Monitoring solution
    - One click patching

Database as a Service – Rationale

- Provide flexible and cost effective Database as a Service
  - Owners are grouped by a mailing group (access authorization)
  - Owners receive full DBA privileges on their instances
  - Owners are responsible for ensuring that their systems, and the use of their systems, are fully compliant with the Rules of CERN Computing Facilities (including security)
  - The “Database on Demand” (DBoD) service – OS administration and providing support for self-service portal functionality
  - The DBoD service does not provide DBA or application support

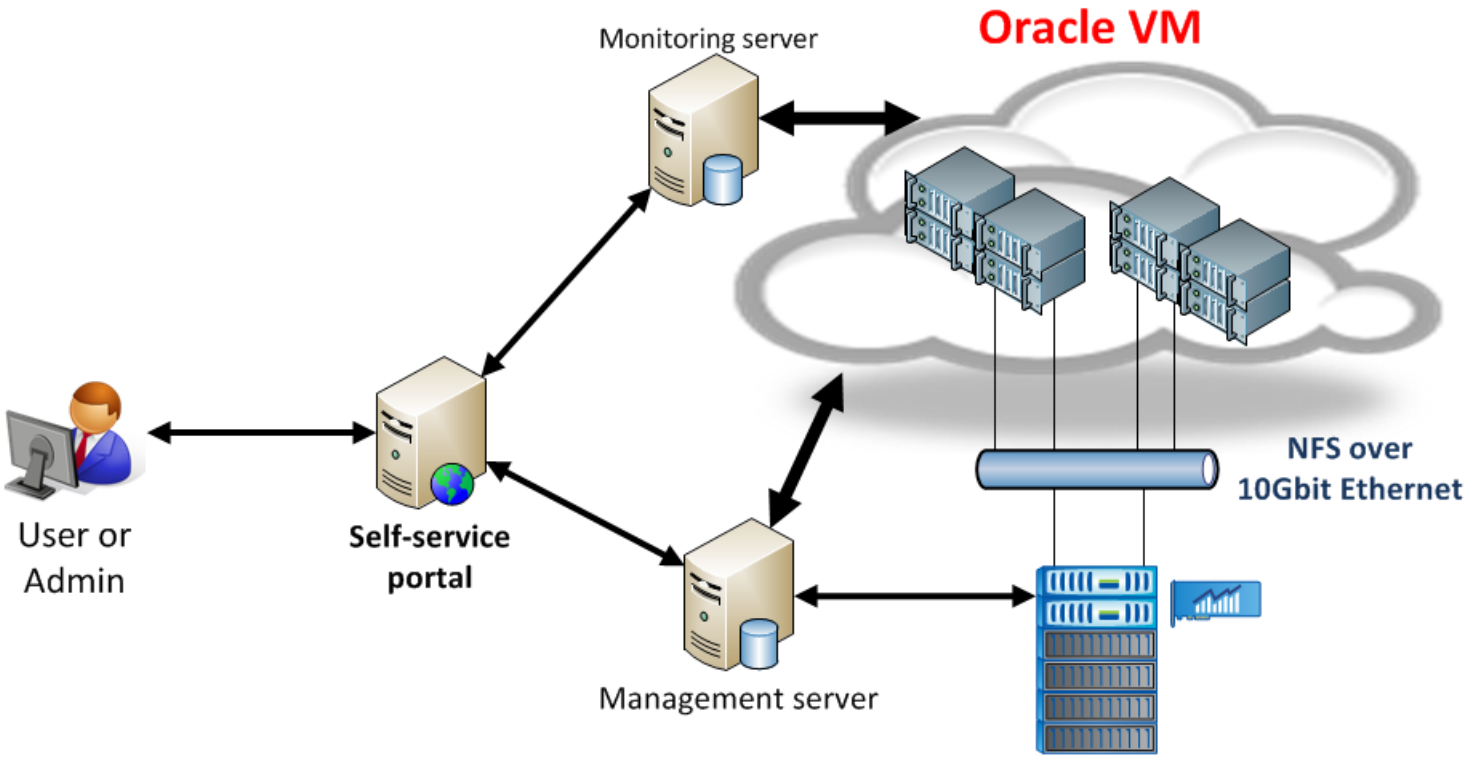
Private cloud model

- Reuse existing virtualization infrastructure and know-how – cost efficient
- Improve operations
  - Standardization
  - Aim to consolidate – migrate existing DBs to DBoD service
  - Reuse tools and management frameworks
- HA via virtualization (live migration)

