NSLS-II IT SYSTEMS

IT REVIEW JANUARY 2018

ROBERT PETKUS

OVERVIEW

- NSLS-II IT environment – scope
- Network upgrade
- Current projects
- Future plans

NSLS-II IT ENVIRONMENT

The IT group maintains a large and robust **private** infrastructure comprising:

- >450 computer systems (~70 VMs and containers)
- >250 multilayer network switches
- ~3.5PB distributed storage

Consistent, homogeneous standards are applied across the facility:

- Debian Linux
- HP servers, workstations
- Extreme (nee Brocade) network hardware
- NetApp storage
- GPFS distributed filesystem

Which provides the following benefits:

- Deep working knowledge
- Interchangeable components
- Targeted development

IT SCOPE

Systems are grouped based on function:

- Remote access (e.g., SSH, NX, data transfer)
- Infrastructure, automation, and security (e.g., Puppet, DNS, GIT, monitoring)
- Web applications
- EPICS softIOCs and gateways
- Beamline data capture, computation, and analysis
- Storage and filesystem services
- Backup and recovery

ACCELERATOR IT

- The accelerator network is segmented by subsystem with a global controls subnet for channel access across the facility:
  - RF, vacuum, power supply, diagnostics, timing, etc.
- This network has a hybrid topology (hierarchical + mesh) with distribution nodes organized by region:
  - Mezzanine (cells 1-30), ISB, RF, etc.
  - All distribution nodes connected to each network core (MCT + lag)
- All internal services and many softIOCs are centrally located
  - Storage, archiving, database, gateways, etc.
  - PS and diagnostics softIOCs are distributed locally (only required for PS)
- Remote access to accelerator controls:
  - On campus: workstation → NSLS-II gateways
  - Off campus: workstation → BNL SSH/VPN → NSLS-II gateways

BEAMLINE IT

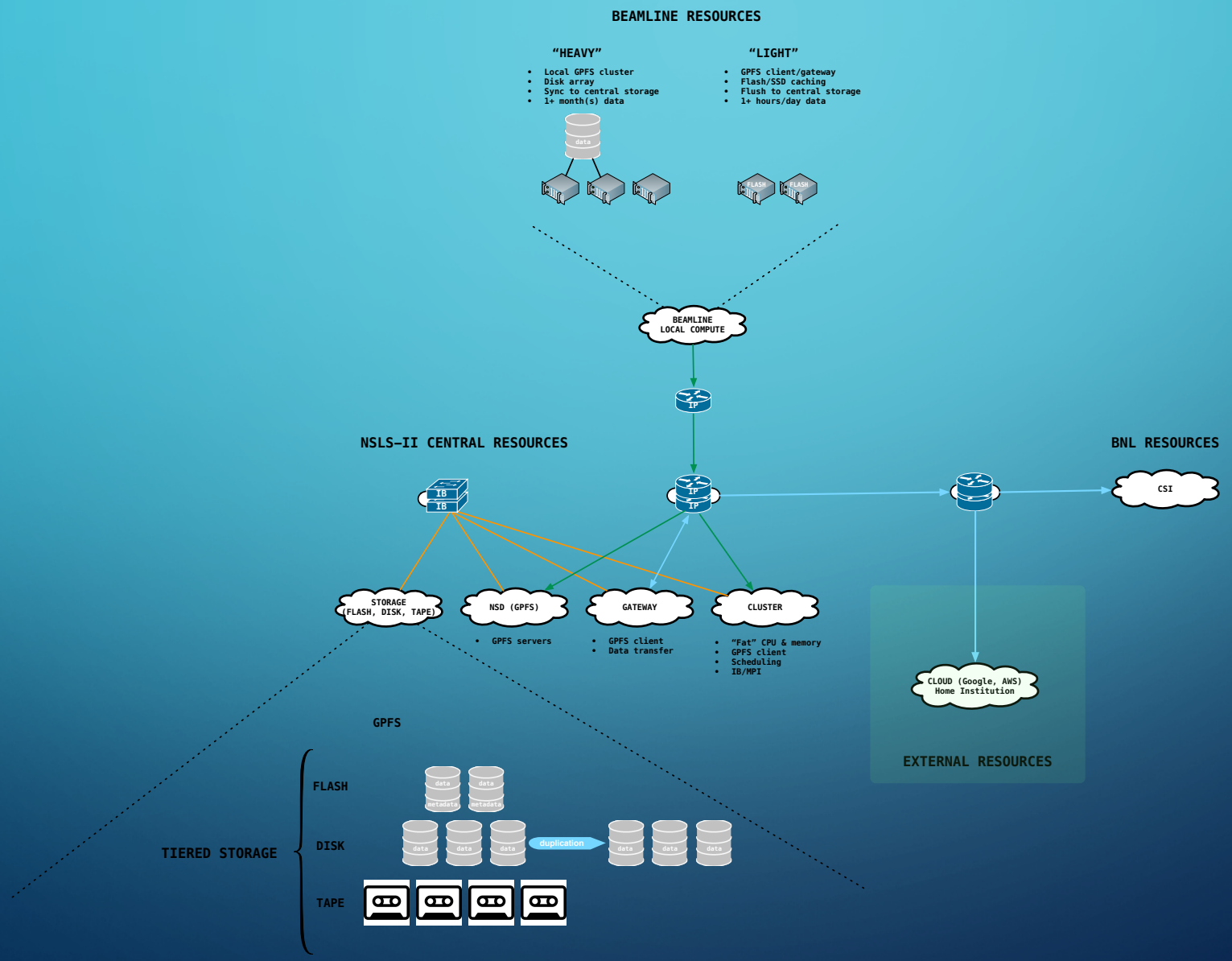
All beamlines share a basic architecture with minor variation:

