



Report of the Conditions DB WG

Paul Laycock, on behalf of
the HSF Conditions DB WG

Charge and Scope

- Scope: **Conditions data** includes any **ancillary data** associated with primary data taking such as **detector configuration, state or calibration** or the environment in which the detector is operating. In any non-trivial experiment, conditions data typically **reside outside the primary data store** for various reasons (size, complexity or availability) and are usually **accessed at the point of event-data processing or analysis** (including for Monte Carlo simulations). The ability of any experiment to produce correct and timely results depends on the **complete and efficient access** of the necessary conditions for each stage of data handling.
- Charge: This group should **evaluate all elements of the infrastructure required for the access and management of conditions data in HEP for the coming 5-10 years**. By looking at representative use cases, successful architectural patterns that can be applied to different experiments should be examined. **Where possible the group should study the possibility to develop common solutions and make recommendations.**

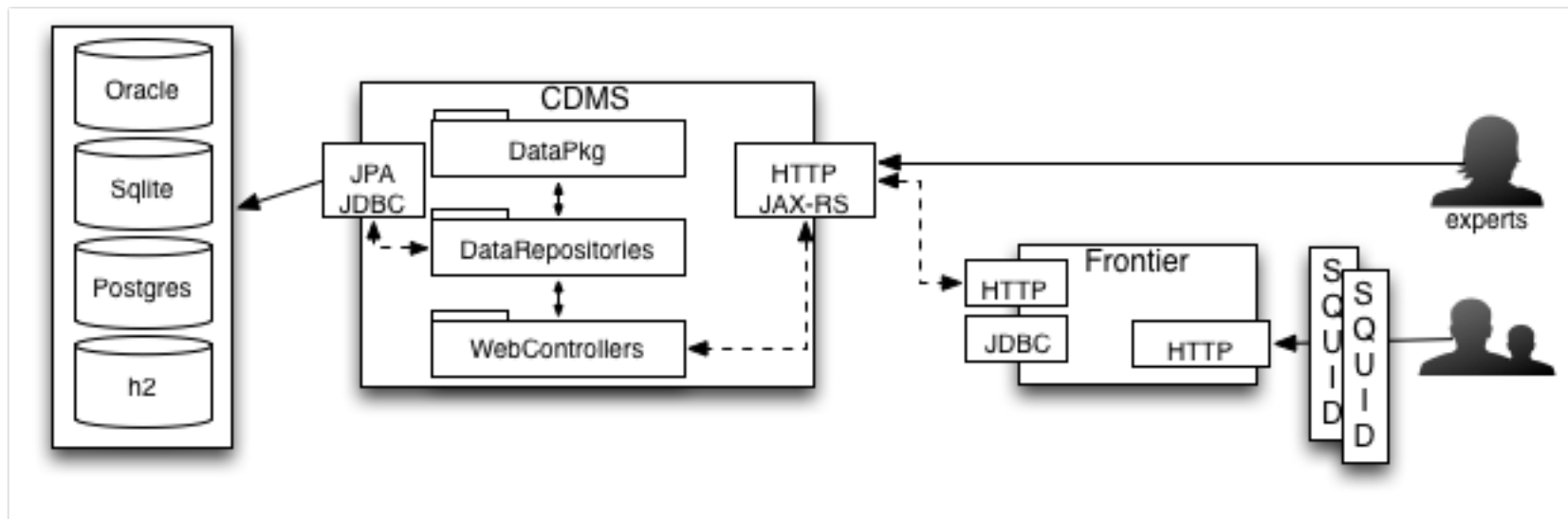
Key challenges and opportunities

- + We're ahead of the curve, thanks to active collaboration between CMS and ATLAS. How far could we go towards consensus on "a" conditions database management model?
- + New experiments (e.g. BelleII, NA62) are very keen to profit from LHC experience, use this as an opportunity to document useful experience
- A small community of individuals reflecting the level of effort the community puts into this.
- + A really nice bunch of people who know how to collaborate really well.

