

Assigned Action Items

LSST Database

For most recent action items, see the trac tickets.

2008-01-23, DataAccWG

- document formal requirements related to querying catalogs at Main Archive by Telescope Control System group [Tim]
- suggest baseline for the above issue [Jacek, KT, Andy]
- document use cases that will drive use of provenance [Celeste]
- improve comments in schema [Jacek at al]

~~2007-11-30, DataAccWG~~

- ~~• document relevant queries from WFCAM Archive paper in trac [Jacek/KT]~~
- ~~• mark in trac queries that need to be rewritten [Jacek]~~
- ~~• understand how EA supports functions/procedures [Jacek]~~
- ~~• check with Tony about queries that involve processing images [Tim]~~

~~2007-11-16, DataAccWG~~

- ~~• document results from decimal compression tests [KT]~~
- ~~• schedule meeting DataAcc ApplicationWG [Jacek]~~
- ~~• come up with specific deliverables for DC3-scalable architecture [Jacek]~~
- ~~• research hadoop [Andy]~~

2007-07-11, DataAccWG

- Document issue of generating templates per airmass on the fly [Robert]

2007-07-03, DC2

- Research linking in c++ code into mysql [Jacek/KT]
- test various scenarios related to where we do computation for Association Pipeline [Serge]
 - ◆ simple cross match in db + other computation in application
 - ◇ application code run on client machine
 - ◇ application code run on server machine
 - ◆ all in application

2007-06-29, DataAccWG

- look into pipeline/harness [KT]
- document Alert email discussion [Lynne]

~~2007-06-27, DataAccWG~~~

- ~~• image file granularity? follow up with Robert/Tim? [Jacek]~~
- ~~• fixing systematic errors in fits header follow up with Robert/Tim? [Jacek]~~

- ~~storage for calibrated images: last 30 days or more? — follow up with Tim [Jacek]~~
- ~~check storage estimates (xls and documentation) and provide references [Kem, due July 1st]~~

~~2007-06-20, DataAccWG~~

- ~~schema change: flagging provisional objects [Jacek]~~
- ~~follow up with transient working group [Lynne, Serge]~~

2007-06-13, DataAccWG

- update Source Classification table [Kem]
- redo schema for Source Classification table, add probability [Jacek]

DBMSStorage [K-T]

- ~~Create realistic (but not fully real) test cases. Probably Exposure and DIASource for output; maybe DIASource or Object for input.~~
- ~~Import code into SVN.~~
- ~~Rewrite code to accommodate boost::serialization, if that is chosen.~~
- ~~Investigate hacking Coral to allow setting multiple fields in one call.~~

~~2007-05-25, DataAccWG~~

- ~~propose metadata definition [Tim, by June 1st]~~

~~2007-05-23, DataAccWG~~

- ~~find why we need historical DIASource for Association Pipeline [Jacek]~~
- ~~merge MOPS schema with Precursor Schema [Jacek, Franceseo]~~

2007-05-16, DataAccWG

- Check with Deborah about QA [Serge]
- run cross-match tests with Serge's DIASources using SQL Server [Maria]
- follow up on slow cross match query with MySQL [Jacek]

DC2 plan

- coral decision [due May 15, all 4]
- partitioning tests done [due May 31, Serge]
- cross matching tests done [due June 15, Serge]
- dbms prototype, incl bulk insert, bulk read and things we discover during Serge's tests [due June 30, K-T, Serge]
- catalog ingest service, including integrating with dbmsStorage [due July 20, K-T, Serge]
- db software installation on nesa hardware: Coral, MySQL [due June 30, Jacek]
- performance tests of nesa hardware [July 15 Serge]
- deployment of CIS, DBMSStorage [July 31, K-T]
- db and underlying file system tuning (nesa hardware) [Aug 30, all]
- integration of schema with application code and input image data [Aug 30, Jacek]

Schema related changes for DC2 and beyond [Jacek]

