

# Database Meeting 2016-03-02

## Date

02 Mar 2016

## Attendees

John Gates, Andy Salnikov, Fritz Mueller, Brian Van Klaveren, Vaikunth Thukral, Serge Monkewitz, Fabrice Jammes, Unknown User (kelsey), Jacek Becla, Unknown User (npease)

## Discussion items

### General Scheduling

- Now starting a short, 2-week sprint, which will be tacked on as an "afterburner" sprint to the W16 cycle. Focus on wrapping up any remaining work attached to W16 epics.
- Will be followed by an additional short, 2-week, intercalatory (Unknown User (kelsey)) sprint to finish out the month. This will be the first sprint of the X16 cycle, likely focused on documentation tasks.
- Following that will be two of the usual-length 1mo. sprints, carrying us through April and May to the end of X16 and the start of the F16.

### How are our KPI deliverables impacted?

- Rule of thumb (per Jacek, with caveats) is to push back all previously scheduled deliverables by 3mo. to account for the new X16 cycle; this would include for our purposes KPI deliverables.
- HOWEVER, overall project timelines are in flux right now somewhat due to the ongoing replan.
- Also, we need to separately stay on top of our FY16 deliverables to SLAC, still due October 1, which will not be adjusted. For the qserv group, this entails a demonstration of effective shared scans (see



DLP-649 - Jira project doesn't exist or you don't have permission to view it.

for details).

### PanStarrs

- Fabrice mentioned that IN2P3 has expressed concern about data transfer time for PanStarrs, and would like to start planning ahead for this (perhaps several dedicated machines on boundary of the cluster that could be trickle-loaded ahead of deployment).
- Vaikunth is also very eager to get PanStarrs data for xswap tests.
- Data is not yet available; probably arriving sometime this summer.

### BLOBs and partitioner

- The partitioner as currently implemented parallelizes work on line boundaries, defined by newlines. This is incompatible with BLOBs as dumped directly from MySQL, which may contain embedded newlines. We have a new integration test that fails because of this.
- In the short term, we will rework the test to use a BLOB without newlines. In the longer term, we must schedule work to address this properly for users.

### Unique query-id

- Vaikunth really needs this available to log messages.
- Should properly be a cluster-unique (unique across multiple czars) monotonic 64-bit unsigned integer, allocated by the qmeta service.
- Work scheduled for this: Andy will do the qmeta allocation part, John will plumb through qserv internals.

### Shared-scans

- John has completed his worker implementation, but needs MemManReal implementation changes from Andy Hanushevsky per previous design discussion before it will work as intended with memory management.
- In the meantime, John will make sure the IN2P3 cluster is ready-to-go with the high-volume test script.

### New Job Req

- Responses to our existing req have dwindled down. Jacek would like to try a new strategy: open a new req for somebody with primarily Python and web-service chops. We have a lot of this sort of work, and maybe this different sort of req will generate a new burst of candidates.
- Nate, Brian, Fritz to assist Jacek with job description.

### Education

- We've spent a lot of time trying to hire additional DB-expertise. Time to book up! There was broad interest from John, Fritz, Nate, Vaikunth in taking advantage of good DB classes available down the hill at Stanford.
- Probably we would audit and do it study-group / reading-group / seminar style.
- Brian recommends CS-{245, 345, 346, 347} as good, relevant courses.

## Serialization for geometry structures

- Serge has a need for this, in order to store various geometry structures in database columns. Was eliciting recommendations from the group.
- No slam-dunk winner. Serge had found CBOR (<http://cbor.io/impls.html>) which seems pretty reasonable.
- Might be more expedient to just roll own serialization methods in the style of CBOR directly onto geom objects without going full-general-purpose. If done this way, must be called "Sergialization".

## Action Items

- ✓ (Jacek Becla) Adjust Brian's allocation in resource loading spreadsheet to .1 discussion, .5 ADQL/VO
- ✓ (Jacek Becla) Hit the "Start Sprint" button
- ✓ (Fritz Mueller) Find out why can't hit "Start Sprint" button
- ✓ (Brian Van Klaveren, Unknown User (npease), Fritz Mueller) Provide job description draft to Jacek for new req.