Large Synoptic Survey Telescope

Data Management Monthly Report

September 2017

High-level Summary

DM successfully completed the Joint Status Review during the week of Sept 4 in Tucson. DM status was also presented to the Aura Management Council for LSST the following week.

Following discussion at the LSST2017 Project & Community Workshop, we are investigating retaining all raw images on disk. In support of this, a working group to study lossy compression of image data has been convened with Robert Gruendl (NCSA) as chair, and we are exploring the possibility of sharing images with the EPO subsystem. Successful completion of these actions will address issues surrounding poor user experience when accessing images.

The DM Science group completed a draft report on requirements and the status of the implementation of DM systems needed for special programs was prepared, uncovering a number of potential gaps and areas where DM's designs and plans need to be refined. Addressing these issues will continue to be a major focus through November.

The Architecture team overhauled the DM build system, which is now completely based Python 3.6, with the last components which require Python 2 having been removed. Automatic Python standards checking has also been included in the development process.

The JupyterLab-based environment, which forms part of the Science Platform, was deployed on a test Kubernetes cluster at NCSA and integrated with the Data Facility authentication system ("CI logon").

The Qserv database system was upgraded to support disconnected/asynchronous queries, including query submission, monitoring, result retrieval, cancellation.

The Science Pipelines group completed initial benchmarks of the proposed alert distribution system, demonstrating its basic capability to scale to operational levels, and published technical note DMTN-061, describing the current state-of-the-art in terms of image differencing within the DM codebase. Further, new algorithm that more efficiently and effectively rejects artefacts when creating coadds has been made available in the Stack. This algorithm is still undergoing detailed validation at the time of writing.

The Summit Facility and Summit Computer Room will not be available until end of October and full occupancy not until the end of the year. Planning of the network ducts and power in the computer room and cable trays in the Telescope building is being finalized. At the

advice of the LSST Architect, in order to secure site occupancy from Besalco as soon as possible, these elements will be installed by another contractor after Besalco has granted full occupancy.

The bulk of the REUNA DWDM (Dense Wavelength Division Multiplexing) and AURA DWDM equipment was received. We will do First Optical Light demonstrations of >40Gb bandwidth from La Serena to NCSA in October, and from Cerro Pachon to NCSA in December.

Risk Management

The DM Risk Register was reviewed in the monthly process. No new risks were added and no significant changes to existing risk exposure were made.

Milestone summary

Achieved Milestones

ID	Description	Due Date
	International Communications and Base Site	
DMTC-7400-2330	LSST and AURA summit to base DWDM equipment	2017-09-29
DMTC-8100-2090	100G ring operating at 97.0% availability	2017-09-29
DMTC-8100-2290	FIU-CIARA Management Contract Delivery #3	2017-09-29
	System Management	
DLP-558	LSST Software Release 7.1 Complete (Stack Release 14)	2017-08-31
	LSST Data Facility	
DLP-412	Monitor development cluster	2016-11-30
DLP-847	Meet camera DAQ/OCS milestones for FY16	2016-11-30

Delayed Milestones

After the replan we still carry some DLP milestones which will be completed in FY17. It will take a few months to get the LCRs baselined and the milestones cleaned up.

ID	Description	Due Date

	DM System Architecture	
DLP-538, DLP-539, DLP-541, DLP-544	Assorted ICDs updated to Phase 3 *Work is in progress to remove all TBDs and TBRs from these ICDs and achieve "substantial completion" by the end of the calendar year.*	2017-05-31
	Science Data Archive & Application Services	
DLP-802	AP/L1 Database Design Functional prototype has been delivered to AP team. Further scale/performance experiments pending deployment of Oracle consolidated db systems at NCSA.	2016-11-30
DLP-472	Qserv Data Distribution Delayed due to personnel reassignments. Gaponenko assigned to this full-time; currently on track for completion in F17.	2017-05-31
	LSST Data Facility	
DLP-366	Mid-scale processing of eligible camera data *This was postponed due to SuperTask framework status. Upcoming milestone LDM-503-2 will demonstrate mid-scale processing of eligible camera data using an alternate production-capable framework. Expected completion date is November 30 2017.*	2016-11-30
	International Communications & Base Site	
DMTC-7400-2420	REUNA La Serena - Santiago DWDM equipment *Delayed due to vendor delay. Expected completion November 1, 2017*	2017-07-31
DMTC-7400-2090	Report on functional fiber connections, including AURA equipment *Delayed due to REUNA DWDM equipment delay. Expected completion November 15, 2017*	2017-08-18
	System Integration & Test	
DLP-579	Usability and developer efficiency *This is currently being held up by Orchestration and Supertask framework. Several other blocking issues have been resolved.*	2016-08-31

Detailed Project Progress

1.02C.01: System Management

Current accomplishments

DM Project Manager and Deputy Project Manager successfully guided DM through the Joint Status Review Sept 6-8th 2017 in Tucson. We mainly looked at the responses to the July review recommendations.

DM status was presented to the Aura Management Council for LSST (AMCL) in Tucson. All subsystems presented status post JSR.

A workshop was held in Tucson to discuss how NOAO, STScI and LSST might go forward on on providing science exploitation to end users in a more uniform way. This was quite valuable and will form the basis of ongoing work. There is clear scope for this to be relevant both to the development of the LSST Science Platform and also to our adoption of widely used tools and libraries such as Astropy.

