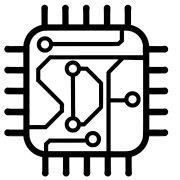


Unix Town Hall

Scientific Computing Services

August 27th, 2020

U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

Unix Town Hall Meeting

Objectives:

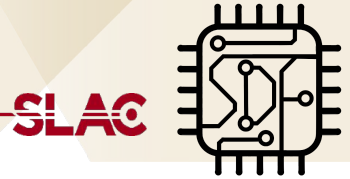
- Communication
- Collaboration

Join our mailing list: unix-community@slac.stanford.edu

email to: listserv@slac.stanford.edu

subscribe unix-community

Scientific Computing Services



unix-admin@slac.stanford.edu

support/questions

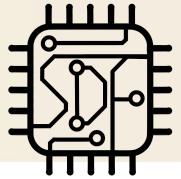
yemi@slac.stanford.edu

650-926-2863

Scientific Computing Services

- New production version of Confluence is coming in September
- Available for testing today:https://confluence-uat.slac.stanford.edu/

Scientific Computing Services



WE ARE HIRING!!

Job #4118: High Performance Computing Administrator - UNIX Clusters

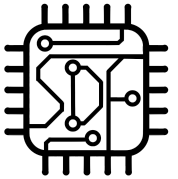
Job #3970: Unix Scientific Computing Specialist

Unix Town Hall Meeting

Agenda:

- Welcome our new CIO Jon Russell
- Partnership & Engagement: 12-month recap (Yemi)
- SDF Beta Environment (Yemi)
- Active Directory accounts for Identity Management (Karl)
- SDF User Experience (Yee)
- NERSC (Debbie)
- Intermission
- SDF Filesystem Update (Lance)
- Batch Compute Update (Yemi)
- Tape Storage Roadmap (Guangwei)
- Storage Update (Lance)
- CentOS / RHEL Platform Update (Karl)
- Next Steps for SDF (Yemi)
- Cyber Security (Olga)
- Questions/Discussion

Jon Russell
SLAC CIO



U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

A little about me…

- 20+ years experience in IT
- 6 Years in Stanford University IT
- 10 Years with the DOD
- 2nd Time as a CIO

SLAC IT Focus Areas

Improve the Customer Experience

Connect to the Research Mission

Improve Communication & Transparency

Build Strength through Partnerships

Modernize Infrastructure & Core Services

Issues of Potential Interest

- Improving the level of communication
- Identity and Access Management (IAM) Roadmap
- Operationalize Shared Data Facility (SDF)
- SRCF-II
- DOE GCP Contract
- Unix Account CS100 Requirement

Partnership & Engagement: 12-month recap

Yemi Adesanya

August 27th 2020, Unix Town Hall

U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

Partnership & Engagement: 12 month recap

Q4 FY19

- SDF core networking and storage components jointly funded by LCLS and SLAC IT
- Machine Learning funds GPU cluster (~$250K) along with additional storage capacity and performance (~$240K)

Q1 FY20

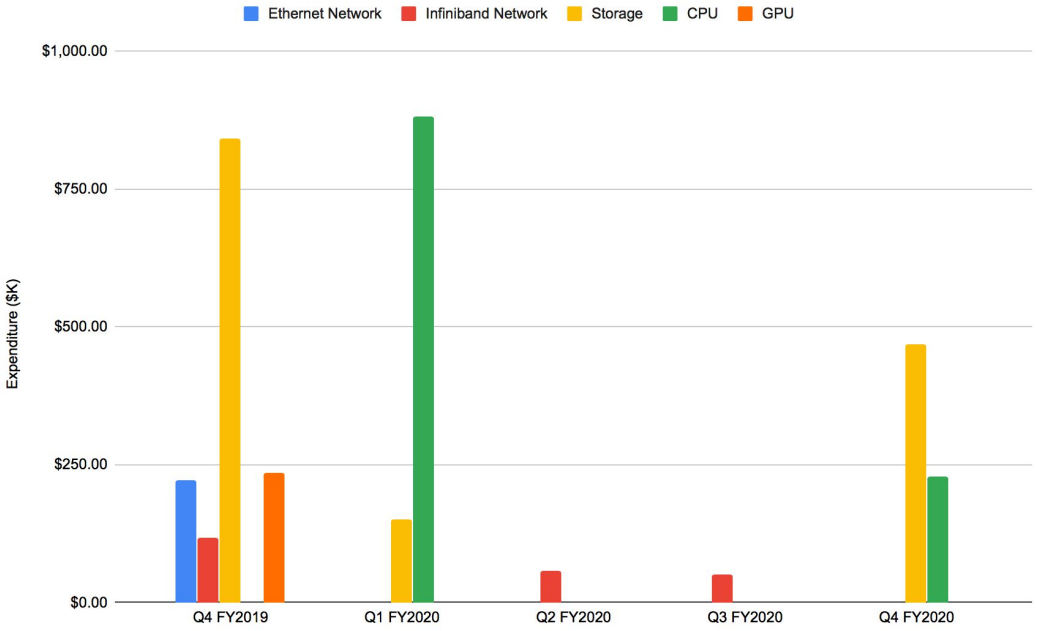
- CryoEM purchase drive capacity
- 11264-core AMD “Rome” CPU cluster purchase with funding from LCLS, HEP (Fermi, HPS, SuperCDMS), CryoEM, SUNCAT (BES)