Architecture

- Virtualization
  - Oracle VM (2.2, 3.0.2 and 3.1.1) on Linux x86_64
  - Typical VM size: 2 cores, 8 or 16GB RAM
- Storage
  - NFS over 10 Gigabit Ethernet
- Provisioning
  - Open-source Quattor Toolkit
  - CERN is currently adopting Puppet
- Management framework
  - Syscontrol – developed at CERN
- Web self-service portal

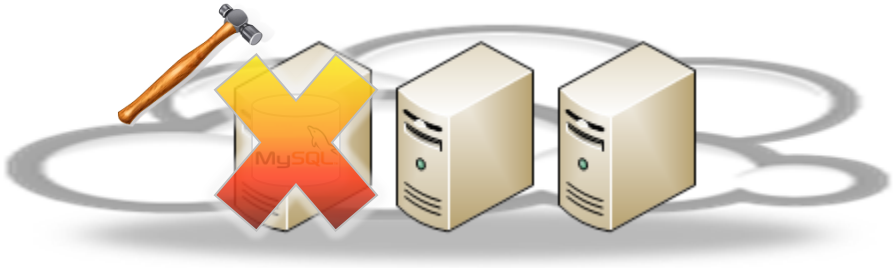
Architecture



Oracle VM

ORACLE VM

- CERN has more than 2 years of experience of running Oracle VM in production
  - Easy scale-out of VM Pools
  - 10 Gbit Ethernet – one network only (NAS over NFS)
  - Currently testing OVM 3.1.1 and 3.2.1
  - Live migration for HW interventions and host OS upgrades

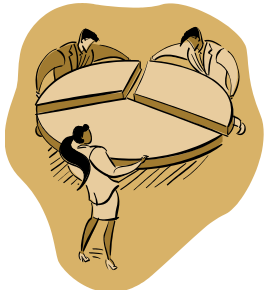


MySQL

- Currently running MySQL Community Edition 5.5
  - InnoDB as the preferred storage engine – backup & recovery
  - Binary logs enabled
    - ACID (atomicity, consistency, isolation, durability) – *innodb_flush_log_at_trx_commit* = 1, *sync_binlog* and *innodb_flush_method* = O_DIRECT
  - Using *innodb_buffer_pool_size* of ~5GB
  - Using thread cache (big gain for some clients)
  - Using query cache (*query_cache_size = 768M*)
  - Performance schema is enabled by default

Shared Instances

- DBoD supports more than one MySQL instance on one VM
  - Sharing CPU
  - Sharing MySQL binaries
  - Separate buffer pools (pre-allocated memory)
  - Separate NFS volumes
    - Independent backup and restore



Empowering users

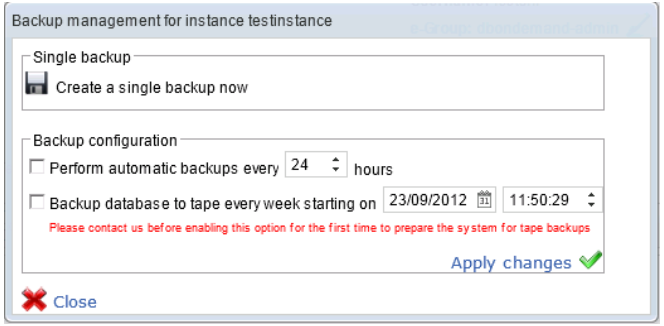
- Self-service portal
  - Instance administration (status, start, stop)
  - Manage configuration and logs
    - MySQL: download/upload my.cnf, download slow queries log
  - Set up backups (automatic or manual) or command a restore
  - One button instance upgrade
  - Access to monitoring information

