DM Leadership Team Face-to-Face Meeting, 2019-10-28 to 30

Logistics

Date

28–30 October 2019.

The meeting will start at 09:00 on 28 October. Non-local attendees should plan to travel the day before.

The meeting will finish by lunchtime on 30 October; feel free to arrange travel for that afternoon.

This meeting will be followed by a DM-SST meeting on the afternoon of 30 October. SST members should not plan to leave before 17:00. Interested T /CAMs are welcome to attend.

Accommodation

- A room block has been reserved at the Creekside Inn, 3400 El Camino Real, Palo Alto, California 94306.
- **Use this link to book.**

Location

- SLAC National Accelerator Laboratory. (B51/ Kavli 3rd floor conference room)

BlueJeans

- The meeting will be broadcast on BlueJeans at https://ls.st/jmc (the usual DMLT BJ connection).

Slides

- Please upload slides to Confluence in advance of your presentation.

Participants

- Leanne Guy
- Unknown User (gcomoretto)
- Unknown User (mbutler)
- John Swinbank
- Eric Bellm
- Robert Gruendl
- Fritz Mueller
- Simon Krughoff
- Tim Jenness
- Wil O'Mullane
- Jeff Kantor
- Kian-Tat Lim
- Jim Bosch
- Frossie Economou (Remote)
- Gregory Dubois-Felsmann (Remote)
- Colin Slater
- Unknown User (npease)
- Robert Lupton

Agenda

<table>
<tr><td colspan="5" align="center">Day 1: 2019-10-28</td></tr>
<tr><th>Time (Local!)</th><th>Topic</th><th>Chair</th><th>Discussion Topics</th><th>Notes and action Items</th></tr>
<tr>
<td>09:00</td>
<td>Welcome</td>
<td>Wil O'Mullane</td>
<td>

Confirm agenda.
Review action items.

</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Project news & Data Facility updates</td>
<td>Wil O'Mullane</td>
<td>

It's not yet clear what we'll be able to say about changes (or not changes) to the LDF by the time of this meeting, but we should give what status update we can and allow time for discussion.
Also any other project current events.

</td>
<td>

DOE:

  Had one telecon with each of the DMLT labs.
  All "enthusiastic and very nice".
  Have provided labs with assessment criteria.
  Will be doing a test deployment of a Kubernetes based service.
  First site visit is SLAC, this week (following the DMLT/DM-SST meetings); BNL and Fermilab over the following couple of weeks.
  Then we expect labs to submit documentation, and hope to open this to wider DMLT review.

Naming:

  Lots of uncertainty!
  There might be an announcement before AAS. Or there might not.

Things seem to be returning to normal for AURA staff in Chile.
Steve Kahn suggests that we should anticipate a subsystem technical meeting in ~February to discuss how to complete construction.
Steve adds that we are expecting NSF to fund Alison Rose's film; details still being worked out.

  But nobody is forced to participate!

</td>
</tr>
</table>

09:45	Middleware status	Fritz Mueller	• Latest updates on generation 3 middleware.
			• Timeline for dropping support for gen 2.
			• Slides
				• Major blockers are multi-user registries and repo stability: these seem to form the irreducible core of Jim Bosch's work.
				  ○ Can we use a "friendly user" shared schema to avoid the multi-user issue?
				  ○ High risk of data loss, etc.
				  ○ But could do ad-hoc backups, etc.
				  ○ And it's not clear how this could work in DACs.
				  ○ Jim thinks that the friendly user mode could still be a big advantage, in particular to free up his time before the Algorithms Workshop, although we would ultimately need to sort out the multi-user registry.
				• Does Jim really need to be the person working on multi-user registries? Surely this just needs database expert?
				  ○ Risk of delegation is that new developers would have unpredictable velocities.
				  ○ Try to separate the issues of design and implementation, and have Jim focus only on the former.
				• Need to develop a technical plan to deploy a system without a multi-user registry; would still be significant work.
				  ○ There are multiple DMLT pundits suggesting how this could be done.
				• Do users have to share the same registry as production?
				  ○ If they don't, there's a lot of duplication.
				
