COOL

Conditions Database for the LHC Experiments

Development and Deployment Status

<u>Andrea Valassi (CERN IT-DM)</u>

M. Clemencic (CERN - LHCb)

S. A .Schmidt, M. Wache (Mainz - ATLAS)

R. Basset, G. Pucciani (CERN IT-DM)

IEEE-NSS 2008, 23rd October 2008

CERN IT Department

Outline

- **Introduction**
- **Development activities**
  - Maintenance and code consolidation
  - Functionality enhancements
  - Performance tests and optimization
- **Deployment-oriented activities**
  - Scalability tests with simulated data
  - Support of actual deployment with real data
- **Conclusions**

What is COOL

- **Software for LHC 'conditions data' access**
  - Time variation (validity) and versioning (tags)
  - Offline (calibration, alignment) and online (DCS)
- **Common project of Atlas, LHCb, CERN IT**
  - Atlas and LHCb store conditions data using COOL
  - Persistency Framework of LCG Application Area
- **Collaboration with other LCG AA projects**
  - CORAL for C++ access to SQL on relational DBs
  - ROOT/Reflex for Python bindings (PyCool)
- **Support for several relational backends**
  - Oracle, MySQL, SQLite, Frontier (all via CORAL)

COOL development overview

- **Mature functionality and code base**
  - First release in April05, latest (2.5.0) in June08
  - Test-driven development, automatic nightly tests for all supported relational database backends

- **Maintenance and code consolidation**
  - Internal refactoring of existing functionalities
  - New platforms (osx/Intel, gcc43, VS9, SLC5…)
  - New versions of external software
  - Fix bugs/issues identified in real-life deployment

- ***Not yet fully in maintenance mode***
  - *Functionality enhancements*
  - *Performance optimization*

Functionality enhancements
(work in progress)

- **Tagging enhancements**
  - "Partial tag locking" (prevent tag modifications)
- **Data retrieval enhancements**
  - Payload queries (fetch time for given calibration)
    - Default use case: fetch calibration at given validity time
- **Database connection enhancements**
  - User control over database transactions
  - DB session sharing between COOL sessions