Backup solution

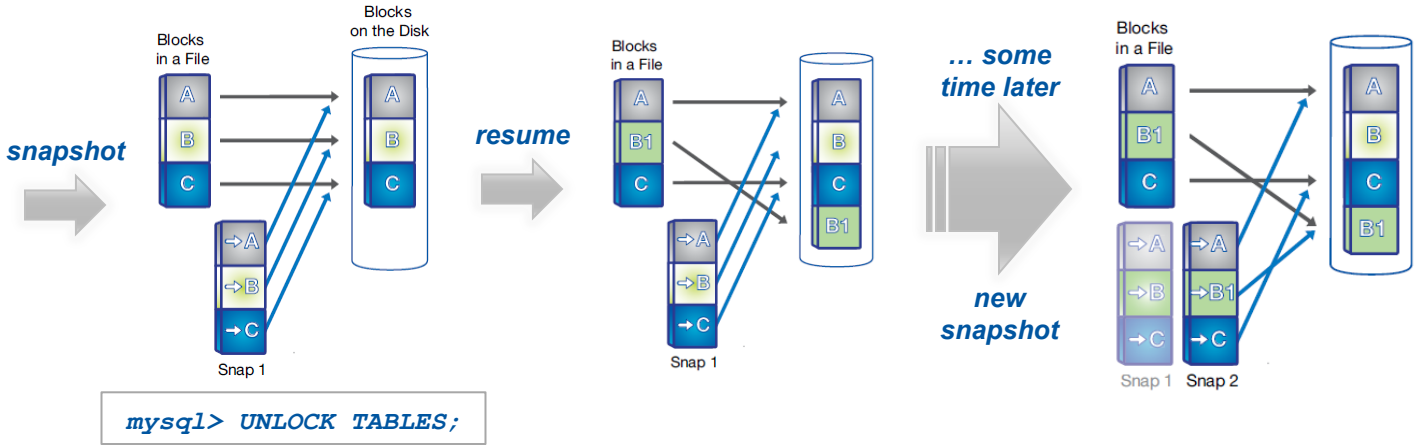
- MySQL instances running in binary log mode (InnoDB recommend)
- Oracle instances running in archivelog mode
- Backups based on storage snapshots
  - Full online DB backups done just in a few seconds
  - Manual and/or automatic (scheduled)
  - Small storage overhead (depends on instance activity)
  - Point-in-time recovery – easy with snapshots and binary/archive logs
  - Snapshots can be configured to be sent to tape (DR)

Backup management

- Backup configuration panel
- Backup procedure:



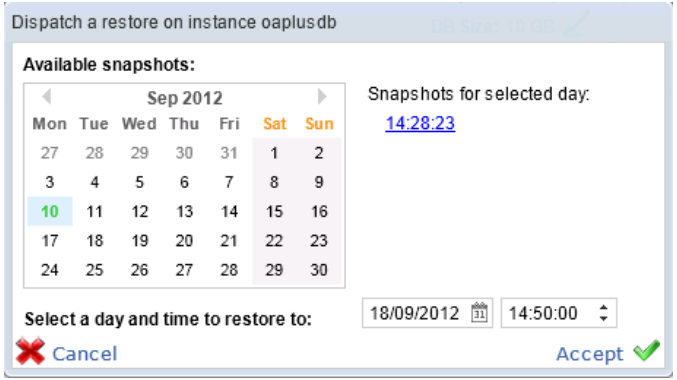

```
mysql> FLUSH TABLES
       WITH READ LOCK;
mysql> FLUSH LOGS;
```



