Future Database Requirements in the Accelerator Sector

Ronny Billen

Beams Department
Controls Group
Data Management Section

Database Futures Workshop – 6-7 June 2011

Outline

- Data Managed in the Accelerator Sector
- Past Accelerator Logging Experience
- Current Accelerator Logging Experience
- Future Logging Requirements
- How Realistic is this Future?

Data managed in the Accelerator Sector

	Configuration Data	Logging Data
Purpose	Reality modeling for exploitation	Live data tracking of over time
Data Model	Complex / Very complex > 100 objects	Simple: time series Values may be complex (events)
Data Interdependency	High Many relations, constraints	Low Few relations, constraints
Data Evolution	Quite static History of changes	Very Dynamic Continuous growth
Data Volumes	Small < 10GB	Huge > 100TB
Topics	Hardware Installations Controls & Communications Operational parameters Alarm definitions	Hardware/Beam Commissioning Equipment monitoring Beam measurements Post-Mortem events
Data Criticality	Low – High – Very High Data integrity is essential	High Data correctness not guaranteed
Current Implementation	Oracle RDBMS	Oracle RDBMS

Continue to implement this way

Main question & worry for the future

Past Accelerator Logging Experience

- LEP Logging 1992-2000
  - Purpose: Centralized storage and accessibility of data acquisitions of interest over time (beam and equipment)
  - Initial idea to keep *one year* up to *a few years* of useful data
  - Initial estimation of the *very large database*: 8GB/year
  - Implementation started with Oracle 6
  - Provided a generic GUI to *visualize* and *extract* data
  - Pushed by the end-users, this evolved into:
    - Short-term Measurement DB
    - Long-term Logging database
    - Spinoff LEP RF Measurement database
  - Data exploited several years after the LEP stop
  - The grand total of 266GB of data still available

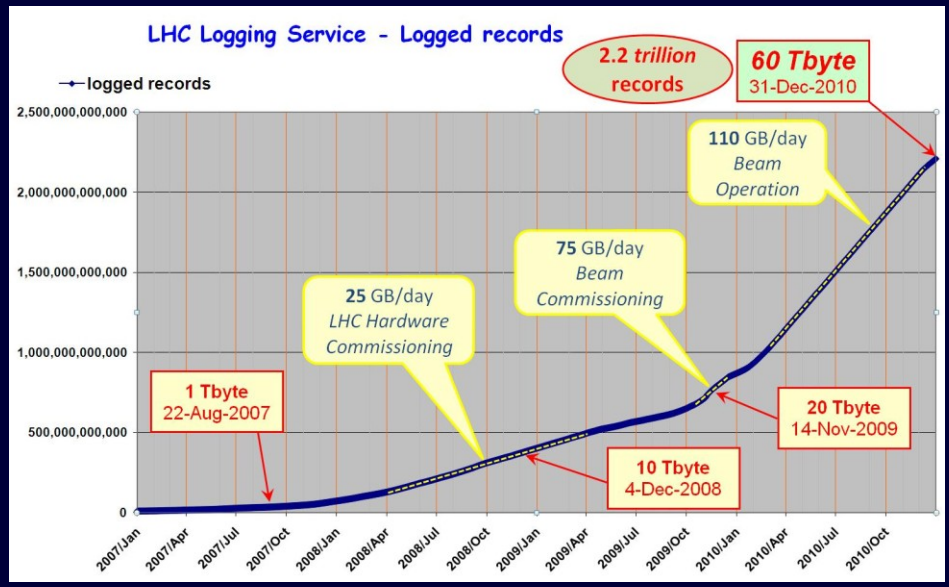
R. Billen

Current Accelerator Logging Experience

- LHC Logging
  - Project start in 2001
  - Design based on *experience* and *technical progress*
  - Initial idea to keep useful data *on-line* for the lifetime of LHC
  - Provided a generic GUI to visualize and extract data
  - Provided Java API, used extensively for data analysis
  - Evolution of estimations for LHC steady-state

Year of Estimation	Estimated volume
2003	1 TB/year
2006	5 TB/year
2008	10 TB/year
2010	50 TB/year

  - Today's outstanding request: *Store more, analyze faster*



R. Billen

Future Logging Requirements

- The 3 TeV CLIC example
  - 50km installation, ~6 10^6 data acquisition channels
  - 50Hz synchronous data (20ms cycle base)
  - Raw data +100GB/s i.e. per year in the Exabyte range
- Estimated yearly storage requirement in the Petabyte range
- Hypothesis ⇦ *Can be challenged!!!*
  - 200 days/year
  - Standard reduction by removing redundancy and zero suppression
  - Average 100 byte/channel
  - Several logging policies
  - Sample size is equipment dependent (RF equipment, Beam Instrumentation,…)

Courtesy M. Jonker

Logging Policy	Retention Time	Filtering function	Samples
Very Short Term	7000s	All pulses	350'000
Short Term	4 days	1s averages	350'000
Medium Term	40 days	10s averages	350'000
Long Term	2 years	1.5 min averages	350'000
Very Long Term	20 years	15 min averages	350'000
Post-Interlock Snapshots Short Term	4 days	All interlock events	700'000
Post-Interlock Snapshots Long Term	20 years	1/100 interlock events	700'000

R. Billen

How realistic is this future?

- Storage requirements in the range Petabyte-Exabyte
- Extrapolating from the past ⇨
- Will we be able to handle this?
  - ⇨ Data storage
  - ⇨ Data retrieval
  - ⇨ Data Backup
- Technical Solutions
  - ⇨ Database powered
  - ⇨ Experiment-type file-based solutions
- The issue is not to be addressed in 10 years…but right now…
  - …which brings me to the important part of this presentation:

The Discussion

Year	Accelerator Operation	Required Storage
2000	LEP	250 GB
2010	LHC	50 TB
2020	HL-LHC	10 PB
2030	CLIC HE-LHC	2 EB

x 200
x 200
x 200

R. Billen