Q2,Q3 FY20

- B050 Datacenter deployment with further investment from HEP and SUNCAT to complete the SDF networks
  - Jointly funded by LCLS, Machine Learning, SLAC IT.
  - *Special thanks to Networking and Datacenter Teams for working under challenging social distancing conditions* 🙏🙏

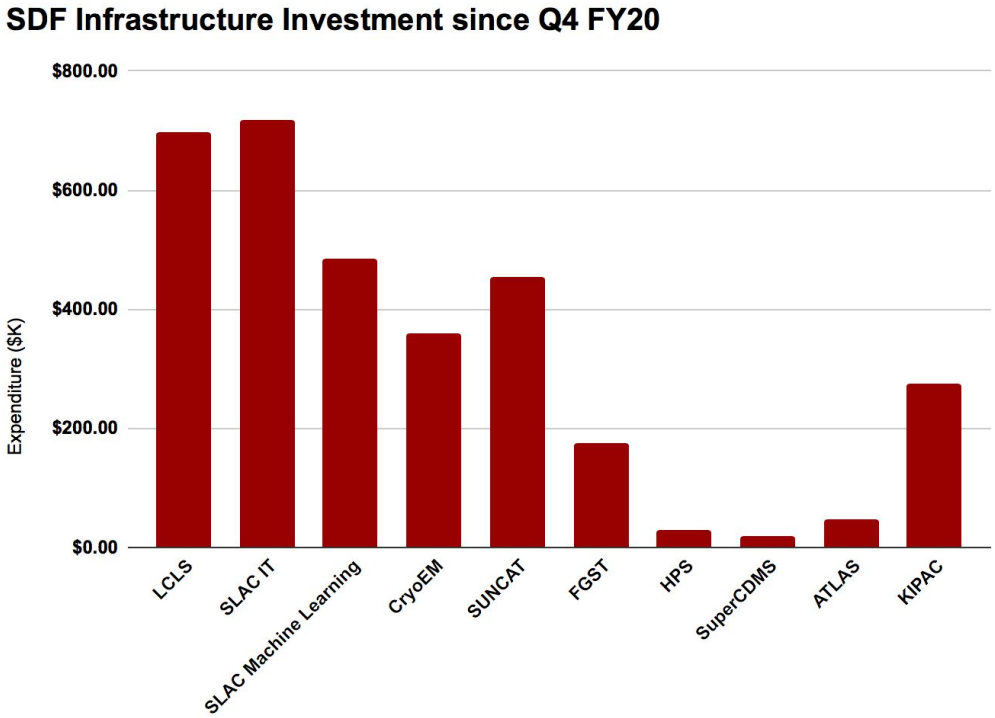
Q4 FY20

- Another major storage expansion purchase funded by ATLAS, CryoEM, KIPAC, SLAC IT
- Additional 2560 AMD cores for KIPAC



Partnership & Engagement: 12 month recap

- SLAC has committed more than $3.2M committed to SDF network,storage and compute since Q4 FY19
- SDF = “Shared”
  SDF != “Siloed”
- A single, integrated facility for data analytics
- Maximize utilization by leveraging idle cycles (opportunistic)
- Multi-tenant filesystems with scalable capacity and performance
- Drive for a sustainable Baseline with continual support and funding from SLAC IT (indirect).



