FABRIC and FAB update

Research and Development
LHCOPN-LHCONE meeting #50
FZU, Prague CZ
April 18-19, 2023

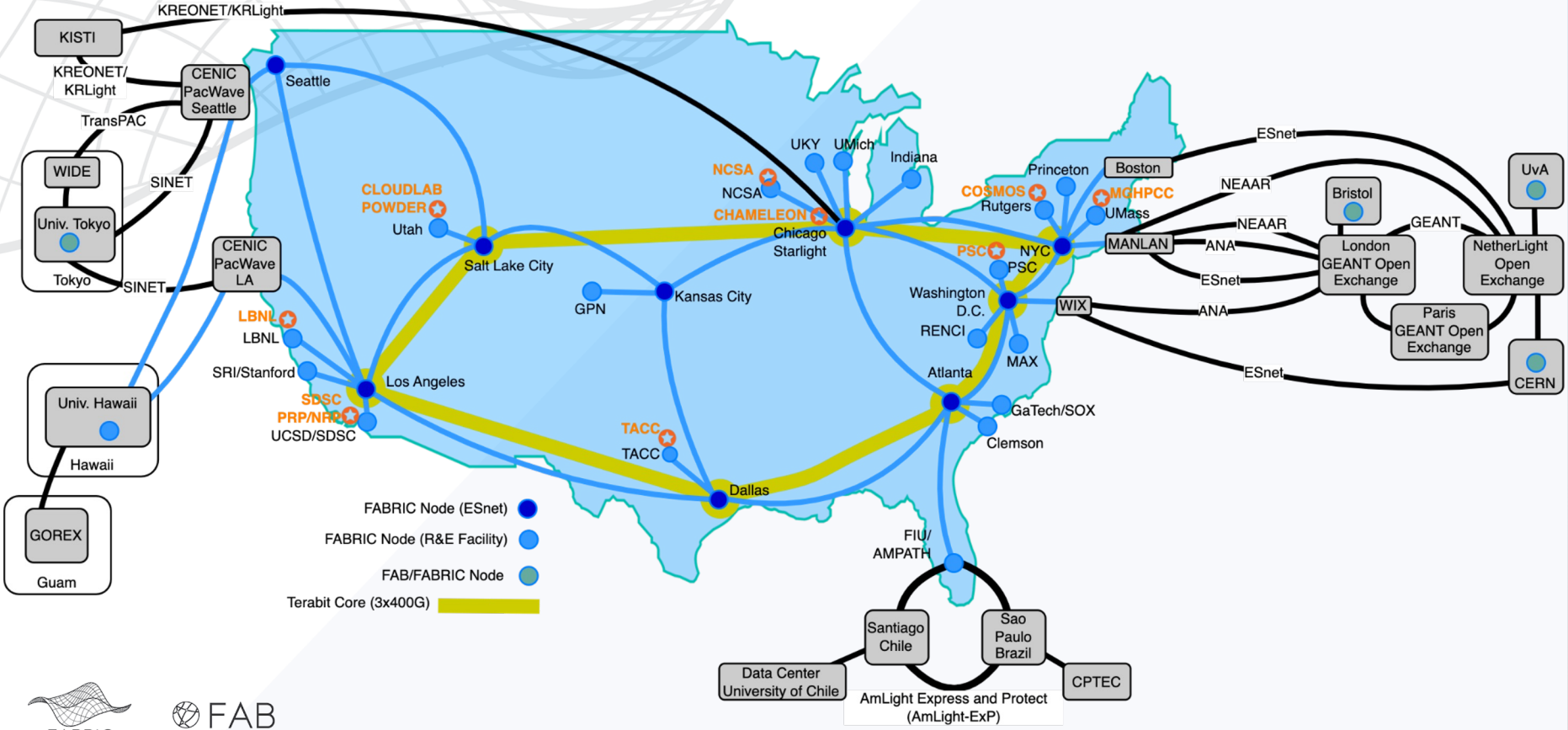
What is FABRIC?

- NSF-funded MSRI-1 (Mid-scale Research Infrastructure)
- Led by RENCI/UNC Chapel Hill
- 5 core team members: University of Kentucky, University of Illinois at Urbana-Champaign, Clemson University and ESnet
- Many other partners, including Internet2
- $20M budget for construction, separately-funded operation phase expected
- Started in 2019, expected completion 09/2023

