

Issues to Discuss

LSST Database

This is a place for dumping issues that we identified, that we will eventually need to take care of. In no particular order.

Next several telecons

Related tickets

Latest tickets are listed first.

- [#300](#) - Do we need x,y,z components of (ra, decl) in Source table?
- [#299](#) - Quantities in Object Catalog to describe time behavior
- [#297](#) - precision required by MOPS
- [#296](#) - generating templates per airmass on the fly
- [#295](#) - updating object if corresponding DIASource was deleted
- [#294](#) - How many moments should we persist in Source table?
- [#293](#) - persisting photometric flags
- [#292](#) - keeping DIASources in sync with template images
- [#291](#) - tracking systematic errors in unreleased fits headers
- [#290](#) - granularity of image metadata
- [#289](#) - persisting extended parameters
- [#288](#) - cross-release queries
- [#287](#) - persisting full color images

All other topics

Infrastructure

- MySQL access control
- mysqld config
 - ◆ we should put it under versioning control: need to discuss exact procedure
 - ◆ need to understand potential issues with configuration required by MOPS
- We need to start planning for creating a testbed for large scale database tests
- deployment of lsst database servers

Classification

- Do we carry objects with unknown classification?
- Do we setup a threshold: if an object is classified with probability below that threshold then we don't assign classification at all
- Do we prune objects with classification probability below certain threshold? Who does the pruning? Deep Detection?
- Do we classify only high level types (star, galaxy etc), or we go much deeper (spiral galaxies... galaxies with bluer centers...)
- Should we persist probabilities of each match (Source --> Object)?

MOPS

- we need to better understand the "third pass": matching unmatched MOPSPreds against DIASources
- Will MOPS be integrated into the 3-phase approach? (preprocess before exposure taken, real time, postprocess):
 - ◆ will prediction be done during prepare phase
 - ◆ will MOPSPred be in-memory which implies preloading (prepare phase) and flushing to disk (post-process phase)
 - ◆ will the MOPS database/tables be driven by CIS (or equivalent mechanism)?
- need to understand implications for db if we only do position predictions at the Base Camp

Nightly processing

- review: - cross matching in database or in application?
- revisit if it is more cost-effective to use one fat server or several smaller
- how to implement hot fail-over: through mirroring or replication?
- revisit periodically if all columns from Object table needed for AP
- revisit if we should we keep VarObj catalog at the Base Camp
- should we push the computation which is not needed for alert generation into main archive? If we do that, should we transfer the updates from the Archive back to Base Camp?
- What follow up alerts after an initial alert are we required to send
- should we recalculate values for objects based on a new DIASource if changes are small and they don't justify sending an alert?

Association Pipeline

- What computation is needed in AP beyond cross-match?
 - ◆ recalculating averages?
 - ◆ updating scalegram coefficients?
 - ◆ Mapping columns: DIASource to Object - how?
 - ◆ Do we calculate magnitude? How? average? Weighted
 - ◆ Can muRA, muDecl and parallax be computed based on a single observation in one filter?
 - ◆ Calculating classification related probabilities in AP or DP?
- What to do if a DIASource is associated with an object which is classified differently than the DIASource? Should this be done in AP or alert generation?

Provenance

- how to combine information about software version, with the actual code used? Example: needed to reconstruct source classification related information.
- how to track in provenance the deletion of objects that were reclassified (was variable object, became moving object)
- As soon as we start throwing out DIASources (at Base Camp), provenance for Object starts to make no sense. We need to think about capturing what gets removed...

Unclassified

- allowing to telescope control system group on the mountain to query most recent data at the mountain top AND historical data at the Main Archive.
- integrating db queries with pixel data analysis. It has been mentioned by some people including Tony that we should integrate database with images. We have some ideas how this could be implemented,

- see PixelProcessing. We need to understand the exact requirements.
- Which pipeline will associate sources from different filters and build objects?
- Persisting DIASources
 - ◆ Will we keep all DIASources at main archive? If not, we need to revisit dbQuery019?
 - ◆ We decided we would keep up to 10 most recent DIASources per object at the Base Camp. When are we going to trim the oldest DIASources? Every day?
- Storage for calibrated images: current estimate assumes last 30 days? We should probably revisit it.
- How to persist Source Classification? Options:
 - ◆ completely generic, normalized, full flexibility, complex schema
 - ◆ simple, but adding new attribute will require adding new column to a very small table
- Will QA write to database?
- revisit disk IO at the Base Camp
- taking advantage of input from community about sent alerts - should we persist this input in database?
- all-in-database, computation-in-application or all-custom
- Implications of partitioning
- schema evolution
- regression tests
- postage-stamp cut-outs
- Precision. [?url](#)
- Database file sizes. [?url](#)
- SQL Query loggin. [?url](#)
- Issue of one object becoming two (and vice versa.), see: Object Split
- mysql + parallel file systems (lustre, ibrix, maybe gpfs)
- Default values: NULL vs out of range value like -0.99999x1010 ? See **?WFCAM Archive paper**