Database Futures Workshop

Monday 6 June 2011 - Tuesday 7 June 2011

CERN

Book of Abstracts

Contents

Welcome & Introduction 1

Development of a noSQL storage solution for the Panda Monitoring System 1

Future Oracle use by CMS offline dataflow and workflow management 1

ATLAS DDM/DQ2 & NoSQL databases: Use cases and experiences 1

Overview of Data Management solutions for the Control and Operation of the CERN Accelerators 2

Future Database Requirements in the Accelerator Sector 3

Experience with the CMS online DB and prospects for the future 3

CMS experience with offline Conditions Database and prospects for the future 4

CMS Offline experiences with NoSQL data stores 4

CMS requirements on CERN/IT provided NoSQL data stores 5

Frontier and HTTP caching in the future 5

Enhance the ATLAS database applications by using the new Oracle 11g features 6

Database choices for CERN Drupal service 6

NoSQL Databases and Monitoring 6

Database services for ALICE Detector Control System 7

Use of MySQL in the ALICE data-acquisition system. 7

LHCb Databases - Present and Future 7

Evolution of Databases for the ATLAS experiment 7

Future plans for CORAL and COOL 8

Administrative & Engineering Requirements 8

Databases for the ATLAS Detector Control System: experience and future requirements 8

Qserv: Distributed Shared-nothing from MySQL and Xrootd 9

Rapid Summary 9

0

Welcome & Introduction

Implementations II / **1**

Development of a noSQL storage solution for the Panda Monitoring System

Author: Maxim Potekhin[1]

[1] *Brookhaven National Laboratory (BNL)*

Corresponding Author: maxim.potekhin@cern.ch

For the past few years, Panda Workload Management System has been the mainstay of computing power for ATLAS experiment at the LHC. Since the start of data taking, Panda usage gradually ramped up to 840,000 jobs processed daily in the Fall of 2010, and remains at consistently high levels ever since. Given the upward trend in workload and associated monitoring data volume, the Panda team is facing a new set of challenges in the areas of database scalability and efficiency of its monitoring system. These challenges are being met with a R&D effort aimed at implementing a scalable and efficient monitoring data storage based on a noSQL solution (Cassandra). We present our motivations for using this technology, as well as data design and the techniques for efficient indexing of the specific data, which have been tested in two different hardware configurations.

Requirements III / **2**

Future Oracle use by CMS offline dataflow and workflow management

Author: Tony Wildish[1]

[1] *PRINCETON*

Corresponding Author: tony.wildish@cern.ch

We describe the current use of Oracle by CMS offline dataflow and workflow components (T0, PhEDEx, DBS).

We consider how the database use is expected to evolve over the next few years, in terms of both data-volume, data-structure and application-use

Proposed speaker:

Tony Wildish

Implementations II / **3**

ATLAS DDM/DQ2 & NoSQL databases: Use cases and experiences

Authors: Angelos Molfetas[1]; Donal Zang[2]; Gancho Dimitrov[3]; Luca Canali[4]; Mario Lassnig[1]; Vincent Garonne[5]

[1] *CERN-PH-ADP-CO*

[2] *IHEP*

[3] *BNL*

[4] *CERN-IT-DB*

[5] *Conseil Europeen Recherche Nucl. (CERN)-Unknown-Unknown*

Corresponding Author: vincent.garonne@cern.ch

The Distributed Data Management System DQ2 is responsible for the global management of petabytes of ATLAS physics data. DQ2 has a critical dependency on Relational Database Management Systems (RDBMS), like Oracle, as RDBMS are well-suited to enforce data integrity in online transaction processing application. Despite these advantages, concerns have been raised recently on the scalability of data warehouse-like workload against the transactional schema, in particular for the analysis of archived data or the aggregation of data for summary purposes. Therefore, we have considered new approaches of handling vast amount of data. More specifically, we investigated a new class of database technologies commonly referred to as NoSQL databases. This includes distributed filesystem like HDFS that support parallel execution of computational tasks on distributed data, as well as schema-less approaches via key-value stores, like HBase, Cassandra or MongoDB. These databases provide solutions to particular types of problems: for example, NoSQL databases have demonstrated that they can scale horizontally, deliver high throughput, have automatic fail-over mechanisms, and provide easy replication support over LAN and WAN.
In this talk, we will describe our use cases in ATLAS, and share our experiences with NoSQL databases in a comparative study with Oracle.