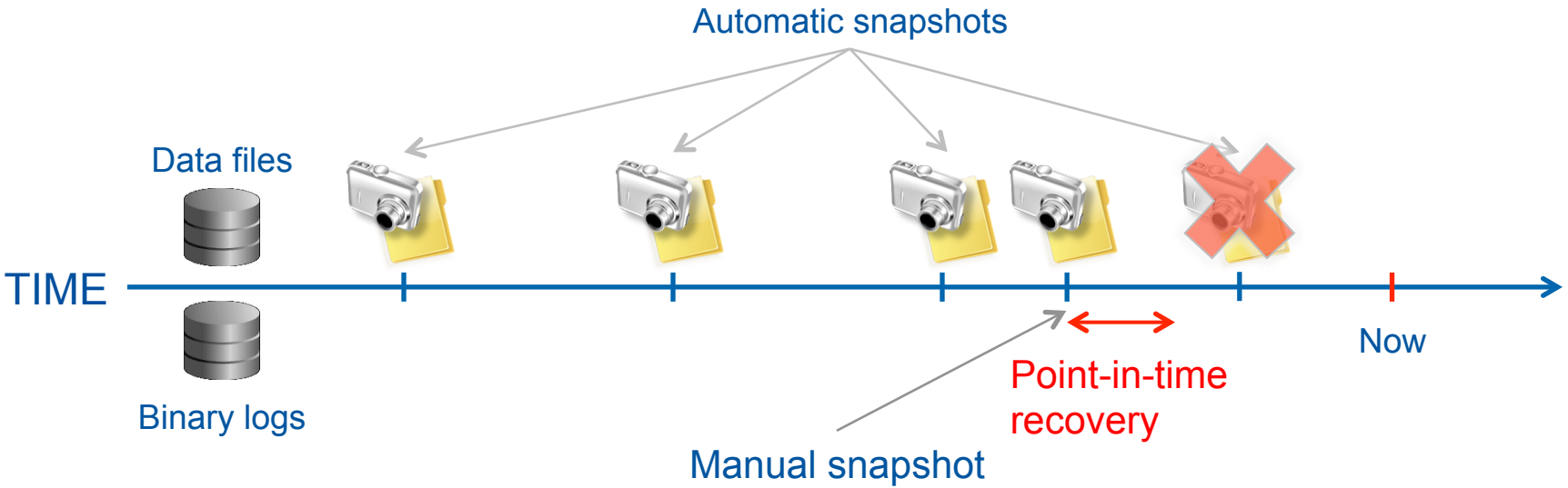

```
mysql> UNLOCK TABLES;
```

Instance restore

- Owners can submit point-in-time recoveries
- Full restore takes just a few seconds
- Recovery time depends on number of binary logs/redo logs to replay/apply
- Warning: snapshots taken after the one used for recovery are lost