Planned activities

The DM Project Manager will attend the ADASS conference in Santiago, Chile, and will prepare for and attend the Directors Review of the Operations Proposal.

Interviews for both the Release Manager and Subsystem Scientist posts will also take place during October.

1.02C.02.01: Data Management Science

Current accomplishments

Graham finalized revisions and re-submitted the photo-z paper (PUB-44) and continued to investigate the impact of IR photometry on LSST photo-z. Major updates were made to the Special Programs document[1] to incorporate information from the LSST 2017 Project & Community Workshop. Graham delivered a remote talk at the LSST:UK meeting on DM data pipelines and products. She also began developing the “Science > Q&A” community.lsst.org framework for DM-community communication.

Slater worked on the remaining details of the paper on star-galaxy separation, addressing comments and suggestions from Ivezic. He also investigated how to benchmark database processing with Spark in order to understand the performance improvement from adding

[1] https://github.com/lsst-dmsst/DM_SP_study/blob/master/ms.pdf

synchronization (shared scans) to Spark task scheduling, using elastic AWS resources. A proof-of-concept implementation of shared scans has been demonstrated, and will be written into a report following the submission of the star-galaxy manuscript.

Suberlak continued the research with Slater in support of making decisions on expected performance and processing constraints in crowded fields. He processed DECam images for areas representative of source density levels with the LSST Stack, and verified relative photometric and astrometric accuracy with color-magnitude diagrams.

Juric implemented the decisions of the LSST 2017 working meeting on retention of processed images (RFC-325). A design change to retain all raw images on disk will be proposed (DM-11880; Petravick), a working group to study lossy compression has been convened (Gruendl; DM-11819, RFC-339, and https://confluence.lsstcorp.org/display/DM/Lossy+Compression+WG), and possibilities to share images with EPO will be explored (DM-11881; Dubois-Felsmann). Convened a meeting with the Directorate to organize the update of the LSST Data Access White Paper (Document-5373); the Deputy Director agreed to convene and chair a working group to define the committees and process necessary to resolve outstanding data rights issues. Developed proposal for Operations-era data products nomenclature, to supersede the Level 1, 2, 3 nomenclature currently in use (proposal committed as Document-27013). Juric also presented the LSST Science Platform to the LSST Science Advisory Committee (Presentation-451), and provided answers to written follow-up questions (Document-27012).

Planned activities

- Advance and possibly finalize the technical note on Special Programs, spawn change requests and shepherd their progress. (Graham)
- Roll out the “Points of Contact” and “Data Q&A” DM-community communications frameworks, per Presentation-546. (Juric)
- Work out the framework for data submission and cooperation with the Minor Planet Center in Operations. (Juric)
- Contribute to working groups for broker policy document, DM communications, and data rights access. (Graham)
- Make the star-galaxy separation paper to be ready for submission. (Slater)
- Continue crowded field performance investigation by processing StarFast-simulated images. (Suberlak)
- Juric will step down as DM Subsystem Scientist and transition to DM SST Coordinator role.

1.02C.02.02: DM System Architecture

Current accomplishments

The DM Architecture Team supported the Joint Status Review, outlined the documentation for an upcoming Header Service design review, and prepared for an Early Integration exercise and Interfaces working session. In the area of software, the team removed the vestiges of Python 2 in the Python 3 build (which is now standardized on Python 3.6), integrated Python standards checking, and prepared for test coverage reporting. The first drafts of the Data Access Use Cases and Butler Requirements were prepared with the Butler Working Group.

Planned activities

The DM Architecture Team will assist with Early Integration Exercise #5a, participate in the Interfaces working session, help define and delineate the "Special Programs" capabilities of the DM system, continue to facilitate experiments with using SQL Server for certain database uses, continue leading the Butler Working Group, and attend ADASS and IVOA meetings at which a poster on the use of Python 3 within LSST will be presented.

1.02C.03: Alert Production

Current accomplishments

02C.03.00 – Alert Production Management Engineering and Integration

- Travel & meetings:
    - Bellm travelled to Pasadena during the week of 25 September. He will be working on commissioning the ZTF system through October. This work is not funded by LSST, but the experience gained with ZTF will directly feed back into the LSST construction effort.
    - Parejko (virtually) attended the Fall 2017 Astropy Coordination Meeting. Notes from this meeting are available on the LSST Community Forum. [DM-11968, DM-11969]
- Construction of an "end-to-end" Alert Production test system [DM-9676, DM-10770, DM-10771, DM-10773, DM-10775]:
    - The prototype end-to-end AP pipeline is now scripted to run within the alert production verification framework. [DM-11390]

02C.03.01 – Single Frame Processing Pipelines

- Replacement of XYTransform [DM-10086]:

    - The camera geometry system has been upgraded to use the new, AST-backed, transformation system rather than the older, deprecated XYTransform. [DM-5922]

02C.03.02 – Association Pipelines

- Prototype source association system [DM-10768]:
    - It is now possible to save DIASources (detections on difference images) and DIAObjects (collections of DIASources corresponding to particular astronomical objects) to persistent storage. [DM-11040]
    - A simple matching algorithm for assembling DIASources to create DIAObjects has been completed and is now available within the DM Stack. We expect this work to undergo further refinement through construction, but this provides a firm basis from which to work on the end-to-end alert production test system. This algorithm will be encapsulated in the AssociationTask for easy reuse. [DM-11041]