				☑ Fritz Mueller — consider mechanisms by which Robert Lupton and other science users can provide feedback on user-facing middleware issues to the development team. 18 Dec 2019
				
				☑ Jim Bosch — identify a design which enables hand-off of middleware implementation to alternative developers ASAP, probably including "friendly user" shared registries. 18 Nov 2019

10:30	Break (Refreshments Provided)			
11:00	Future middleware development plans	Fritz Mueller	• Staffing needs and management plan for middleware development from the start of calendar year 2020 for the remainder of construction. • Slides	• "Get Jexit done", thank you Kian-Tat Lim. • Note that converting Science Pipelines code to PipelineTasks is not the same as working on middleware development itself. • Concerns about addressing batch processing workflows through notebooks, interface with CAOM model and IVOA services, virtual data products, etc. — when can the middleware handle this? At what level is it an overriding priority for commissioning? ☑ Leanne Guy — identify a product owner for middleware (or, possibly, a science owner and a technical owner). They should not be (a) member(s) of the development team. 18 Nov 2019 ☑ Fritz Mueller — arrange a DMLT-level demo of the current state of the middleware. 18 Nov 2019 ☑ Wil O'Mullane — identify the long-term *management* structure (as opposed to product owner) for middleware. **DM-22393** - Getting issue details... STATUS 18 Nov 2019

| 11:30 | SDM standardization & database ingest | Yusra AlSayyad & Colin Slater | Update on the status of SDM-standardization efforts. Should include: A reminder of the "big picture" plan, including an overview of all the moving pieces (required pipeline outputs, Felis, database ingest) and the schema(s) they support/understand/map to each other. Results of SDM-standardization work carried out on the AP and DRP pipelines in the F19 development cycle. Status of database ingest. How close are we to trivially having the results of HSC processing published through the LSP? Slides | Ultimately, expect Qserv to be able to ingest directly from Parquet (not TSV). But it shouldn't really matter as long as users don't have to see. How do we allow users to access this data while being mindful of Qserv constraints? Automated QC should be on Parquet files. Human poking can be on either Qserv or Parquet. Should aim for continuous integration: load after every bi-weekly processing run. A smaller Qserv system is being set up for commissioning. Gregory Dubois-Felsmann suggests that we can alternate periods of stability and instability on the main 30-node Qserv system. Can we scale up the small Qserv instance at NCSA (from 6 to 30 nodes)? Nobody seems to quite know why not, but also nobody quite seems to know if it's necessary. When do we need something like Qserv for commissioning? When ComCam goes on sky, which is likely mid-2021. However, we need HSC data in Qserv at scale well before that, to make sure it's ready to receive the commissioning data. The Commissioning Team are keen to have tooling that can work against Qserv as soon as possible. They need a small, stable Qserv with HSC and Gaia loaded. Do not expect that individual developers will ever run database ingest on their small runs, so QA tooling will have to work with Parquet. Expect to have to support incremental ingest to Qserv, which was not part of the original design. Expect to have to use cloud resources for scale testing Qserv. Need work to reconcile the pipelines outputs with what is specified in the DPDD. Request DM-SST help with this. This means e.g. adding more information to the tables which is not currently specified in the DPDD. We regard the DPDD as a minimum, but note that it has impacts on the sizing model. No objections to putting some of QA in a separate table which needs joining against. There is a row-by-row calibration in the object table to take account of spatially varying calibrations. A per-object zeropoint. But we do store nJy, so this calibration has already been applied. It is essential that all calibration can be "undone" if required by science users. (Context of this question was: will users have to perform JOIN queries to get DPDD-ish, calibrated catalog data? Current answer appears to be: no.) ☑ Unknown User (mbutler) — arrange for procurement of a small Qserv cluster, per the FY20 procurement plan, available in ~January. 06 Jan 2020

☑ Colin Slater , Leanne Guy , Yusra AlSayyad — start a process for reconciling the DPDD with the required information for pipeline QA. 18 Nov 2019 Ticket:
DM-22078 - Getting issue details...
STATUS |
| | | Yusra AlSayyad & Colin Slater | | |

