Report on Operations Rehersal #1

William O'Mullane and Robert Gruendl and Robert Blum

2019-07-13

1 Introduction

LSST Data Management (DM) conducted the first operations rehearsal from May 7^{th} to 9^{th} 2019. The plan for the rehearsal was outlined in LDM-643, the main principle was to simulate nominal operations with CCD data flowing from Chile, being processed (calibrated) and having some quality assurance done on it. LDM-643 details the procedure and personnel involved, hence in this document we give only a brief summary of what happened in the rehearsal.

2 The rehearsal #1

There was a short prep meeting on the Monday before the rehearsal started to make sure everything was in place.

2.1 Setup and limitations

A set of raft data (simulated data from DC2) which intersected intersect tract 4849 patch 2,2. The datasets for each night are comprised of "observations" from two bands (night1: r, i-bands; night2: z, y-bands, night3: u, g-bands).

Processing is to run singleFrameDriver.py - this is still Gen2 butler.

The three Dark Energy Science Collaboration (DESC) DC2 data sets were transferred to a Data Transfer Node (DTN)[1] in Chile to act as the acquired data. A cron job was installed to start a script[2] to transfer the files to the receiving node in NCSA [3]. Secure copy (`scp`) was used to make the transfer under the user `womullan`.

In operations the transfer nodes should be automatically transferring data which shows up in an operations directory. The script was only for this exercise. The DTN which was used is

[1] lsst-user@139.229.127.99
[2] `https://github.com/womullan/opsforwarder`
[3] xfer.ncsa.illinois.edu

an Florida International University (FIU) machine not one of our final nodes and hence is not set up as we will for operations.

Ingest, processing and Quality Assurance (QA) were run on the full dataset once transferred not as each file arrived.

2.1.1 Communications

A Slack channel #ops-rehearal-1 was created for minor communications.

Daily telecon was held using bluejeans at 11:00PST with the agenda:

- Serio Status on mountain (can tailor for what is actually happening right now and we can discuss how this looks in commissioning and operations). Andrew gives ”mock” night report.
- Gruendl Processing summary. How did ingest go. How did processing go.
- MacArthur, Slater Metrics/QA, what's in the logs, what can we say about the data. Summary plots and metrics. What is missing in our view. What can we add for next night?
- Morganson Issues, resolution. Did anything happen that we can fix for the next night
- All Plan for next night. What do we need to do to make sure next night works better (or as well).

2.2 Day 1

The daily meeting took place as planned at 11:00 Project Science Team (PST).

It was noted the transfer took longer than expected - a potential network problem was suspected. The data transfer script has a 5 second delay (which should have been removed) that made the transfer take far too long. To keep the process running the dataset, which was already at NCSA, was copied into the incoming folder.

This was ingested into `/project/OpsRehearsal_1/night1` and processing (`singleFrameDriver.py`) ran smoothly averaging 1.5 CCDs/s (running with 24 cores x 3 nodes), taking a total of 48 minutes.

29 "FATAL ' errors (w/58 Runtime Errors) were noted among the 1,997 CCDs processed There were issues with N stars and Point Spread Function (PSF) build and Flux limit (linked to a a known problem).

The odd number of CCDs (not a multiple of 9) was immediately remarked upon and investigated.

Morganson found it was missing one exposure from the 01687569 series using : `for exp in` `ls | cut -c 1-8 | sort -u`; do nexp=`ls $exp* | wc -l`; echo $exp, $nexp; done This was not a transfer problem: the simulation was made with one CCD missing in one exposure.

QA scripts were kicked off by MacArthur including `visualizeVisit.py`[4], the latter did not work and needed a patch. Plots were accessible `https://lsst-web.ncsa.illinois.edu/~lauren/OpsRehearsal_1/attempt1/plots/`. Some issues became clear perusing some of the plots. As an example, the upward tilt towards bright magnitudes in Figure 1 is a red flag for the "Brighter Fatter" issue.

