2017-12-04 to -08 Science Platform Detailed Design and Engineering Workshop

Attendees

1. Gregory Dubois-Felsmann (local)
2. Unknown User (xiuqin) (local)
3. David Shupe (local)
4. Trey Roby (local)
5. Unknown User (cwang) (local)
6. Loi Ly (local)
7. Tatiana Goldina
8. Fritz Mueller
9. Frossie Economou
10. Simon Krughoff (split with DESC?)
11. Adam Thornton
12. Angelo Fausti
13. Brian Van Klaveren
14. Unknown User (npease) (Tuesday through Thursday)
15. Kian-Tat Lim
16. Kenny Lo
17. Igor Gaponenko
18. Vaikunth Thukral
19. Andy Salnikov
20. John Gates (Monday through Wednesday)
21. Unknown User (cbanek) (in-person), Unknown User (jmatt) / Unknown User (bemmons) (remote)
22. Unknown User (awithers) (probably Monday through Wednesday)
23. Steve Pietrowicz (Monday through Thursday)
24. Arfon Smith (STScI Monday through Wednesday ?) dial in potentially Mike and Iva
25. Colin Slater
26. ? (please fill in)

Remote attendance will be possible but in-person participation will be strongly encouraged.

Remote access

We will use SUIT BlueJeans for this conference.

IP: 199.48.152.152

code: 319856717

URL: https://bluejeans.com/319856717/

Block Agenda

Note: there are many questions called out below as needing answers. We will try to pre-discuss a number of them and present a proposed answer at the workshop for confirmation. We don't expect to discuss everything "live" at length.

We are currently planning for relatively long breaks, to encourage spontaneous followup discussions, and we are not planning "working lunches" in a meeting of this many days - we hope stretching our legs will help us think better!

Monday 04 Dec 2017

Time	Location	Topic

Time	Location	Topic
13:30-15:00	MR102	**Review basic documentation** (not fully approved after the DM review)

13:30-14:30	770 S Wilson Ave., Pasadena	(1) LSP requirements document LDM-554:
		
		• Incorporate API Aspect requirements
		   ○ Requires pre-meeting preparation
		• Incorporate Notebook Aspect requirements
		   ○ Mostly done but not folded in yet? Pre-meeting preparation
		• Identify work required on performance requirements - expected to be needed to drive testing
		• Identify any missing items and associated groups

14:30-15:00		
15:00-15:30		break
15:30-17:00	MR102	Continued discussion on requirements
15:30-16:00		
16:00-16:20		
16:20-16:40		
16:40-17:00		

Tuesday 05 Dec 2017

Time	Location	Topic
08:30-10:00	KS410	**Summary reports from the Aspect groups** (1) Current development, test, and deployment plan ("d.t.d.") for SUIT / Portal Aspect (2) Current d.t.d. plan for JupyterHub/Lab / Jellybean / Notebook Aspect (3) Current d.t.d. plan for Data Access / API Aspect (These are meant to be "as of start of workshop" and it is definitely OK for them to be modified by what is discussed / decided at the workshop. All reporters are encouraged to comment on areas where they think there is a lack of clarity on scope or on assignment of scope to a group. All reporters are encouraged to comment on any substantial technology choices or design decisions that remain to be made.)
10:00-10:30		break

10:30-12:00	KS410	**LSP design discussion: connection among three aspects**
		
		• workspace interfaces in each Aspect
		   ○ "file workspace", i.e., VOSpace/WebDAV (confirm development responsibilities)
		   ○ "database workspace" / "MyDB"
		   ○ How are these accessed in each Aspect? E.g., is the VOSpace workspace mounted as a filesystem visible in the Notebook Aspect?
		• cross-aspect connections for queries:
		   ○ by user (i.e., by being able to ask the API aspect for "my previously run queries")
		   ○ by query ID returned from an asynchronous query request
		   ○ UI actions for transferring queries between Portal and Notebook
		   ○ use of the workspace for query results

12:00-13:30		Lunch on our own (campus cafeteria or Lake Ave. restaurants)

13:30-15:00	KS410	**Authentication and authorization**
		
		• login, token passing across Aspects
		• granularity of access rights
		  ○ This is about the "toolkit" we give to the operations team, not about our making policy decisions in this workshop.
		  ○ Should we implement separate rights for each Aspect (e.g., "may use the Notebook Aspect")? This may be useful for managing abusive users.
		  ○ Should we plan for the possibility of data rights separated by Data Release? (E.g., if you had rights to DR3 because you were a grad student of someone with data rights, and then you move back to your home country, can you retain DR3 access for some time, perhaps while you finish a paper, without being granted DR4 rights?)
		  ○ What is the granularity of control we give to users to manage access to user-created ("Level 3") data?
		• what is the programming model for testing whether a user has a specific right? Just by attempting the operation desired, or is a "pre-verification" possible?
		• user management
		  ○ What is the API and/or UI for self-service group management? (E.g., for users to create collaborations that can have rights?)
		• user profiles
		  ○ Quota management
		  ○ Does NCSA provide a flexible profile service (e.g. a key-value store for each user)?