02C.03.03 – Alert Generation Pipelines

- Testing prototype alert distribution system [DM-7355]:
    - The initial set of benchmarks of the prototype alert distribution system have been completed. These tests are described, and the results tabulated, in DMTN-028. We record network throughput as well as memory and CPU loads on both the alert production and consumer systems and total latency. We also check that all events are transmitted successfully, with none being lost in transit.
    - Although these benchmarks are early, and carried out under artificial conditions, our results show the promise of the Apache Kafka-based distribution system being developed for LSST. [DM-10407]

02C.03.04 – Image Differencing Pipeline

- Include decorrelated Zackay et al. (2016) and Alard & Lupton (1998) image differencing algorithms in the LSST stack [DM-9989]:
    - DMTN-061, a technical note describing the current state of image subtraction in the DM Stack has been published. This provides a convenient summary of the achievements of David Reiss, who left the project at the of September. This will form a convenient reference document for continued development of the image differencing codes after David's departure. [DM-11813]
    - A number of minor issues and improvements were made to the image differencing codes during the course of writing DMTN-061. [DM-11951]
- Incorporate Differential Chromatic Refraction (DCR) in the DM Stack [DM-9613]:
    - A suite of visualization tools to aid in understanding the behaviour of DCR correction algorithms and their possible failure modes have been produced. [DM-6904]

    - DCR template generation codes were updated to support Python version 3 (which will be adopted as the LSST baseline in early 2018) and to interoperate with the latest versions of the Data Butler I/O abstraction system. [DM-11459, DM-11966]
    - Work on the DCR correct algorithms is being recorded in technical note DMTN-037.

02C.03.05 – Application Framework for Exposures

- Emergent work [DM-10068, DM-10774, DM-11798]:
    - A number of minor bugs and issues were addressed across the codebase. [DM-72688, DM-9817, DM-11516, DM-11650, DM-11971, DM-11957, DM-12061, DM-12085, DM-12102]

02C.03.06 – Moving Objects Pipeline

- No work was undertaken in this WBS during this month.

02C.03.07 – Photometric Calibration Pipeline

- A description of Jointcal, the simultaneous astrometric and photometric fitter covering multiple exposures, is now available through technical note DMTN-036. [DM-11790]
- Work is now ongoing to perform a comparison between the results of Jointcal and meas_mosaic, the older, Hyper Suprime-Cam package which it is slated to replace (but which is currently required for testing). We are optimistic that meas_mosaic can be dropped from the LSST codebase later in this cycle.

02C.03.08 – Astrometric Calibration Pipeline

- No work was undertaken in this WBS during this month.

Planned activities

02C.03.00 – Alert Production Management Engineering and Integration

- Work will continue on the development of the end-to-end testing system.

02C.03.01 – Single Frame Processing Pipelines

- No work is planned in this WBS during this month.

02C.03.02 – Association Pipelines

- The association algorithm to be run as part of the (initial) end-to-end system will be finalized and implemented.

02C.03.03 – Alert Generation Pipelines

- The early alert system benchmarks reported above will be scaled up to measure performance under more demanding and realistic conditions.

02C.03.04 – Image Differencing Pipeline

- Work will continue on developing and documenting the DCR correction capabilities of the DM Stack. This will include converting existing prototype code to the standard Task convention used within DM.

02C.03.05 – Application Framework for Exposures

- We will continue to service technical dept and emergent requirements.

02C.03.06 – Moving Objects Pipeline

- No work is planned in this WBS during this month.

02C.03.07 – Photometric Calibration Pipeline

- Comparison of Jointcal with meas_mosaic will continue.

02C.03.08 – Astrometric Calibration Pipeline

- No work is planned in this WBS during this month.

Recruitment update

- David Reiss left the AP team at the end of September. A search is underway for his successor. David made very substantial contributions to our image differencing codes: we thank him for his contributions — he will be sorely missed.
- John Swinbank will move to the University of Washington (from Princeton University) in October 2017. He will serve as T/CAM for both the AP and DRP teams (02C.03 and 02C.04) and as Deputy Data Management Project Manager.

1.02C.04: Data Release Production

Current accomplishments

02C.04.00 – Data Release Production Management Engineering and Integration

- Travel & meetings:
    - Swinbank participated in the LSST Joint Status Review 2017 in Tucson, AZ during the week 4 September. [DM-11888]
- Management & planning [DM-10816]:
    - All milestones for the F17 cycle were audited for status and completion date

to ensure that the (old) DM plan will transition smoothly to the replan as the latter is baselined. [DM-11535, DM-11536]
    - The budget and statement of work for Princeton, the lead institute for DRP, were agreed between the university and AURA. [DM-11887]
- Forward Global Calibration Method (FGCM; Burke et al., 2018) [DM-10584]:
    - The command-line task that will provide an entry-point for the FGCM code within the context of the DM Stack is now operational. However, work continues on optimizing it and better integrating it with regular Stack conventions. This work is ongoing into October. [DM-11313]
- Emergent work and pipeline support [DM-10382]:
    - A number of minor issues and bugs have been addressed across the codebase. [DM-9579, DM-10183, DM-10907, DM-11519, DM-11521]
- Pipeline QA [DM-10571]:
    - The QA framework was re-engineered to be substantially faster when accessing and assigning columns in tabular data. [DM-11873]
    - New QA plots, showing only objects used for photometric and/or astrometric calibration, were added to the QA suite, together with new options to use more appropriate units when making some sorts of plot. [DM-11322, DM-11544]
    - Work continues on an upgraded, more flexible, visualization and QA suite based on the Bokeh platform. In particular, this month we have focused on using the HoloViews platform to make analysis of extremely large volumes of data possible.