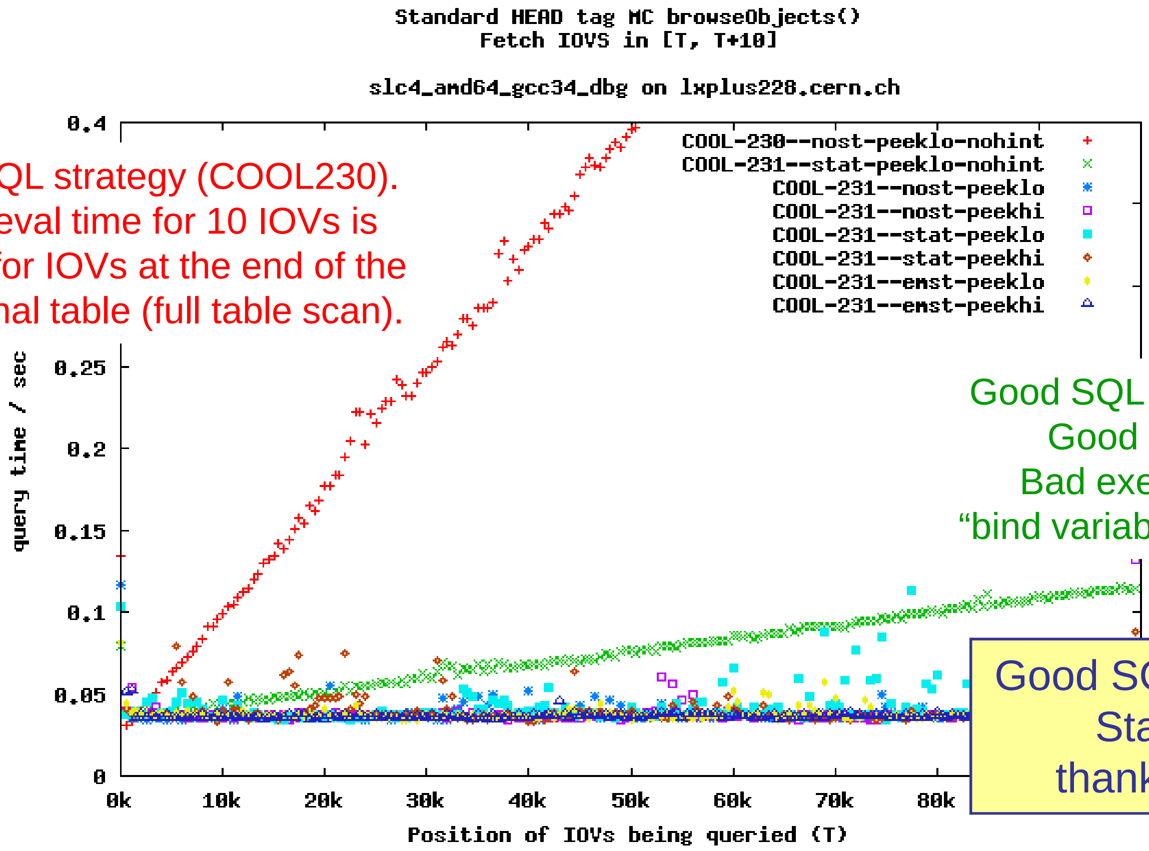
Performance optimization

- **Main focus: performance for Oracle DBs**
  - Master Tier0 database for both Atlas and LHCb
- **Proactive performance test on small tables**
  - Test main use cases for retrieval and insertion
  - Response times should not increase as tables grow larger (indexes instead of full table scans)
- **Oracle performance optimization strategy**
  - Basic SQL optimization (fix indexes and joins)
  - Use hints to stabilize execution plan for given SQL
    - Instability from unreliable statistics, bind variable peeking
    - Determine best hints from analysis of “10053 trace” files

Performance optimization example

CERN IT Department

- **Systematic tests of known causes of instabilities**
  - 6 plots: bind var. peeking (2) x fresh/stale/no statistics (3)
    - Such instabilities were actually observed in the Atlas 2007 tests
  - Stable performance after adding Oracle hints



Bad SQL strategy (COOL230).
Retrieval time for 10 IOVs is
larger for IOVs at the end of the
relational table (full table scan).

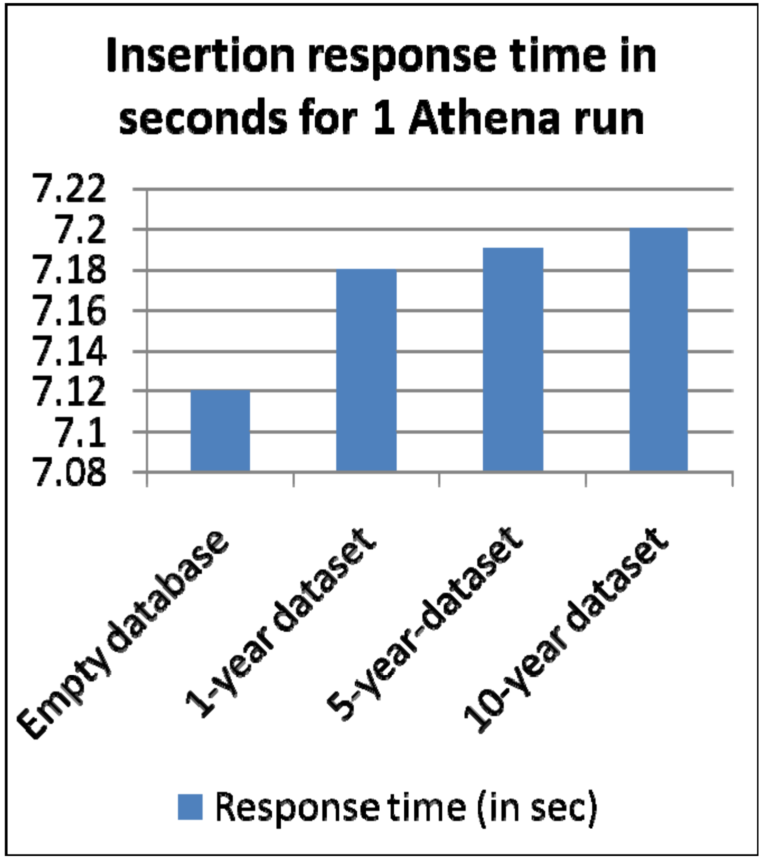
Good SQL strategy (COOL231).
Good Oracle statistics.
Bad execution plan due to
“bind variable peeking” (no hints).