SDF Beta Environment

Yemi Adesanya

August 27th 2020, Unix Town Hall

U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

SDF Beta Environment

- SDF is online, but we're not ready to declare full production status
- Beta testing group of users have started building analysis environments and submitting jobs

- Ongoing Datacenter work
  - Network, power distribution and load-testing

- Filesystems are mounted and DDN we've been performance tuning
  - Distributed Namespace, Data-on-Metadata, Progressive File Layout

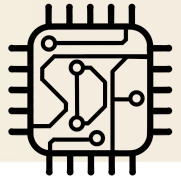
- Wanted: users who can withstand some rapid changes and provide feedback
  - Be prepared for unscheduled outages and configuration changes

- Why does SDF initially appear somewhat bare-bones?
  - This really is our attempt at a new compute facility for SLAC
  - The goal is to minimize dependencies on the existing Unix services
  - Add packages and services to the software stack "as needed" rather than pull in "everything" by default

SDF Beta Environment

- "/sdf" is the primary filesystem
  - All home directories under "/sdf/home/"
  - Shared project/group directories under "/sdf/group/"
  - Lustre project quotas applied to subdirectories
- Limited number of GPFS filesystems available
  - StaaS, Fermi, CryoEM
- No SLAC Automount map
  - The goal is eventual migration to native SDF storage
  - Contact unix-admin if you need help moving data to SDF
- "/sdf" is mounted on a select number of existing non-SDF systems
  - centos7*, rhel6-64*, bullet*, bubble*, kiso*, deft*
  - Helping users transition existing workloads to SDF compute environment
- Environment Modules interface for selecting installed software packages https://en.wikipedia.org/wiki/Environment_Modules_(software)
  - OpenMPI, FFTW, HDF5, etc
  - Provide access to the latest releases in the Red Hat Software Collections distro
- **SDF Beta is now open to everyone. There are no login restrictions. We welcome your input**

SDF Beta Environment



Questions?

Active Directory accounts for Identity Management

Karl Amrhein

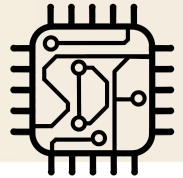
August 27th 2020, Unix Town Hall

U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

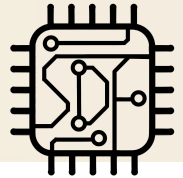
Active Directory accounts for Identity Management

SLAC



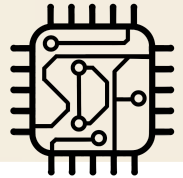
- For our discussion today, what is “Active Directory”, also known as, “AD”?
  - It is a Kerberos authentication server (username and password, is account disabled?, etc.)
  - It is an LDAP server (linux openldap tools can be used for queries, same way you query Unix LDAP).
    - An example of something stored in LDAP is: Unix POSIX groups (primary and supplementary).
- Why are we doing this?
  - Reduce duplication of effort - do we really need Kerberos and LDAP servers on Unix and Windows?
    - Managing parallel infrastructures for Kerberos and LDAP is expensive and time consuming, and this critical infrastructure takes extra special effort manage securely. “keys to the kingdom”
  - Moving towards a single SLAC Identity (AD) means one less account to manage; one less password.
    - Identity and Access Management (IAM) project goals include improved single sign on. Using a single account (AD, or “SLAC ID”) for authentication and authorization moves us in that direction.
  - SDF “greenfield” is a unique opportunity to modernize our authentication and authorization, and make some transformational changes. Using AD for authentication is a common, modern practice documented by Red Hat, and other Linux distributions.
    - https://access.redhat.com/documentation/en-us/red_hat_enterprise_linux/7/html/windows_integration_guide/
    - https://sssd.io/docs/users/ad_provider.html