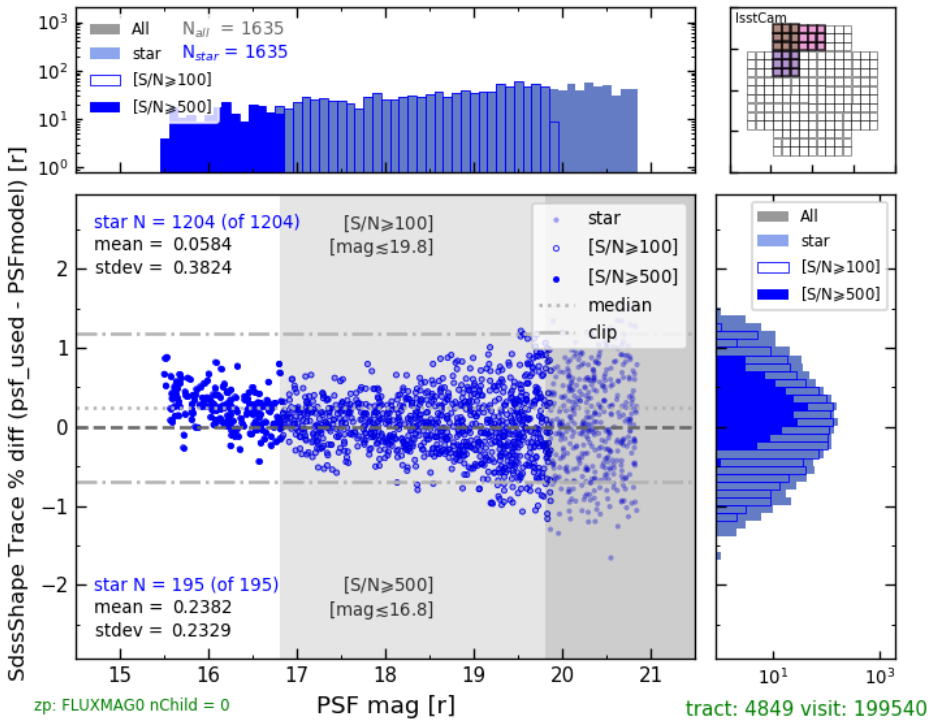

FIGURE 1: Brighter fatter effect evidence

[4] `https://community.lsst.org/t/y-band-stray-light-correction-for-hsc/2517`

2.2.1 Discussion

We discussed how to implement change in code with tight turn around during actual operations (need sign off from SciOps Associate Director (AD)). How should one decide to rerun - we need policies to affect or not nightly or daily changes in case of failures. Probably there is some percent level of problems we would accept, one percent seems ok two starts to seem a lot. It was noted that problems stemming from a new software version should always have the possibility to revert to a previous stable version.

The need for some sort of rolled up QA status were discussed - MacArthur came up with some summary files an example of which `visitAnalysis_OpsRehearsal_night1_i.shortSum.txt` is in Appendix A.

2.3 Day 2

The daily meeting took place as planned at 11:00 PST.

Transfers were initiated earlier than Day 1 with the 5 second delays removed, and all data had arrived at the LSST Data Facility (LDF) by 7:30 am. Data ingestion `/project/OpsRehearsal_1/night2` and processing were initiated shortly thereafter and were completed within roughly 1 hour.

No processing errors were reported, so examination of less severe issues were undertaken. Noted were WARN-level problems that revolved around reference catalogs being in an outdated format and the lack of zeropoint information for z- and Y-bands.

2.3.1 Discussion

There were discussions about how the current pipeline, operating on DC2 data might compare with what might be needed during commissioning and operations. It was noted that because the rehearsal was using a processing pipeline (and QA tools) more amenable to Data Release Processing (Data Release Production (DRP)) that QA products available might not be good comparisons to that needed when working with prompt processing. Also, it was noted that prompt processing QA might form a basis for later selection of inputs to DRP.

Another discussion revolved around whether WARN-level diagnostics might be dealt with in Operations/Commissioning. Mostly it was felt that these fell into two classes: those that were indicative of poor data (which might not require any intervention) and those that indicated software bugs (which should be tracked/resolved through tickets to the DM developers).

2.4 Day 3

The daily meeting took place as planned at 11:00 PST.

Transfers were initiated similar to Day 2 and all data had arrived at the LDF by 7:30 am. Data ingestion `/project/OpsRehearsal_1/night3` required roughly 20 minutes. Processing was subsequently initiated but after 30 minutes it was found that jobs had never been submitted to the compute resource. That resource (a reserved allocation) was found to still be running QA from Night 2. An alternate compute resource was identified and jobs re-submitted, finishing ~35 minutes later.