Proposed speaker:

V.Garonne

Requirements II / 4

Overview of Data Management solutions for the Control and Operation of the CERN Accelerators

Author: Ronny Billen[1]

Co-authors: Chris Roderick [1]; Zory Zaharieva [1]

[1] *CERN*

Corresponding Authors: zornitsa.zaharieva@cern.ch, chris.roderick@cern.ch, ronny.billen@cern.ch

The control and operation of the CERN accelerator complex is fully based on data-driven applications. The data foundation models the complex reality, necessary for the configuration of the accelerators controls systems and is used in an online and dynamic way to drive the particle beams and surrounding installations. Integrity of the data and performance of the data-interacting applications are key requirements and challenges that have been satisfied.
This presentation will give an overview of what is currently in production, from the mission-critical data (controls configuration, operational settings, alarms, logging,...) to the closely related offline information (layout, equipment details,...) and the need that all of this has to fit together (relationally). Figures of complexity and performance will be given, also indicating the means that we have put in place to monitor, diagnose and track the usage of our data management services.

Proposed speaker:

Zory Zaharieva & Chris Roderick

Requirements II / **5**

Future Database Requirements in the Accelerator Sector

Author: Ronny Billen[1]

[1] *CERN*

Corresponding Author: ronny.billen@cern.ch

Since more than two decades, relational database design and implementations have been satisfying data management needs in the CERN Accelerator Sector. The requirements always covered a wide range of functional domains from complex controls systems configuration data to the tracking of high-volume data acquisitions. The requirements to store large data sets have increased by several orders of magnitude between the consecutive epochs from SPS to LEP to LHC. So far, scalability has been ensured by following the hardware and software technology. Looking ahead –towards CLIC for example-, will we still be able to continue the same route or will this strategy fail eventually? This presentation will outline some of these issues in the different domains and raises the questions that have to be addressed in due time.

Proposed speaker:

Ronny Billen

Requirements III / **6**

Experience with the CMS online DB and prospects for the future

Authors: Andreas Pfeiffer[1]; Francesca Cavallari[2]; Frank Glege[1]; Mindaugas Janulis[3]

[1] *CERN*

[2] *Univ. + INFN Roma 1*

[3] *Vilnius University*

Corresponding Authors: frank.glege@cern.ch, francesca.cavallari@cern.ch

CMS has chosen to use an online DB located at IP5
both for security reasons and to be able to take data
even without GPN connection.
The online DB (OMDS) is accessed by various applications for
data acquisition configuration (through OCI libraries via TStore),
detector slow control (via PVSS) and monitoring
via java or c++ libraries.
It also contains offline conditions data which are needed for high
level trigger
system which is running a simplified version of the event
reconstruction program on a cluster of few hundreds of machines.
A caching system based on Frontier allows to
reduce the load on the DB for this application similarly to what is
used in the offline DB.
A web based monitoring allows to display the run list and
most of the monitoring information. This tool makes use of caches

in order to reduce the load on the DB.
Many other applications rely on the DB: storage manager,
elog, access control packages.
Streaming is used to duplicate data for analysis
access via lxplus for the detector experts.

So far the OMDS has collected abut 1.5 TB of data per year.
Heavy use of the query optimization through appropriate indeces and
partitioning is used in the largest accounts.
Partitioning will allow archiving of old data if space limitation
or performance become an issue.

The experience with the online DB during 2010 data-taking is discussed and prospects for the future.

Proposed speaker:

Frank Glege

Requirements III / **7**

CMS experience with offline Conditions Database and prospects for the future

Authors: Andreas Pfeiffer[1]; Antonio Pierro[2]; Francesca Cavallari[3]; Giacomo Govi[4]; Salvatore Di Guida[1]; Vincenzo Innocente[1]

[1] *CERN*

[2] *INFN Bari*

[3] *Univ. + INFN Roma 1*

[4] *Fermilab*

Corresponding Authors: giacomo.govi@cern.ch, francesca.cavallari@cern.ch

CMS experiment is made of many detectors which in total sum up to
60 million channels. Calibrations and alignments are fundamental to
maintain the design performance of the experiment. The conditions
database contains the alignment and calibrations data for the various
detectors.
Conditions data sets are accessed by a tag and an interval of validity through the offline reconstruction program CMSSW, written in C++.
Performant access to the conditions data as C++ objects is a key
requirement for the reconstruction and data analysis.
About 200 types of calibration and alignment exist for the various
CMS sub-detectors. Each set is grouped in a so-called “global tag”
which is valid for a given period of data-taking and for a given data
set (Monte Carlo events or collisions data).
Only those data which are crucial for reconstruction are inserted into the offline conditions DB. This guarantees a fast access to conditions during reconstruction and a small size of the conditions DB.