- Distribution-access hierarchical network
- 10-40G redundant core uplinks
- Dedicated VLANs and IP address space
  - Controls, camera/detector, instrumentation
- Channel access gateway
- 2+ softIOC / areaDetector systems
- 1 archiver / mongoDB appliance
- 2+ analysis / GPFS NSD systems + storage

Or

- 2+ analysis / GPFS caching systems

BEAMLINE RESOURCE OVERVIEW

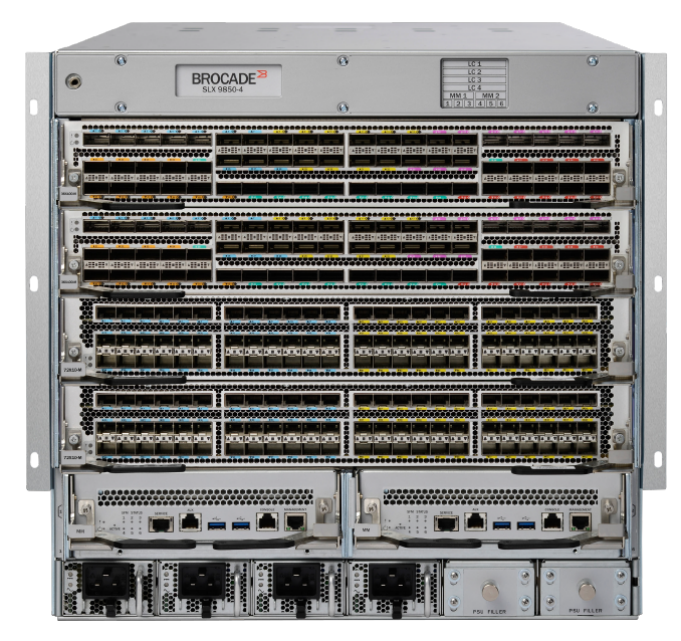


NSLS-II CORE NETWORK

- The NSLS-II enjoyed a long, uninterrupted period of reliable core network services:
  - >1,100 days w/out reboot
  - Outstanding, close engineer and developer support
  - Vertical integration with distribution and edge network
  - Expert in-house competencies
- However:
  - 2018 marked the End of Life (EoL) of our platform
    - Zero support afterwards
    - No bug/security/feature updates
    - Replacement components only available used (auction sites)
  - Lacked future scalability
    - 40 / 100G speeds not supported
    - Poor port density

CORE NETWORK UPGRADE

- The IT group sought a replacement that would:
  - Match and exceed existing features and reliability
  - Offer a platform that would scale well into the future
    - modular, high-perf, high-speed
  - Vertically integrate with existing infrastructure
  - Support modern standards/tools (e.g., virt., SDN, TRILL)
  - Able to be mastered with a reasonable learning curve
  - Have a distant EoL



- Completed arduous vendor evaluation (Brocade, Arista, Juniper, Cisco, HP)
- Based on overwhelmingly positive past experience, the Brocade SLX9850 was chosen
  - + Best-in-class features and performance
  - - New platform

CORE NETWORK "TEETHING ISSUES"

Although the units were tested and *burnt-in* beforehand, we suffered set-backs:

- Broadcom broke Brocade up and sold the pieces
  - Core router business sold to Extreme Networks
  - Exodus of Brocade employees – change in support channel and contacts
  - Limited EST access to expert engineers on our platform
- Incomplete and incorrect documentation
  - Questionable *Best Practices*
- New firmware not available on download site
- Protocol incompatibilities and quirks between other Brocade products!
- Oversight on our submitted configuration
- Hardware failure

CORE NETWORK CURRENT STATE

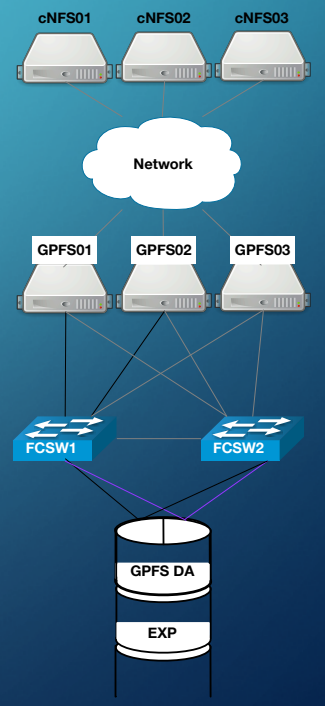
After a painful journey we believe all outstanding issues have been resolved in time for operations to begin anew

- The network is now *battle-tested* and running smoothly
- Marked performance increase across network – esp. GPFS access
- Monitoring tools are updated
- Ready to begin high-performance testing between NSLS-II and ITD for offsite data transfer and access to lab-wide computing resources

CURRENT PROJECTS

The IT group is working on several on-going projects:

- Replacement of existing NFS user filesystem (old system, poor performance, SPF) to a new architecture
  - Single NFS server → GPFS cluster + clustered NFS
- Migrate to lab account management and AA
- Cybersecurity vetting and policy documentation
- Tiered central storage expansion (SSD → HDD → tape)
- GPFS over infiniband
- Beamline distributed storage standardization and data policy
- NX Cloud Server performance tuning
- Offsite data transfer



FUTURE PROJECTS

- External SSH gateways
- Dedicated data transfer nodes
- Data transfer to central lab compute resources
- Data transfer to and utilization of cloud compute resources
- Centralized Mongo
- Further containerization of systems and services
- Open-source VM live migration
- Linux operating system revisit (RedHat)

THANK YOU!