FABRIC

FAB
FABRIC Across Borders

FABRIC + FAB Final Topology

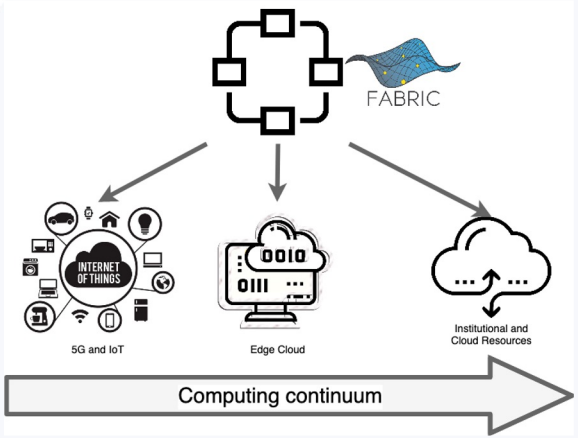


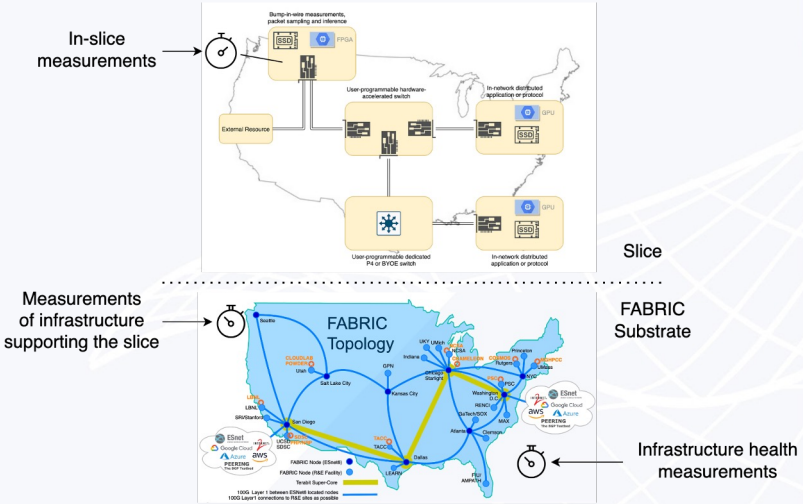
FABRIC

FAB
FABRIC Across Borders

Key FABRIC features

- Network as part of computing continuum
  - ‘Everywhere-programmable’ using different abstractions (P4, OpenFlow, others)
  - Diverse compute, storage capabilities in places where routers typically reside today
  - Dedicated 100G optical links between many sites
  - Support new paradigms in network aware applications and protocols
  - Ability to peer with Internet at IPv4 and IPv6
- Network as a scientific instrument
  - Pervasive measurement collection capabilities in- and outside the slice available to researchers
  - GPS-disciplined PTP clock sources at every site
- Serve a broad range of scientific domains and applications
  - Concerned with data transport for big-data science, cyber-security, terrestrial and 5G hybrid network architectures, federated ML/AI, Internet measurements and many more

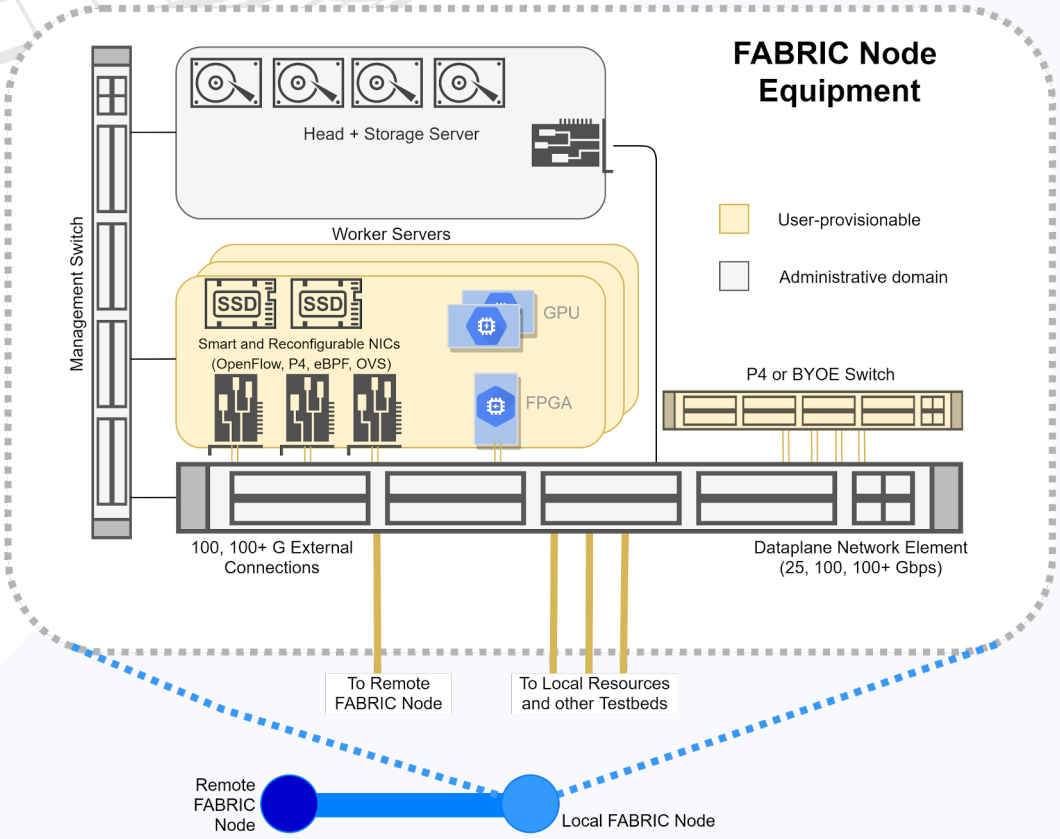




FABRIC

FABRIC “HANK”

Hank: a measured unit of coiled or wrapped yarn or twine



FABRIC

FAB
FABRIC Across Borders

FABRIC Nodes - Network, Compute, Storage

- Interpose compute and storage into the path of fast packet flows
- Rack of high-performance servers (Dell 7525) with:
  - 2x32-core AMD Rome and Milan with 512G RAM
  - GPUs (NVIDIA RTX 6000, T4, A30), FPGA network/compute accelerators
  - Storage - experimenter provisionable 1TB NVMe drives in servers and a pool of ~250TB rotating storage at each site.
  - Network ports connect to a 100G+ switch, programmable through control software
  - Tofino-based P4 switches (4 sites)
- Reconfigurable Network Interface Cards
  - FPGAs (U280 XILINX with P4 support)
  - Mellanox ConnectX-5 and ConnectX-6 with hardware off-load
  - Multiple interface speeds (25G, 100G, 200G+(future)
- Kernel Bypass/Hardware Offload
  - VMs sized to support full-rate DPDK for access to Programmable NICs, FPGA, and GPU resources via PCI pass-through

