SLAC Input to HEP-CCE

HEP-CCE All Hands Meeting
July 2024

U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

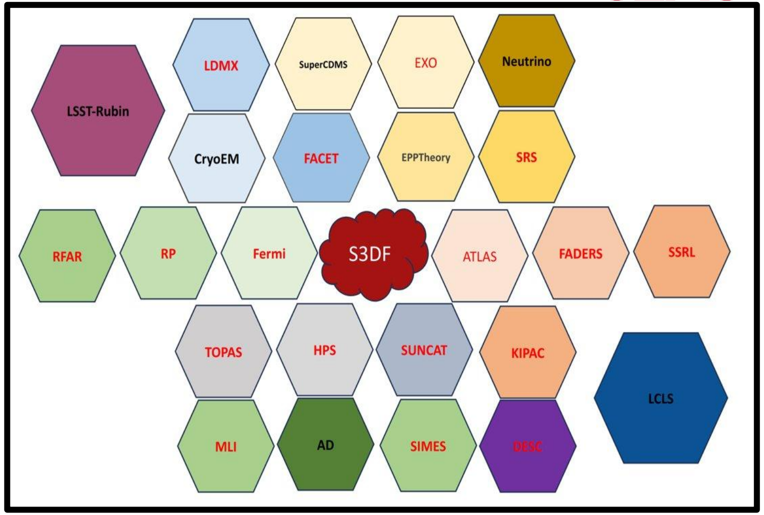
Overview

- Introduction to HEP computing at SLAC
  - SLAC Shared Science Data Facility (S3DF)
  - SLAC HEP Computing
- Possible SLAC tasks within HEP-CCE

What is the SLAC Shared Science Data Facility (S3DF)?

SLAC

- **Common shared scientific computing infrastructure** to tackle the challenges of massive throughput data analytics at SLAC
- Supports several key mission areas:
  - **Critical, data-heavy, scientific computing workflows** for all science at SLAC - including its User Facilities (LCLS, Rubin, UED, CryoEM, SSRL)
  - Baseline scientific computing resources for ALL SLAC users
- This infrastructure offers:
  - “Ever-green” computing and storage platforms (disk and archival storage)
  - Broad set of capabilities and services to support scientific computing
  - Interactive access
  - Batch computing
  - Service Compute
  - User-defined Kubernetes-based services for databases, web-based portals, data management, etc.
  - Jupyter notebook portals



Key use case characteristics

- High-throughput data production
- Large datasets and complex data management workflows
- High-throughput computing
- Data-intensive computing

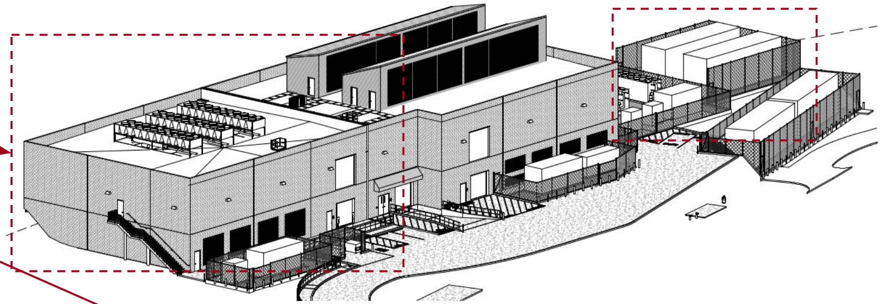
Key S3DF capabilities

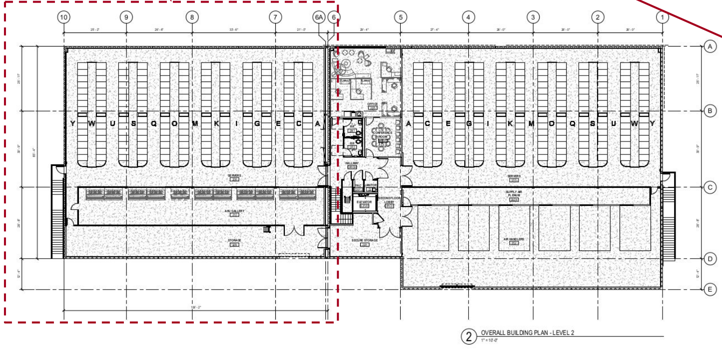
- Systems Integration
- Broad range of services for interactive analysis, workflow control, batch system access, and user management
- Support complex data storage, access, and management needs

