

6 May 2014 CERN openlab IT Challenges workshop, <u>https://indico.cern.ch/event/297559/</u>

CERN and Data Analytics Challenges

Kacper Szkudlarek, CERN Manuel Martin Marquez, CERN Eric Grancher, CERN

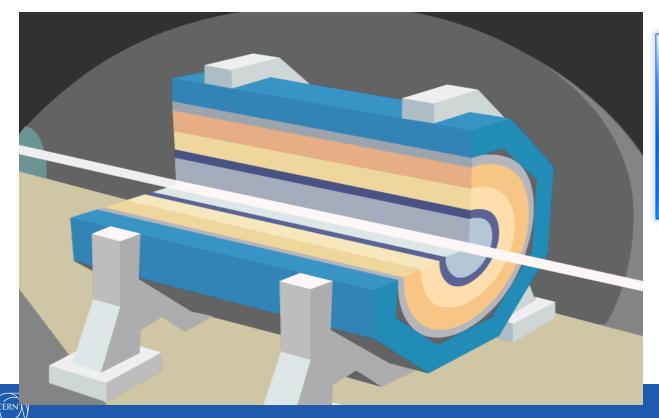


Outline

- CERN and openlab
- Analytics challenges
 - Analytics investigation
 - Fault prediction
 - CERN openIab analytics challenges
 - Technology evolutions
 - Analytics as a service



Events at LHC



Luminosity : 10³⁴cm⁻² s⁻¹

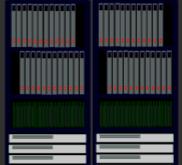
40 MHz – every 25 ns

20 events overlaying

Trigger & Data Acquisition

~ 300.000 MB/s from all sub-detectors

Trigger and data acquisition



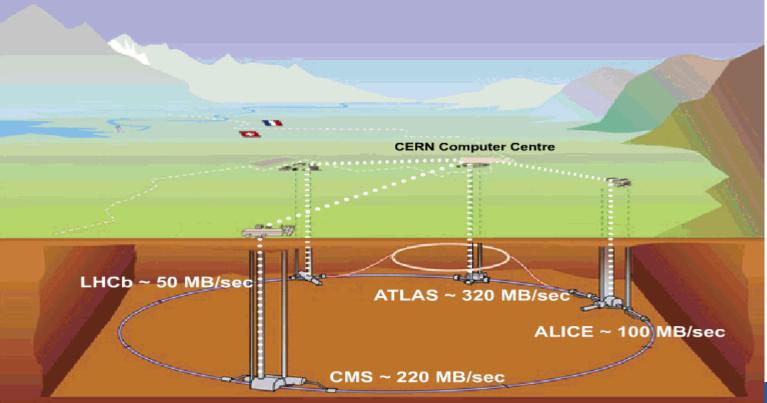
~ 300MB/s Raw Data

Event filter computer farm





Data Recording





# of cores:	99,061
# of disks:	80,754
# of processors:	18,633
# of 10GB NICs:	3,218
# of 1GB NICs:	19,334
# of servers:	10,809
Disk space (TiB):	129,455
RAM memory (TiB):	349
CERN	

Credit: Alberto Pace

Distributed analysis in the different sites (« LHC Computing Grid »)



CPUs



Network



Databases



Storage



Infrastructure



openlab (1/3)

- Public-private partnership between CERN and leading ICT companies, currently in fourth phase (started in 2003)
- Its mission is to accelerate the development of cutting-edge solutions to be used by the worldwide LHC community
- Innovative ideas aligned between CERN and the partners, for products "you make it, we break it"



Partners



ORACLE

SIEMENS

Contributors



Associates





openlab (2/3)

- Many successes (DB competence center/Oracle):
 - RAC on Linux x86 (9.2 PoC, 10.1 production with ASM),
 - Additional required functionality (IEEE numbers, OCCI, instant client, etc.),
 - WinCC OA and RAC scalability,
 - Monitoring with Grid Control,
 - Streams world wide distribution,
 - Active DG, GoldenGate,
 - Analytics for accelerator, experiment and IT,
 - Etc.





openlab (3/3)

- Publications (web, paper) and presentations of results, visitors
- Maaike Limper, best poster award at The International **Conference on Computing** in High Energy and **Nuclear Physics 2013**

An SQL-based approach to Physics Analysis H. Linger CERNoperlab ELECT HIGGS



Introduction: As part of the CERN openiab collaboration an investigation has been made into the use of an SQL-based approach for physics analysis. Currently, physics analysis is done using data stored in centrally produced root-intuples that are accessible through the LHC computing grid. We'll present an alternative approach to physics analysis where analysis data is stored in a database. This would remove the need for customized stuple production, and allows some of the calculations that are part of the analysis to be done on the database side.