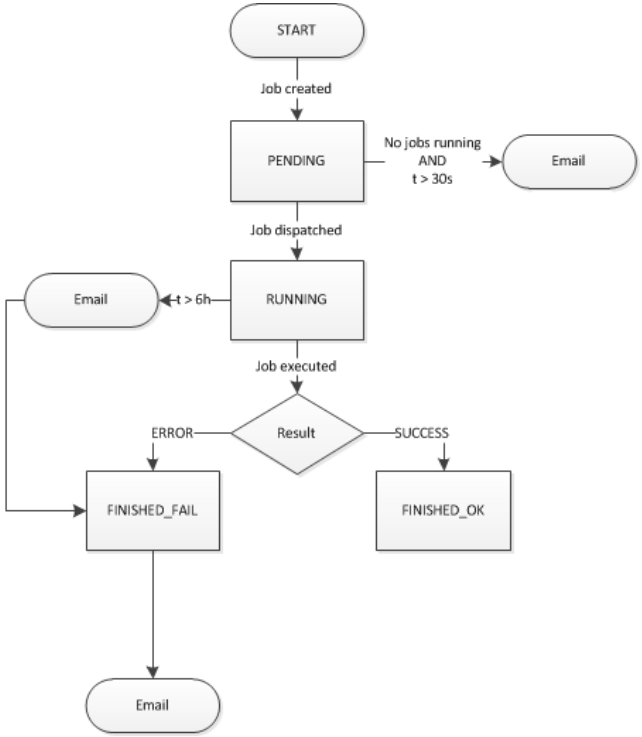


Instance restore



Framework Monitoring

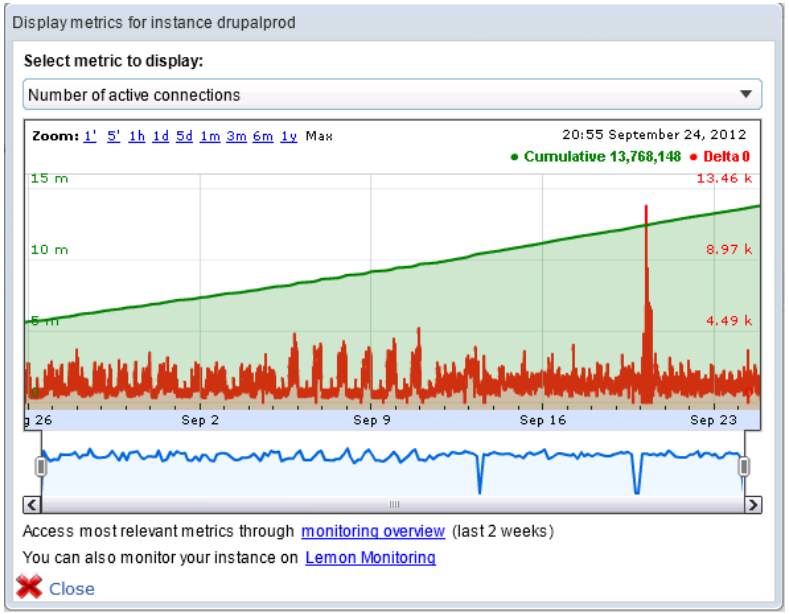
- Management server
  - Queries its jobs table regulary
  - Informs admins in case of:
    - Pending jobs not executed
    - Timed out jobs
    - Failed jobs



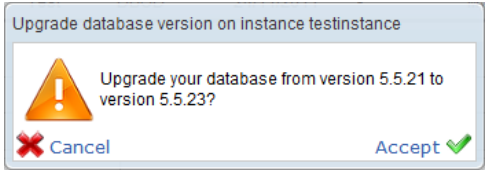
Instance Monitoring

- Evaluated several monitoring products
  - OEM with Pythian plug-in
  - MySQL Enterprise Monitor
- Monitoring server runs RACMon (in-house developed)
  - Availability and performance monitoring system for Oracle DBs, MySQL, NAS storage and VM infrastructure
  - ~30 MySQL metrics stored in monitoring DB
    - `mysql> show status`
  - Selected AWR metrics stored for Oracle instances
  - Admins are notified via email (and SMS if needed) about
    - Availability problems
    - Performance issues (OS level and DB checks)

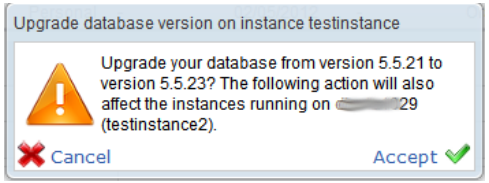
Monitoring interface



One button upgrades



- DBoD admins prepare upgrade scripts
  - Complete upgrade process is scripted and tested
  - Upgrades of one minor version or several minors possible
- Owners can decide to upgrade at their convenience - one button upgrade
  - Instance is stopped
  - Binaries are upgraded
    - shared instances must be upgraded at the same time
  - Instance is restarted
  - All post-installation tasks executed



Important clients

- Hosting ~30 instances at CERN for IT and experiments
  - Drupal content management system
  - BOINC – LHC@home
  - CERN document server
  - Audio video conferencing and webcasts service
  - HammerCloud for experiments
  - Piwik (open source web analytics software)
  - OpenStack Nova
  - Trac for subversion

Summary

- Many lessons learned during the design and implementation of the DBoD service
- Building Database as a Service helped CERN DB group to
  - Gain experience with MySQL
  - Improve tools and operations
  - Standardize on tools and frameworks
  - Consolidate

Acknowledgements

- Other members of Database on Demand
  - Ignacio Coterillo
  - Ruben Gaspar
  - Daniel Gomez

www.cern.ch