Cohesity SmartFiles: Simplify Splunk Data Management

Many data sources, including machine and artificial intelligence, are driving unstructured data growth. As unstructured data volumes continue to grow, knowing what data exists and where it resides is critical. So is how to search, analyze, and visualize the unstructured and machine-generated data generated by websites, applications, and IoT devices, which can be analyzed for more meaningful insights at different stages of the data lifecycle. That's why Cohesity, Splunk and Cisco are so valuable; combining compute, storage, and analytics at scale.

Cohesity and Splunk

Cohesity SmartFiles, deployed on Cisco Unified Computing System (UCS), simplifies the management of warm and frozen Splunk data at scale, particularly in Splunk SmartStore-enabled environments. Smartstore decouples compute and storage, allowing indexers to use local storage as a cache for active data while the bulk of data resides in a remote object store i.e. Cohesity SmartFiles deployed on S3-compatible storage.

The Splunk data lifecycle

Splunk helps administrators gain operational intelligence from machine data improving customer service and business decision-making. Events in Splunk are stored as raw, compressed data with indexes that make them searchable–collectively forming the Splunk Enterprise Index also called Indexer.

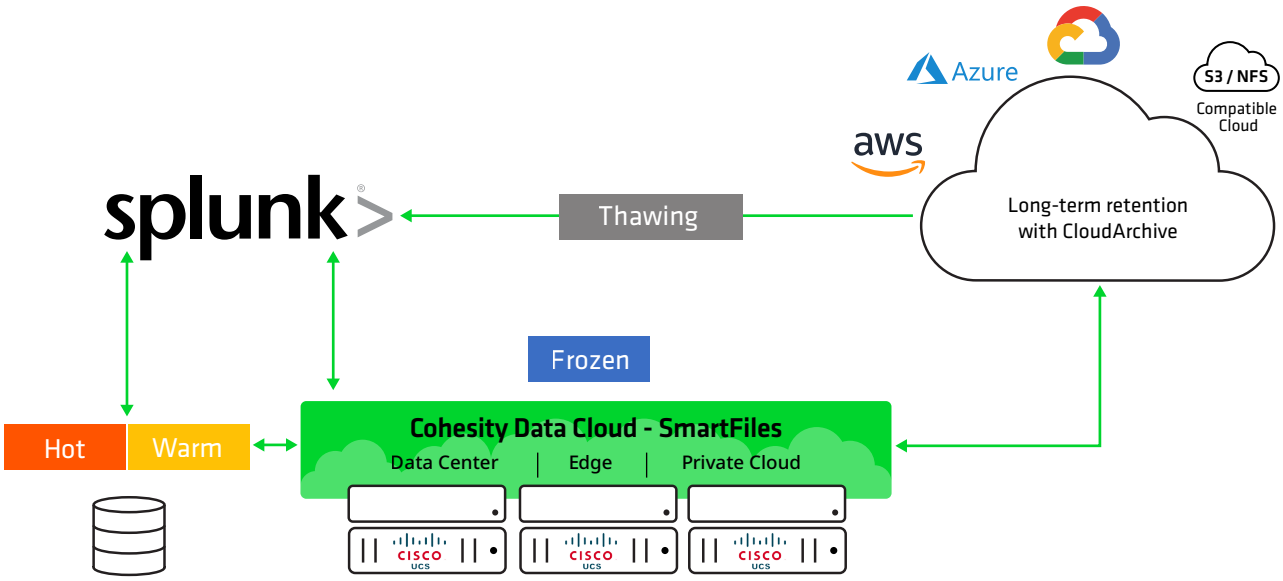
The Splunk Enterprise index typically consists of many buckets, organized by age of the data. Splunk organizes indexed data into directories called "buckets" based on data age. As data grows, the indexer automatically manages it through a policy-driven lifecycle across four stages:

Bucket Stages	Descriptions	Searchable?
Hot	Contains newly indexed data. Open for writing. One or more hot buckets for each index.	Yes
Warm	Data rolled from hot. There are many warm buckets. Data is not actively written to warm buckets.	Yes
Frozen	Data rolled from warm in object store. The indexer deletes frozen data by default, but enterprises can choose to archive it instead. Archived data can later be thawed.	No
Thawed	Data restored from an archive. If enterprises archive frozen data, they can later return it to the index by thawing it.	Yes

Key Benefits

Cohesity SmartFiles with Splunk:

- Reduces storage costs for large SmartStore Repositories with a space and cost-efficient storage tier
- Scale-out platform with unlimited scalability, non-disruptive upgrades, and pay-as-you-grow scalability
- Multiprotocol access supporting NFSv3, SMB2.x, SMB 3.0, and S3
- Global deduplication and compression, with Google-like search across all file and object metadata
- Public cloud integration–Amazon Web Services, Microsoft Azure, and Google Cloud Platform–for archival, tiering, and replication
- SEC 17A4 WORM Compliance with DataLock



Single storage domain

Ever increasing regulatory and industry requirements surrounding data retention mean that enterprises indexing large amounts of data cannot simply allow default processes to proceed as scheduled. Instead, enterprises must carefully plan and orchestrate specific aging policies for data depending on the bucket stage of the data in the Splunk platform. However, customization of the indexer flow presents IT teams relying only on Splunk for data storage with a new challenge: where to store data for varying lengths of time and how to quickly retrieve it when needed?

Cohesity SmartFiles running on Cisco UCS is an ideal data management platform for caching hot buckets and offloading warm, and frozen buckets, ensuring low latency indexing and search performance. As data ages and moves to frozen stages, organizations can **offload to Cohesity SmartFiles, running on UCS for lower TCO, freeing up valuable resources for active workloads.**

Storing frozen Splunk buckets in Cohesity SmartFiles

SmartFiles addresses long-term storage challenges with a webscale platform that eliminates siloed infrastructure. IT teams can configure paths for each Splunk bucket stage (hot, warm, frozen), allowing them to store data in the most appropriate location based on cost, performance and compliance needs.

Hot and Warm buckets continue on high-performance primary storage within SmartStore indexers. This gives Splunk administrators full visibility while optimizing storage cost and performance across the data lifecycle. Built on the Cohesity Helios platform, SmartFiles offers globally distributed NFS, SMB, OpenStack Swift, and S3 object storage. Provisioned as "Views," these volumes leverage capabilities such as scale-out architecture, global variable-length deduplication and compression, and unlimited snapshots and clones,

Cohesity SmartFiles capabilities for storing Splunk buckets:

Capacity optimization	Globally, variable-length deduplicated data is distributed across all nodes
Global search	Powerful indexing of all files and object metadata to enable global Google-like search
Quotas	Volumes, file shares, and object buckets
Security	Software-based encryption
Cloud Integration	Native support for Amazon Web Services, Microsoft Azure, and Google Cloud Platform for policy-based archival, tiering, and replication

Finally, as warm data programmatically moves in the Splunk index to the frozen bucket, the data will move to a Cohesity View with low QoS settings, which can still be accessed via S3 endpoints. Splunk uses a cache manager to retrieve noncached buckets from SmartFiles when required for search i.e. data in a frozen bucket cannot be searched and requires manual intervention. Enterprises opting for archival can take advantage of Cohesity CloudArchive for long-term retention and archival to reduce reliance on tape and lower TCO, while also gaining an easy way to retrieve data back on-premises or recover data to a different site.

By combining Cohesity SmartFiles, Splunk, and Cisco UCS, enterprises can uncover more value from all their data while achieving data retention goals and meeting regulatory requirements.

Cohesity can also send the alerts and logs to Splunk and thus ensure smooth integration with Cohesity.

If your enterprise is already using Splunk, SmartFiles is the ideal complementary storage solution.

Splunk SmartStore datasets, Cohesity can also be a source of Splunk data using the Cohesity add-on for Splunk, which is available here.

COHESITY

cohesity.com
1-855-926-4374
2625 Augustine Drive, Santa Clara, CA 95054

3000061-004-EN 5-2026