The talk describes the experience with the offline reconstruction
conditions database during 2010 and prospects for the future.

Technologies I / **8**

CMS Offline experiences with NoSQL data stores

Authors: Simon Metson[1]; valentin kuznetsov[2]

Co-authors: Dave Evans [3]; Lassi Tuura [4]

[1] *H.H. Wills Physics Laboratory*

[2] *cornell*

[3] *FNAL*

[4] *FNAL/CERN*

Corresponding Authors: valentin.kuznetsov@cern.ch, simon.metson@cern.ch

The CMS Offline project has been developing against "NoSQL" data stores since 2009 and have experience with three projects in particular; CouchDB, Kyoto Cabinet and MongoDB. We present how these tools are used in our software, why they were chosen and lessons we've learnt along the way."

Technologies I / 9

CMS requirements on CERN/IT provided NoSQL data stores

Author: Simon Metson[1]

Co-authors: Andreas Pfeiffer [2]; Dave Evans [3]; Francesca Cavallari [2]; Ian Fisk [3]

[1] *H.H. Wills Physics Laboratory*

[2] *CERN*

[3] *FNAL*

Corresponding Author: simon.metson@cern.ch

We discuss potential future requirements for CERN/IT managed/provided "NoSQL" data stores, and provide some high level observations based on our experiences with these technologies.

Technologies II / 10

Frontier and HTTP caching in the future

Author: Dave Dykstra[1]

[1] *Fermilab*

Corresponding Author: dwd@fnal.gov

Frontier has been successfully distributing high-volume, high throughput, and long-distance data for CMS for many years and more recently for ATLAS, greatly reducing the expectations on the WLCG database servers. This talk will briefly describe the present status and cover the expected changes coming in the future. No major changes are foreseen, but improvements in robustness and security, and increased numbers of user projects are expected. Even more applications are expected for HTTP caches, which will require enhancements to the configuration and monitoring of the WLCG squid network.

Proposed speaker:

Dave Dykstra

Implementations I / **11**

Enhance the ATLAS database applications by using the new Oracle 11g features

Author: Gancho Dimitrov[1]

[1] *BNL*

Corresponding Author: gancho.dimitrov@cern.ch

It is planned that in the beginning of 2012 all ATLAS
databases at CERN will be upgraded to the Oracle 11g Release 2. In the light of making the ATLAS DB applications more reliable and
performant, we would like to explore and evaluate the new 11g database features for development and performance tuning. In the talk will be described the expected benefits of having some of the Oracle 11g enhancements in place and typical ATLAS use cases for which they would suit best.

Requirements I / **12**

Database choices for CERN Drupal service

Author: Tim Bell[1]

[1] *CERN*

Corresponding Author: tim.bell@cern.ch

CERN is deploying a new content management approach based on Drupal (http://drupal.org) for the main www.cern.ch site, departments and experiments. This talk will review the requirements and options for the database part of the deployment to create an infrastructure capable of supporting millions of hits per day.

Technologies I / **13**

NoSQL Databases and Monitoring

Author: Jerome Belleman[1]

[1] *CERN*

Corresponding Author: jerome.belleman@cern.ch

Monitoring typically requires to store large amounts of metric samples which are recorded at a high rate. These samples must then be massively read back and reprocessed for analysis and visualisation purposes. For the past few years, different monitoring systems have been developed on top of NoSQL databases for the scalability they provide. Likewise, a monitoring system for the batch service is currently being developed at CERN and the use of a NoSQL database as one of its components is under investigation. This talk describes to what extent NoSQL databases are suitable for working with monitoring information, as opposed to SQL databases.