Dataset and database design: the bootboark analysis was tested using a subset of ATLAS experiment data from mol-staples that were cantrally produced for the ATLAS top-physics group. Root etaples store data relation wise, while Oracle process all related attributes together by row. A database dealers was chosen where phasics objects were stored in separate tables.

Gola combilency is guaranteed through the PrimaryKey constraint on the Ranificander, EventNumber attributes in the aventData-table which is referred to by all other tables through a ForeignKey constraint The table as the right shows the volume of the test-data in our database, extracted from a subset granger me tz tr of 127 stuples containing a total of 7.2 million events, with 4000 analysis variables per event.

NAMES AND TAXABLE POST OF TAXA 288 1883 1884 100 discourse in 204 100 states and Concerns of the local division of the local

implemented as a benchmark. In this case the original "RootCore"-packages used by ATLAS are compared to a modified set of packages that 271 388 383 14.4 1005 1018 ANT ANTING SC AN AN 20 22 14

Reiskel and pr. Tantif' taking

Property lies

104000

Physics Analysis in SQL: the sqL-version of the bonchmark analysis is built through a series of select statements on each eleject-table, each with a WHERE chapse to apply salection orthants. Object-selection can be danie via temporary tables asing the WETH-AS statement: WITH goodinuons AS ISELECT ... FROM muss WHERE pt-25.) or by explicitly creating a table holding the objects. Hadestalized views can be used in define convexon selection criteria. For example, the benchmarks used a materialized nine to define the good karningsity-block selection. At the end of the query, 30th statements on the Excitoration, EventNumber attributes are used to put information from the different selections together: SELECT ... F4YOM gend_maxees INNED JOBV good_takin USING Gardwater, Everthanter) WHERE (profession Mr.2.4MD) goodiats N-2

The user might be not be able (or willing) to re-write all analysis code in SQL Eimple cabidations can be written in PL/SQL and one can also call existing C++ Shiarles from inside the SQL-query. In this case the external libraries ment to be uploaded to the DB reachines and linked to PL/SQL functions. For excepts, one of the external C++ Brarles used by the benchmarks was used in the b-jet identification to re-calculate the b-tagging likelihood.

CPU usage: The suprod-dated was used to study CPU usage. Here, the sluple version was executed using 60 root-jobs (1 per disk). The plots on the right show that the Higgs+2 benchmark was fastest with the situple version and both version were licelled by lowall. The libar outflow beechmark was faster with the DB-version as the RootCorepackages were limited by CPU.

Hadoop: On the mapred-charter the text dataset was also stored as comma delimited tout-files in the hadron filesystem (hulfe). The Hadosp system was configured to have 40 task slots (8 per node) to match the number of cores in the system. The Higgs+2 benchmark analysis was reproduced using MapReduce-code written in java. The Higgs+2 analyti In Hadoop used a relatively large amount of CPU and was alower than both the staple and DB-version.

retrieve data from the DB via an EQL-gamy. This more realistic analysis involves 319 variables, and used data from the same tables as the Hope+2 benchmark as well as data from the photon-table. Test setups: two types of test satups were used. The 'test?'-satup used 2 machines with network based file storage (NPS) accessible from all reades. The 2nd test salue, "mapred", was designed to run either Hadeop or Oracle RAE and was optimized for fast L/O using 5 machines connected to 5

disk arrays holding a total of 60 disks. On this test-setup the Oracle database used the Autoesalic Storoge Hanageneest Feature, and Hadoop used its fully filesesters, to corned Technetica - "Inco"/ Chapter the data scenty over all devices. R 100 000 2 4 For the comparison with root on Build speed - 240 ABIN 2400 ABIN the manual choice. The situation users distributed every over all disks. Inte CPE name 12 42

Benchmarks: A mothed version of the

as an SQL-matrix. This assatysis returns the

In addition, a cuttless analysis for the bas-out

production cross-section measurement was

40 variables.

search for the Higgs in association with a 2 boson

invariant mass of the lepton- and let-oalr and uses

was implemented, both as a single root-matere and

Parallel execution: An SQL gamy can be executed in serial or is parallel and the degree of parallelism can be set on the table or by a hint inside the quary. For the relapie analysis, parallelium was receiched by running multiple simultaneous root-tobe, each enalysing a subset of files. The rustersion gained more from parallelium than the D8-variant of the analysis. This is because the DB-version is limited by 1/0 speed as it needs to read many columns in the table to Find the relevant variables.