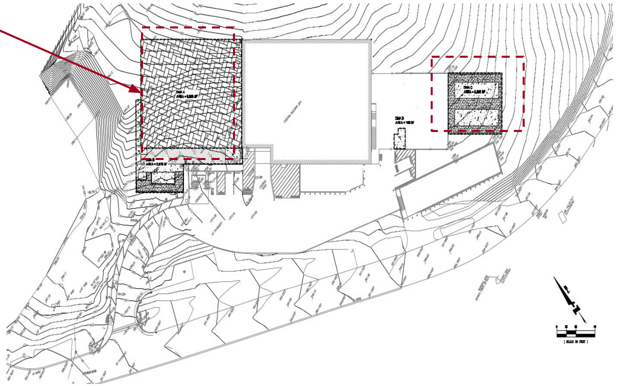
S3DF located at the SRCF Data Center at SLAC

SLAC

SRCF

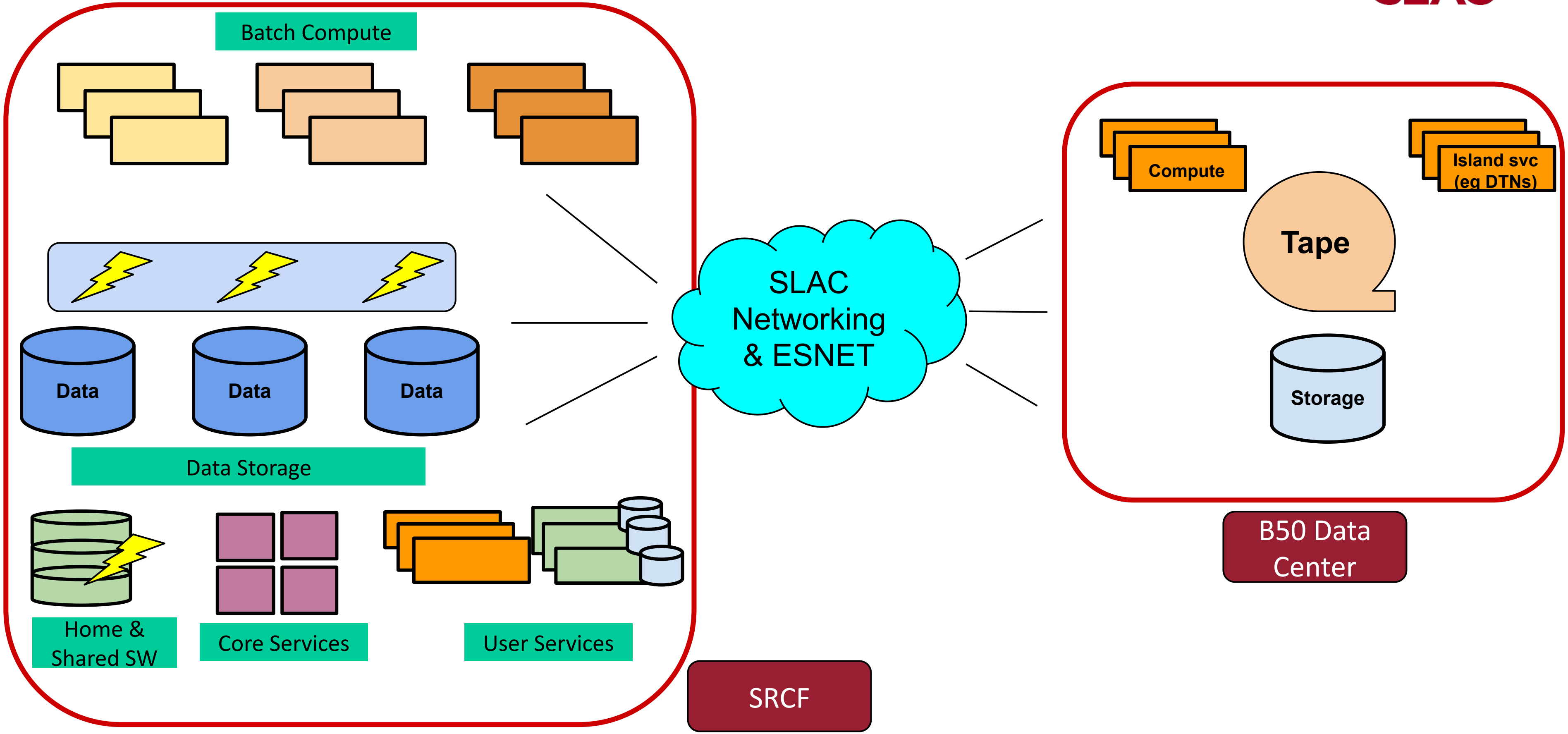






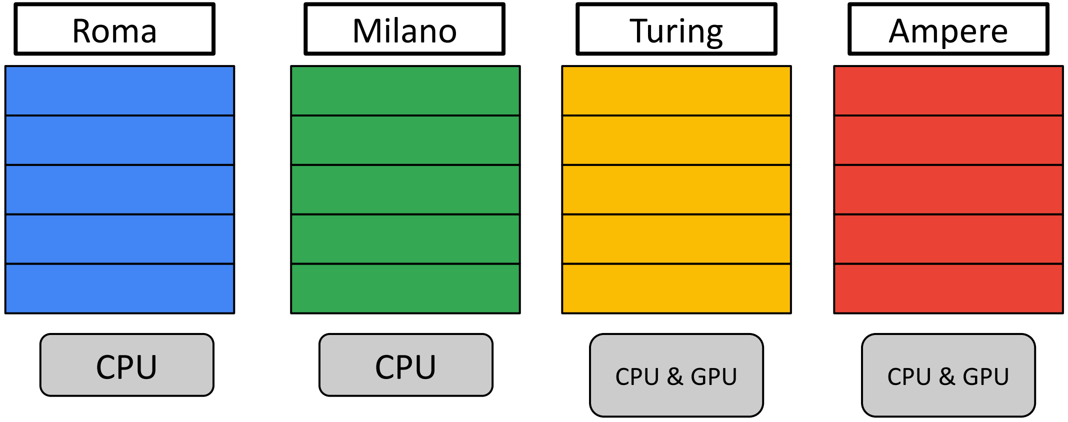
- 6MW Facility with air cooling
- Flywheel + Generator allows for resilient power
- Space for over 100 racks
- 400 Gbps Networking to SLAC backbone