15:00-15:30		break

15:30-17:00	KS410	**LSP architecture and deployment**
		
		• Complete discussion of the role of each of the five (six?) LSP instances
		  ○ rough feature-deployment schedule and cycle for all the LSP instances
		  ○ when / where is the first integration of the Notebook Aspect with the others (on PDAC?)
		  ○ 2018 priority: Science Validation LSP deployment planning: schedule, expectations, datasets, etc.
		• NCSA hardware plans
		• deployment architecture: K8s, docker
		• Plan for completion of documentation of full LSP deployment architecture
		  ○ Define all sub-components and identify internal interfaces
		  ○ Identify specific architectural issues for follow-up
		  ○ SysML?

Wednesday 06 Dec 2017

Time	Location	Topic	Possible breakout
08:30-10:00	KS410	**API Design and AAIM discussion** • Authentication and security for all exposed APIs ○ API for workspace(VOSpace?) access, including security token pass ○ API for "next to DB" data processing ○ How does each Aspect run code as the logged-in user? (I.e., how do the Aspect do "setuid-like" things when needed?) A/k/a "userid mapping". Do all users actually have separate NCSA Unix/NIS-type identities? ○ Programming model for 3rd-party tool developers (e.g., TOPCAT) (time permitting) • AAIM team delivery schedule (Unknown User (awithers) leaves at the end of this session)	
10:00-10:30		break	
10:30-12:00	KS410	**Deployment issues** • Needs for each aspect: hardware, system access • Commissioning Cluster LSP issues ○ When will the computer room be ready? What are the plans for installation of hardware? ○ What are the first things its users will expect to be able to do? ○ Database and DAX planning for the CC deployment ○ Access to EFD data in the Commissioning Cluster, relationship with EFD-reformatter service ■ Is access to the "raw" EFD required? If so, how? ■ What is the DAX interface to the EFD? Where does the table and column metadata come from? • Cross-aspect access to computing resources (e.g., for Portal extensions that need to run Python code as the user)	
12:00-13:30	KS410	Lunch on our own (campus cafeteria or Lake Ave. restaurants)	

Time	Location	Topic	

13:30-15:00	KS410	**LSP Data Model**
		
		• How do Science Platform outputs (e.g., `afw.table` FITS files) become data visible in the LSP?
		  ◦ Is this round-trippable? What is the Python API for access to the DPDD data products in the Notebook Aspect? Is round-tripping only possible in the Notebook Aspect when run inside a DAC?
		  ◦ What is the life cycle of data model metadata (e.g., column metadata like units, UCDs, VO-DML descriptions)? Is it created "at birth" in the Science Pipelines code? Or at ingest? How do database columns get their released names? Are these all in the table-creation code in `afw`?
		• **[RFC-243](#)** - Getting issue details... `STATUS` - when will we start getting deliberate prototypes of the actual DPDD data products?
		• Butlers in the Notebook Aspect and in Portal extensions
		  ◦ How do users get Butlers that provide access to the released data products?
		  ◦ How does a Butler user follow the release of new Level 1 / nightly data?
		• Support for "older releases" (see LCR-908) (probably will not get to this this time)	

15:00-15:30		break	

15:30-17:00	KS410	**Detailed DAX API discussion**
		
		• What is the full envelope of VO interfaces we'll provide? What interfaces do we need for which no satisfactory VO interface exists?
		  ◦ ADQL support level. Any limitations?
		  ◦ VOSpace. Any limitations? Also WebDAV? Third-party transfers?
		• extra API or special DB access needed by portal (besides metaServ, DBServ, and ImgServ), for example certain data exploration flow sequence,
			K8s and notebook deployment

Thursday 07 Dec 2017

Time	Location	Topic	Possible breakout
08:30-10:00	KS410	**Further discussion of deployment** • K8s, docker, more details • What is the batch / parallel computing model that will be exposed to Notebook users? • Next-to-DB processing architecture	Meta data for data holdings (continued from Data Model session)
10:00-10:30		break	