12:15	Verification Control Document	Unknown User (gcomoretto)	• Overview of LDM-692.	• Test cases should be as generic as possible, but we acknowledge that they do occasionally have dependencies on particular datasets, etc. • Note that only a very small fraction of currently defined test cases have been passed. • This tool is useful for the mechanical view of which requirements have been verified, but not (in general) for the scientific validity. • The DM-SST, and in particularly Jeff Carlin, will reach out to the owners of LDM-503 milestones to ensure that the appropriate test cases are incorporated into their milestones. • Milestone owners should coordinate with Jeff Carlin about the contents of their milestones. ☑ Leanne Guy — update LDM-503 to reflect policy around which requirements can be verified within the DM subsystem, based on available data. 18 Nov 2019 Ticket: **DM-22089** - Getting issue details... STATUS
12:30	Lunch (**not** provided — SLAC cafeteria, use your per diem)			

13:30	Test Datasets for Scientific Performance Monitoring	Leanne Guy	• The SST proposal for test datasets for scientific performance monitoring.
			• See DMTN-091.
				• Processing does not currently include SDMification; that will be forthcoming.
				• How long should old test datasets be preserved?
				  ○ It's useful to preserve some months of medium sized processing.
				  ○ Should be no need to preserve CI-processed datasets.
				• Request for an API to trigger processing from CI.
				• No rush to define a "less large large" dataset until a need is actually shown.
				• Can an SDMified dataset be produced frequently, not just following large datasets?
				  ○ Yes; it should be run at least after RC2 data processing.
				  ○ Indeed, we so no reason why it shouldn't run after *everything*, as long as it's fast enough.
				
				☑ Kian-Tat Lim & Leanne Guy — confirm HSC test dataset volume. 11 Nov 2019
				
				☑ Leanne Guy — include DESC DC2 data in DMTN-091. 11 Nov 2019
				
				☑ Leanne Guy and at least Unknown User (mbutler) and Yusra AlSayyad to determine for how long processed test datasets should be preserved. 18 Nov 2019
				DM-22528 - Getting issue details...
				STATUS
				Ticket:
				
				☑ Leanne Guy — make sure that DMTN-091 is clear that it is not *necessary*, rather than not *feasible*, to test DIA processing on "large" datasets. 11 Nov 2019
				
				☑ Leanne Guy – Add note to DMTN-091 about running SDM-standardization to generate DPDD for MEDIUM datasets as well. 11 Nov 2019 Ticket:
				DM-22088 - Getting issue details...
				STATUS

14:00	Status of reprocessing systems at NCSA	Unknown User (mbutler)	- Given the discussion on datasets (above) and the ongoing and emergent requirements of the construction and commissioning projects, we should understand the current capabilities and future schedule for reprocessing efforts at NCSA. In particular: - Following the departure of Hsin-Fang, what staff are assigned to this? - How ready is the NCSA team to support regular test processing as requested by the SST (above)? - How are issues reported and acted upon? - Is tracking metrics with SQuaSH adequate? - Do we also need Jira tickets? - Who is responsible for filing/triaging /resolving issues? - What work needs to be done, either by NCSA or by other teams, to make this process as automatic as possible? - Can we read this last item as "what is the current status of the Batch Production Service"?	- Each dataset being processed should have an owner (e.g. Yusra for RC2/DC2 data processing) who agrees plans for processing with the LDF team. - There should be some default agreed configuration for processing, and a process by which the owners can change that configuration on demand. - Expect the LDF to ultimately aim to become familiar with the warnings issued by the pipeline and to filter out the ones which are unimportant. In the short term, though, there may be some elevated level of warnings issued by the LDF team. - Success criteria: - Comparison with number of files in previous run. - Look for errors being logged. - There is a generic problem that the Pipelines are poor at capturing errors in a coherent way. - This will be easier in Gen 3 (we're assured). - Where appropriate, success can be seen because relevant values are stored in SQuaSH.
14:30	Workflow	Unknown User (mbutler)	(In so far as it is not covered by the item above), please provide an overview of the current design and implementation of the various services which require workflow management. - In particular, following surprise /confusion at the June 2019 DMLT F2F, we should establish whether tools like Pegasus are required: - For the Prompt or Batch Production Services; - For user-triggered processing from the Science Platform; - To respond to the recommendation from the Directors' Review that "more effort should be spent on refining diagnosis and recovery from processing errors as this will be critical for operating at scale."	- Design will be DMTN-123 when it's ready. - Is there a clear division of responsibility in terms of error reporting between PipelineTasks and the workflow system? - In general, PipelineTasks should be atomic (either they throw, or they complete). - There is broad agreement in the room that Pegasus is not a requirement at this stage, although we note that it could be re-added to the system at the appropriate time — compatibility will be maintained. ☑ Unknown User (mbutler) — ensure that Pegasus support is tested in CI. 25 Nov 2019
15:00	Break (Refreshments Provided)			