FABRIC

FABRIC Network Services

- A rich set of L2, L3 and specialized services to aid the experimenters
- L2 services provide Ethernet service between experimenter topology interfaces
  - L2Bridge - local to individual site
  - L2STS - two sites, any number of interfaces
  - L2PTP - two sites, two interfaces, QoS guarantees
- L3 services provide IPv4 and IPv6 services with an option to connect to the outside world
  - FABNetv4 - FABRIC-routed IPv4 service, using RFC1918 addresses
  - FABNetv6 - FABRIC-routed IPv6 services using FABRIC's IPv6 allocation
  - Both have externally-connectable counterparts (IPv4 variant uses FABRIC's limited IPv4 allocation)
- Specialized services
  - Port mirroring - currently mirrors a specific physical port in the topology

FABRIC

FABRIC External Connections Overview

- FABRIC experiments (slices) can run in an isolated manner within FABRIC Infrastructure, and isolated from external networks.
- Slices can also utilize FABRIC's external connections to access a variety of external experimental and production resources.
- These external connections are organized as follows:
  - Layer 3 IPv4 and IPv6 public connectivity on demand (with policy restrictions) via peering points provided by ESnet and Internet2
  - Layer 2 Services Peering (aka Facility ports)
  - Public Cloud Connections via CloudConnect

FABRIC

FABRIC By The Numbers

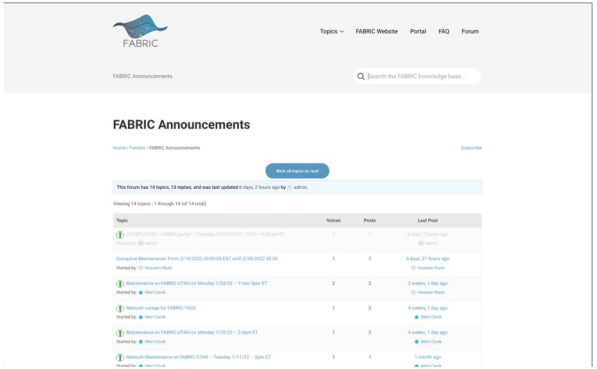
- FABRIC Resources:
  - 27 sites deployed: (3 in dev ring, 21 in production topology available to experimenters, 3 waiting for new dataplane connection)
  - 6 more sites to deploy
  - Multiple 100G segments available in the core via ESnet6
  - Multiple 400G DWDM segments deployed, working connecting to FABRIC dataplane elements
    - 3x400G, 1.2TB FABRIC Terabit Core ring
  - Multiple 100G L1 segments regional connections (UEN, NCSA, MERIT, GPN, IU)
    - Several in the works - CENIC, MAX, NJEdge, FIU
- FAB Resources:
  - CERN deployed and part of FABRIC Production Topology
  - U.Bristol, U.Amsterdam, U.Tokyo equipment shipping underway

FABRIC

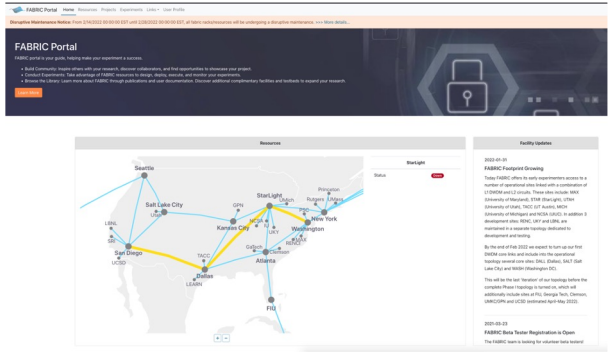
FABRIC UI

- FABRIC Portal, Jupyter Hub, User Knowledge base all operational
- Integrated with CI Logon for federated identity
- Accessible and used by early experimenters