Good SQL strategy (COOL231).
Stable execution plan
thanks to the use of hints.

Scalability tests

- **Proactive performance test on large tables**
  - Stable insertion and retrieval rates (>1k rows/s)
  - Simulate data sets for 10 year of LHC operation
- **Test case: Atlas**
  - Largest data set: DCS
    - 1.5 GB (2M IOVS) / day
    - From PVSS into COOL
- **Work in progress: Oracle partitioning**
  - For data management
    - Performance impact?

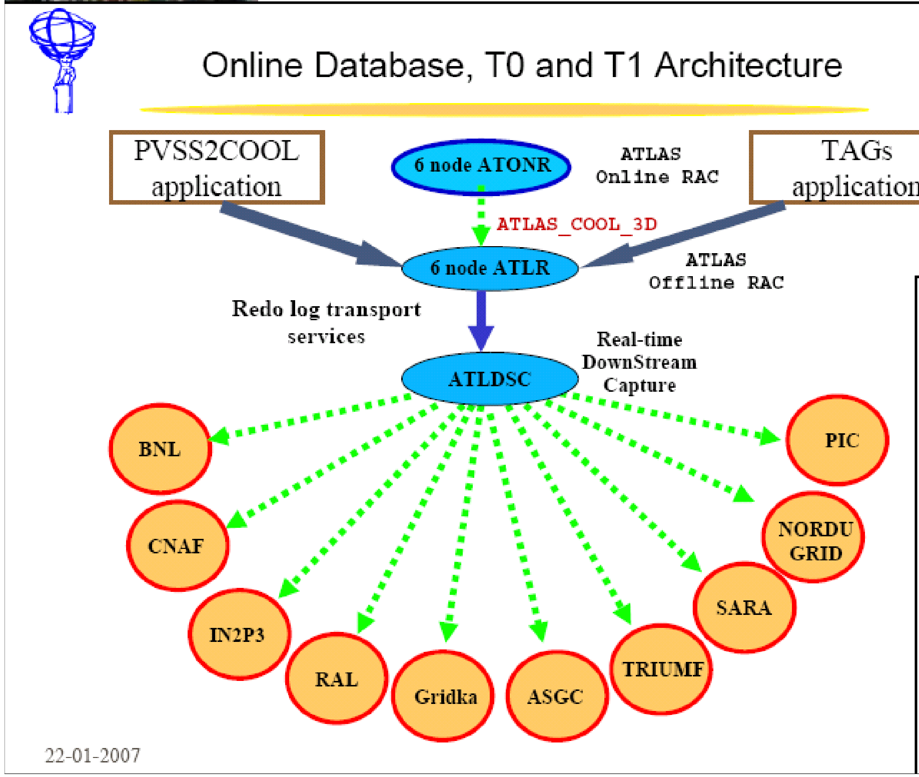
Romain Basset (DCS data)



COOL deployment overview

- **Similar Oracle setups in Atlas and LHCb**
  - Two separate servers at CERN (online, offline)
  - Distributed replicas at the experiment Tier1 sites
  - *Replication via the Oracle Streams technology*