02C.04.01 – Application Framework for Catalogs

- Middleware and framework development [DM-10586]:
    - The DRP team has continued to play an active role in the Data Butler Working Group. This has included developing use cases and requirements for the redesigned Data Butler (which provides I/O abstraction for the DM Stack). [DM-11723, DM-11732]
    - In addition, the DRP group has spearheaded the creation of a initial design sketch and prototype for the form that the redesigned Butler might take. [DM-11751]
- Emergent work and reduction of technical debt [DM-10383]:
    - Significant work this month included a substantial performance improvement to operations involving SpanSets (which correspond to pixel regions on images). [DM-11894]
    - A number of other minor issues were addressed and documentation improvements made throughout the codebase. [DM-11433, DM-11744, DM-11882, DM-11894, DM-11911]

02C.04.02 – Calibration Products Pipeline

- Auxiliary telescope development [DM-10581]:
    - Following staff leave, work on this activity resumed during September. These efforts focused on testing of the obs_ctio0m9 camera support package, which is being used to reduce spectroscopic data taken on the 0.9m telescope at CTIO which serves as a prototype for the LSST Auxiliary Telescope. [DM-10964]
- Processing camera test stand data [DM-10897]:
    - The obs_comCam camera support package, which is used to access test stand (and ultimately ComCam) data, has been updated to support Python 3 and included in the DM continuous integration system. This work was completed but not fully merged at the end of September; it will be marked as done in October. [RFC-391, DM-11926, DM-11653, DM-12070, DM-12098]
    - Work is now continuing on using the DM Stack to derive calibration products from test stand data. [DM-11347]

02C.04.03 – PSF Estimation

- Wavefront measurement and PSF reconstruction [DM-10355]:
    - The “donut” model being used to characterize out-of-focus PSF images has been validated by comparison to theoretical predictions made using the Zemax optical system design software. [DM-11043]
    - We infer the pupil obscuration from flat-field images taken through a pinhole filter. This is an approximation, since the pupil should be defined with respect to a plane wave rather than a flat field. We have used ray tracing software to demonstrate that the impact of this approximation is not significant in the context of the current work. [DM-11672]
    - ProcessDonutTask, the driver for the donut fitting process, is now available as command-line task. This provides a more convenient entry-point and user interface to the code, especially when experimenting with multiple configuration options. [DM-11701]

02C.04.04 – Image Coaddition Pipeline

- Warped image comparison [DM-8290]:
    - CompareWarpAssembleCoaddTask, the new robust coaddition algorithm, is now available in the standard DM stack distribution. This provides an effective way to remove artefacts (moving objects, cosmic rays, etc) from coadds while they are being constructed. [DM-11432]
    - Work will continue on testing & characterizing this work as well as providing full end-user documentation.

02C.04.05 – Object Detection and Deblending

- Deblender development [DM-10353]:
    - The new NMF ("non-negative matrix factorization") deblender now includes the capability of fitting multiple components to a single peak. This enables us to better handle, for example, spiral galaxies with resolved bulge+disk structure, or extended galaxies with color gradients. [DM-10614]
    - Work continues on integrating the NMF deblender fully into the LSST stack to enable large-scale testing of the algorithm under realistic conditions.

02C.04.06 – Object Characterization Pipeline

- Experiments in shear measurement on coadds [DM-10579]:
    - Work has continued in assessing the Metacalibration and BFD (Bayesian Fourier Domain) approaches to shear measurement. This work was delayed by problems encountered when adding more complexity to simulated objects. These issues, which were due to incorrectly seeding the random number generator, have now been overcome, and the comparison is now proceeding as planned. [DM-11311, DM-11907]
- Improved galaxy flux measurement algorithms [DM-10580]:
    - Work has continued to focus on integrating the Synpipe synthetic object generation system with the DM Stack. A number of minor issues with the Synpipe codebase, combined with parental leave for the member of staff assigned, have delayed this activity which is now behind schedule. We expect to complete this work during October. [DM-10849, DM-11318, DM-11898]

Planned activities

02C.04.00 – Data Release Production Management Engineering and Integration

- Swinbank, T/CAM for DRP, will switch employer from Princeton to the University of Washington in mid-October. He will continue to act as DRP T/CAM, in addition to serving as T/CAM for Alert Production and as Deputy Data Management Project Manager.
- Preparation for the early-November DMLT face-to-face meeting.
- Preparation for the end-of-F17 LDM-level milestones.
- Begin planning for the S18 cycle.

02C.04.01 – Application Framework for Catalogs

- Finish the merger of the new afw::math::Statistics code, following extensive review.
- Continue working with the Data Butler Working Group to refine Butler requirements and develop the Butler design sketch.