Active Directory accounts for Identity Management



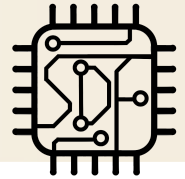
- How does this actually work?
  - ssh to sdf-login01 or sdf-login02. Enter your SLAC username as always, but your Windows password.
  - Kerberos works the same as you might be used to:
    - type “klist” and you will see a Kerberos Ticket Granting Ticket (TGT).
    - except you’ll see it say “username@WIN.SLAC.STANFORD.EDU” (note the “WIN”)
    - “kinit -R” will renew your TGT before it expires, without typing a password.
  - Your username, UID, and GIDs are the same inside SDF as outside SDF (eg, rhel6-64, centos7)
    - Your AD account object includes two attributes which allow linux login: uidNumber and gidNumber
    - on sdf-login01 or 02, type “id” or “id [username]” to see your UID and GIDs, or someone else's
    - Configuration on linux host for database lookup order: /etc/nsswitch.conf -> SSSD -> AD
    - standard commands such as “getent” still work (display entries from databases in nsswitch.conf)
      - eg, “getent password [username1 username2 …]”, or “getent group [group1 group2 …]”
  - openldap client tools can be used to query Active Directory. openldap command lines are unwieldy, so we have provided a couple of scripts on sdf-login01 and 02 which can list all usernames and all POSIX groups defined on the system: “**listusers** [username1 uid2 …]”, “**listgroups** [groupname1 gid2 …]”, “listusers” (to view all users), “listgroups” (to view all groups).

Active Directory accounts for Identity Management

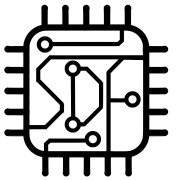


- What is not changing:
  - usernames and UIDs, group names and GIDs
    - we are keeping the legacy group names and GIDs to ease storage migration and keep the same access to it (eg, existing file and group ownership and permissions work in SDF)
  - the management of unix groups is not changing: use ‘ypgroup’ (for now) on rhel6-64.
    - the larger Identity and Access Management (IAM) project will modernize this process
- What if a SLAC user has a unix account but does not yet have an Windows/AD account?
  - self-service web portal to create an AD account for SDF login access, if you already have unix account
    - https://ad-account.slac.stanford.edu

Active Directory accounts for Identity Management



Questions?



SDF User Experience

Yee-Ting Li

August 27th 2020, Unix Town Hall

General

- compilers etc.
  - gcc still generally recommended, newer versions (7+) will be provided
  - intel to be added
- openmpi 4.0.4 as standard
- modulefiles as standard now
  - free to continue to install software as is; but it does…
    - provide a layer of portability (just modify modulefile instead of changing hard coded paths)
    - visibility for other users who may also want to run same software
  - build your software in docker/singularity containers
    - allows portability and better application management
    - wrap a modulefile around it!
- will have examples in documentation site...

Better Documentation!

Previously:

- numerous (and old) documentation around
- groups would often write their own documentation
- pages are on html files or restricted write confluence pages/other wiki
- not searchable in google etc.
- status and performance information fragmented and difficult to customise

Goal:

- provide easy to maintain, community driven documentation on everything SDF
- User guides, tips, FAQs etc.
- Easy for anyone to contribute (both inside SLAC and outside)
- Reduce the duplication and redundancy of information
- Put somewhere where people can find and read it!

Web based tools for SDF

- https://jupyter.slac.stanford.edu will be deprecated
  - jupyterhub (at least the way its been implemented) doesn't meet the needs of scalable, integrated and reliable infrastructure.
  - pros:
    - simple interface
    - (convoluted) bring your own jupyter
  - cons:
    - use of docker containers for runtime meant integration had to be reviewed for cyber security reasons
    - limited resources: wasn't integrated into slurm, so separate k8s nodes were required
- web-based access
  - integrate monitoring; easily check up status
  - file upload/downloads

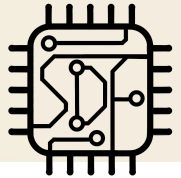
Single Frontend

- In development currently (lots left to write)
  - https://ondemand-dev.slac.stanford.edu
- backed with github and markdown files
  - users can submit changes through github and we can review with ‘pull-requests’ to merge changes
- will be moved to http://sdf.slac.stanford.edu

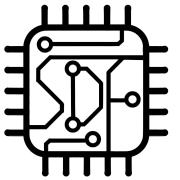
Demo

- https://ondemand-dev.slac.stanford.edu

SDF User Experience



Questions?