***Atlas* (G. Dimitrov, F. Viegas)**



Online Database, T0 and T1 Architecture

PVSS2COOL application

6 node ATONR

ATLAS Online RAC

TAGs application

ATLAS_COOL_3D

6 node ATLR

ATLAS Offline RAC

Redo log transport services

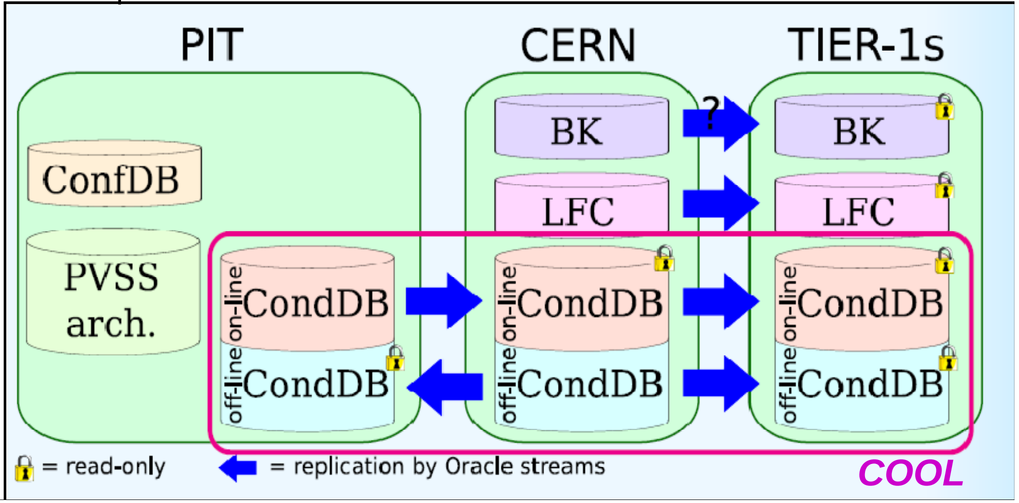
ATLDSC

Real-time DownStream Capture

BNL, CNAF, IN2P3, RAL, Gridka, ASGC, TRIUMF, SARA, NORDU GRID, PIC

22-01-2007

***LHCb* (M. Clemencic)**



PIT | CERN | TIER-1s

ConfDB

PVSS arch.

