High Performance Storage System

Incremental Scalability

Based on storage needs and deployment schedules, HPSS scales incrementally by adding computer, network and storage resources. A single HPSS namespace can scale from petabytes of data to exabytes of data, from millions of files to billions of files, and from a few file-creates per second to thousands of file-creates per second.

|			
Incremental Scalability	Performance & Efficiency	Availability & Durability	Delivery & Support

---

About HPSS : HPSS for Spectrum Scale (formerly GPFS)

Overview

- Not only is HPSS a highly scalable standalone file repository, a single HPSS may also provide disaster recovery and space management services for one or more Spectrum Scale file systems.
- Spectrum Scale customers may now store petabytes of data on a file system with terabytes of high performance disks.
- HPSS may also be used to backup your Spectrum Scale file system, and when a catastrophic failure occurs, HPSS may be used to restore your cluster and file systems.
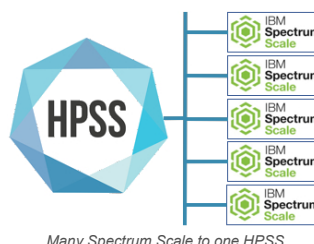
HPSS for Spectrum Scale
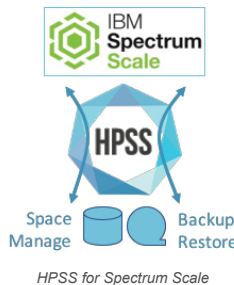
HPSS for Spectrum Scale key terms

- Spectrum Scale: A proven, scalable, high-performance data, and file management solution (based upon IBM General Parallel File System or GPFS technology). Spectrum Scale is a true distributed, clustered file system. Multiple nodes (servers) are used to manage the data and metadata of a single file system. Individual files are broken into multiple blocks and striped across multiple disks, and multiple nodes, which eliminates bottlenecks.
- Spectrum Scale ILM policies: Spectrum Scale Information Lifecycle Management (ILM) policies are SQL-like rules used to identify and sort filenames for processing.
- Disaster recovery services: Software to capture a point-in-time backup of Spectrum Scale to HPSS media on a periodic basis, manage the backups in HPSS, and restore Spectrum Scale using a point-in-time backup.
- Space management services: Software to automatically move data between high-cost low-latency Spectrum Scale storage media and low-cost high-latency HPSS storage media, while managing the amount of free space available on Spectrum Scale. The bulk of data are stored in HPSS, allowing Spectrum Scale to maintain a range of free space for new and frequently accessed data. Data are automatically made available to the user when accessed. Beyond the latency of recalling data from tape, the Spectrum Scale space management activities (migrate, purge and recall) are transparent to the end-user.
- Online storage: Data that are typically available for immediate access (e.g. solid-state or spinning disk).
- Near-line storage: Data that are typically available within minutes of being requested (e.g. robotic tape that is local or remote).

Many to one advantage of HPSS for Spectrum Scale

- A single HPSS may be used to manage one or more Spectrum Scale clusters.
- ALL of the HPSS storage may be shared by all Spectrum Scale file systems.
- One Spectrum Scale file system may leverage all HPSS storage if required.
- The Max Planck Computing and Data Facility (MPCDF, formerly known as RZG) is space managing and capturing point-in-time backups for seven (7) Spectrum Scale file systems with ONE HPSS.

Many Spectrum Scale to one HPSS

What's New?

DOE Announces HPSS Milestone - Todd Heer, Deputy Program Lead, Advanced Simulation and Computing (ASC) Facilities, Operations, and User Support (FOUS), announced that DOE High Performance Storage Systems (HPSS) eclipse one exabyte in stored data.

Atos Press Release - Atos boosts Météo-France's data storage capacity to over 1 exabyte in 2025 to improve numerical modeling and climate predictions. Want to read more?

HPSS 9.2 Release - HPSS 9.2 was released on May 11th, 2021 and introduces eight new features and numerous minor updates.

HPSS 9.1 Release - HPSS 9.1 was released on September 24th, 2020 and introduces a few new

What HPSS software is installed on the Spectrum Scale cluster?

- HPSS Session software directs the space management and disaster recovery services.

- HPSS I/O Manager (IOM) software manages data movement between HPSS and Spectrum Scale.

- HPSS Session software is configured on all Spectrum Scale Quorum nodes but is only active on the CCM (Spectrum Scale Cluster Configuration Manager) node.
- The CCM may failover to any Quorum node, and the HPSS Session software will follow the CCM.

