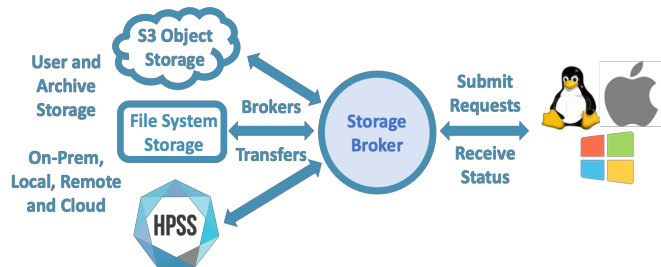# HPSS Storage Broker

Storage Broker is a new HPSS interface built to improve movement, organization, and management of long-lived data across HPSS and hybrid multi-cloud environments. Administrators, owners, and consumers of archive data will enjoy the benefits of Storage Broker to increase the availability, durability, movement, and sharing of digital assets. Tape storage allows Storage Broker to further protect archive data against malicious attacks by introducing an air gap, and tape storage cuts the carbon footprint of archive data.

Hybrid multi-cloud storage

Storage Broker delivers hybrid cloud storage by integrating multiple private or public cloud storage services (like Amazon S3 and IBM Cloud Object Storage) with on-premises storage (including HPC clustered file systems, S3 object stores, Network Attached Storage and HPSS disk and tape).

Simplifying data transfers

Storage Broker streamlines the transfer of archive data using a unified interface to archived data across multiple storage systems and locations. Storage Broker collects user transfer requests and efficiently processes them asynchronously. Users continue working on other tasks while the Storage Broker schedules and completes transfers on their behalf.

Organize data as preservation objects

Storage Broker protects long-lived data against threats, including corruption of the digital content, organizational changes, and obsolescence of hardware and software. Storage Broker employs Open Archival Information System (OAIS) concepts for digital archives by storing archived data as self-describing, self-contained, data containers using the SNIA Self-contained Information Retention Format (SIRF). Storage Broker SIRF-based preservation container format and archive strategy allows institutions to organize their long-lived data for future generations.

Value of data containerization

Storing and recalling a dataset comprised of many objects as a preservation container allows Storage Broker to reduce the physical number of data files and objects placed into archive storage, which reduces the burden on the underlying storage system (e.g., a billion-file archive is reduced to a million-container archive). Additionally, large streaming transfers are best for disk and tape storage performance, and Storage Broker preservation containers increase the size of data transfers thereby improving the transfer performance to the underlying storage system. Thus, preservation containers allow Storage Broker to meet transfer requirements with less hardware (e.g. fewer tape drives and fewer storage units).

Project data management and sharing

Project ownership of data improves HPSS project resource management and simplifies archive data ownership and sharing. Administrators assign project owners to a project. The project owners grant and revoke access and assign user roles which gives project owners control of project data. Data written by a user is owned by the project. When a user leaves, a project owner simply changes the user's access and roles.

Data placement, protection and retrieval

Project storage policies direct where project data are stored. Users select the data to be archived and the project storage policy, and Storage Broker does the rest.

Project storage policies also direct how data are protected. Storage Broker replicates data by placing copies on local, remote and cloud storage. Compared to making a second or third copy of data, employing a project storage policy that enables erasure encoding protects data and cuts redundancy costs by 50% and more! Storage Broker automatically rebuilds missing or corrupt data on recall using the erasure codes. Timely data recovery improves user satisfaction.

Project storage policies also control how and from where archived data are recalled. Project owners assign policies for primary recall copies and copies used for failover purposes. Optionally, project owners add an approval process for recalls. For example, additional fees may be incurred for recalls and an additional project approval is desired.

Minimizing transfers cuts storage expenses

Storage Broker efficiently transfers archive data between high performance user storage (HPC storage and Campaign storage) and lower cost archive storage (tape and cloud). Storage Broker copies data from user storage directly to archive storage without transfers through middle storage tiers. Direct transfers eliminate tiers of storage which saves money.
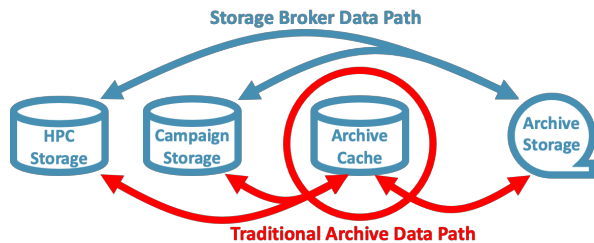
Archive storage independence

Storage Broker provides a unified interface to shareable data containers of files and objects across storage systems. Users easily archive data from one storage system and recall directly to another storage system. For example, archive from S3 to HPSS tape and recall from HPSS tape directly to an HPC file system. Storage Brokers data transfers are streamlined compared to tightly coupled file system and archive storage solutions that use "stubs" require data to be recalled through the file system before being copied to the new destination.

Value of Storage Broker database

Storage Broker provides a single interface to archive storage that is enhanced by using a robust high-performance database. This database contains all details of the preservation containers stored across the archive storage endpoints for all projects. Storage Broker also mirrors digital archive metadata to this high-performance database to improve user request throughput, search time, and preservation object metadata and namespace recovery time without impacting archive storage.

Storage Broker is the competitive advantage

Storage Broker encapsulates a set of differentiating strategies that modernizes data storage techniques leading to a competitive advantage for archive storage:

- Hybrid multi-cloud storage support.
- Simplified high performance data transfers.
- Improved project data management through assigned storage policies and user access and roles.
- Embracing a preservation approach for storing data.
- Supporting techniques for archive data durability.
- Optimizing transfers.
- Promoting archive storage independence.

Let's talk about your archive storage needs

Ramin Nosrat   Project Executive   ramin@us.ibm.com
Jim Gerry   Architect/Consultant   jgerry@us.ibm.com