Lustre @ SDCC

Vincent Garonne, Jason Smith, Jerome Lauret, **Qiulan Huang**, Zhenping Liu, Douglas Benjamin

Liaison Meeting, 04-06-2023

Lustre status @ SDCC

- Lustre storage: **~64.12PB** (Second biggest disk storage @SDCC)
- 8 production Lustre instances

  - **sPHENIX*, NSLS-II***, STAR/EIC, ATLAS HPC, BNLBOX (CSI), CFN

* High priority

sPHENIX (sphnx02)	sPHENIX (sphnx01)	NSLS2 (data2)	NSLS2 (data)	STAR/EIC (lustre02)	ATLAS HPCl (atlas01)	BNLBOX (CSI)	CFN (cfnlstr1)
5.0PB	38.2PB	3.5PB	3.5PB	8.7PB	3.0PB	219.0TB	2.0PB

Brookhaven National Laboratory

Lustre support @ SDCC

➜ Decrease of Lustre core expertise has initiated the (necessary) involvement of more SDCC teams (with storage expertise) for Lustre support

- Teams: IT Services, Services & Tools and Experimental computing Support
➜ Bi-weekly Lustre meeting has been established to work on a solid documentation / knowledge base set of recipes and best practice for knowledge retention

- Topics

  - Service operations & incidents

  - Tasks, plan & priorities

  - Expertise sharing,e.g., deployment model, settings

  - Synergy work on common components, e.g., upgrade procedure, configuration management, testsuite, cf., gitea repo

➜ Support: StorageManagement RT

The strategy is to gradually leverage the shared expertise in other instances, involve new members(e.g.) and iterate on the operational model. E.g., Teams/Names per instances, 1st/2nd/3rd level support

Brookhaven
National Laboratory

3

Lustre operations @ SDCC

A task-oriented approach inspired by the agile Kanban methodology

- sPHENIX priorities tasks
  - Support to testing & commissioning activities
  - Deployment strategy and extension of capacity (38.2 PB → 65.4 PB)
- NSLS-II priorities tasks
  - Upgrade plan to RH8+Lustre-2.15 #35177 — prerequisite for validation on prepro instance Possible(is undergoing)
  - Some broken directories cf. #35168, #35190
- Upgrade other instances upgrade to RH8+Lustre-2.15

Brookhaven
National Laboratory

Preproduction Instance

- To validate any changes/improvements/new Lustre versions prior to any changes in production
- Allowed the clients to connect to receive the compute load
- 4 Servers working for Lustre OSS with identical HW specifications

Server HW specifications
- 256GB RAM, 48 CPU cores
- Network - 2 x 25 Gbps = 50Gbps
- One JBOD per server
    a. 84 x 8TB drives
    b. ~500TB available
- HA configuration: Group the headnodes to let one JBOD attached to the pair nodes

Lustre Disk Organization
- 10 x (6+2) RAID 6 LUNs
- One LUN one OST
- ~ 500TB per OSS