An Integrated Compute, Storage & Services Architecture



S3DF Batch Systems & Services

SLAC

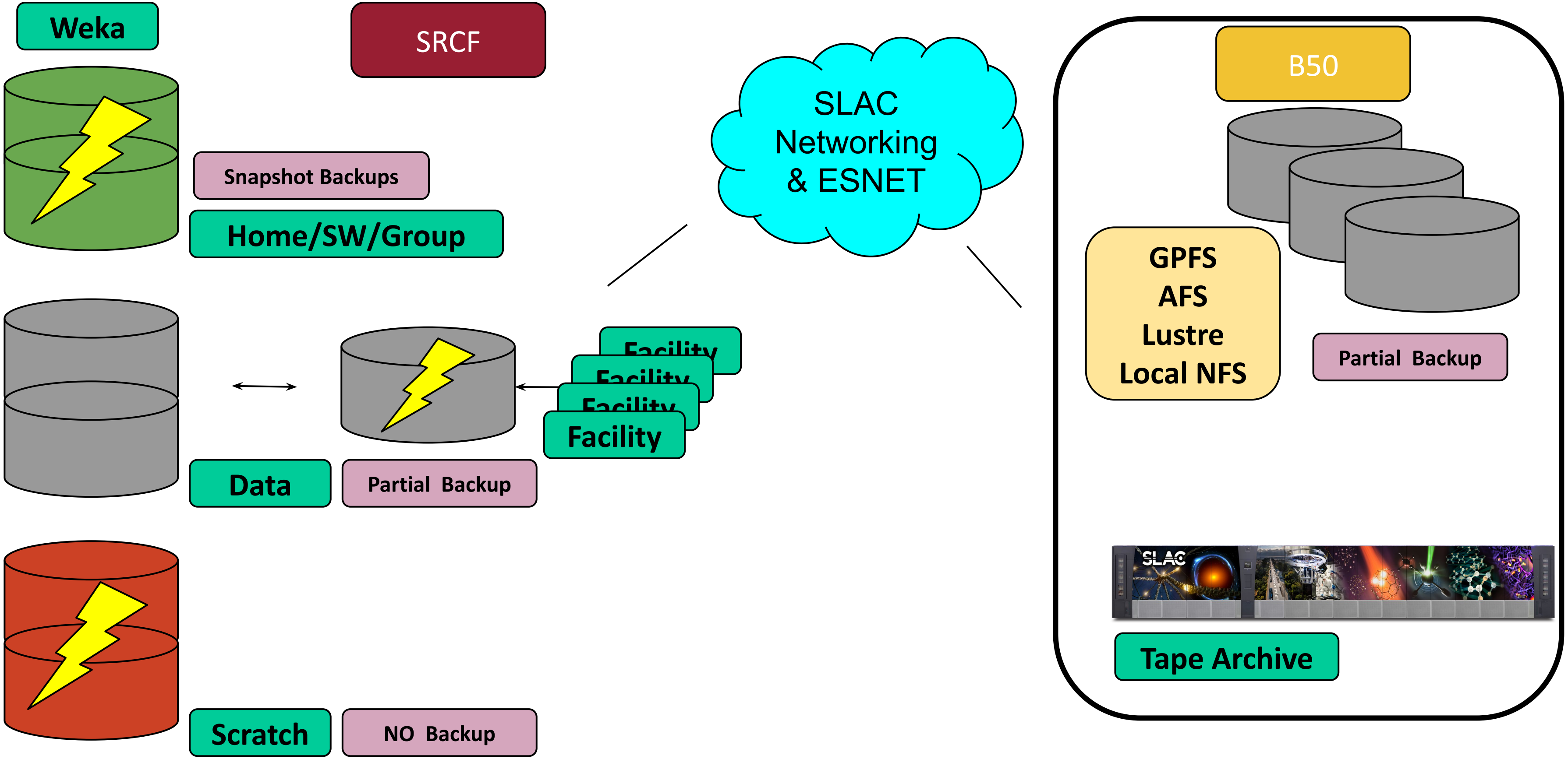


- Compute Clusters (Partitions) are available to all Facilities
- User access granted per project/experiment to system resources