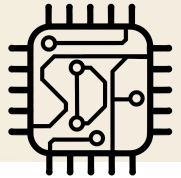
NERSC Update

Debbie Bard

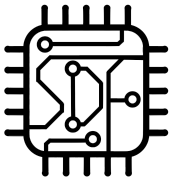
August 27th 2020, Unix Town Hall

U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY



Intermission

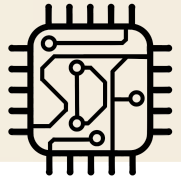


SDF Filesystem Update

Lance Nakata

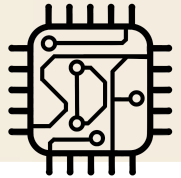
August 27th 2020, Unix Town Hall

SDF Filesystem Update



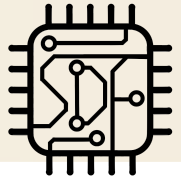
- Two DDN ES18KE subsystems, fully redundant
- Internal SAS SSDs for metadata targets (MDTs)
- 10 external disk trays, ~7.3PB /sdf, ~24TB /scratch
- Plan to add 10 more disk trays during Q1FY21 and perhaps double /sdf capacity
- Runs DDN's EXA5 software, based on Lustre 2.12.3 with DDN extensions
- Disks use Declustered RAID (DCR) software for protection and faster rebuilds

SDF Filesystem Update (cont'd)



- Distributed NamespacE Phase 2 (DNE2): distribute a directory's contents over multiple MDTs. Used for directories with (potentially) large subdirs within them
- Data on MDT (DoM): write first 1MB to MDT for better small file I/O performance
- Progressive File Layout (PFL): automatically adjust striping policy based on filesize
- Block and inode quotas on all directories (mostly based on projectIDs)
- Default home directory quota: 25GB and 500K inodes

SDF Filesystem Update



Questions?

Batch Compute Update

Yemi Adesanya

August 27th 2020, Unix Town Hall

U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

Batch Compute Update

Rome CPU Cluster

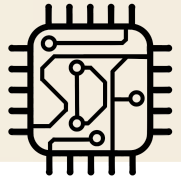
- AMD Rome EPYC 7702 CPUs @ 2.0GHz
- 4GB RAM per core
- 100Gb Infiniband
- 11264 total cores or 176 TFLOPs
- Currently fully-funded via science directs
- 5-year hardware lifecycle
- Partially deployed as of 8/2020
- Datacenter is working hard to resolve B050 power distribution challenges
- *Additional 2560 cores coming for KIPAC

Project	# Cores
LCLS	2816
Fermi	2048
SUNCAT	5376
CryoEM	384
HPS	384
SuperCDMS	256
KIPAC*	2560

Batch Compute Update

- Access via slurm job scheduler
  - We have a developer support agreement for the SDF CPU & GPU clusters
- We must ensure paying stakeholders get instant access to their dedicated partition
  - Delegate an owner (POC) for each partition that can manage their group membership
- A shared queue is available for opportunistic job submissions
  - Run the risk of preemption (job termination)
  - Recommend checkpointing your code
  - Can we justify indirected funded cores? How many?
- TO DO: Generate periodic utilization reports (monthly or quarterly) for each partition
  - Help stakeholders plan for lifecycle and support their grant proposals

Batch Compute Update



Questions?

Tape Storage Roadmap

Guangwei Che

August 27th 2020, Unix Town Hall

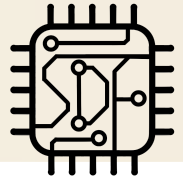
U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

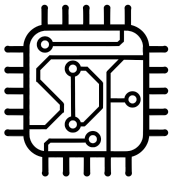
Tape Storage Roadmap

- We currently own two Oracle SL8500 tape libraries
- Oracle stopped enterprise tape drive development in 2017. Growing concern about Oracle's tape commitment plus increasing library maintenance and media costs
- Potential tape library replacements: IBM TS4500 and Spectra Logic TFinity
- Uncompressed tape library capacity of 120-160 PB, expandable to over 1 EB via media frame additions and drive upgrades
- Tape drive choices: TS1160 (20TB) or LTO-9 (18TB)
- Upgrading tape drive technology to enhance data migration rate to 6 GB/s or above
- Exploring solutions to archive SDF cold data to tape to reduce disk storage expense
- Planned tape library deployment: FY21

Tape Storage Roadmap



Questions?