Brookhaven National Laboratory

Preproduction Instance: Deployment

- Lustre MDS - Lustre v2.15.2
  - One VM - 1TB **HDD disk**, 16 cores, 64GB RAM
- Single Lustre file system constructed from 4 OSS servers (previous slide)
  - 2 PB total capacity (4 x 500TB)
- Configure Failover for Lustre OSS(High Availability)
  - Each JBOD attached to the pair nodes
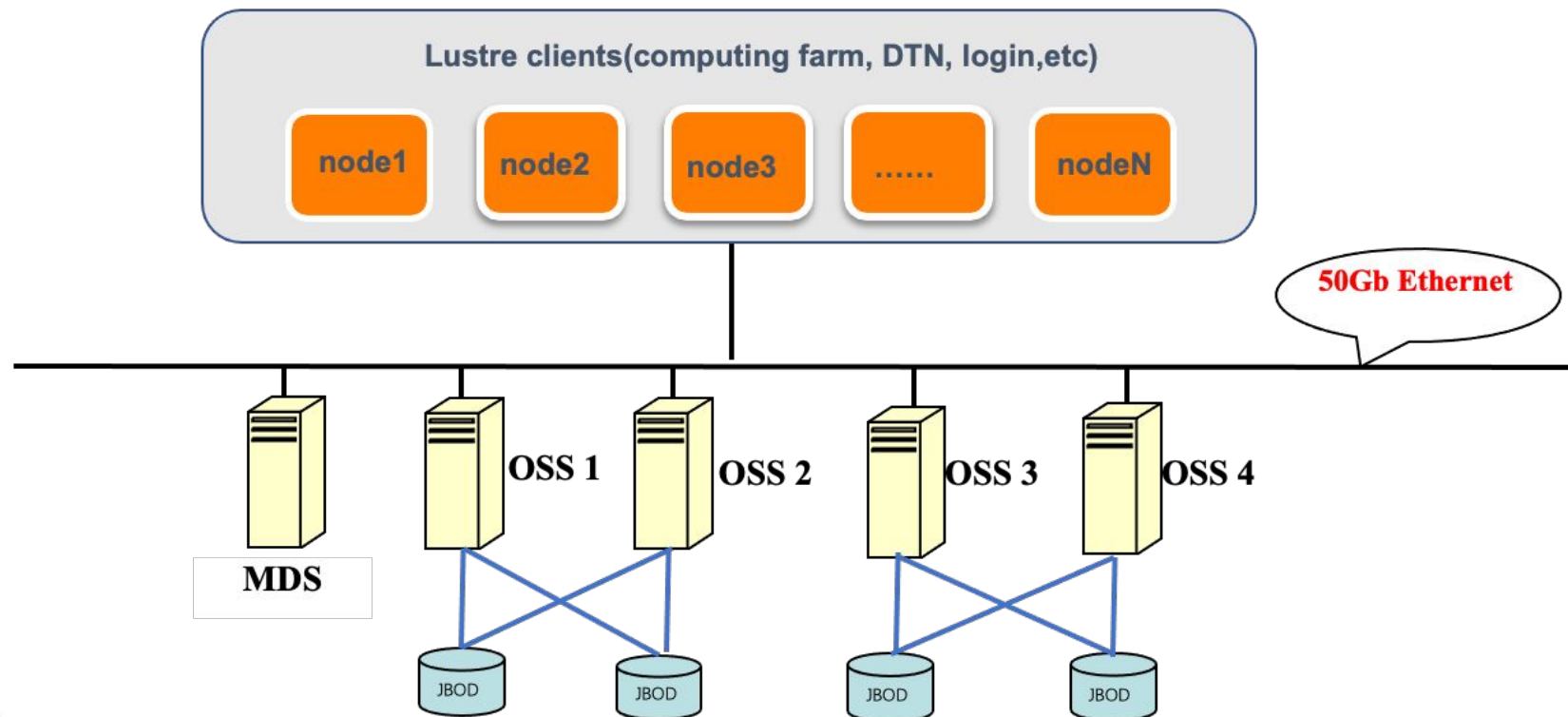
Continues Improvement/Delivery(1)

- High Availability
  - Lustre Failover solution
  - Each storage target (MGT, MGS, OST) must be associated with a second node to create a failover pair
  - External HA software(pacemaker, corosync and resource fencing)
    - node monitoring, failure detection, take care of the storage target once a failure detected
  - **Benefits: Hardware fault/Intervention on OSS or MDS can be transparent to clients except rebuilding the OS kernel**
- **Status**
  - NSLS2 is working with HA
  - Consider to extend to other instances

Brookhaven
National Laboratory

Continues Improvement/Delivery(2)

- Quota management
  - Quota is the most basic but useful storage management mechanism
  - Too many files inside one entry, which is not recommended. It would degenerate Lustre performance
- Quota for user, group, and project
  - user block
  - user inode
  - group block
  - group inode
  - project block
  - project inode

Brookhaven™
National Laboratory

Continues Improvement/Delivery(3)

- Performance monitoring
  - ○ Lustre specific: collect + Graphite+ Grafana
  - ○ General metrics: Ganglia
- Log collector and Filtering
  - ○ Grafana Loki
- Data popularity and performance optimization
  - ○ Introduce AI/ML solution
- Kernel-level iusse troublesthooting
  - ○ gdb, crashdump,SystemTap
- Deployment and configuration
  - ○ Puppet, ansible-playbooks

- Alarm
  - Nagios, email, etc
- IO benchmark
  - Metadata: mdtest
  - I/O throughput: fio, IOZone
  - Real experiment jobs
- Management scripts
  - /proc: Jobstat, extend_stat, offset_stat …

Monitoring and storage analytics

- ◆ Provide a uniform monitoring interface for all Lustre instances
- ◆ Help troubleshooting and performance optimization