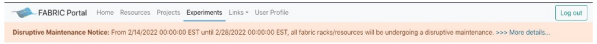
FABRIC Knowledge Base with User Forums



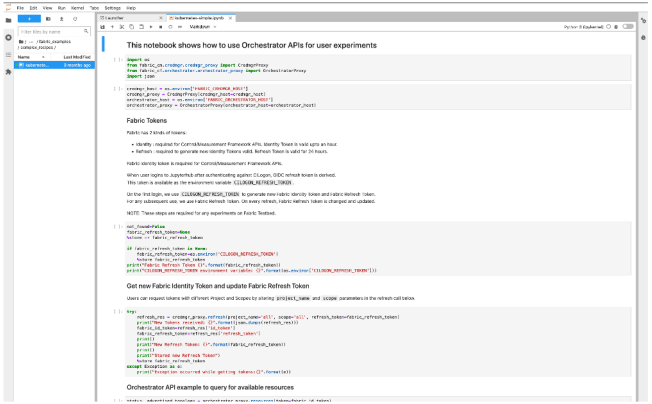
FABRIC Portal Home Page

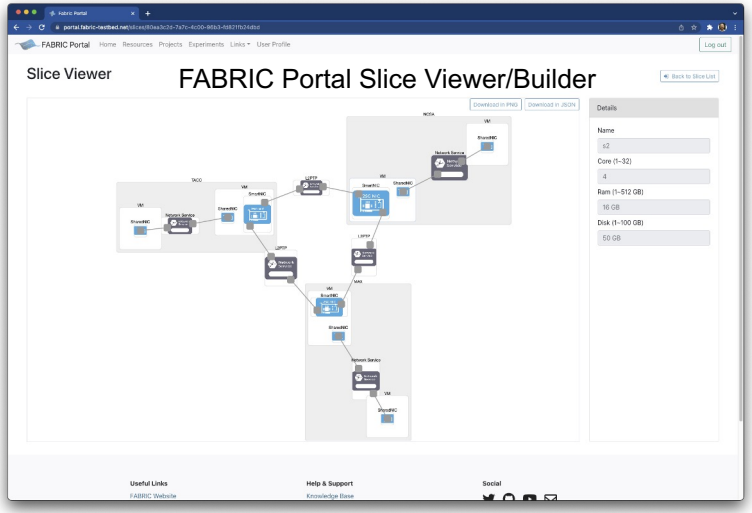


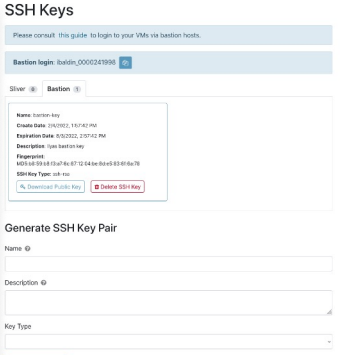
FABRIC Portal Experiments page



FABRIC Jupyter Hub





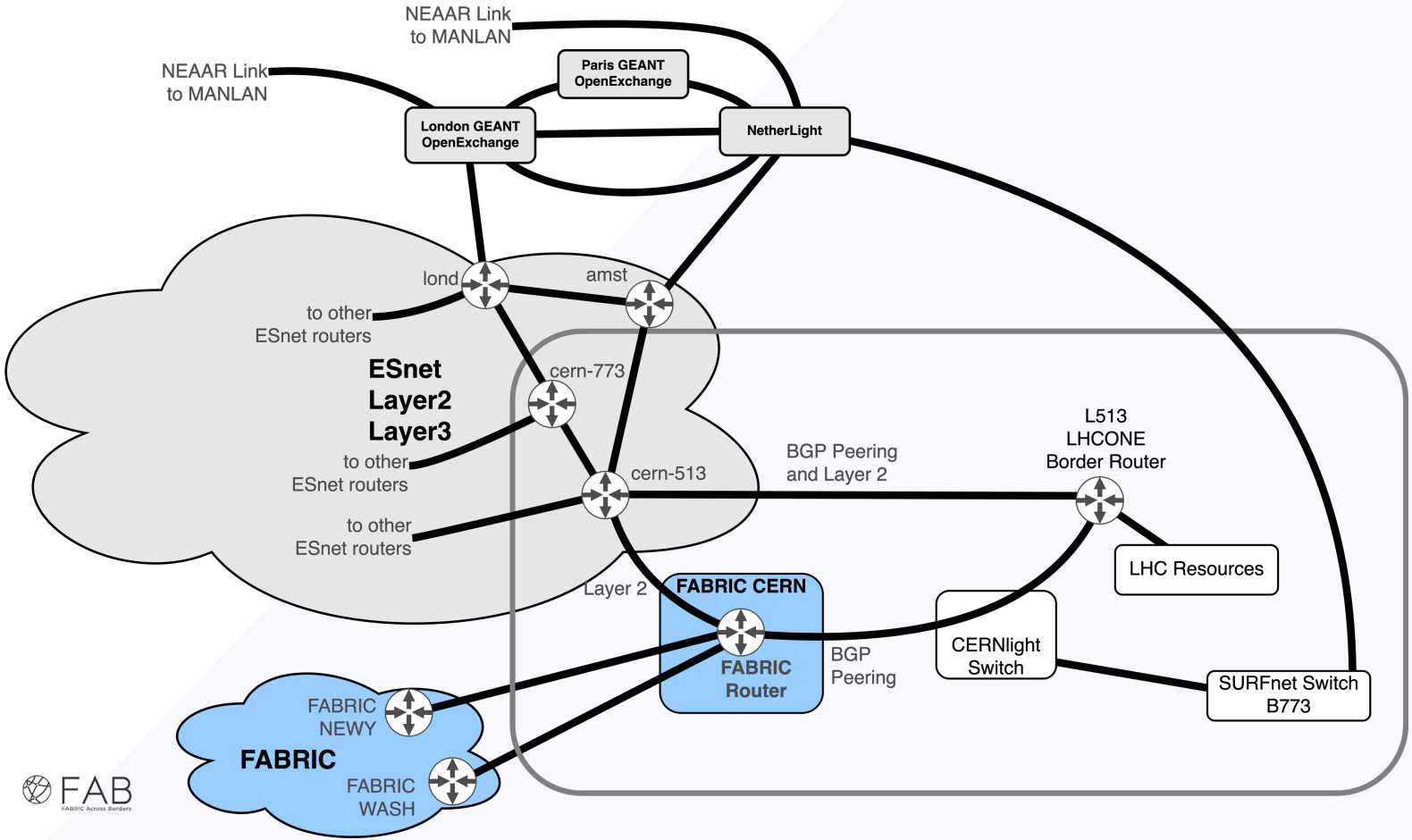


FABRIC features

- **Facility Ports** - ability to add external facilities to slices using on-demand L2 connections
- Mirror Ports - ability to mirror traffic from the dataplane switch into slice
- Support for on-demand public connectivity
  - Slices with L3 IPv4 or IPv6 can connect on-demand with public internet
  - This is in testing
- L3VPN service + CloudConnect
- **Persistent storage** for slices - get storage allocation at multiple sites for your project
- **In-slice measurement framework** - instrumentize your slice to get data about its performance
- Support for P4 Tofino switches in topologies (future)
- Support for P4 workflows on top of U280 FPGAs (in collaborations with OCT/NorthEastern and ESnet) (future)
- Inter-testbed federation features, more Facility Ports