15:30	Commissioning (and Control) Dataflow and Processing	Kian-Tat Lim	Review of DMTN-111 highlighting missing pieces; confirmation of timelines for delivery of them.	- In the RHL ideal world: write a notebook, have it execute on some big cluster without him worrying about the details. - We will need a lot of compute, and a lot of flexibility; should not be locked into particular modes of operation. - Should be possible to perform full real-time reductions of LSSTCam data, e.g. by transport to the base, or by providing sufficient computing at the summit. - We need to come up with a by which notebook users can call out to back-end execution services, which execute code from the notebook. - PipelineTask, in and of itself, does not fill this complete role, but might be part of the solution.
16:15	Influx EFD	Frossie Economou	- Demo with questions. Simon Krughoffplans to present this. - And a status update of who is responsible for delivering and running which services where (in particular, what is an LDF responsibility, what is a SQuaRE responsibility, what is a T&S responsibility).	- Handled 50Hz on M1M3. - Doesn't matter whether CSCs are using SalObj; everything using SAL can be ingested. - Is the EFD "reliable". - It's been live for some time now. - But not using versioned schemas since T&S can't currently support those. - Some issues with T&S provided timestamps. - Occasional downtime, but no data is lost. "Construction era production" quality. - Can we use the EFD to recreate SAL messages, enhancing the reliability of the SAL system? (I don't follow the technical details of this) "Would not be the weakest link in the chain". - This at least seems worthy of investigation; more effort would be necessary to determine if it's really practical. - InfluxDB 2.0 will change the way annotations are handled. The SQuaRE team are engaged with the InfluxDB authors, attempting to get them to provide an API for adding annotations. "Measurements" in InfluxDB are "topics" in SAL/DDS/Kafka. "Dead man's switch" sends an alarm when no data is received. - This should also be implemented in the Watcher. - What's the path for making this available to "the rest of us"? - When it is replicated to the LDF. Don't want everybody hitting the deployment in the lab. - A few weeks away from exposing data from the summit to the LSP. ☑ Frossie Economou — discuss with Russell Owen & Tiago Ribeiro what service monitoring is being carried out by the Watcher and make sure functionality isn't being duplicated. 18 Nov 2019
17:00	Close			
Day 2: 2019-10-29				

09:00	Networks Status and Planning	Jeff Kantor	- Brief overview of Summit, Summit — Base, and Base — LDF Network status and schedule - Update on networks Verification and Validation plan - Q&A	- Shared Chicago-Atlanta is FY20 only; will become dedicated next FY. - Jacksonville–Atlanta shared link will remain shared through construction; become dedicated in Ops. - Can use secondary paths for anything, but expect they are chiefly there for failover from the primaries. - V&V covers networks on the summit, which are T&S deliverables, as well as DM networks. - If we run AuxTel, can we guarantee networks uptime? - Yes. By June. - Modulo power issues on the summit & other telescope AIV activities causing problems. 80% for general internet & 90%+ for image data. - Failover to microwaves from the summit is working 9 out of 10 times. - There will be an IP address reassignment sometime before ComCam observing, but there is no date for this yet.
09:30	Ops Rehearsal #2	Robert Gruendl	- While a basic outline has been set the devil is in the details and these need to be discussed as they will set a timeframe where this rehearsal can occur - Also need to discuss whether or not verification tests should be convolved with the rehearsal	- Expect OR#2 to run for ~1 week. - Timescale under debate, depending on what instrumentation is available and what is actually required. - The primary aim is to exercise people, not hardware. - When ComCam is on the summit, all data should be transported to NCSA. - Discussion about whether the ops rehearsals should focus on hardware and service delivery and integration events, or on training the ops team. - We should engqge with the SIT-COM team to determine how they can be involved in using the Ops Rehearsals to demonstrate successful functioning of the observatory, rather than just successful function of the operations team. However, this may be part of OR#3, rather than #2. - OR#2 will be based on AuxTel, in February, processing will be based on the summit. - Robert Gruendl is empowered to ask other people for help in fleshing out the documentation for the OR. ☑ Wil O'Mullane — coordinate with T/CAMs and Bob Blum to identify staff who will be involved in the operations rehearsal. 18 Nov 2019 Set up on Operations Rehersal#2
10:30	Break (Refreshments provided)			
11:00	Product tree	Unknown User (gcomoretto)	Overview of DMTN-104, the "extended" version of the product tree.	- Agreed that this document should be an LDM-level document, and hence will be reviewed by the DM-CCB when it is ready. - Currently document is still in draft, and it is not yet ready for wide review. The Architecture team will call on other members of the project to review when they are ready. - Document should be done in 6 months to ensure that it is ready for use in reviews next summer. ☑ Unknown User (gcomoretto) & Architecture team — ensure a version of the detailed DM product tree (LDM-ized version of DMTN-104) is baselined. 28 Apr 2020