Storage Update

Lance Nakata

August 27th 2020, Unix Town Hall

Storage Update

- Storage as a Service (StaaS)
  - StaaS disk hardware is 90% full and entering retirement
  - No additional space planned for this service
  - Will be transitioning experiments to SDF storage starting FY21

- Hardware Lifecycle
  - FY21 and beyond: retire legacy Sun, LSI, and Dell storage
  - Groups can migrate data to SDF storage, delete it, or archive it to tape (additional cost)
  - You will receive migration (downtime) notices for data moves

Storage Update

Questions?

CentOS / RHEL Platform Update

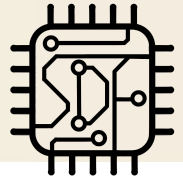
Karl Amrhein

August 27th 2020, Unix Town Hall

U.S. DEPARTMENT OF ENERGY
Office of Science

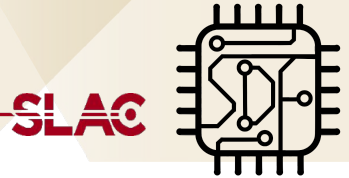
SLAC NATIONAL ACCELERATOR LABORATORY

CentOS / RHEL Platform Update



- Red Hat Enterprise Linux Life Cycle: https://access.redhat.com/support/policy/updates/errata
- Red Hat Enterprise Linux 5
  - Extended Lifecycle Support (ELS) ends November 30, 2020
  - No updates at all after that date.
- Red Hat Enterprise Linux 6
  - “Maintenance Support 2” ends November 30, 2020
  - Extended Lifecycle Support (ELS) starts after that.
- Recommended Operating System / Linux Distribution
  - For servers: CentOS 7 or RHEL 7
  - For desktops/laptops: Ubuntu LTS (recommended) or CentOS 7
- Support for CentOS 8 coming next year, after RHEL 5 systems have been retired

CentOS / RHEL Platform Update

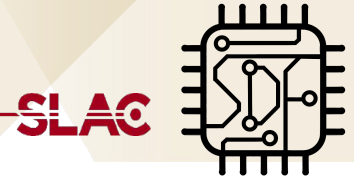


- Configuration Management
  - Taylor for RHEL 5 and RHEL 6
  - Chef for CentOS 7, RHEL 7, Ubuntu LTS

- SLAC Chef cookbooks (configuration management code) on github:
  - https://github.com/SLAC-CHEF/
  - Send your github username to unix-admin and we will add you to the SLAC-CHEF organization
    - Then you can request/recommend code changes

- Unix Platform work in progress
  - Support of SDF, especially regarding changes to authentication
  - Support of IAM project and integration of the InCommon Trusted Access Platform suite
    - https://spaces.at.internet2.edu/display/ITAP/InCommon+Trusted+Access+Platform+Library
  - Remaining RHEL 5 servers need to be replaced with CentOS 7 or RHEL 7, so we can focus on RHEL 8
  - Retire OpenStack cluster: migration to VMware or Containers instead

CentOS / RHEL Platform Update



Questions?

Next Steps for SDF

Yemi Adesanya

August 27th 2020, Unix Town Hall

U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

Next Steps for SDF

- Short-Term
  - Comprehensive user guide for SDF on the github site
  - Complete the CPU cluster deployment in B050 datacenter
  - Begin migration of all GPU nodes into SDF
  - Establish the base compute software stack via Environment Modules interface
  - Complete the recently purchased storage expansion - includes capacity for CryoEM, ATLAS and KIPAC
  - New Tape Library (Phase 1)
  - Assemble a steering group to review our current/proposed SDF operational policies and usage quotas

- Longer-Term
  - Datacenter strategy to support SDF growth beyond B050 (SRCF-II,.....)
  - Cost model and budget projections for Computing (indirect) funding to deliver baseline SDF capabilities
  - Align to SLAC IAM roadmap and consider Federated Identity models
  - Facilitate NERSC for SLAC computing
  - Opportunity for SDF cloud Proof Of Concept