USDF Infrastructure

Omar Quijano
Department Head, SCS-ISS

Riccardo Veraldi
Lead Network Engineer

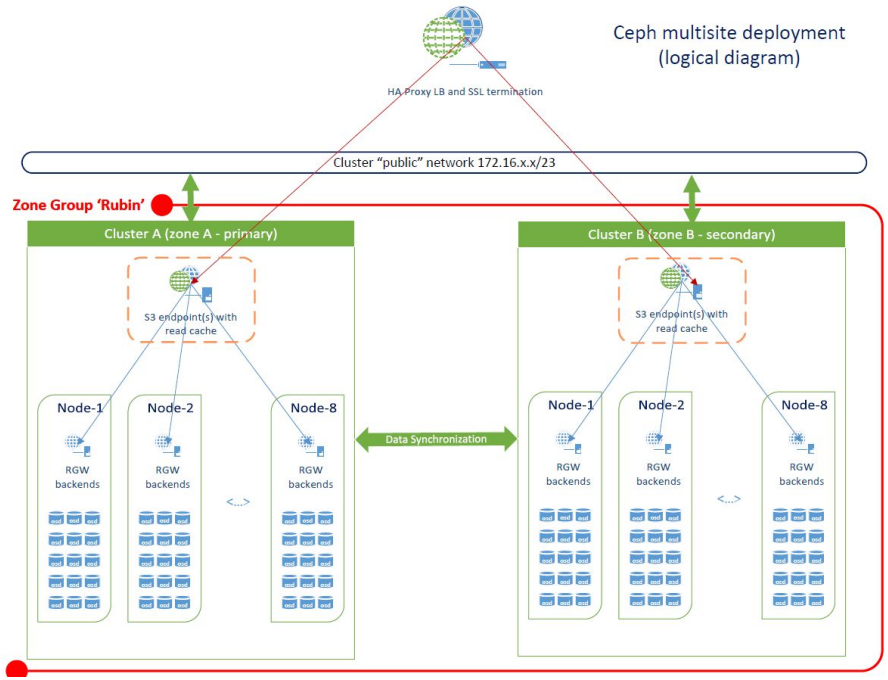
U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

Agenda

- Background
  - Embargo layout
    - Initial Design
    - Current Design
  - Network layout
    - Embargo Changes
    - S3DF Architecture
- PerfSONAR Deployments
- Work in Progress
- Q & A

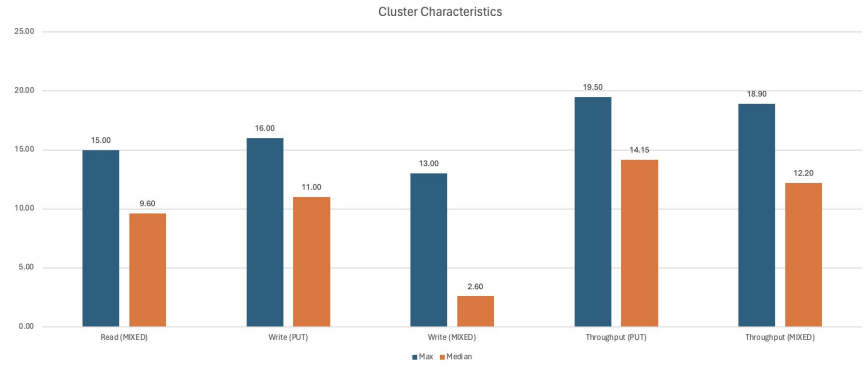
Embargo Layout: Initial Layout

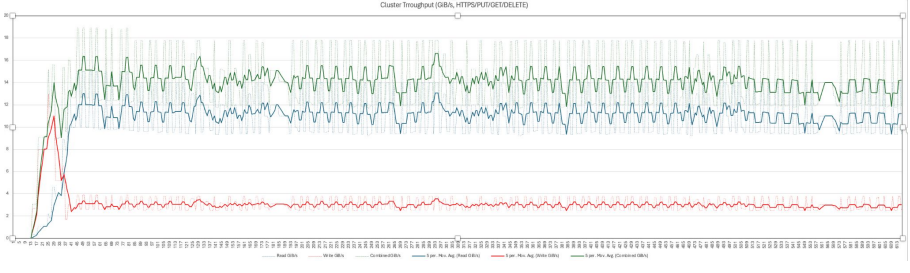


- Requirements
  - Resiliency (Maximize Uptime)
  - Steady State Capacity: 800 TiB
  - Encryption
  - Performance: 7s Latency
- Delivered
  - Usable 1.3 PiB

	T1		T2	
42	sdfemb101	42	sdfemb201	42
41		41		41
40	sdfemb102	40	sdfemb202	40
39		39		39
38	sdfemb103	38	sdfemb203	38
37		37		37
36	sdfemb104	36	sdfemb204	36
35		35		35
34	sdfemb105	34	sdfemb205	34
33		33		33
32	sdfemb106	32	sdfemb206	32
31		31		31
30	sdfemb107	30	sdfemb207	30
29		29		29
28	sdfemb108	28	sdfemb208	28
27		27		27
26	sdfembs3[101-102]	26	sdfembs3[201-202]	26
25		25		25