11:15	Image Display Working Group	Yusra AlSayyad	• Summary of WG results and timeline towards final report.	• Document is still in progress because writing sucks! • HSC quick-look tool has not been evaluated. Details on this seem to be scarce. ◦ But it does pre-process the data. • Concerns raised about scalability of Camera Image Viewer to widespread use by DM team. • Is it necessary to have both DS9 and JS9? • There is a programmatic interface to JS9 through a node.js backend.
11:45	Lunch (**not** provided — SLAC cafeteria, use your per diem) Visit to the Camera			

13:45	Configuration Management	Wil O'Mullane	• We are already at the stage that we are deploying real, operational services; over the next few years, both the number of capabilities being deployed and the number of users will increase substantially.
			• We have paid lip-service to the idea of configuration management, but we've not taken many concrete actions.
			• How do we arrive at a system that satisfies our need for rapid development and deployment while provide an adequate level of configuration control?
				• "Configuration management lets us know what we have; configuration control lets us know when it changes".
				• Refer to SQR-035.
				  ◦ Frossie suggests this describes tooling to implement whatever configuration control procedure is required.
				• Lacking resources to properly maintain two fully functional LSP instances, one for the "neophiles" and one for the "stabilityphiles".
				  ◦ It may be possible to deploy new *stack* features, without deploying new *Jupyterl ab* features.
				  ◦ There may be only a few people who really need new Jupyterlab features.
				• There should be a controlled update cycle to lsp-stable, driven by need and signed off by representatives of the users.
				• There is currently little downtime on lsp-stable. The issues more seem to be related to robustness of the service as a whole, rather than configuration control.
				• The SQR-035 "principles" could be applied to Pipelines, if the workflow system was ready to deploy them from Docker containers.
				• The release management process is the *input* to the above process: somebody needs to determine what changes go into new containers.
				• How do we drive the technology/process suggested by SQR-035 into a unified, cross-LSST system?
				  ◦ SQuaRE have engaged with T&S, but they are struggling with the release management process.
				  ◦ There is work ongoing there.
				  ◦ Wil O'Mullane and Frossie Economou are responsible for making sure T&S don't create a parallel system.
				• The DMLT seems happy with the above technology stack, but it's not clear at what level the DM-CCB (or some other body) will actually sign off on which changes.
				• Discussion of whether a "canary" model can apply to DM services; it's not clear we have enough users to make this worthwhile.
				  ◦ Perhaps this depends which of the various DM services we care about.
				• Assertion that the DMLT "doesn't care" about control of stack containers going to the LSP; all that matters is about the containers defining the Juypyterlab service.
				• Frossie is skeptical that the CCB-as-gatekeeper would add to the process that she already goes through; Wil reckons that the point is a sanity check on Frossie's decision.
				• Worry that LSP developers are feeling pressure to support services when they break; more configuration control might help with this. Frossie suggests the SQR-035 model will help address this.
				• Conclusion is that Frossie Economou will remain in her current role of "gatekeeper", with no DM-CCB or other direct oversight, until the SQR-035 plan has been fully implemented. At that point, we should review.
				  ◦ And that there should be a more rigorous system for managing access to commissioning data, possibly involving input from Bob Blum.