HPSS Software

Where HPSS software runs in the cluster

- HPSS IOM software may be configured to run on any Spectrum Scale node with an available Spectrum Scale mount point.
- There are five processes that comprise the HPSS Session software: HPSS Process Manager, HPSS Mount Daemon, HPSS Configuration Manager, HPSS Schedule Daemon, and HPSS Event Daemon.
- There are three processes that comprise the HPSS IOM software: HPSS I/O Manager, HPSS I/O Agent, and HPSS ISHTAR.
- HPSS I/O Agents copy individual files while ISHTAR copies groups of files (ISHTAR is much like UNIX tar, but faster, indexed, and Spectrum Scale specific).
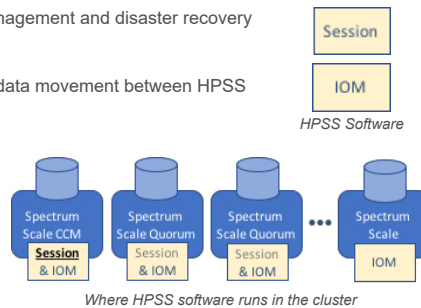
Additional Spectrum Scale cluster hardware

- HPSS IOMs are highly scalable and multiple IOMs may be configured on multiple Spectrum Scale nodes for each file system.
- Bandwidth requirements help determine the expected node count for a deployment.
- New Spectrum Scale quorum nodes for HPSS may need to be added to the cluster.
- HPSS for Spectrum Scale is typically deployed on a set of dedicated nodes that will become the Quorum nodes.
- HPSS Session and IOM software are configured on the new Quorum nodes.

Dedicated HPSS nodes becomes Quorum nodes
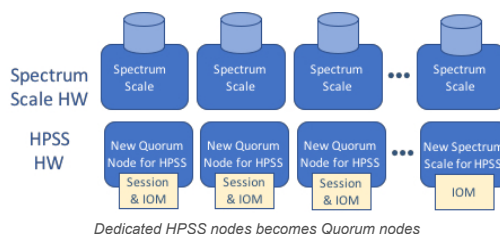
Space manage Spectrum Scale with HPSS

- Periodic ILM policy scans are initiated by the HPSS Session software to identify groups of files that must be copied to HPSS.
- The HPSS Session software distributes the work to the I/O Managers. Spectrum Scale data are copied to HPSS in parallel.
- The HPSS advantage is realized with two areas: (1) high performance transfers; and (2) how data are organized on tape.

IOMs send data to HPSS Movers periodically

Data flows to HPSS on migration

- When Spectrum Scale capacity thresholds are reached (the file system is running out of storage capacity), unused files are purged from Spectrum Scale, but the inode and other attributes are left behind.
- All file names are visible, and the user may easily identify which files are online and which files are near-line.

- The HPSS Session software will automatically recall any near-line file from HPSS back to Spectrum Scale when accessed.
- Tools to efficiently recall large numbers of files are also provided.

IOMs receive data from HPSS when files are recalled back to Spectrum Scale

Data flows to Spectrum Scale on recall
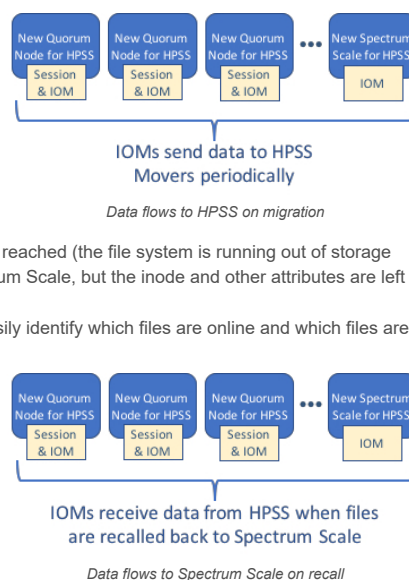
Backup and disaster recovery with HPSS for Spectrum Scale

- The backup process captures the following data:
  - File data - the space management process (discussed above) is the process used by HPSS to capture the data for each file.
  - Namespace data - the Spectrum Scale name space is captured using the Spectrum Scale *image backup* command.
  - Cluster configuration data - the cluster configuration is saved to HPSS to protect the cluster configuration.

- The HPSS disaster recovery processing minimizes data movement and is ideal for high performance computing (HPC) environments where the goal is to bring the namespace back online quickly.

< Home

Home | About HPSS | Services | Contact us