02C.04.02 – Calibration Products Pipeline

- Demonstrate successful reduction of CTIO 0.9m data based on the DM Stack.
- Demonstrate successful derivation of detector gains based on test stand measurements processed with the DM Stack.

02C.04.03 – PSF Estimation

- Complete a technical note ("DMTN") describing the work which has been performed to date on PSF estimation.

02C.04.04 – Image Coaddition Pipeline

- Complete validation, test and characterization of the new robust coaddition system.

02C.04.05 – Object Detection and Deblending

- Work towards at-scale testing of the NMF deblender.

02C.04.06 – Object Characterization Pipeline

- Complete integration of Synpipe with the DM Stack.

Recruitment update

- Sophie Reed joined the group as a postdoctoral scientist in mid-September. Sophie recently completed a PhD on the discovery of high-redshift quasars in optical and IR surveys at the University of Cambridge.
- During this month applications for the post of project manager for the DM group in Princeton were assessed. Interviews have been scheduled for October.
- Perry Gee retired from the DRP team at the end of this month. Thanks are due to Perry for many years of important contributions to LSST.

1.02C.05: Science User Interface & Tools

Current accomplishments

02C.05.00 – Management, Leadership & Other Costs

- Conducted employee annual review.
- Prepared FY18 budget for SUIT group.
- Gregory attended the EPO review as review board member for three days.

02C.05.01 – Basic Archive Access Tools

- Bug fixes and code refactoring. (DM-11609, DM-11602)
- Provided the input for LAAIM system for other DM members to comment on. (DM-3637)
- Firefly document update. (DM-11045)

02C.05.02 – Data Analysis and Visualization Tools

- New functions
    - Support image drawing layer sub-grouping. (DM-5794)
    - Create a framework to support multiple image sources, enabling future grouping and filtering of the image sources. (DM-12011)
    - Make plot.ly plotting api to handle unrecognized chart types cleanly. (DM-11477)
- Improvement for plotting API. (DM-11601)

Planned activities

02C.05.00 – Management, Leadership & Other Costs

- Prep work for lSST Science Platform workshop to be held in IPAC.
- Prep work for DM allhands meeting in March 2018, possibly hosted by IPAC.
- Gregory and Xiuqin will attend ADASS in Chile.
- Gregory will attend IVOA interoperability meeting in Chile.

02C.05.01 – Basic Archive Access Tools

- Bug fixes and improvements.
- UI elements for access user workspace. (DM-10855)
- Finish the change to use embedded DB for table data support. (DM-11814)
- Work with DAX on ImageServ and MetaServ API v1.

02C.05.02 – Data Analysis and Visualization Tools

- New functions:
    - Use Embedded DB in the table data manipulation. (DM-11814)
    - Use the multi-trace chart architecture for tri-view. (DM-10833)
    - Publish Firefly to npm in support of future JupyterLab widgets development. (DM-11528)
- Upgrade the nom.tam.fits java package in Firefly. (DM-11883)
- Finish Firefly_client Python documents and examples, publish the doc.

Recruitment update

- None.

1.02C.06: Science Data Archive & Application Services

Current accomplishments

02C.06.00 – Science Data Archive and Application Services Management Engineering and Integration

- Team continued improving robustness of DAX services and otherwise supported SUIT integration efforts in the PDAC as necessary.
- Mueller, Pease, and Hanushevsky had vacations this month.

02C.06.01.01 – Catalogs, Alerts and Metadata

- No work was carried out in this WBS element this month.

02C.06.01.02 – Image and File Archive

- Lo continued work on initial implementation of image service supporting image cutouts via IVOA SODA protocol.

02C.06.02.01 – Data Access Client Framework

- Van Klaveren participated throughout the month in the Data Butler Working Group, helping to develop and refine data butler use cases.
- Pease implemented db storage infrastructure for afw BaseCatalog reader and writer for Butler [DM-11449, DM-11450].
- Pease resumed work on S3 storage interface for Butler [DM-11741, DM-11811, DM-11884, DM-11934].
- Pease addressed obs_base build failure on NFS-mounted systems [DM-11771].
- Pease addressed exception thrown from SqlRegistry due to missing super.__del__ [DM-11861].
- Pease addressed incorrect behavior within writeFitsCatalogStorage [DM-11862].

02C.06.02.02 – Web Services

- Gates and Lo corrected an incorrect bounding box test within metaserv, discovered in PDAC testing [DM-11596].

02C.06.02.03 – Query Services

- Salnikov implemented KILL QUERY support for asynchronous/disconnected queries in Qserv [DM-11729].
- Salnikov implemented improvements to MySQL Proxy timeout behavior and error

reporting for Qserv [DM-11655].
- Jammes completed refactorization of Qserv containers so that MariaDB may be run in its own container under Kubernetes [DM-11126, DM-11893].
- Jammes implemented improvements to Qserv container build process under CI [DM-8914].
- Gaponenko to continued work on Qserv data distribution/replication framework.
- Hanushevsky continued work converting czar code to use the new XRootD SSI v2 API for content-addressed messaging between czar and workers (only update of unit tests remains to be done.)

02C.06.02.04 – Image Services

- No work was carried out in this WBS element this month.

02C.06.02.05 – Catalog Services

- No work was carried out in this WBS element this month.