				☑ Wil O'Mullane — ensure that a standardized deployment mechanism is documented and required across subsystems. 25 Nov 2019 this is now in **DM-22416** - Getting issue details... STATUS ☑ Frossie Economou — develop configuration management systems based on SQR-035, and report on progress at the next DMLT F2F meeting. 24 Feb 2020 ☑ Kian-Tat Lim — ensure that A&A systems are managed following the SQR-035 plan. 25 Nov 2019 **DM-22368** - Getting issue details... STATUS
		Frossie Economou	• Configuration Management and Deployment Infrastructure for Kubernetes-based services (sqr-035 teaser) • What model is appropriate to our current level of development and mode of operation (priorities, userbase) • LSP-specific issues in the imperative->declarative transition	
15:00	Conda packaging updates	Kian-Tat Lim	• We've been hearing for a long time about plans to move third party packages to Conda, to adopt a Conda-based toolchain, etc. Let's have a summary of the work which has been performed to date, and a summary of and timeline for future plans.	• All LSST patches, modulo pytest-flake8 and eigen, are no longer necessary. • This proposal would require LSST to set up and host in perpetuity a web-facing Conda channel. • Conda development patterns: "we do not (yet) understand what it would do to people's everyday lives". • When can we switch all the third parties to Conda packages? As soon as the scipipe_conda_env becomes an EUPS package. There is currently no timescale. • It is not clear which group would be responsible for maintaining a Conda channel (and, indeed, SQuaRE would like to get out from under supporting stack builds in general). ☑ Wil O'Mullane — clarify who is responsible for developing **and maintaining** services in support of stack build and deployment. 25 Nov 2019
15:30	Break (Refreshments Provided)			

| 16:00 | Size and cost model updates | Kian-Tat Lim / Unknown User (mbutler) | • Status update on the work to produce revised (and simplified!) sizing and cost models. | • New model is not yet done, but is ready for a status update.
• No compute is currently reserved for staff (ie, for ad hoc QA, etc).
• Currently assumes 2 months of LSSTCam in FY22, then full operations from FY23 onwards.
• Consider making "additional DRP steps" parameter into separate per visit & per object steps.
• Model does not currently include daytime solar system processing; zeroth-order assumption is that daytime processing can simply use the (idle) AP infrastructure (but these have not been shown to match up).
• Model assumes full AP in LOY1. But this is a relatively small part of the compute budget.
• Total spend to end of FY23 is around double what the initial estimate of $14M.
• **BUT we should not get hung up on these numbers for now** — there are still huge uncertainties in this process, which need to be resolved quickly.
• This model does not yet account for two data release productions in LOY1.
• And should account for 10% of storage for users (as well as 10% of compute).
• And does not yet account for IN2P3.
• DMTN-135 is not yet complete, but it is ready for comments on the text from DMLT members.

☑ Leanne Guy & the DM-SST — Compare contents of LSE-81/82 (science inputs to sizing) with results from the HSC processing (NB this is also a risk mitigation). 25 Nov 2019

Ticket
DM-22082 - Getting issue details...
[STATUS] |
| 17:00 | Close [Adjourn to Alpine Inn beer garden, 3915 Alpine Road, Portola Valley for dinner and refreshments] | | | |

<div align="center">

Day 3: 2019-10-30

</div>

09:00	Current status of the Prompt Products Database	Fritz Mueller / Colin Slater	• How close are we to hitting (or missing?!) our performance requirements?
			• What investigations have been performed? What's our currently-favoured solution?
			• What is the timeline for this converging?
			• ...and what's the latest on plans for public PPDB releases, renaming to the APDB, etc?
			  ○ This was proposed during the June Community Broker Workshop, but I'm not aware of any follow-through.
			• Slides
				• Metric is time taken to select all DIASources, DIAObjects, DIAForcedSources for a visit.
				• RDBMS (both Oracle and Postgres) performance is off by a factor of ~ a few to meet requirements.
				• Cassandra TBD; making back-up plans in case it doesn't get us there.
				• Potential technical mitigations around avoiding reading the object history by querying the database, e.g. by precomputing and storing in a blob.
				  ○ Could break the problem into a "quickly changing part" and a "slowly changing part"; would enable using Qserv, filesystem, etc for slowly varying part.
				• Bottleneck is predominantly IO rather than indexing, but fundamentally is a combination of factors.
				  ○ Cassandra addresses this by using multi-node parallelism, clustering, smarter caching.
				• There is a spatial indexing package for Oracle, but not clear it buys us anything over the existing HTM indexing.
				• Would it help to reduce the width of the DIASource table?
				  ○ Unclear; would need experimentation.
				• Relaxing latency requirement would not help throughput issues.
				  ○ (But there doesn't seem to be huge pressure to keep the 60s requirement)
				• What is necessary to adopt the naming scheme?
				  ○ Update the product tree
				  ○ Update the glossary
				  ○ Rename the dax_ppdb repository, and any related code artifacts
				