10:30-12:00	KS410	Detailed schedule and planning
		
		• Role of LSP in operations rehearsals
		  ◦ Need for simulations of real LSST datasets
		  ◦ Successively approach real data model
		• Generally review datasets to be handled (HSC, ZTF(?), Gaia, LSST-CatSim, LSST-PhoSim)
		  ◦ for PDAC, SV
		• Operations concept for transition of a data release from production through science validation to public release
		
		  ◦ How does the SV environment for DR(N+1) get access to the released data for DR(N)?
			Table data formats:
			
			• Streaming-compatible formats?
			• Support for very large tabular results
			• VOTable usage - do we want the header but send the body a different way?
			• Options previously mentioned: FITS binary table, SQLite files

12:00-13:30		Lunch on our own (campus cafeteria or Lake Ave. restaurants)	

13:30-15:00	KS410	**Review basic documentation & group summaries**
		
		LSP design document LDM-542:
		
		• Agree on a framework for detailed design documentation (in LDM-542 or in subsidiary documents)?
		• Confirm the list of instances of LSP:
		  ◦ Original 5: integration/PDAC, science validation, comm'g cluster, US DAC, Chilean DAC
		  ◦ Do we need a "development environment" instance?
		
		LSP verification and test plan
		
		• Aspect-level verification and testing
		• Integrated test of the LSP environment
		  ◦ Distinguish "verification" from "user testing" - both are needed
		  ◦ Annual "user testing" plans
		• Discuss needs for datasets for verification and testing
		  ◦ HSC, Gaia, LSST-Phosim, LSST-Catsim? ZTF (data rights?)?
		• Review all existing LSP Level 2 and Level 3 milestones	

Time		Topic	
15:00-15:30		break	
15:30-17:00	KS410	To be identified by previous discussion	

Friday 08 Dec 2017

Time	Location	Topic
08:30-10:00	KS410	Session 5A Wrap-up of long-range planning
10:00-10:30		break
10:30-12:30	KS410	Session 5B Wrap-up of short-range planning and action items
12:30-14:00	Capital Seafood, Arcadia	Dim-Sum lunch. Please let Xiuqin know if you would like to come.

Documents

- LSP Requirements: LDM-554
- LSP Design: LDM-542
- Data Access White Paper: Document-5373
- For reference: original text file with agenda contents draft for this workshop

Logistics

Hotel

- Accommodations have been secured at: The Hilton Pasadena, 168 South Robles Avenue, Pasadena, CA 91101   Tel: 626-577-1000
- **Travelers are responsible for making their hotel reservations on-line using this link:** http://www.hilton.com/en/hi/groups/personalized/P/PASPHHF-LSST12-20171204/index.jhtml?WT.mc_id=POG
- **Reservations must be made by November 24 to ensure the negotiated rate of $173**
- Group Name is: LSST Science Platform Detailed Design and Engineering Works

Nearby restaurants

For lunch within walking distance: (of course Yelp provides more choices)

- On campus cafeteria: Chandler
- **Abricott**

  $$ Cafes, Asian Fusion, French
  238 S Lake Ave
  Pasadena, CA 91101
  Phone number (626) 796-1613

- **Mediterranean Cafe**

  $ Middle Eastern, Greek, Mediterranean
  273 Shoppers Ln
  Pasadena, CA 91101
  Phone number (626) 793-8844

- **Pie 'n Burger**

  $$ Burgers, Breakfast & Brunch,Sandwiches
  913 E California Blvd
  Pasadena, CA 91106Phone number (626) 795-1123

- **The Counter Pasadena**

  $$ Burgers
  140 Shoppers Ln
  Pasadena, CA 91101Phone number(626) 440-1008

- **Lemonade**

  $$ American (New), Sandwiches

146 S Lake Ave
Pasadena, CA 91101Phone number(626) 304-7700

- **Noda Sushi**

  $$ Sushi Bars
  546 S Lake Ave
  Pasadena, CA 91101Phone number(626) 793-2600

- **Veggie Grill**

  $$ Vegetarian, Vegan
  200 S Lake Ave
  Pasadena, CA 91101Phone number(626) 535-9561

- **Wahoo's Fish Tacos**

  $ Seafood, Mexican
  264 S Lake Ave
  Pasadena, CA 91101Phone number(626) 449-2005

- **Chipotle Mexican Grill**

  $ Mexican, Fast Food
  246 S Lake Ave
  Pasadena, CA 91106
  Phone number(626) 229-9173