Planned activities

02C.06.00 – Science Data Archive and Application Services Management Engineering and Integration

- Team continued improving robustness of DAX services and otherwise supported SUIT integration efforts in the PDAC as necessary.
- Mueller, Hanushevsky, and Pease to attend XLDB 2017 in Clermont-Ferrand, France.
- Van Klaveren and Lo to attend ADASS XXVII and IVOA Interop meetings in Santiago, Chile.
- Salnikov has vacation scheduled during this month.

02C.06.01.01 – Catalogs, Alerts and Metadata

- No work is planned for this WBS element this month.

02C.06.01.02 – Image and File Archive

- Lo to complete initial implementation of image service supporting image cutouts via IVOA SODA protocol.

02C.06.02.01 – Data Access Client Framework

- Van Klaveren to continue participation in Data Butler Working Group.
- Pease to complete afw BaseCatalog reader and writer for Butler.

02C.06.02.02 – Web Services

- Van Klaveren to begin working on support for asynchronous queries at the web service layer.

02C.06.02.03 – Query Services

- Gates to work to synthesize “KPM30” Qserv test dataset (sized to 30% DR1) for use in upcoming KPM scale tests.
- Hanushevsky to complete converting czar code to use the new XRootD SSI v2 API for content-addressed messaging between czar and workers.
- Salnikov to address cleanup of stale qmeta entries in Qserv after czar crash/restart.
- Salnikov to improve/cleanup multi-worker integration tests in Qserv.
- Gaponenko to continue work on Qserv data distribution/replication framework.
- Jammes to continue adapting Qserv for deployment within Kubernetes.

02C.06.02.04 – Image Services

- No work is planned for this WBS element this month.

02C.06.02.05 – Catalog Services

- No work is planned for this WBS element this month.

Recruitment update

- No recruitment activity this month.

1.02C.07: LSST Data Facility

Current accomplishments

02C.07.00 – Processing Control and Site Infrastructure Management, Engineering and Integration

We began working on articulating requirements documents and defining verification tests consistent with the Data Facility and overall DM roadmap, and began breaking down replan planning packages into work packages for the next planning cycle. We continued articulating mid-level architecture of LSST Data Facility in the Archi modeling tool, including creation of WBS layered views of logical service constructions and cross-cutting views to show service dependencies.

We finalized work for the FY18 contract, including review of the FY18 staffing plan, modification of the statement of work for annual contracts to align with the replanned WBS, and development of the FY18 acquisition strategy plan and budget for provisioning new hardware capabilities.

We finalized plans and provisioning schedule for the remaining FY17 hardware procurements, including the emergent need for Qserv head node expansion, planned GPFS storage capacity expansion, and the raft-scale consolidated database system.

We participated at the NSF Large Facilities Cyberinfrastructure workshop as a member of the organizing committee.

We held initial discussions with LSST EPO about potential technical and operational considerations of hosting the EPO infrastructure at NCSA and providing services to the department during LSST operations.

Robert Gruendl began chairing and leading activities for the Lossy Compression Working Group (RFC-399).

02C.07.01 – Processing Control

We continued to enhance the DESDM-based processing service based on the need for a stable production-like system in the near term. We continued planned reprocessing of HSC data based on DM needs, to verify the software stack and to test the development of science algorithms. We also documented our investigation of Pegasus failure cases based on operational use cases.

We continued refining the archiving and prompt processing Level 1 services, using the established Early Pathfinder Software Integration Activity schedule to drive development priorities. Specific activities in September included demonstrating readout of the Camera DAQ system, and preparing for the October early pathfinder activity which demonstrates response of the OCS “next visit” event. We also planned initial deployment of the newly acquired expansion to the L1 Complete Test Stand integration system, which will be used in the near term to support testing of the Spectrograph Archiving Service.

We drafted documentation of the header client/service and reviewed it internally in preparation for a design review in October.

02C.07.02 – Infrastructure Services

We continued investigating methods for data movement over a wide-area network (WAN) to support archiving of data from the Spectrograph test stand in early 2018, as well as long-term solutions for data distribution in the Data Backbone. We finished an initial investigation into the Rucio framework and catalog with production file management of the central archive.

We stood up a prototype development-level Kubernetes service using the NCSA Nebula OpenStack system to host a JupyterLab environment for the SQuaRE team, including filesystem access and access control.

We continued efforts to develop an operations-ready deployment procedure for managing containerized applications, including potential use of the OpenShift container application platform. We concluded that the OpenShift layer is not needed for the planned FY18 Kubernetes capability deployment, supporting development, integration and production use cases. The FY18 capability will be a raw Kubernetes install while we continue to investigate the utility (e.g., security and management controls) of OpenShift.

We continued participation in the Butler Working Group, including contributions to requirements and design specifications.

02C.07.04 – Site Infrastructure

We continued provisioning activities for the remaining FY17 hardware acquisitions, including the expanded L1 Complete Test Stand integration platform, a 3PB GPFS storage

expansion, a refresh of the developer database (i.e., lsst-db) platform, and setup and testing of the Base Authentication and Authorization (AA) system at NCSA prior to shipping to Chile.

We began work in consolidation of ITC management systems, investigating the use of packages such as Foreman, Pakrat, and Katello to provide maintainable processes and procedures for managing multiple enclaves at NCSA, at the Base, and the network-based security systems at the Summit.