				☑ Fritz Mueller — Report on APDB on Cassandra progress at the February DMLT vF2F. 24 Feb 2020
				
				☑ Fritz Mueller — Update the glossary, product tree, and code to reflect proper nomenclature for the AP/PP DBs. 25 Nov 2019

09:30	F19 retrospective / S20 plans	John Swinbank	Each team to give a brief (~15 minute) overview of: • Highlights from F19 (only a couple of slides) • Plans for S20 In this order: • Architecture (Kian-Tat Lim) • DM Science (Leanne Guy) • Alert Production (John Swinbank) • Data Release Production (Yusra AlSayyad) • DAX (Fritz Mueller) • Data Facility (Unknown User (mbutler)) • SQuaRE (Frossie Economou) (We assume networks are adequately covered by Jeff Kantor's talk the previous daty	☑ John Swinbank — convene a telecon of at least Robert Lupton, Yusra AlSayyad, Merlin Fisher-Levine to plan for DM-Camera interactions in S20. 25 Nov 2019
10:30	Break (Refreshments Provided)			
11:00	Plans for S20 (continued)	John Swinbank	Discuss outstanding issues, refine timelines, resolve cross-team issues.	

11:30	Review action items & plans for future meeting	Wil O'Mullane	- Next meeting is a Virtual F2F, 24–27 February 2020 - Other meetings in 2020: - Seattle, 11-14 May 2020 - Tucson, 9–12 November 2020 - Consider a F2F in La Serena in 2021.	- Please remember to add your slides to this page! - Suggestion to augment DMLT meetings with more frequent, more focused topical discussions. - Provenance WG good to go based on draft charge; Wil will make it official shortly. - Please ensure construction papers are included in cycle planning. - Should go ahead and book a room in Seattle for the May 2020 UW, although we might later release it if we don't feel there is a pressing need for the meeting. - Hoping for a JTM in 2021 in La Serena, TBD.
12:00	Lunch (Provided)			
12:30	SST meeting. Folks not involved with the SST are free to leave.			

Attached Documents

File	Modified
JPEG File alpineinn.jpeg	Oct 10, 2019 by Fritz Mueller
Microsoft Powerpoint Presentation current %22re%22processing at NCSA.pptx	Oct 28, 2019 by mbutler
Microsoft Powerpoint Presentation current reprocessing at NCSA final.pptx	Oct 28, 2019 by mbutler
PDF File LDM-692 Verification Control Document DMLT-F2F Oct 2019.pdf	Oct 28, 2019 by gcomoretto
Microsoft Powerpoint Presentation workflow environment .pptx	Oct 28, 2019 by mbutler
Microsoft Powerpoint Presentation Kantor Networks.pptx	Oct 28, 2019 by Jeff Kantor
PDF File DMTN-104 Detailed Product Tree DMLT-F2F Oct 2019.pdf	Oct 29, 2019 by gcomoretto
PDF File DMLT-F2F-20191029-OPSRehearsals.pdf	Oct 29, 2019 by Robert Gruendl
JPEG File DMTL-SLAC-OCT-2019.jpg	Oct 29, 2019 by Wil O'Mullane
File Arch S20 Plans.key	Oct 30, 2019 by Kian-Tat Lim
File New Sizing Model.key	Oct 30, 2019 by Kian-Tat Lim
File Conda Packaging Status and Plans.key	Oct 30, 2019 by Kian-Tat Lim
File DMTN-111 Status and Plans.key	Oct 30, 2019 by Kian-Tat Lim
PDF File SST Plans DMLT-F2F 2019-10-28.pdf	Oct 30, 2019 by Leanne Guy
PDF File 2019-10-30 — AP S20.pdf	Oct 30, 2019 by John Swinbank
JPEG File DMTL-SLAC-OCT-2019b.jpg	Oct 30, 2019 by Wil O'Mullane
Microsoft Powerpoint Presentation next 3 months [Autosaved].pptx	Nov 13, 2019 by mbutler
PDF File Middleware Staffing, 2019-10-28.pdf	Nov 13, 2019 by Fritz Mueller
PDF File Middleware Status, 2019-10-28.pdf	Nov 13, 2019 by Fritz Mueller
PDF File DAX Plans, 2019-10-30.pdf	Nov 13, 2019 by Fritz Mueller
PDF File PPDB Status, 2019-10-30.pdf	Nov 13, 2019 by Fritz Mueller
PDF File dmlt_2019_10.pdf SQuaRE slides	Nov 13, 2019 by Frossie Economou