Requirements I / 14

Database services for ALICE Detector Control System

Author: Peter Chochula[1]

[1] *CERN*

Corresponding Author: peter.chochula@cern.ch

We describe the architecture and implementation of the ALICE DCS database service. The whole dataflow from devices to the ORACLE database as well as the interface to online and offline data consumers is briefly overviewed.
The operational experience with the present configuration as well as future plans and requirements are summarized in this talk.

Requirements I / 15

Use of MySQL in the ALICE data-acquisition system.

Author: Sylvain Chapeland[1]

[1] *CERN*

Corresponding Author: sylvain.chapeland@cern.ch

MySQL has been in use to store and access structured information for the ALICE data-acquisition since 2004. It copes with the implementation of 9 distinct data repositories (configurations, logs, etc) for the online subsystems implemented at the experimental area, all of them having different I/O patterns and requirements.
We will review the architecture, performance, features, and future needs of our online database systems. Feedback about our positive experience with this tool will be given.

Implementations I / 16

LHCb Databases - Present and Future

Author: Marco Clemencic[1]

[1] *CERN PH-LBC*

Corresponding Author: marco.clemencic@cern.ch

Several database applications are used by the LHCb collaboration to help and organize the day-to-day tasks, to assist the data taking, processing and analysis.
I will present a brief overview of the technologies used and the requirements for the long term support of both the current database applications and the possible future ones.

Implementations I / 17

Evolution of Databases for the ATLAS experiment

Author: Dario Barberis[1]

[1] *CERN*

Corresponding Author: dario.barberis@cern.ch

The use of databases in ATLAS is going through a continuous process of development, deployment and optimisation, in order to cope with the increasing amounts of data and new demands from the user community. In 2011 and 2012 work will concentrate on two major lines, namely the transition to Oracle 11g and re-optimisation of the existing database in Oracle, and the study of new technologies (NoSQL databases) for specific applications. This talk will give an overview of these activities and an introduction to specific talks.

Technologies II / **18**

Future plans for CORAL and COOL

Author: Andrea Valassi[1]

[1] *CERN*

Corresponding Author: andrea.valassi@cern.ch

This presentation will report on the current plans for the future meintenance and development of two Persistency Framework packages used by several LHC experiments for accessing Oracle databases: CORAL (the generic RDBMS access layer, used by ATLAS, CMS and LHCb) and COOL (the conditions database package used by ATLAS and LHCb). It will also cover the status and plans for the CORAL Server, the middle tier technology (similar to Frontier/Squid) used by ATLAS online for the configuration of the High Level Trigger.

Requirements II / **19**

Administrative & Engineering Requirements

Authors: Christophe Delamare[1]; Derek Mathieson[1]

[1] *CERN*

We present the range of Administrative and Engineering applications together with expectations for future developments, growth and requirements.

Implementations II / **20**

Databases for the ATLAS Detector Control System: experience and future requirements

Author: Stefan Schlenker[1]

Co-author: Viatcheslav Khomutnikov [2]

[1] *CERN*

[2] *PNPI Petersburg*

Corresponding Author: stefan.schlenker@cern.ch

The ATLAS detector control system (DCS) archives detector conditions data in a dedicated Oracle database using a proprietary schema (PVSS Oracle archive) and representing one of the main users of the ATLAS online database service. The contribution will give an overview about the database usage and operation experience, e.g. with respect to data volume, insert rates, and pending issues. Constraints and ideas for future requirements in the view of experiment operation and upgrades are discussed.

Where Next? / **21**

Qserv: Distributed Shared-nothing from MySQL and Xrootd

Author: Daniel Wang[1]

[1] *SLAC National Accelerator Laboratory*

Corresponding Author: danielw@slac.stanford.edu

The LSST catalog of celestial objects will need to answer both simple and complex queries over many billions of rows. Since no existing open-source database efficiently supports its requirements, we have are developing Qserv, a prototype database-style system, to handle such volumes. Qserv uses Xrootd as a framework for data-addressed communication to a cluster of machines with standalone MySQL instances. Xrootd provides fault-tolerance and replication support, while MySQL provides a basic SQL execution engine. Using a spatial spherical partitioning approach, Qserv fragments queries and aggregates results scalably even for expensive spatial self-joins.

Where Next? / **22**

Rapid Summary

Corresponding Author: tony.cass@cern.ch