- Access to compute resources is gated by Slurm Batch System

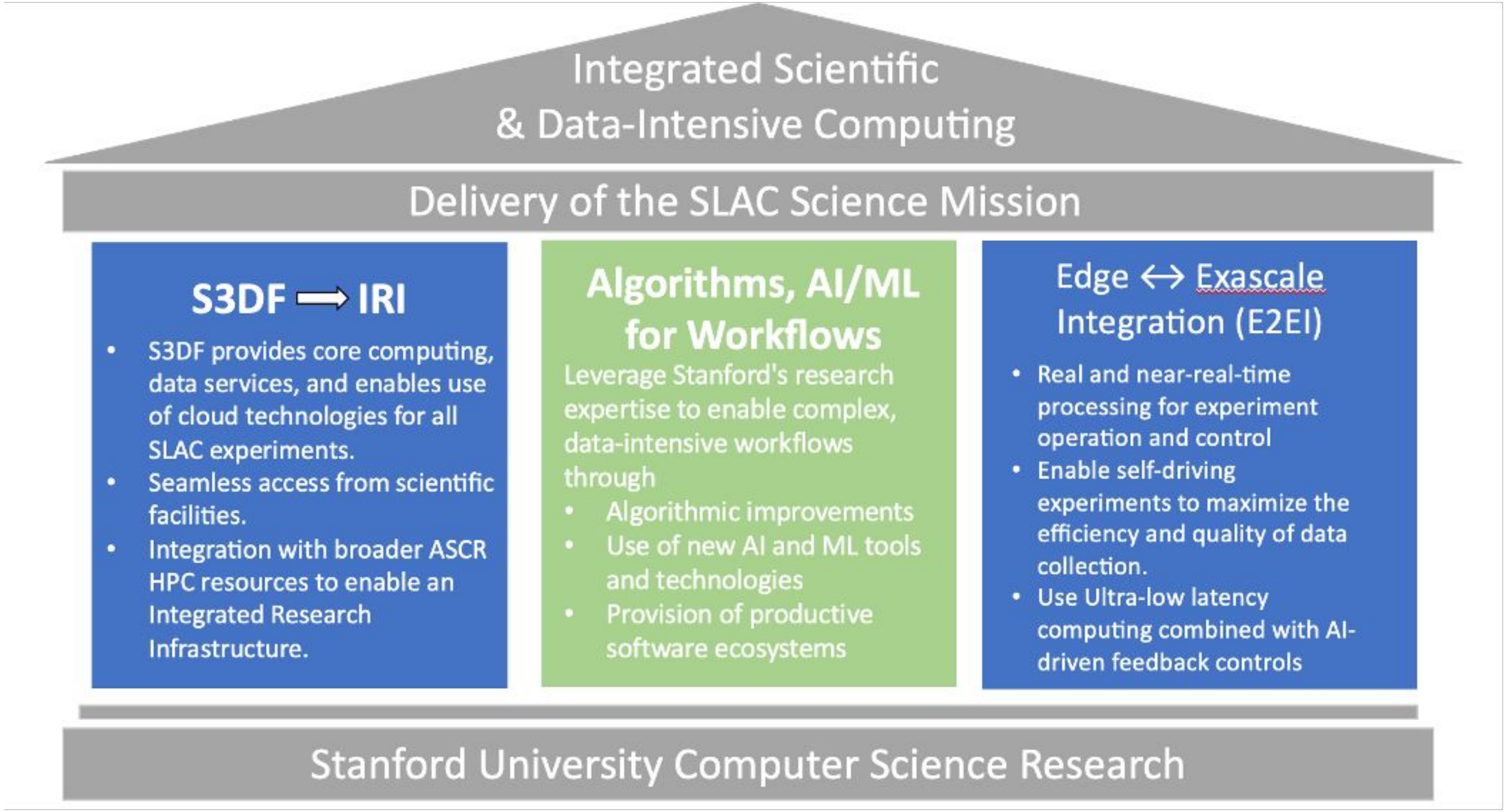
- Systems refreshed yearly with new technology
- System evolves with the needs of Science experiments