FABRIC

FABRIC CERN



FABRIC

FAB
FABRIC Across Borders

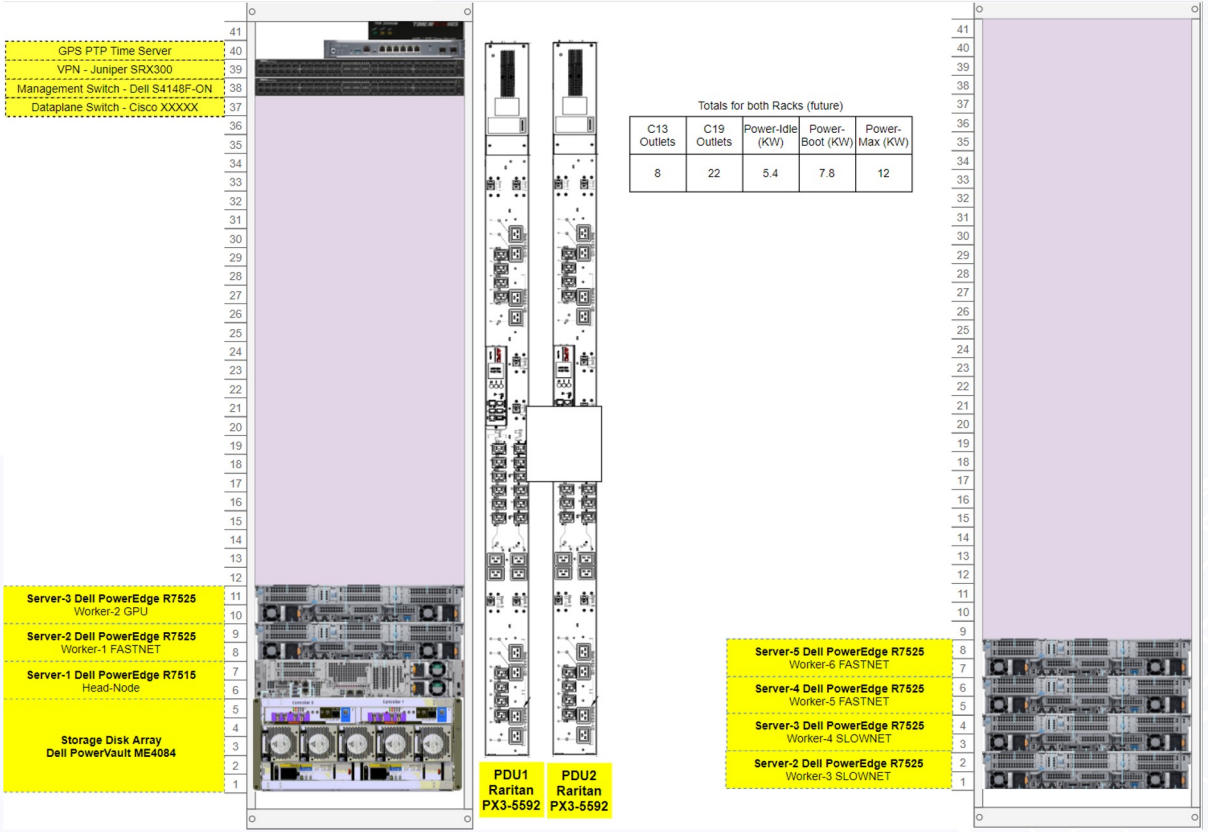
CERN FABRIC/FAB Rack Details

Rack 1

GPS PTP Time Server
VPN - Juniper SRX300
Management Switch - Dell 4148S-ON
Dataplane Switch/Router - Cisco NCS 5700
SLOWNET - PowerEdge R7525
SLOWNET - PowerEdge R7525
SLOWNET - PowerEdge R7525
FASTNET - PowerEdge R7525
FASTNET - PowerEdge R7525
HeadNode - PowerEdge R7515

Rack 2

SLOWNET - PowerEdge R7525
SLOWNET - PowerEdge R7525
SLOWNET - PowerEdge R7525
FASTNET - PowerEdge R7525
FASTNET - PowerEdge R7525
GPU - PowerEdge R7525

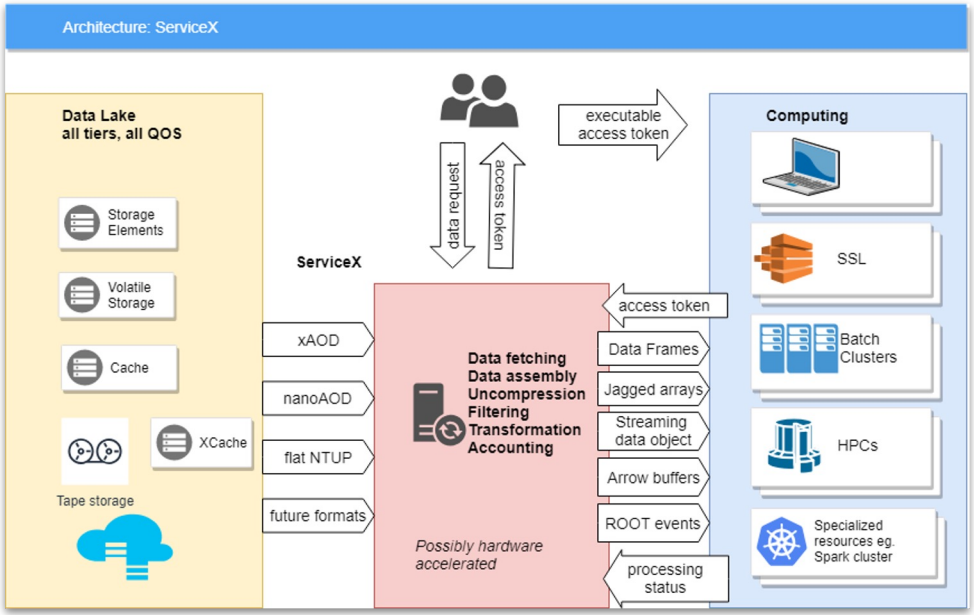


FABRIC

FAB
FABRIC Across Borders

What is ServiceX?