Download All

Action Item Summary

Task report

Looking good, no incomplete tasks.

Pre-Meeting Planning

Topic	Requested by	Time required (estimate)	Notes
Verification Control Document	Unknown User (gcomoretto)	15 minutes	overview of LDM-692
Product Tree	Unknown User (gcomoretto)	15 minutes	overview of DMTN-104, as an extended version of the product tree
Middleware development status	John Swinbank	30 minutes	How close are we to "one Butler for Christmas"?
Middleware staffing transition	John Swinbank	30 minutes	Who will be working on middleware during 2020? How can we free up folks — in particular Jim Bosch — to focus on other tasks? Fritz Muellerhas agreed to come up with a plan which he will discuss at this meeting.
Status update and timeline for SDM standardization	Leanne Guy		What is the status of SDM standardization (previously DPDD-ification) and having the outputs of HSC reprocessing in Parquet/Qserv and queryable via the LSP?
Automation of HSC (and other) reprocessing	Leanne Guy		The current monthly reprocessing of HSC data is still a very manual process run by Hsin-Fang. As we move towards the end of construction, running these re-processings more frequently and on different datasets is essential to understand the performance of the pipelines. This will not happen unless we automate the process.
Image Display WG	Leanne Guy		I'd be interested to hear from the image display WG, **if** they have anything to report yet?
Regular processing at the LDF	John Swinbank	30 minutes	For the last few years, Hsin-Fang has provided an invaluable service to the DRP team by regularly reprocessing the HSC RC2 dataset every few weeks (initially fortnightly, currently monthly) and reporting issues. As Hsin-Fang has on from NCSA, and as we move closer to commissioning/science validation, we should review whether this is still the most effective way to proceed. Specifically: - How much of a resource drain is this on NCSA? - Can it be automated (see also the discussion topic above)? - Which other datasets need to be reprocessed? - DESC DC2? - Something for AP? - Something for Science Verification & Validation? - How should issues be flagged / addressed?
Networks Status and Planning	Jeff Kantor	30 minutes	- Brief overview of Summit, Summit - Base, and Base - LDF Network status and schedule - Update on networks Verification and Validation plan - Q&A
What's the story with workflow?	Leanne Guy		Where are we with workflow management?
Test Datasets for Scientific Performance Monitoring	Leanne Guy	30 mins	I will present the SST proposal for Test Datasets for Scientific Performance Monitoring
Third party packages in Conda	Leanne Guy / John Swinbank	15 mins	What's the current status of third party packages in Conda? What are the outstanding issues? When do we get to use it?
OPS Rehearsal #2	Robert Gruendl	30 minutes	- While a basic outline has been set the devil is in the details and these need to be discussed as they will set a timeframe where this rehearsal can occur - Also need to discuss whether or not verification tests should be convolved with the rehearsal
Influx/EFD	Frossie Economou	30 minurtes	Demo with Questions.

Commissioning (and Control) Dataflow and Processing	Kian-Tat Lim	30 minutes	Review of DMTN-111 highlighting missing pieces; confirmation of timelines for delivery of them.
T/CAM Plans for Next Cycle	Kian-Tat Lim	Wed AM?	Traditional sketch of what is to come
Current status of the PPDB	Leanne Guy / John Swinbank	30 mins	We know a traditional RDBMS has been investigated, and there is work ongoing with Cassandra.. but what's the current status? When do we expect this to converge? What's the risk that we will simply be unable to hit performance targets?