Key results - we converged

- High degree of separation between client and server, client-side is simple but takes care of (de)serialisation
- Conditions accessed via a REST interface
- Caching must be built in, good experience using web proxy technologies. Clients should be able to deal with multiple proxies and servers
- Relational DB for data model is preferred
- File-system approach also interesting and can be very useful (though not generally considered to be the master payload storage)

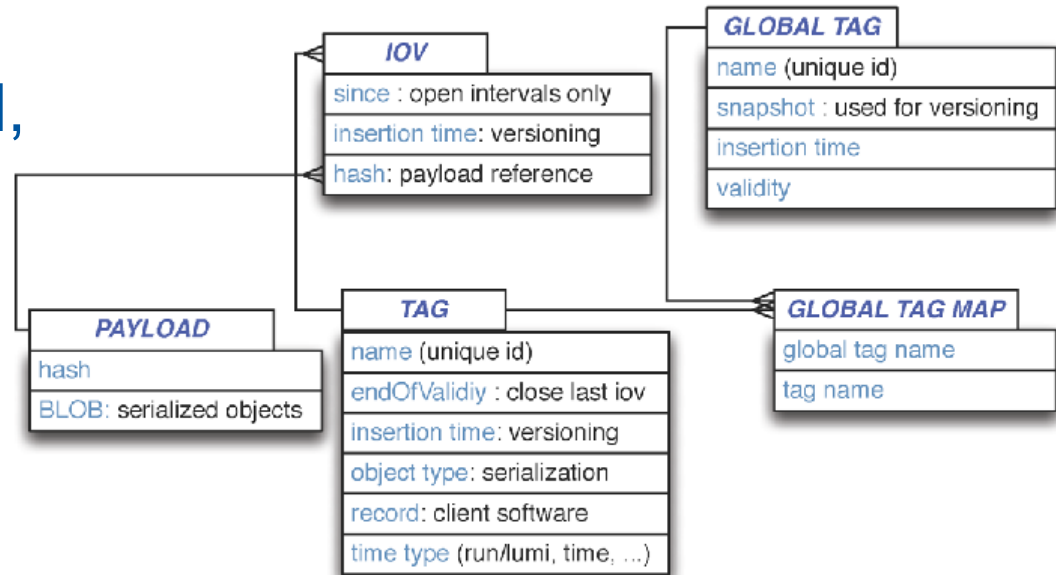
Conditions@REST



- REST communication between client and server
- Ensures simple interfaces
- Benefit from industry-standard solutions
- Loose coupling to allow component replacement

Conditions Metadata

- Data Model: relational DB
- Single tables for payload, tags, IOVs
- IOVs and payloads resolved independently
- Cache-friendly design



Supporting Diversity

- File-system approach is appealing, e.g. NA62 is migrating from files-in-the-release to a file-system approach:

/cvmfs/<experiment>/conditions/<gtag>/<system>/RunXYZ/payload

- Staged approach, first to a file-system, then to database (running experiment)
- git also discussed as an interesting approach used by LHCb
- Applicability of the solutions to end-user analysis

Practical considerations

- To a good approximation, nobody works on conditions
 - Cross-experiment collaboration good, need support from the wider community as this will likely (?) get worse
- The good news is that we have good solutions
- These ideas are being tested by CMS, ATLAS, LHCb, BelleII, NA62, now and in the coming years
- Aim for loose coupling of simpler systems, REST interfaces, web proxies, more adoption of industry-standard tools rather than home-grown solutions, where possible

Across WGs

- Analysis WG already showed interest in handling of analysis conditions
- The tools are ignorant of payload content by design, so “new” use cases like ML should (?) not effect anything
- Must ensure the assumptions about data access patterns are in line with other WGs

CWP Status

- Draft is in good shape and ~complete from the WG perspective
- Would appreciate feedback from other groups, particularly those who think they may have surprises regarding data access patterns, etc.
 - <https://docs.google.com/document/d/1yTcw51TOc68DCZQ4AO7o1hBdkPbN5l52ysJgJXNnJl8/edit>
- Would appreciate feedback rather than rebooting

Auxiliary material

- Main WG organiser: Paul Laycock
- Main work: everyone else

- 1/2 day April workshop:
 - <https://indico.cern.ch/event/633275/>
 - Attendees: Paul Laycock, Andrea Formica, Giacomo Govi, Dave Dykstra, Marko Bracko, Lynn Wood

- Thanks also to Hadrien Grasland, Andreas Pfeiffer and Marco Clemencic for useful discussions