PDAC Meeting 2018-04-05

Date

05 Apr 2018

Attendees

- Gregory Dubois-Felsmann
- Kian-Tat Lim
- Fritz Mueller
- Simon Krughoff
- Fabio Hernandez
- Tatiana Goldina
- Unknown User (xiuqin)
- Unknown User (mbutler)
- Unknown User (npease)

Goals

- Advance the progress of the Prototype Data Access Center and the deployment of the Science Platform

Discussion items

Time	Item	Who	Notes
	Kubernetes Commons status	Unknown User (mbutler), Simon Krughoff, Gregory Dubois-Felsmann	• Kubernetes Commons is available for experimentation • Jellybean (Notebook Aspect) is working on the system, integrated with NCSA authentication • Only thing missing is persistent storage and integration with identity management. Looking at incorporating Samba-over-GPFS into the mapping of users' GPFS space into the Jellybean containers. This has been discussed on #dm-kubernetes as well as in private chats. ☑ Unknown User (mbutler) Make a short technical description of the planned GPFS/Samba/NFS architecture available for the next meeting. 18 Apr 2018 • Plan for user home directories within the Notebook Aspect is evolving. Current version is as shown here (and later updated here) on #dm-infrastructure: ◦ LSST staff and other "internal" users have a "traditional" home directory for sessions reached via ssh to, e.g., lsst-dev. This will not be an entitlement of general LSST science users. ◦ All of a user's Notebook Aspect sessions' Unix processes (e.g., the JupyterLab Python kernel, or JupyterLab terminal sessions) will share a separate "Jupyter Home" directory. ▪ For internal users, this directory will be sym-linked/linkable as a subdirectory of their "traditional" home directory. ▪ (Gregory Dubois-Felsmann comment post-meeting) The same home directory should probably be used for the user-specific "Firefly Python micro-services" envisioned in the LSP design. ◦ The "user file workspace" available for a user via VOSpace (and possibly also WebDAV) is yet another directory tree. It is not the same as the "Jupyter home" directory, in order to avoid making possibly security-sensitive "."-files and directories (e.g., ".git") visible via VOSpace and providing an additional attack surface. The user's workspace will be a (possibly symlinked) subtree of their "Jupyter Home" directory. • A future meeting will need to address the details of the plan for deploying databases, DAX services, and Portal services in the SV environment. There is no immediate plan for a Qserv deployment in the SV environment.
	Expanding focus of PDAC meetings to cover all the development LSP deployments	Gregory Dubois-Felsmann, all	• Motivated by the above discussion, we discussed whether we should formally accept the *de facto* notion that the "PDAC meetings" should expand to cover at least the Science Validation LSP deployment as well. ◦ Makes sense because there will be a lot in common between deployments, particularly regarding Kubernetes configuration, etc. ◦ General agreement that this makes sense. ☑ Gregory Dubois-Felsmann Describe proposed modification of PDAC meeting series scope at the 09 Apr 2018 DMLT meeting.