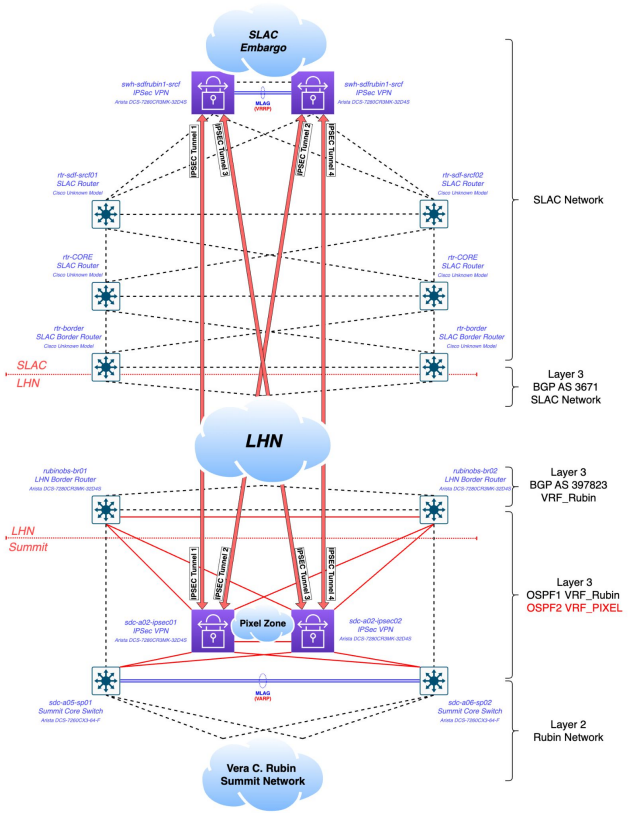
Embargo Layout: Existing Configuration





- Requirement
  - Increase Capacity to 3 PiB
  - No Cost (Current Hardware)
  - Maintain Performance
- Delivered
  - Usable
    - HDD: 4.7 PiB
    - NVMe: 2.1 PiB
- Security
  - Data encryption: LUKS
  - Key encryption: TPM
- Access Controls
  - Security Measures
    - Audit Login: Splunk
    - EDR: CrowdStrike
  - Hosts are managed by Ansible-Pull to ensures uniformity and consistency across servers.

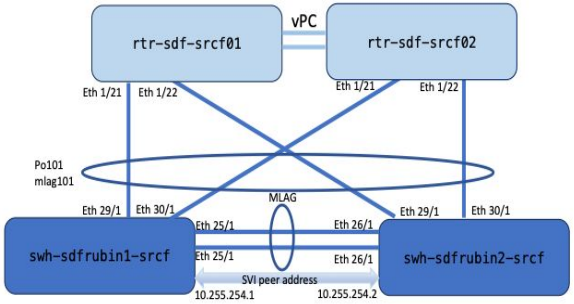
Network Layout: Embargo



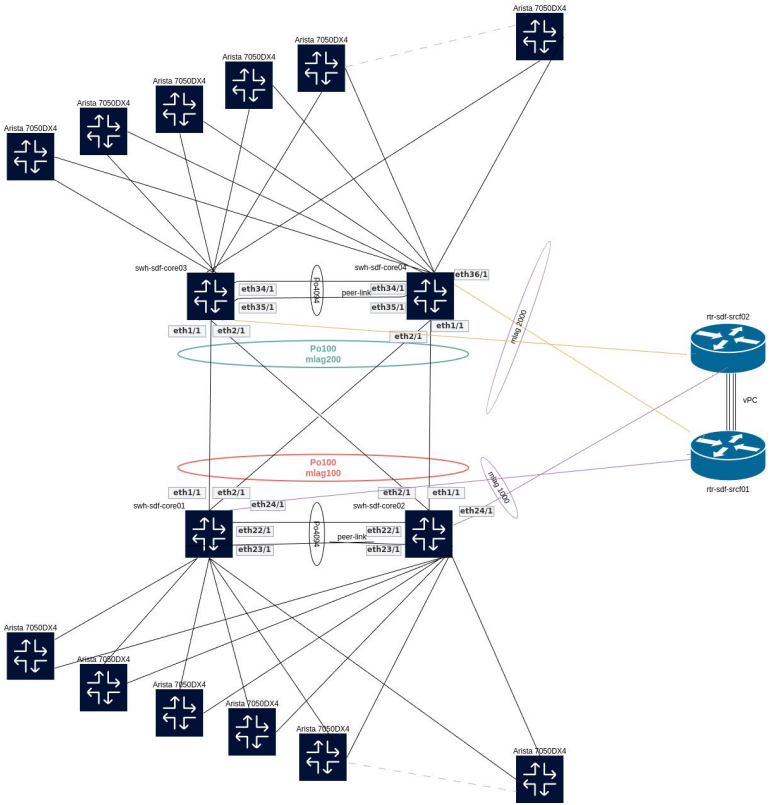
- Requirement
  - Secure
  - Performance

[SUM] 0.00-7.00 sec 4.42 GBytes 5.43 Gbits/sec 35492 sender
[SUM] 0.00-7.23 sec 4.01 GBytes 4.77 Gbits/sec receiver
iperf Done.