Conclusion: Physics Analysis using SQL on data stored in a database can provide an alternative way to analyse the large datasets produced by the LHC experiments, Row-based storage in combination with wide tables limits performance by the I/O read speed of the system. Future studies will focus on columnar stores to improve performance



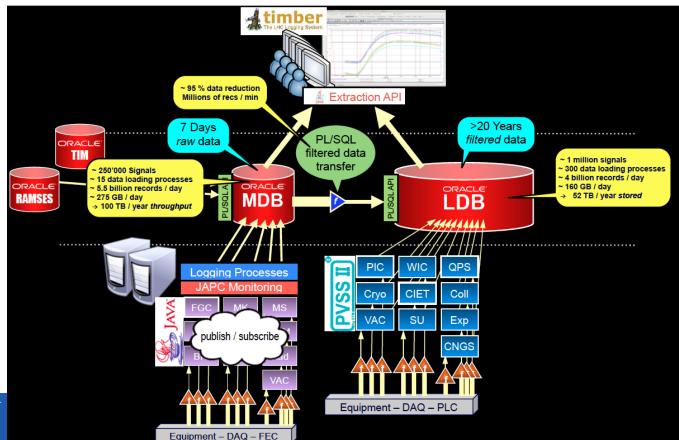
Networked devices

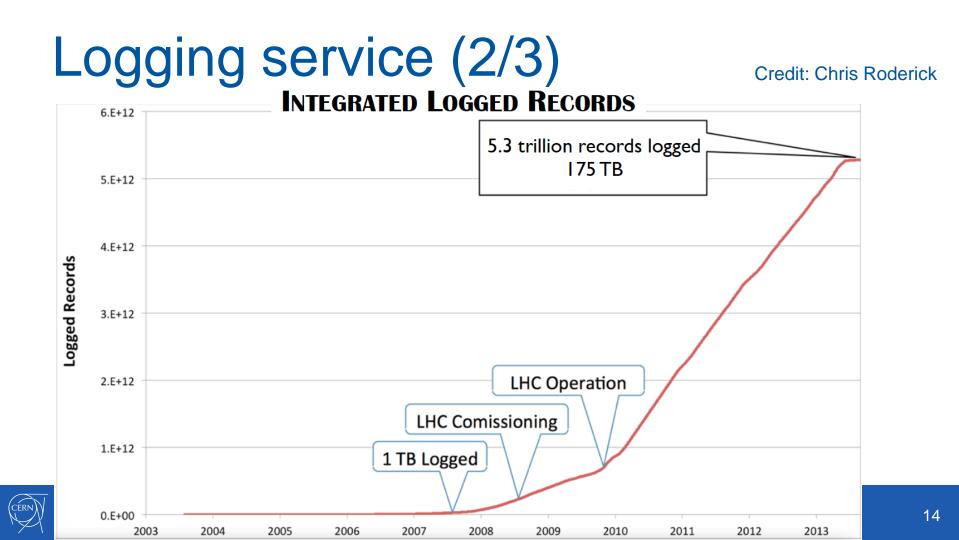
- Large number of control devices, front-end equipment, etc.
- Many critical systems: Cryogenics, Vacuums, Machine Protection, etc.
 - (Also called "Internet of Things")
- Example: LHC logging service (BE-CO)