Similar WARN-level problems were identified prior to the daily meeting as were a small number of failures, similar to night 1 (number of stars resulting in failed PSF modeling).

2.4.1 Discussion

Discussions centered on what information needed to be captured from this rehearsal for future rehearsals. It was felt that this rehearsal proceeded relatively smoothly but did not include processes that could mimic commissioning team's needs (i.e. fast turn around processing) and ad-hoc processing of data taken to address specific tests. The next rehearsal (late 2019/early 2020) is meant to focus toward LSST Commissioning with ComCam and should include Commissioning Team members.

3 Conclusion and lessons learned

Though there were some limitations in out setup (Section 2.1) this was still useful exercise.

One central issue that should be followed-up on in subsequent rehearsals are the QA metrics to be gathered during prompt processing and that utilities to provide rollups/summaries.

Though in the telecon status was discussed it was not properly recorded in the google sheet for the day, there was no designated minute taker an individuals did not necessarily add a summary. In operations such a status report should probably be filled even before the meeting.

A Visit summary

```
# Stars: Mag(Gaussian) - PSFMag (mmag) (commonZP)
#                filter    mean    stdev    median     num      numTot  NumEntries
# straight avg       i    -0.46     6.84    -0.56     51258     52230        83
# weighted avg       i    -0.43     6.67    -0.52     51258     52230        83
#================================================================================
# Stars: Mag(CircAper12pix) - PSFMag (mmag) (commonZP)
#                filter    mean    stdev    median     num      numTot  NumEntries
# straight avg       i    -1.08     9.63    -1.20     51373     52230        83
# weighted avg       i    -1.05     9.57    -1.16     51373     52230        83
#================================================================================
# Stars: Mag(Gaussian) - PSFMag (mmag)
#                filter    mean    stdev    median     num      numTot  NumEntries
# straight avg       i    -0.46     6.84    -0.56     51258     52230        83
# weighted avg       i    -0.43     6.67    -0.52     51258     52230        83
#================================================================================
# Stars: Mag(CircAper12pix) - PSFMag (mmag)
#                filter    mean    stdev    median     num      numTot  NumEntries
# straight avg       i    -1.08     9.63    -1.20     51373     52230        83
# weighted avg       i    -1.05     9.57    -1.16     51373     52230        83
#================================================================================
# Stars:   SdssShape Trace (calib_psf_used): $\sqrt{0.5*(I_{xx}+I_{yy})}$ (pixels)
#                filter    mean    stdev    median     num      numTot  NumEntries
# straight avg       i     1.50     0.01     1.50     39435     39441        83
# weighted avg       i     1.51     0.01     1.51     39435     39441        83
#================================================================================
# Stars:  SdssShape Trace: $\sqrt{0.5*(I_{xx}+I_{yy})}$ (pixels)
#                filter    mean    stdev    median     num      numTot  NumEntries
# straight avg       i     1.51     0.01     1.51     51513     52230        83
# weighted avg       i     1.51     0.01     1.51     51513     52230        83
#================================================================================
# Stars:  SdssShape Trace % diff (psf_used - PSFmodel)
#                filter    mean    stdev    median     num      numTot  NumEntries
# straight avg       i     0.05     0.54     0.06     39411     39441        83
# weighted avg       i     0.03     0.52     0.05     39411     39441        83
```

```
#================================================================================
# Stars: PSF - ref (calib_psf_used) (mmag)
#                filter    mean    stdev    median     num      numTot  NumEntries
# straight avg       i     1.08     7.07     1.12     36861     37100        83
# weighted avg       i     1.04     7.04     1.08     36861     37100        83
#================================================================================
# Stars:  PSF - ref (calib_photom_used) (mmag)
#                filter    mean    stdev    median     num      numTot  NumEntries
# straight avg       i     1.30     7.58     1.26     48957     49377        83
# weighted avg       i     1.27     7.55     1.23     48957     49377        83
#================================================================================
# Stars: PSF - ref (mmag)
#                filter    mean    stdev    median     num      numTot  NumEntries
# straight avg       i     1.26     7.60     1.24     50483     50939        83
# weighted avg       i     1.23     7.56     1.19     50483     50939        83
#================================================================================
# Stars: CircAper12pix - ref (calib_psf_used) (mmag)
#                filter    mean    stdev    median     num      numTot  NumEntries
# straight avg       i     0.41     7.33     0.35     35636     37100        83
# weighted avg       i     0.41     7.32     0.34     35636     37100        83
#================================================================================
# Stars:  CircAper12pix - ref (calib_photom_used) (mmag)
#                filter    mean    stdev    median     num      numTot  NumEntries
# straight avg       i     0.28     7.41     0.22     47018     49377        83
# weighted avg       i     0.29     7.39     0.22     47018     49377        83
#================================================================================
# Stars: CircAper12pix - ref (mmag)
#                filter    mean    stdev    median     num      numTot  NumEntries
# straight avg       i     0.20     7.50     0.12     48218     50939        83
# weighted avg       i     0.20     7.48     0.12     48218     50939        83
#================================================================================
# Stars: matches_distance_calib_astrometry_used (mas)
#                filter    mean    stdev    median     num      numTot  NumEntries
# straight avg       i     7.51     4.51     6.70     50403     50939        83
# weighted avg       i     7.51     4.50     6.72     50403     50939        83
#================================================================================
```

```
# Stars: matches_distance (mas)
#                filter    mean    stdev    median     num      numTot  NumEntries
# straight avg       i     7.51     8.76     6.70     50403     50939        83
# weighted avg       i     7.51     8.76     6.72     50403     50939        83
#================================================================================
# Stars:  $\delta_{Ra}$ = $\Delta$RA*cos(Dec) (mas) (calib_astrom_used)
#                filter    mean    stdev    median     num      numTot  NumEntries
# straight avg       i    -0.02     7.04    -0.08     50781     50939        83
# weighted avg       i     0.01     7.05    -0.06     50781     50939        83
#================================================================================
# Stars: $\delta_{Ra}$ = $\Delta$RA*cos(Dec) (mas)
#                filter    mean    stdev    median     num      numTot  NumEntries
# straight avg       i    -0.02     7.04    -0.08     50781     50939        83
# weighted avg       i     0.01     7.05    -0.06     50781     50939        83
#================================================================================
```