We continued development and improvements to system monitoring in support of service level monitoring, including preliminary monitoring of GPFS filesystems, defining views and roles for display in Grafana, and improving displays by aggregating low-level metrics.

Planned activities

02C.07.00 – Processing Control and Site Infrastructure Management, Engineering and Integration

We will work toward completion of outstanding service management issues, including documenting and socializing incident reporting.

We will prepare for the face-to-face LSST DM Leadership Meeting in Tucson.

02C.07.01 – Processing Control

We will complete outstanding reviews of Level 1 code packages.

02C.07.02 – Infrastructure Services

We will document findings of investigations into moving containerization toward being operationally ready, and finish outstanding issues in providing prototype development containerized application services.

We will complete our commitment to the Butler Working Group, pending feedback from DM leads and overall progress on deliverables.

02C.07.04 – Site Infrastructure

Provisioning of remaining hardware systems from the FY17 acquisition plan will continue through October.

Work on system monitoring will concentrate on what will be needed for supporting the Authentication and Authorization (AA) service and network-based security service endpoints in Chile, planned for deployment in FY18, as well as general incremental improvements.

Work toward consolidation of ITC management systems will continue with evaluation and testing of third-party packages.

02C.07.05 LSST Data Facility Management, Service Architecture, and Project Controls

We will continue articulating requirements documents and defining verification tests consistent with the Data Facility and overall DM roadmap. We will finish breaking down

replan planning packages into work packages for the next planning cycle. We will continue articulating mid-level architecture of LSST Data Facility in the Archi modeling tool.

We will continue work planning for provisioning of FY18 hardware acquisitions, focusing on the L1 system expansion and infrastructure expansion necessary to provision subsequent systems.

We will continue activities leading the Lossy Compression Working Group.

We will continue pursuing incremental improvements to Service Management processes, as well as design of service level monitoring capabilities. This includes preparing for 24/7 incident response as needed to support systems in Chile AA Base.

02C.07.06.01 LDF-offered Services

We will continue periodic reprocessing of datasets in support of stack testing and pipeline development.

02C.07.06.02 Reusable Production Services

We will continue supporting the in-place prototype Kubernetes installation to support JupyterLab development. We will also continue work towards developing a stable containerized management architecture.

02C.07.07 Data and Compute Services

We will continue work investigating file transfer mechanism toward support of the Spectrograph test stand.

02C.07.08 LDF Service Software

We will continue work on L1 Service software, focusing on support for the DAQ, implementing the first version of Spectrograph controller functionality, unit tests for L1 components, and responding to outcomes of integration activities.

We will hold a design review for Header Client software design, and begin enhancements to the software based on recommendations.

We will begin python3 conversion for DESDM framework batch production code, and make changes as necessary to support ongoing dataset reprocessing.

We will prepare software and systems to host and participate in the Pathfinder Integration Activity “Next Visit” milestone and workshop.

02C.07.09 ITC and Facilities

We will begin planning for provisioning of the first phase of implementing the FY18 acquisition plan. The initial focus will be planning general enclave infrastructure expansion, Kubernetes cluster installation, Level 1 Camera Control System (CCS), and Level 1 Integration system expansion. We will make improvements to the existing disaster recovery system. We will begin installing and testing system-level monitoring services for the AA system prior to delivery to Chile, as well as making general improvements and additions to existing system monitoring.

Recruitment update

Christopher Heller joined the project at NCSA as Senior Storage Engineer. Christopher Pond rejoined the project as Senior Database Administrator. A requisition for a full-time database administrator is in progress.

1.02C.08: International Communications & Base Site

Current accomplishments

02C.08.01 – Base Center

- Summit Base ITC Tiger Team: Completed review and adopted standard DNS naming convention for ITC devices at Summit – Base complex.
- Summit and Base Networking and Computing: Firmed up with NCSA that they will come to Chile to install the security system in December, 2017. Equipment list was reconciled. Cables and equipment will be shipped prior to December. We will install some the equipment as it arrives ahead of the NCSA visit. Racks and power will be ready.

  Most of the Cisco summit network equipment arrived. Still missing a few items, namely the ACI cables in order to start sandbox testing and configurations.

  Condition of construction at the telescope indicates we will not get beneficial occupancy until end of October and full occupancy at the end of the year. Planning of the network ducts and power in the computer room and cable trays in the Telescope building is being finalized. At the advice of the LSST Architect, in order to secure site occupancy from Besalco as soon as possible, these elements will be installed by another contractor after Besalco has granted full occupancy.

02C.08.02 – Chilean Data Access Center

- No activity this month.

02C.08.03 – Long-Haul Networks

- Ron Lambert attended three day course on Coriant Network Management System in Santiago.
- Jeff Kantor will travel to Chile on October 17 to attend SAACC and ADASS, then continue on to La Serena on October 27 to coordinate the Summit and Summit - Base installations and configuration and test. He will return to Tucson on January 9, 2018.

02C.08.03.01 – Chile National WAN

- Summits–AURA Gatehouse Network: No activity. This segment is accepted.
- DWDM Equipment: The AURA DWDM installation has been delayed due to the Caseta on Cerro Pachon not having enough backup generator capacity per AURA Operations. AURA Operations has started the process to procure and install a new generator. We will install the AURA DWDM for testing even though the Pachon caseta will not have a new generator until mid December. The majority of the REUNA DWDM equipment was received from Coriant. Installation will start in October in Santiago and complete in early November in La Serena.
- Santiago–La Serena: No activity. This segment is accepted.
- La Serena–AURA Gatehouse: No activity. This segment is accepted.