BK ? BK

LFC LFC

on-line CondDB → on-line CondDB → on-line CondDB

off-line CondDB ← off-line CondDB → off-line CondDB

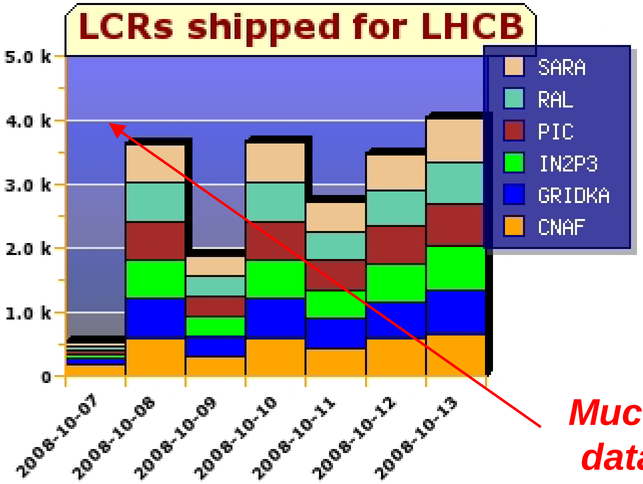
🔒 = read-only

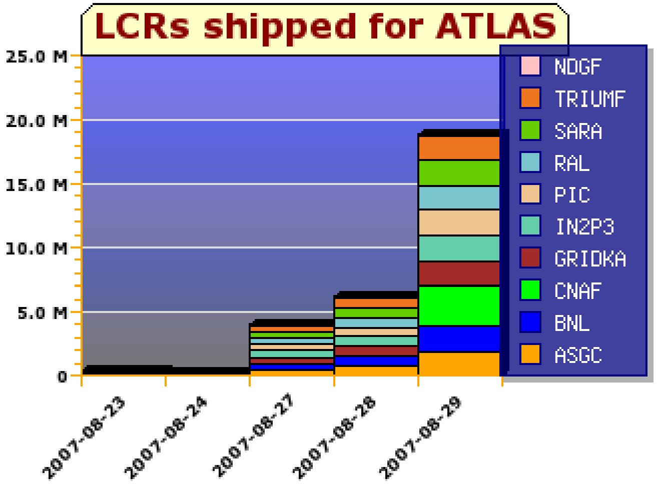
⬅ = replication by Oracle streams

COOL

Deployment status

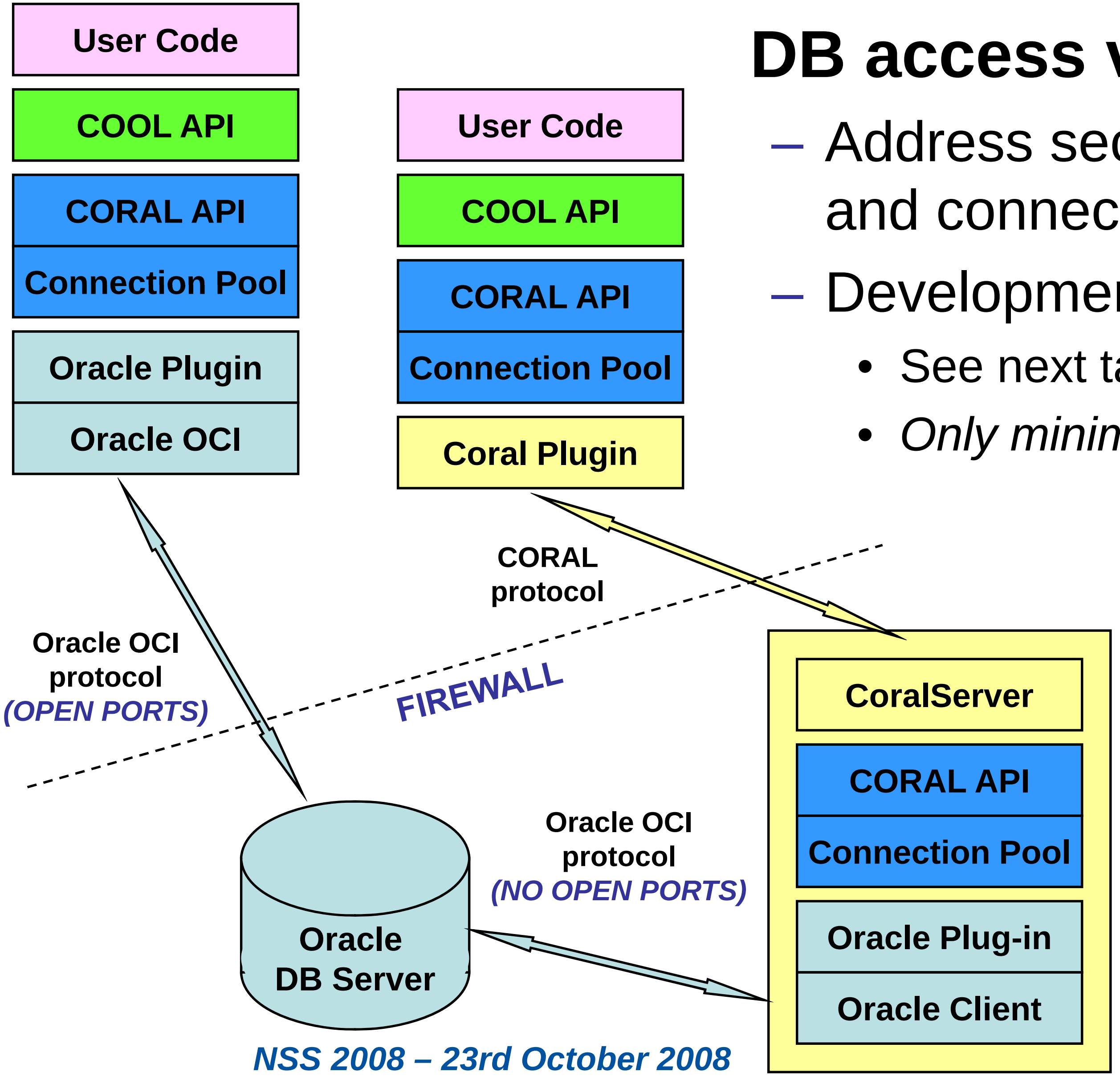
- **Setup is complete for both experiments**
  - T0 online/offline DBs, T1 sites (6 LHCb, 10 Atlas)
- **Distributed tests are very useful for COOL**
  - Several lessons from Atlas tests in 2007 already
    - Most T0 and T1 databases were up by Q4 2006 already
  - New issues identified and addressed in 2008
    - e.g. user-level read access during Streams write activity





Much larger data rates in ATLAS

New deployment model?