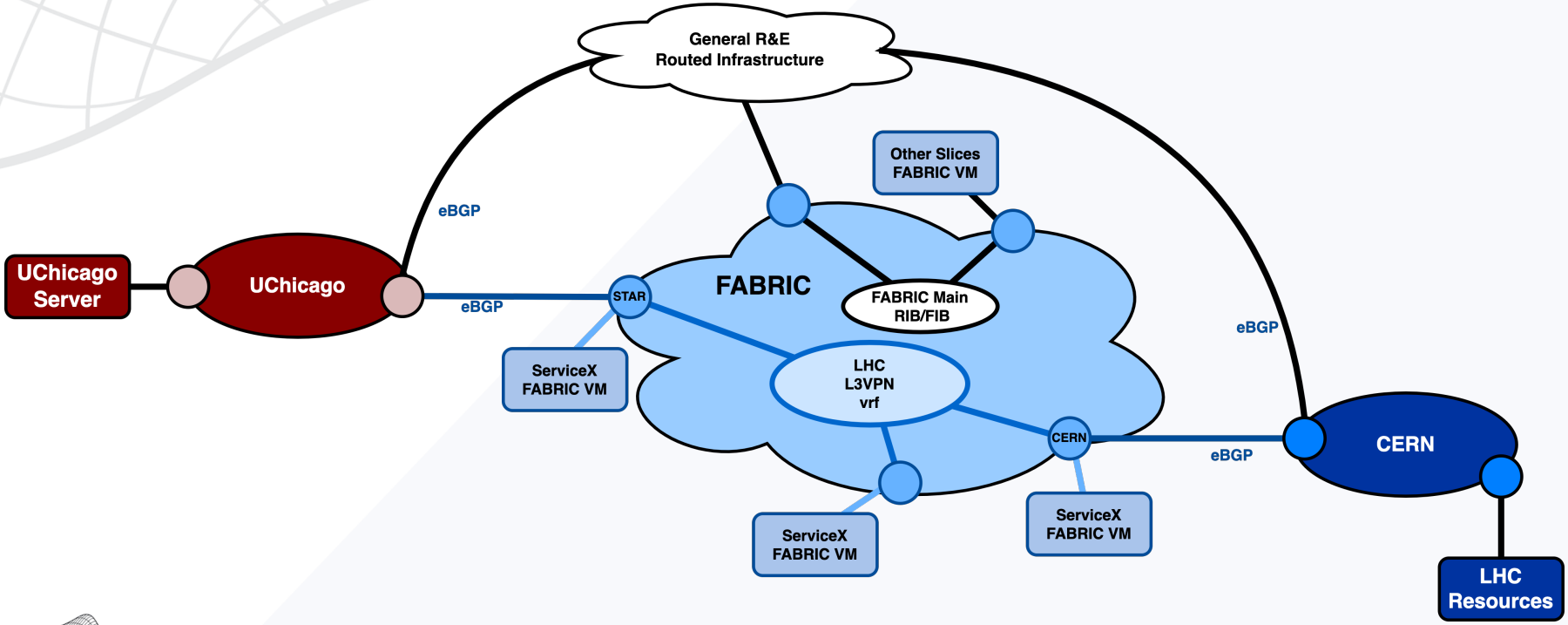
- Developed by IRIS-HEP for **quasi interactive,** high-performance **columnar-based analyses**
- **Performs on-the-fly data access, filtering, derivation and delivery** into variety of data formats
- Containerized for Kubernetes
- ServiceX Project



https://iris-hep.org

R&D for future analysis pipelines at the HL-LHC: c.f. IRIS-HEP Analysis Grand Challenge Tools Workshop

FABRIC Service-X Slice Topology



FABRIC

FAB
FABRIC Across Borders

Thank You!

Questions?

