

Thoughts on a Conditions Database

Online Meeting

Mark M. Ito

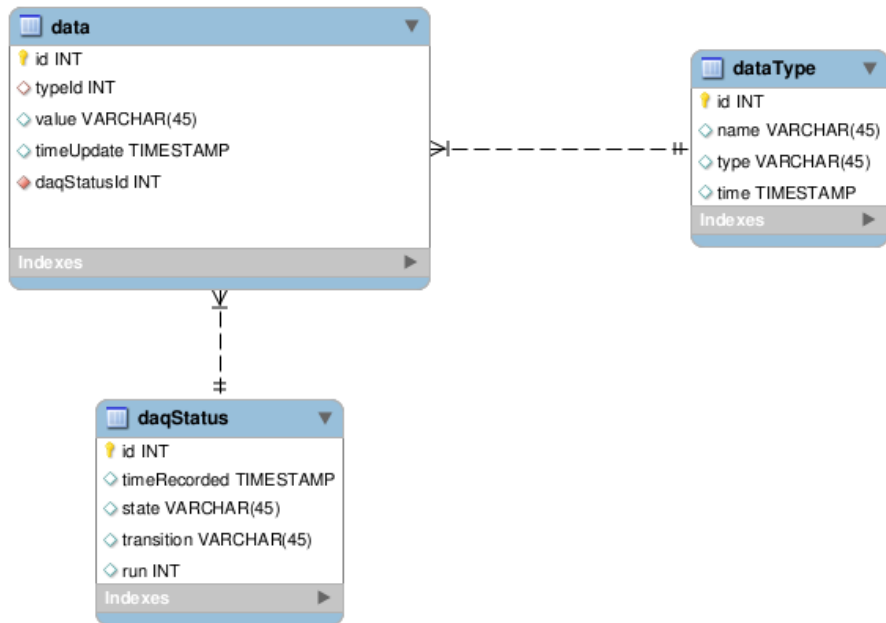
Jefferson Lab

March 27, 2013

The Problem

- During data taking various conditions of the detector, electronics, data acquisition system, hall, and counting room will be recorded.
- These data will be recorded in different ways for different reasons, but one of the main uses is for use in subsequent data analysis.
- Things like magnet power supply currents, temperatures, trigger rate, scalers will all need to be available for offline processing.
- These quantities will need to be identified by time of recording, run number, and perhaps event number.
- These quantities will not have to be recorded on an event-by-event basis; most will only be recorded once per run.

Proposed schema



Features

- Different processes can update quantities for which they are responsible independently.
 - ▶ Special events, e. g., scalers
 - ▶ Alarm systems
 - ▶ Log book
- Central server alone needs to know DAQ status (single number, the ID)
- New quantities can be added without re-writing the schema
- Frequency of update completely determined by the data producing client

Technology Comments

- MySQL ubiquitous, can be exported, transformed
- CLAS experience: base database was INGRES, almost no offline client had access
 - ▶ created copy in MySQL
 - ▶ imported data directly to Map on parallel path
 - ▶ Map conversion to CalDB: online constants went along for ride
- EPICS archiver: offline client access?

Why not index on run number alone?

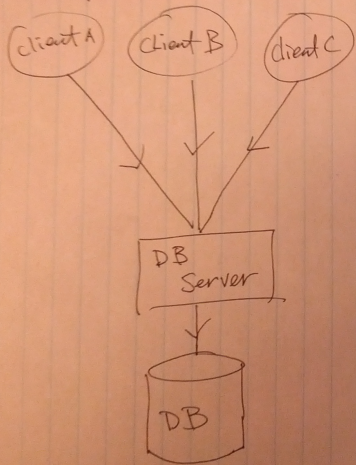
- Not all events are driven by the data acquisition.
- Concept of run number not well defined between run, during shutdowns or hall access
- Data acquisition system may not be running at all: global condition = unknown

Why a global conditions table?

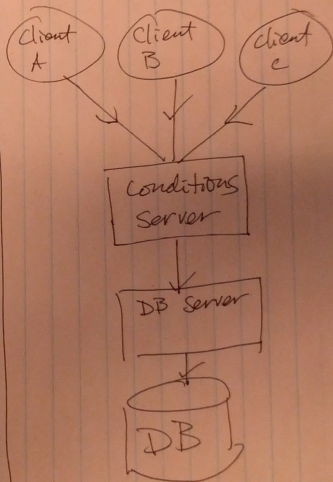
- Question: how tightly to tie data to data acquisition (run number)
- time matching problematic
 - ▶ synchronization, least count difference significant
 - ▶ transactions, need a structure for how to bundle data
- relational constraints in SQL ($>$, $<$): cannot make an index that helps (not an issue with time series)
- build in well known grouping
- CLAS experience: only CODA controlled processes wrote to the online database
- here: anyone can write, server to insert global condition key
- disadvantage of global conditions: intermediate server needed

Architecture Without and With Global Conditions

no global conditions table



with global condition table.



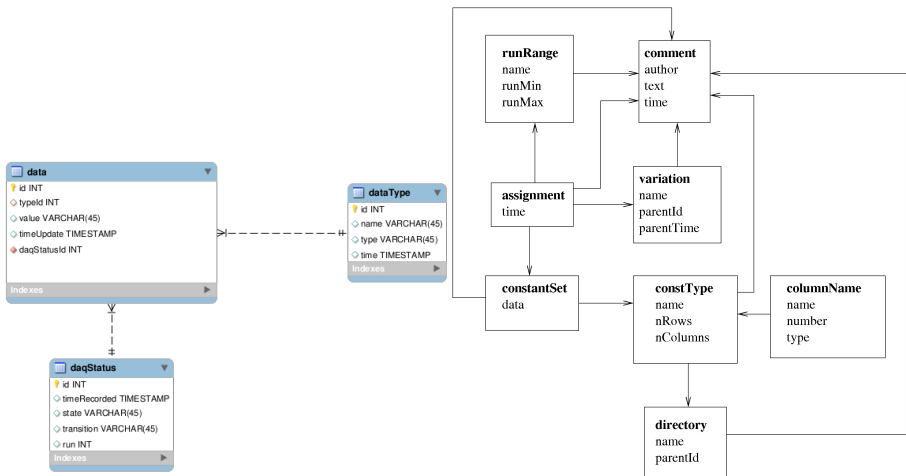
Why is all data stored in a string?

- Disadvantage: will need parsing
- Advantage: do not need a vast universe of data tables to have a vast universe of data types
- Creation of new data type trivial

Why not use calibration database?

- calibration database:
 - ▶ It is 2017. what were the 176 time offset constants we used last year (2016) for analysing the tof in data taken two years ago (2015), specifically from jan 3 (run 10000) through jan 10 (run 10080)?
- conditions database:
 - ▶ it is 2017. what was the solenoid current at the beginning of run 10000 (jan 3, 2015)?
- “Glue” information same size for one number as for arrays of thousands
- rate of conditions much higher than rate of new sets of calibration constants
- conditions completely dominate the glue tables

Conditions vs. Calibration Schema



Web Interface

- Should support web display, CLAS experience, very useful for technicians.
- Plotting one variable vs. another