DB access via CORAL server

- Address secure authentication and connection multiplexing
- Development still in progress
  - See next talk by Zsolt Molnar
  - *Only minimal changes in COOL*

Conclusions

- **COOL: conditions DB for Atlas and LHCb**
  - A joint project with CERN IT and LCG AA
- **Development is mature but not finished**
  - Performance optimization is the highest priority
    - Proactive tests and support for real deployment issues
- **Distributed deployment setup is ready**
  - Waiting for more data from LHC!

Reserve slides

COOL collaborators

Core development team

- **Andrea Valassi (CERN IT-DM)**
  - 80% FTE (core development, project coordination, release mgmt)
- **Marco Clemencic (CERN LHCb)**
  - 20% FTE (core development, release mgmt)
- **Sven A. Schmidt (Mainz ATLAS)**
  - 20% FTE (core development)
- **Martin Wache (Mainz ATLAS)**
  - 80% FTE (core development)
- **Romain Basset (CERN IT-DM)**
  - 50% FTE (performance optimization) + 50% FTE (scalability tests)
- On average, around 2 FTE in total for development since 2004

Collaboration with users and other projects

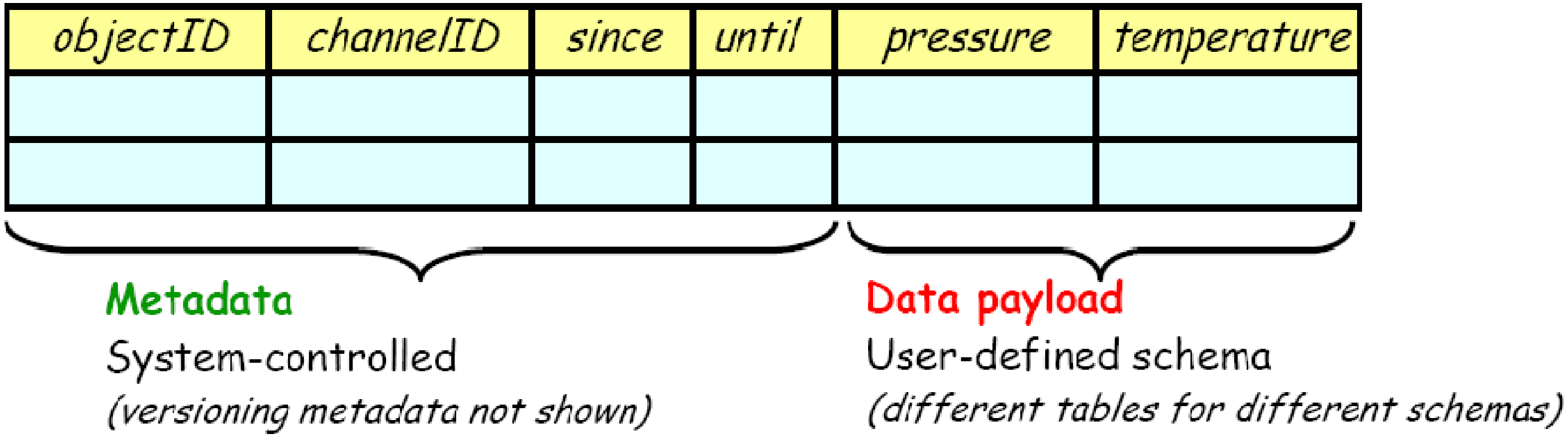
- *Richard Hawkings* and other *Atlas users and DBAs*
- *The CORAL, ROOT, SPI and 3D teams*

Former collaborators

- *G. Pucciani, D. Front, K. Dahl, U. Moosbrugger*

COOL data model

- **Modeling of conditions data objects**
  - System-managed common **“metadata”**
    - Data items: many tables, each with many channels
    - *Interval of validity* - “IOV” [since, until]
    - *Versioning information* - with handling of interval overlaps
  - User-defined schema for **“data payload”**
    - Support for fields of simple C++ types
- ***Main use case: event reconstruction***
  - *Lookup data payload valid at a given event time*



objectID	channelID	since	until	pressure	temperature

Metadata
System-controlled
(versioning metadata not shown)

Data payload
User-defined schema
(different tables for different schemas)