- Changes
  - Arista EOS upgraded to 4.34.0F
  - Added vlan sdf-rubin-pub for allowing ipv4 public address
  - ACLs on the ipv4 tunnel interfaces end-points to allow connections only from Summit VPN peer (drop brute-force attacks)

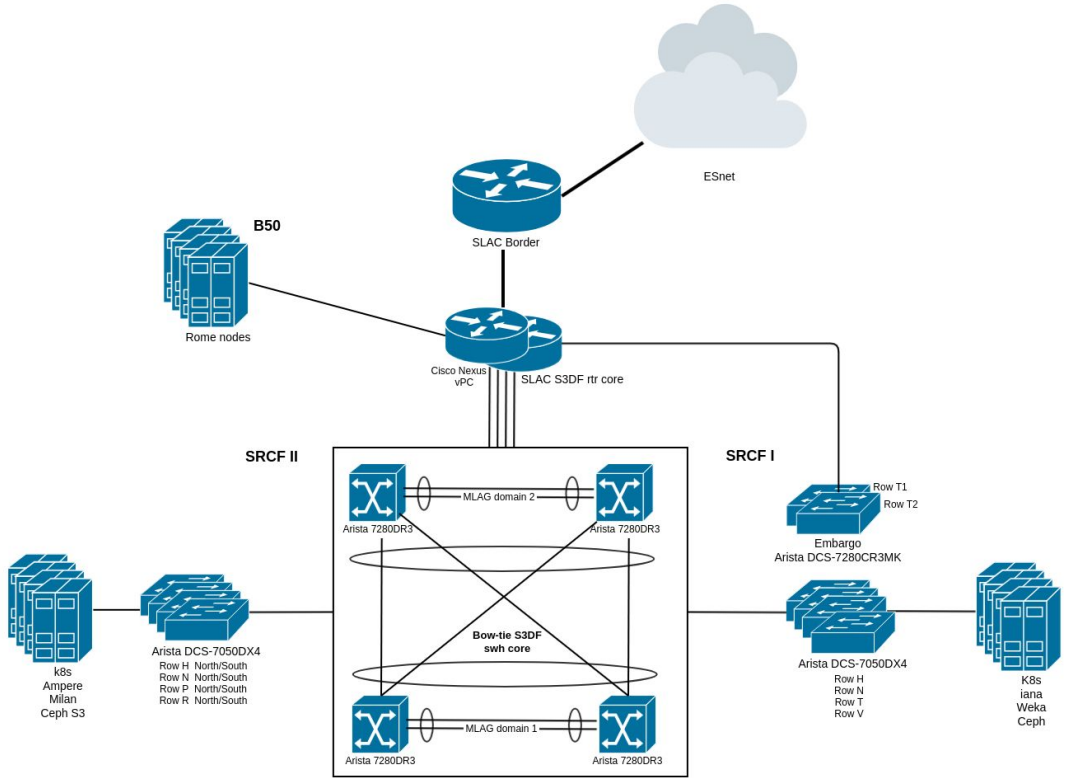


Network Layout: S3DF



- BowTie Model
  - Layer 2
    - Allows connections between two different MLAG switch pairs - 4x400G LACP port channel
  - Layer 3:
    - For each SVI define a **ip virtual-router address**
    - Inter-S3DF VLAN routing will be performed using the virtual ipv4 address (Assigned to all of the 4 cores)
  - Based on VARP
    - Enables both switches in an MLAG pair to route traffic *actively*, improving resilience and performance.
    - Unlike VRRP, VARP avoids extra hops by allowing *local routing*, reducing latency and simplifying the path.
  - Performance
    - 1600Gb backbone across cores
    - Leaf switches will be connected with 400Gb to each core (2x400G)

Network Overview



perfSONAR Deployment

- perfSONAR is a network performance monitoring tool used to detect and troubleshoot network issues across domains.
- Three servers:
  - Routed through LHN to SLAC border
    - 100Gbps capable, currently using rate limiting to 10Gbps
    - Operational
  - Routed through LHN/SLAC to S3DF/USDF (sdfperfsonar001)
    - 100Gbps
    - Operational
  - Routed IPSEC Tunnel (sdfembps001)
    - 2 x 100Gbps
    - In Progress
- Next Steps:
  - Finalize the routing to the IPSEC Tunnel
  - Work with AmLight to finalize integration

Work In Progress

- Improve overall Embargo Storage performance.
- Define better metrics to help understand bottlenecks.
- Visualize files within the hot or cold pool.
- Finalize perfSONAR deployment and integration.
- Improve internal monitoring and alerting within S3DF.
- Define better problem escalation processes.
- Staffing.

Q & A