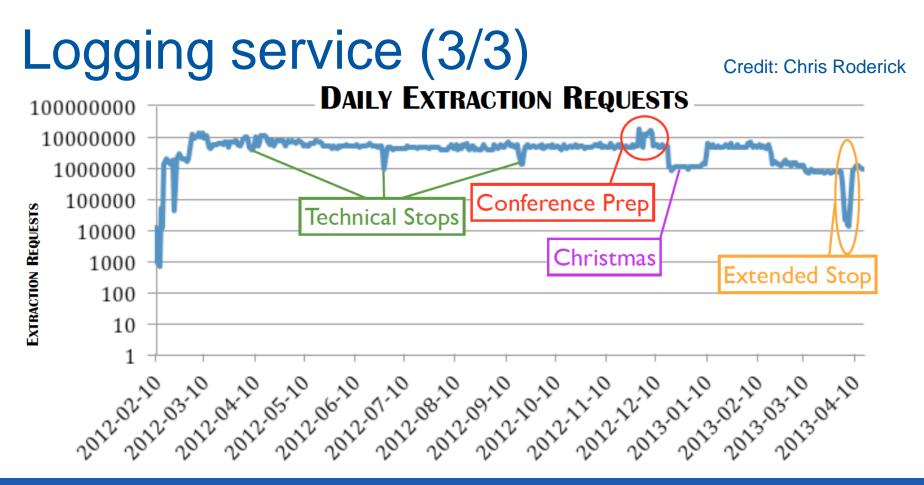


Logging service (1/3)

Credit: Chris Roderick









Improvements for Archiver of WinCC Open Architecture

Report on work by departments: Information Technology Beams Engineering



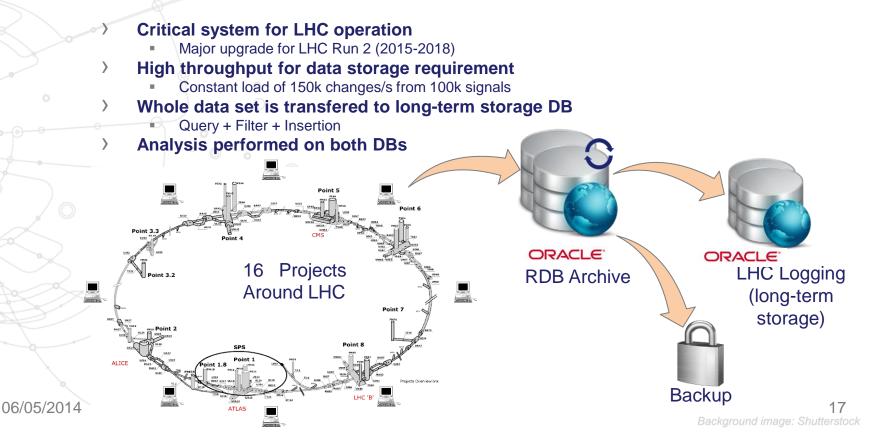
Kacper Szkudlarek

Background image: Shutterstock

CERN openlab & Oracle Analytics Innovation Summit



Use case: Quench Protection System





BE-CO

IT-DB

EN-

ICE

Performance Optimizations

> Use of Oracle Index Organized Tables
> Tuning of DB data queries

- Search predicates, time-based partitioning
- Alignment of data in RAC cluster
- Changes in database schema
 - Focus on redo log and space reduction
- > Tuning of database parameters



Results summary

I/O Megabytes per Second

280

240

160

8D

08:00PM 10:00PM 12:00AM 02:00AM 04:00AM 06:00AM

Nominal conditions

- Stable constant load of 150k changes/s
 - 100 MB/s of I/O operations
- 500 GB of data stored each day

Peak performance

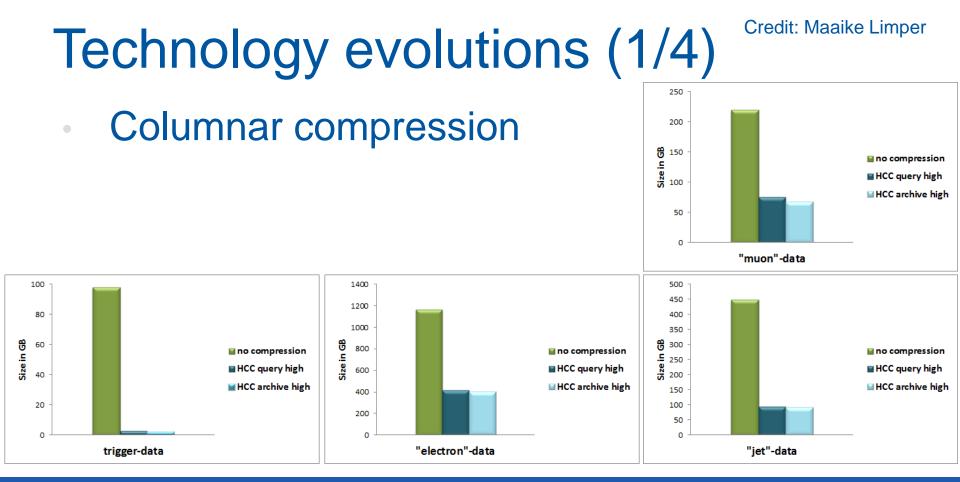
- Exceeded 1 million value changes per second
- 500-600 MB/s of I/O operations