Visit https://whatisfabric.net

Ask info@fabric-testbed.net

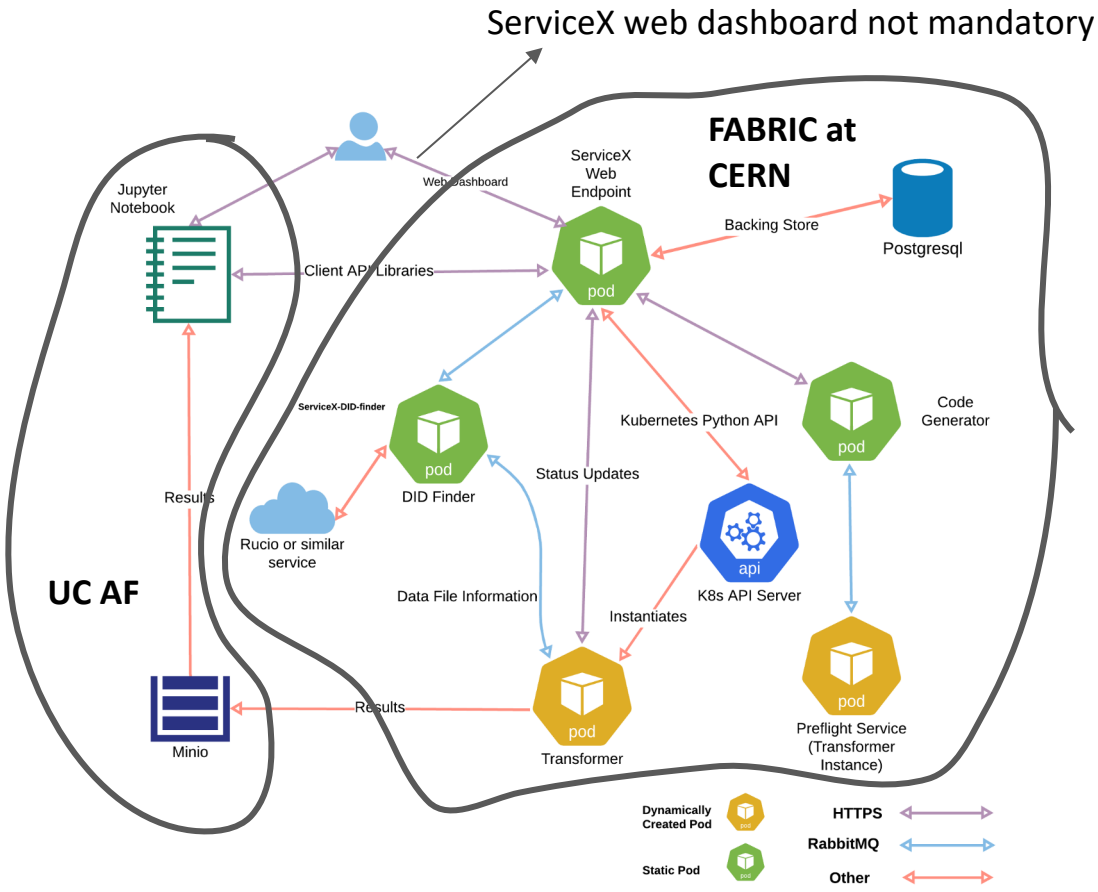
FABRIC Software: https://github.com/fabric-testbed

This work is funded by NSF grants CNS-1935966, CNS-2029261, CNS-2029235, CNS-2029200, CNS-2029261, CNS-2029260

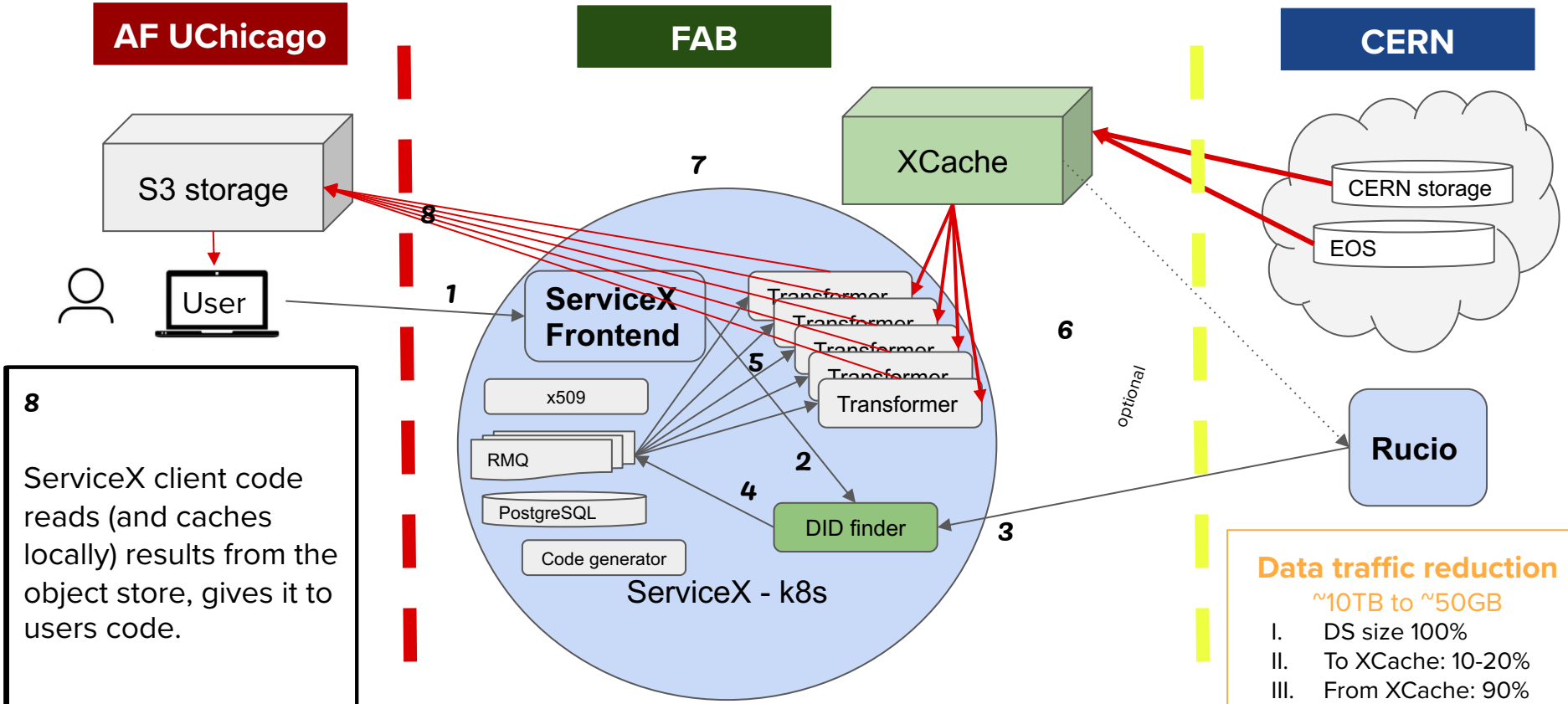
FABRIC

ServiceX - the "application"

- Complex application with **multiple pods** deployed statically and dynamically
- **Flask server** with PostgreSQL for persistent storage
- Users send a “request” to Flask server with an embedded “query” indicating which events to process from a data set
- **Code Generator** translates queries into transformer executables
- **Transformers** are spawned and autoscaled to process each file in a request’s data set
- **Other microservices** find the input files, renew access proxes etc.