02C.08.03.02 – International Chile–US WAN

- *100 Gbps Managed Ring:* No activity. The ring is performing as expected. Equipment was shipped from Ampath to enable a demonstration at bandwidth of >40Gb at the October ADASS meeting. This equipment can be shipped to La Serena when we have the AURA DWDM equipment installed at the summit and base to carry out Cerro Pachon to NCSA tests later in the year.
- *Management and Coordination Contract:* The LSST Network Engineering Team (NET) meeting was held on September 21. Several points concerning the Virtual Network Operations Center (NOC) requirements were discussed: need to specify consistent, general requirements for failure tracking, notification, weather maps, etc. Updates on Quality of Service (QoS0 plan are delayed because of the hurricane Irma, but will be prepared for the next meeting. Details about the shipping the equipment for the First Fiber Light Event planning to Chile were discussed. A monitoring system to track the traffic during the ADASS demonstration was also mentioned. Members of the NET attended conferences relevant to LSST Networks:
    - NSF Large Facilities Cyberinfrastructure Workshop September 5–8
    - Large Scale Networking (LSN) September 17–21
    - Global Lambda Integrated Facility (GLIF) Conference September 24–28
- *Spectrum Contract:* The Angola Cable contract is signed. Activity centered on finishing Operations and Management contract and payment schedule.
- *US National WAN:* ESNet collaboration negotiations with DOE and with FNAL on an operations role continued. A comparison of internet2 only and internet2 + ESnet approaches was documented. FIU continued working with FLR and other fiber providers to determine the most cost effective and greatest throughput for transit from AMPATH north.

Planned activities

02C.08.00 – International Communications and Base Site Management Engineering and Integration

- Several members of the team will attend the October ADASS conference in Santiago.
- The South American Astronomy Coordination Committee (SAACC) meeting will be held at REUNA on October 19, 2017.

02C.08.01 – Base Center

- Summit to Base ITC Tiger Team: Tiger Team completed its last regularly scheduled meeting in CY2017. Will resume with the focus on the Base Facility after Jeff Kantor returns from Chile in January, 2018.
- Summit and Base Networking and Computing: Install the NCSA-provided network security taps at the WAN side of the AURA gateway router. With the arrival of the missing ACI cables we will start working with the ACI architecture in a sandbox in La Serena.

02C.08.03 – Long-Haul Networks

02C.08.03.01 – Chile National WAN

- Summits - AURA Gatehouse Network: No activity. This segment is accepted.
- DWDM Equipment: We expect to do the bulk of the REUNA DWDM and AURA DWDM equipment in October and early November. The LSST DWDM equipment will not be installed until March 2018.
- Santiago-La Serena: No activity. This segment is accepted.
- La Serena - AURA Gatehouse: No activity. This segment is accepted.

02C.08.03.02 – International Chile - US WAN

- 100 Gbps Managed Ring: Two cross-connects were purchased at Level3 Data Center to open the 100G links in Chile. A loaner 100G module was sent to Santiago until AURA provides a confirmation of the solution to be purchased (options provided last month with three scenarios).
- Management and Coordination Contract: Continue to coordinate NET meeting and SAACC meeting. Attend and present at the ADASS Conference in Santiago, Chile in October 22-26, 2017: International Networking in support of Extremely Large Astronomical Data-centric Operations; Challenges of Standardizing and Supporting ITC Services in a Widely Distributed Project: ITC Design of the LSST Summit - Base Complex.
- Spectrum Contract: Continue work on the Operations and Maintenance Agreement.
- US National WAN: Discussions with ESNet for LSST within DOE to define the ESNet

service and costs will continue. FIU will continue working with FLR and other fiber providers to determine the most cost effective and greatest throughput for transit from AMPATH North.

Recruitment update

- Andres Villalobos, the new System Administrator, started work with LSST.
- With the completion of hiring of IT staff in Chile discussion with LSST Tucson commenced as to how to divide the IT Support work between the sites. This includes documentation of a common framework of standards, processes, and tools to be shared by both sites.

1.02C.10: Science Quality and Reliability Engineering

Current accomplishments

Highlights of work completed this month include:

- The JupyterLab-based notebook environment has been deployed on a test Kubernetes cluster hosted on Nebula at NCSA. This gives us confidence that the system can be made available to DM staff once a stable Kubernetes cluster with persistent file storage is made available to us. Thanks to Matias Carrasco Kind (NCSA) for his assistance.
- We have integrated the JupyterLab-based notebook environment with the LSST Data Facility authentication system (CI logon).
- The TE1 and TE2 KPMs were added to validate_drp.
- The default packaging build has been moved to Python 3 and migrated off lsst-dev to improve reliability and performance.
- Metadata from LaTeX source documents curated in LSST-the-Docs is now being extracted.
- SQuaRE-relevant usecases were submitted to the Butler Working Group.
- A Butler bug that affected ci_hsc processing was fixed.

Planned activities

- Deploy the kubernetes-based SQuaSH into production.
- Publish the 14.0 release.

Recruitment update

None planned.