Brookhaven National Laboratory

Activities we have done

- Provide a consistent set of monitoring information (access, performance, etc ...) for all instances
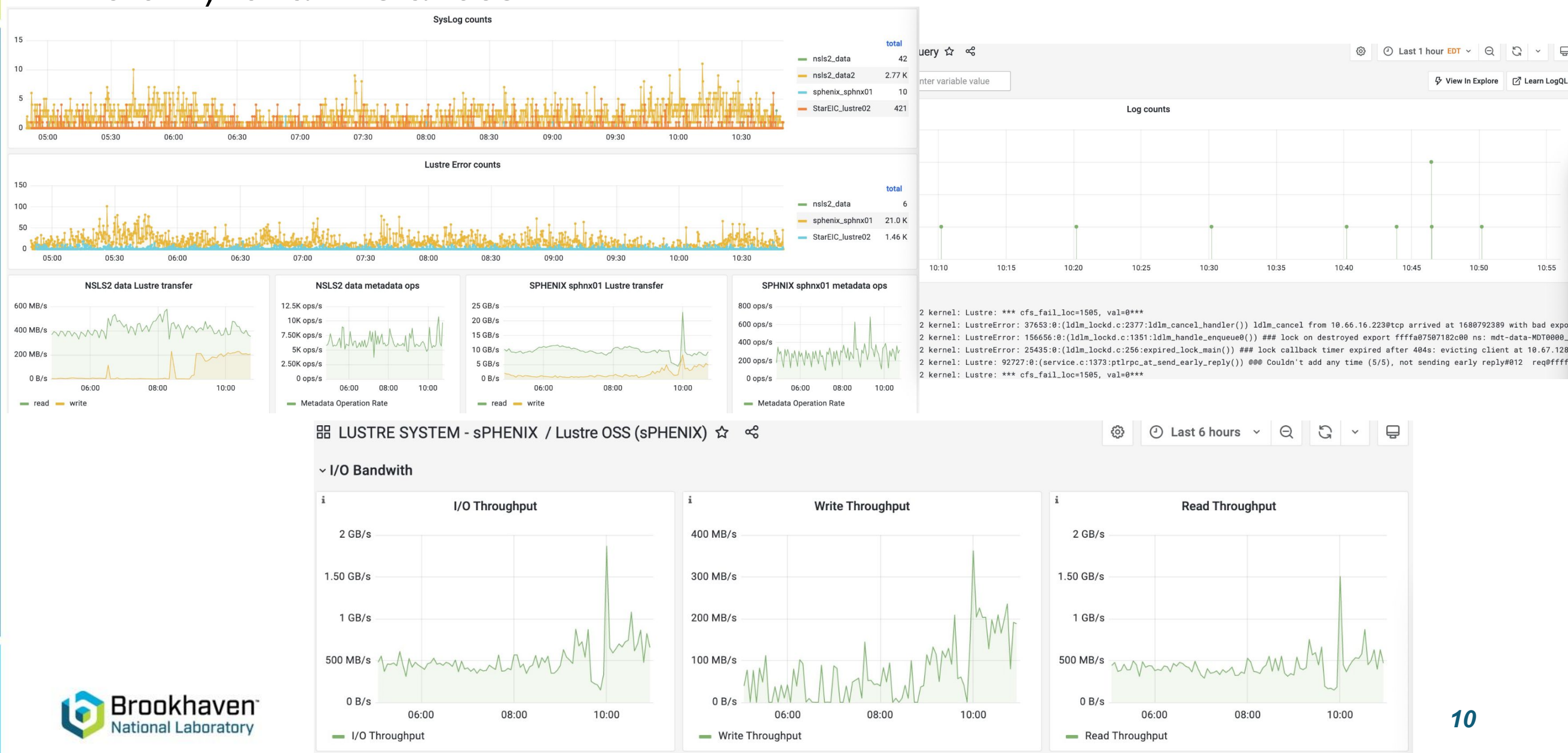
Undergoing activities

- Lustre upgrade testing (NSLS-II high priority task)
  - Hardware prepare and testbed setup
  - Prepare the Lustre upgrade procedure(From RH7+Lustre 2.12 to RH8+Lustre 2.15) for NSLS-II
    - Prerequisite validation for is undergoing
    - Hardware setup
    - Complete the OS and Lustre upgrade
    - Validate the data availability after upgrade(data is accessible)
    - Interoperability testing: Client Lustre 2.12 to access Server with Lustre 2.15
  - High availability(HA, Lustre failover solution) testing
  - NSLS-II upgrade intervention rescheduled to the next downtime in August(It was scheduled in this month)
- sPHENIX performance analysis

Brookhaven
National Laboratory

Enrichment Program

- Lustre training

  - To be more in-depth knowledge through a professional training

  - Take 9 training course modules to understand Lustre internals for the team

  - Held at the end of the month

- Be more active in Lustre community

  - lustre-discuss@lists.lustre.org

  - Lustre User Group

Brookhaven
National Laboratory

Thank you
Any questions?