All CERN production WinCC OA systems (accelerators, detectors and technical infrastructure, 600 servers) will benefit from these optimizations

Next challenge: ~10x increase

Required for next major upgrade (2019-2020)

06/05/2014





Technology evolutions (2/4)

- In-memory databases
 - Commodity servers with 6TB memory or more
 - Together with compression, enable to have (for some applications) the active data set fully in memory
 - Flash with low latency storage



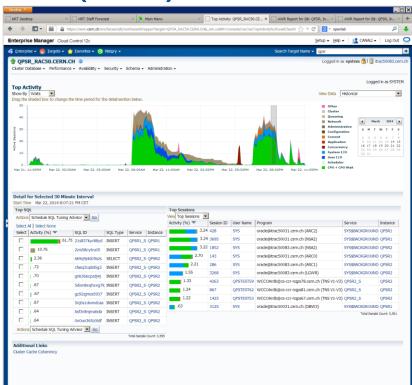
Technology evolutions (3/4)

- Networking
 - 10GbE and more
 - (security...)



Technology innovation (4/4)

Scalable database and storage infrastructure
Powerful analysis packages like ROOT and R



Credit: Luca Canali



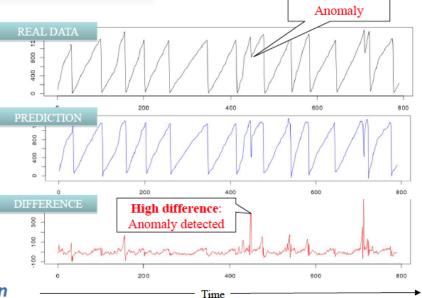
Anomaly detection

Credit: Massimo Lamanna, Sebastien Ponce (IT-DSS), Stefano Alberto Russo (ex IT-DB)



1) Build a **SVM** (*Neural Network like*) **model** - **self trained**

- no supervision
- 2) Predict and compare: Real data Vs Prediction
 - 1) Blindly recognize anomalies
 - 2) No other information required (i.e. thresholds)





Information discovery

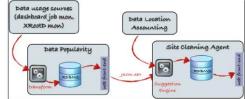
<u>URL link</u>

Manuel Martin Marquez, CERN IT-DB



Experiment data placement

Intelligent data placement models for the CMS and ATLAS experiments



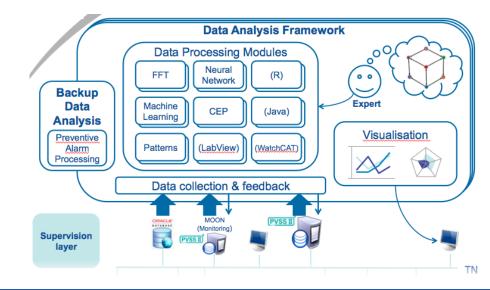
- Need to extract further knowledge from the monitoring data in order to implement an effective data placement
 - Correlate file-access monitoring with site status
 - Readiness, queue length, storage and CPU available
 - Classify analysis activities and needed resources
 - Making recommendations
 - Learn from the past trends and patterns



Analytics as a service

- "Analytics platform" or (Big data) "Analytics-as-aservice" (A³S ?):
- Data fed from multiple sources (live)
- Stored reliably
- Data processing with multiple systems
- Easy access, domain expert natural language (DSL)
- Visualisation

Credit: CERN EN-ICE





Conclusion

- CERN: very diverse and challenging requirements
 CERN openIab IV, a lot done on Analytics
- CERN openlab V, a lot to be done, CERN and others (EMBL-EBI, ESA, etc.), whitepaper) <u>http://cern.ch/openlab/</u>
 - "you make it, we break it"
- Complex integration challenges of rapidly evolving technologies
- Some more about Oracle and CERN... DB12c





www.cern.ch