	Kubernetes conversion of PDAC	Fritz Mueller, Unknown User (mbutler)	• Initial experimentation has turned up some configuration changes required for Docker, etc. • Fritz Mueller has proposed a set of modifications and emailed them to Unknown User (mbutler) for implementation on the DAX node in PDAC for initial testing. The changes will be made the week of 09 Apr 2018 and then evaluated, which will take a few days. ☐ Unknown User (mbutler) Make a (DM?) ticket for the PDAC DAX node configuration changes and post its ID to #dm-pdac. **10 Apr 2018** ◦ Once the DAX-node configuration is found acceptable for Kubernetes federation, Fritz Mueller will request its rollout to the rest of the PDAC cluster after coordination with IPAC. • Fritz Mueller has been tentatively planning to use the DAX node for the Kubernetes head node, but is not really happy with this idea; it is more standard in contemporary Kubernetes practice to have the head node not also used as a pod host. ◦ Others agreed that this doesn't seem like what we want; the DAX node is fairly heavily loaded, too. ◦ So we need another node, not previously foreseen in the PDAC provisioning. ◦ Some discussion that the Portal/SUIT load-balancer's node is over-provisioned for its function; could it be split into two VMs, one of which could be the Kubernetes head node? Also discussion about where the main ingress controller could/should run. The SUIT load-balancing node is a reasonable candidate. ▪ Decided to leave it to NCSA to figure out how to do this additional provisioning. Do we need a ticket for this?
	Local (Docker) registry for LSP at NCSA	Fritz Mueller, Unknown User (mbutler)	• Discussion of the status of planning for a local, NCSA-hosted registry for Docker images associated with the Science Platform deployments. ◦ NCSA is already planning to have a single local registry that is shared by the NCSA LSP deployments (i.e., in particular, by the PDAC and SV deployments). Hardware is already allocated. ◦ Planning to have the image storage on local disk. Currently considering whether this needs to be SSD or whether it can be conventional storage. Gregory Dubois-Felsmann expresses concern that it will be a dependency nexus for a large number of hosts in the LSP deployments and may be under very heavy and difficult-to-serialize load during startups / version changes.
	Addition of Notebook Aspect / Jellybean to PDAC	Simon Krughoff, all	• Initial integration of the Notebook Aspect into the PDAC LSP deployment is a specific goal of the current cycle, i.e., aimed at 31 May 2018. • Gregory Dubois-Felsmann Initially this doesn't need to provide a large-scale service. The May requirement is to demonstrate the functionality of accessing the other PDAC components from the Notebook Aspect and to enable work to start on feature-level integrations. • We need to determine what hardware will be used to support this. Discussion of whether the NCSA cloud infrastructure could be used for this. Should be possible as long as the "walled garden" constraints of PDAC can be met. • Computational requirements for the Jellybean service are on SQR-018. ☑ Gregory Dubois-Felsmann and Frossie Economou will discuss this and make a ticket with the request; we've been talking about this too informally for too long. 10 Apr 2018
	Authentication and security issues	Gregory Dubois-Felsmann	• It has long been desired to make the PDAC available directly on the public Internet - and this will also be applicable to the full SV deployment of the LSP. ◦ This depends on two pre-requisites, **both** of which must be satisfied: a. NCSA and the ISO must be satisfied that the exposed surface of the service (e.g., the Firefly server and the DAX VO services) have been reasonably vetted for security issues. b. The services must have login/authentication capabilities that allow them to be restricted to authorized users only. ◦ In the absence of these, the services must remain behind the NCSA VPN. • Getting this done is not a May milestone but should be completed before the planned science-user testing of PDAC in the fall. • Regarding a), a round of analysis has been done on Firefly and has resulted in some recommendations for changes to the Tomcat server configurations. IPAC has deferred this work until after the PDAC Kubernetes conversion, because it will very likely also involve issues such as the configuration of the ingress controller(s). • Regarding b), IPAC has done some initial work to test out the CILogon authentication mechanism and include a login mechanism in Firefly. However, this did not include the required ability to deny access to unauthorized individuals; this depends on the use of the group-membership service proposed by NCSA. ☑ Loi Ly will review the proposal for the group membership service and provide comments. 18 Apr 2018 ☑ Gregory Dubois-Felsmann link the NCSA group membership service proposal to this page. 10 Apr 2018 ◦ The other Aspects should review this as well so that NCSA's work can proceed. • This sort of work will be needed for the API Aspect / DAX services as well. • At the next PDAC meeting (19 Apr 2018) we will go over a plan for this, so that a properly resource-loaded plan to get this work done early in the next cycle will be ready for the May DMLT meeting.
	Agenda for next meeting		• Review plan for getting PDAC services on the public Internet • Clarify what is required to get Jellybean into PDAC by the end of May • Usual status checks

LSST IAM Group Naming Convention

Action items

☐