- Kubernetes-based (K8s) “Service Compute”
  - Resilient website and web-based applications
  - K8s cluster with access to S3DF storage
  - User defined portals for front-ends to web applications
  - Workflow control services
- Virtual Machine services
- Database services

S3DF Storage Systems



An Integrated effort to support Scientific Computing



SLAC HEP Computing

- SLAC is involved with computing for many HEP experiments
  - Rubin Observatory/LSST
    - US Data Facility (USDF) located at SLAC hosted in S3DF
      - Data transferred to SLAC within 2 minutes of acquisition – real time alert generation
      - Offline processing
    - LSST Dark Energy Science Collaboration – (computing mainly done at NERSC, ANL, Grid)
- Current
  - Atlas , Fermi Gamma Ray Space Telescope, Super CDMS, LZ, DUNE , Theory
- Future
  - CMB-S4, LDMX, DMRadio, …
- Emphasis on support of Cosmic Frontier experiments
  - Including many small/medium size experiments which often need large scale computing with small teams of people
  - Many experiments use HPC facilities as well as international collaborators (grid sites)
  - Goal to leverage HEP-CCE and IRI/HPDF to provide improved support/standardization

SLAC involvement with HEP-CCE (I)

- SLAC is very interested in becoming an institutional member of HEP CCE and participate in phase 2
  - We already have experiment involvement particularly from LZ and LSSTDESC
  - We have encouragement from lab management and DOE
  - Goal is to provide concrete proposal of scope before end of FY 24
- Portable Applications/Portable Workflows
  - Many SLAC experiments have complex workflows (see LZ talk yesterday for examples)
    - Currently different experiments use different tools
    - Increasingly these are hybrid workflows involving traditional compute/HPC
    - Want to understand (and give input) on how IRI/HPDF will handle hybrid workflows
    - Very interested in participating in the taskforce Salman described yesterday
- Optimizing Data Storage
  - Expand on SLAC's experience and support for xrootd
    - Add (remote direct memory access) RDMA support to xrootd
    - Bridge between HPC and the internet by fast-tracking HPC edge services
    - Expand data caching services popularized by Atlas, CMS, ESNET, Pelican, OSDF, PRP
  - Meta-data catalogs and data streaming, as above, this would benefit from combined HEP input

SLAC involvement with HEP-CCE (II)

SLAC

- Scaling up HEP AI/ML Applications
  - Anomaly detection in LZ (+ potentially: XLZD)
  - SLAC has strong AI/ML groups, and several initiatives
    - Definitely an area we are interested in, collecting input from this meeting and will present more detailed proposal in future
- Accelerating simulations
  - High-precision simulations useful to plan future experiments
  - GPU-accelerated simulations (particle tracking) relevant to XLZD, nEXO, DUNE
  - Simulation-based inference actively being pursued for ATLAS and DUNE