Service X - Data flow



8

ServiceX client code reads (and caches locally) results from the object store, gives it to users code.

Data traffic reduction

~10TB to ~50GB

I. DS size 100%
II. To XCache: 10-20%
III. From XCache: 90%
IV. To S3 and client: 5%

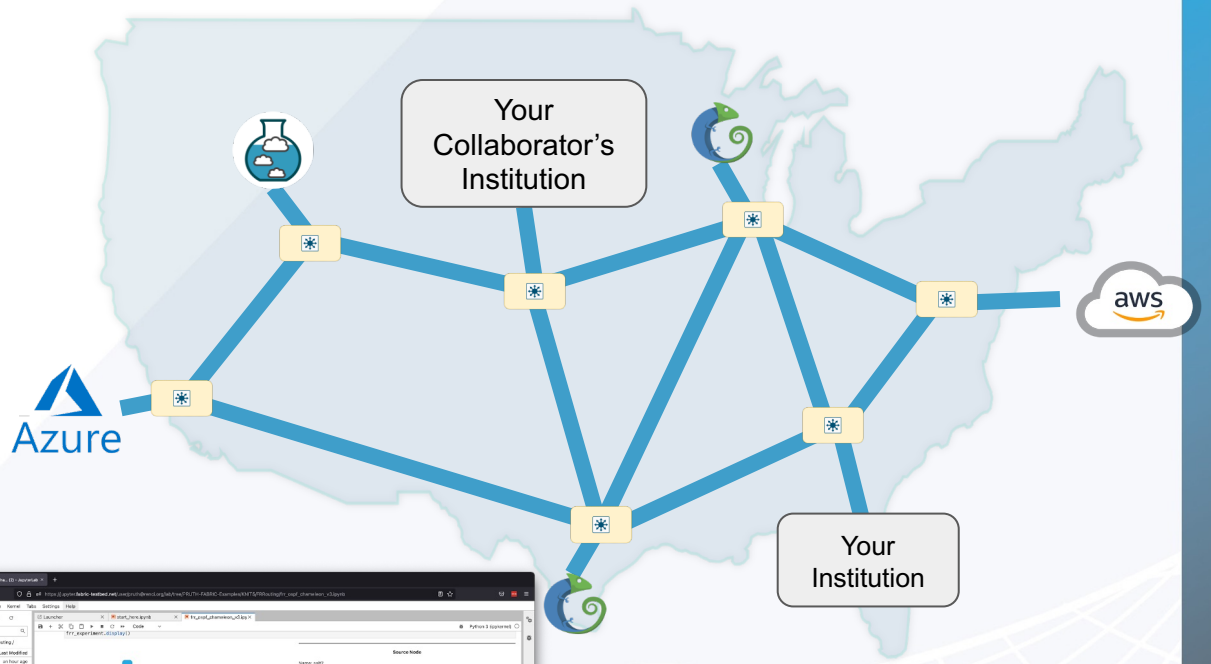
Initial Beta Testers

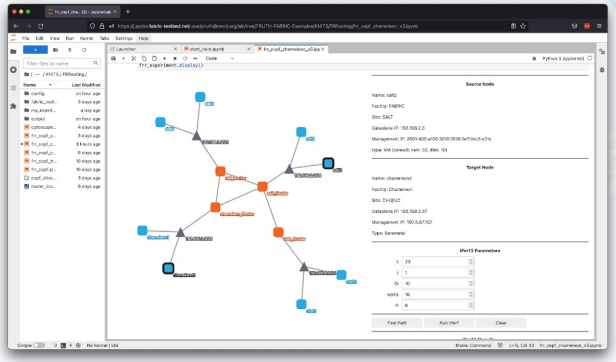
- Autonomous self-defending network control layer (Phil Porras, SRI)
- Reliability in Petabyte Scale File Transfers (Violet Syrotiuk, ASU)
- Anonymization testing (Richard Brooks, Clemson)
- Smart Sensors on the Georgia Coast (Russ Clark, Georgia Tech)
- The Genome Lake (Alex Feltus, Clemson)
- Service-X (Robert Gardner, Shawn McKee, Joe Breen)

FABRIC

Early Experiments

- Users: >250 active (340 registered)
- Projects: >40 (research and education)
- Project topics:
  - P4/SDN
  - Honeypots
  - Named Data Networking (NDN)
  - ServiceX
  - BGP Peering
  - Datacenter protocols
  - Scalable Genome Analysis
  - Fast data transfers
  - Internet Privacy





FABRIC

FAB Science Use Cases & Partners

- Astronomy (Vera Rubin Observatory/LSST, Chile)
- Cosmology (CMB-S4)
- High Energy Physics (CERN ATLAS;UChicago) - Rob Gardner
- Weather/Climate (UMiami & CPTEC, Brazil) - Ben Kirtman, Atmospheric Science & Paolo Nobre
- Urban Sensing/IoT/AI at Edge (UBristol) - Dimitra Simeonidou
- 5G across borders, P4/SDN - (UTokyo) Aki Nakao; KISTI (Korea Institute of Science and Technology Information)