References

[LDM-643], Johnson, M., Gruendl, R., 2019, *Proposed DM OPS Rehearsals*, LDM-643, URL `https://ls.st/LDM-643`

B Glossary

AD Associate Director.

CCD Charge-Coupled Device.

Commissioning A two-year phase at the end of the Construction project during which a technical team a) integrates the various technical components of the three subsystems; b) shows their compliance with ICDs and system-level requirements as detailed in the LSST Observatory System Specifications document (OSS, LSE-30); and c) performs science verification to show compliance with the survey performance specifications as detailed in the LSST Science Requirements Document (SRD, LPM-17)..

Data Management The LSST Subsystem responsible for the Data Management System (DMS), which will capture, store, catalog, and serve the LSST dataset to the scientific community and public. The DM team is responsible for the DMS architecture, applications, middleware, infrastructure, algorithms, and Observatory Network Design. DM is a

distributed team working at LSST and partner institutions, with the DM Subsystem Manager located at LSST headquarters in Tucson..

Data Release Processing Deprecated term; see Data Release Production..

DESC Dark Energy Science Collaboration.

DM Data Management.

DRP Data Release Production.

DTN Data Transfer Node.

FIU Florida International University.

LDF LSST Data Facility.

NCSA National Center for Supercomputing Applications.

Operations The 10-year period following construction and commissioning during which the LSST Observatory conducts its survey.

patch An quadrilateral sub-region of a sky tract, with a size in pixels chosen to fit easily into memory on desktop computers..

pipeline A configured sequence of software tasks (Stages) to process data and generate data products. Example: Association Pipeline..

Project Science Team an operational unit within LSST that carries out specific scientific performance investigations as prioritized by the Director, the Project Manager, and the Project Scientist. Its membership includes key scientists on the Project who provide specific necessary expertise. The Project Science Team provides required scientific input on critical technical decisions as the project construction proceeds.

PSF Point Spread Function.

PST Project Science Team.

QA Quality Assurance.

tract A portion of sky, a spherical convex polygon, within the LSST all-sky tessellation (sky map). Each tract is subdivided into sky patches..

Visit A sequence of one or more consecutive exposures at a given position, orientation, and filter within the LSST cadence. See Standard Visit, Alternative Standard Visit, and Non-Standard Visit,DM TS Sims,, Education and Public Outreach (EPO),The LSST subsystem responsible for the cyberinfrastructure.