- merge Object_photoZ with Object for nightly
- add column to DIASource to keep track of data source (dc2 only)
- flag "dirty" in Object Catalog to mark rows that need to be written from in-memory table to disk
- match table
- need a flag in Object table to mark provisional objects (objects produced at the base camp, not by deep detection)
- schema for DIASourceIdTonight table
- Source table: add scalar value(s) to capture short term variability (discussed at DataAccWG, 2007/07/11)
- rework image metadata for DC2: "flat" model at CCD level, not Amp level
- object table: xyColor should be "NULL", not "NOT NULL"
 - ◆ muRA, muDec and parallax too?
- from Zeljko
 - ◆ DIASource
 - ◇ page 19: ra/dec accuracy 0.0001 OK! errors: can be in arcsec (or milliarcsec), not degs
 - ◇ page 21: it would be good to replace petro mags for DIA sources by some type of model mags (say, adaptive 2D gauss), petro mags make sense for (static) galaxies and, e.g., adaptive 2D gauss provide better SNR properties
 - ◆ MovingObject
 - ◇ page 23: mAmplitude would be better as mRms, since the light curves can have complex shapes, but we may want both; period is also an idealized concept - we need something more general (e.g. time scale obtained by weighting the power spectrum)
 - ◆ ~Object~
 - ◇ page 25: variability flag would be better as variability probability; ideally, we would provide this probability for each band
 - ◇ page 25: need error for parallax
 - ◇ page 26: amplitude and period - see comments under MovingObject. 2D position error: good!
 - ◇ page 27: mTimescale: good! but where does 274 years come from?
 - ◇ page 33: what is a scalegram?
 - ◇ page 33: in Constraints, riColor is listed twice in Columns
 - ◆ ObjectPhotoZ
 - ◇ (page 34) Only two moments for the posterior photo-z probability distribution are probably not enough. At least, we need some information on the non-gaussianity of that distribution. One possibility is to report the fraction of the integral of that distribution outside the $(\text{redshift} \pm 2 * \text{redshiftErr})$ interval. Another concern is that most likely we will have more than one method for estimating photo-z. At the very minimum we should plan for two.
 - ◆ Source
 - ◇ page 35: raErr4wes: 0.04 arcsec may not be accurate enough
 - ◇ page 37: flux has incorrect description; what kind of flux is this
 - ◇ page 37: it is not clear what snr means; for what quantity? We also need some type of model mags (say, adaptive 2D gauss)
 - ◇ page 37: we need error estimates for the moments; why do we go all the way to the fourth moment?
 - ◆ VarObject
 - ◇ page 39: how can a VarObject get access to all individual measurements; it would be good to add the first four moments of the light curve (for prescreening before accessing the full records)
 - ◆ Amp_PSF

◇ page 43: ~~why do we need sky values here?~~ I am not sure that this model allows for spatially varying psf - does it?

◆ Amp_PhotoCal

◇ page 44: photometric zeropoint may be varying on the scale of a single amplifier, we need at least a linear approximation and estimated rms variation around it

- ~~replace "xpos" and "ypos" with "row" and "col" respectively~~
- ~~ex, ey, ez in various tables should be DOUBLE, not FLOAT~~
- add tables for coadded exposures

~~2007-05-11, DataAccWG~~

- ~~document in trac issue about binding queries [K-T]~~
- ~~rebuild CORAL [K-T/Jacek]~~
- ~~document in UML: formatters / thread safety [K-T]~~
- ~~get coral approved [Jacek]~~
- ~~measure CORAL performance overhead [K-T]~~
- ~~check when Maria is back [Ani]~~

~~2007-05-09, DataAccWG telecon about metadata~~

~~For Jeff Bartels, check on May 23~~

- ~~use cases~~
- ~~which columns in current schema belong to metadata~~
- ~~sizes, in particular~~
 - ◆ ~~average number of metadata elements per Object, Source and DIASource~~
 - ◆ ~~distinct values vs values shared between many Objects/Sources?/DIASources~~
- ~~flexibility vs speed~~
- ~~interaction between pipelines and metadata stored in database, in particular during Alert Generation~~
- ~~schema design: what to store in metadata, what to store in main "core" tables~~

Unassigned

- ~~how to efficiently persist vectors in mysql~~
- ~~look into mysql archive engine~~

~~2007-05-08, discussion with Tim~~

- ~~add flag "data source" to Object catalog for DC2 only [Jacek]~~

~~2007-05-04, DataAccWG telecon~~

~~nightly processing~~

- ~~understand poor disk performance during loading chunks from disk to memory [Serge]~~
- ~~try doing cross match using extract from Object table (objId, ra, decl, x,y,z) [Serge]~~
- ~~check with application people about cross matching against var object [Jacek]~~
- ~~try doing cross match first against variable objects first [Serge]~~
- ~~document requirements / assumptions [Jacek]~~
- ~~investigate why dropping in memory table slow [Serge]~~

Assigned Action Items

- ~~test number of tables per db to find sweet spot [Serge]~~
- ~~check with Tim whether proper motion adjustment will be made only during deep detection, or during nightly processing too [Jacek]~~

mysql related

- ~~bad performance of composite index on (zoneId, ra) — follow up with mysql [Jacek]~~
- ~~MyISAM uses 1K disk blocks (8K in new generation myisam) — investigate why it is not tunable [Jacek]~~

Others

- sql stored procedures coding standards [Jeff]
- ~~check existing standards related to stored procedures [Ani]~~
- ~~investigate how easy is to move disks/RAM from lsst-dev03 to lsst-dev02 [Jacek]~~

~~2007-05-02, DataAccWG telecon~~

- ~~define baseline approach how we are going to do association in DC2 [due May 11, Serge, Ani, K-T, Jacek]~~
- ~~ping Ray about cleaning AltFormatter [Jacek]~~
- ~~legal issues — Stanford copyrights [Jacek, Jeff]~~
- ~~DBMSStorage [K-T]~~
 - ◆ ~~cleanup namespaces~~

Unassigned Action Items / Todo's

- schema related (after DC2)
 - ◆ add table for keeping multiple matches (at the moment we only keep the best match)