

## Uncover More Value from All Your Data While Achieving Data Retention Goals

A number of sources, including machine data and artificial intelligence, are driving unstructured data growth. As data volumes continue to increase exponentially, knowing what data exists and where it resides is critical. So is how to search, analyze, and visualize the unstructured and machine-generated data businesses capture from websites, applications, and IoT devices for more meaningful insights at different stages of the data lifecycle. That's why Cohesity and Splunk are so valuable.

### Cohesity and Splunk

Cohesity empowers organizations to simplify the management of all secondary data and apps at web-scale, on a single software-defined platform. Complementary technologies, both Cohesity and Splunk, a purpose-built platform to make machine data accessible, usable, and valuable at scale to businesses, simplify the management of massive amounts of data and help enterprises achieve their data retention goals.

### The Splunk Data Lifecycle

Splunk helps administrators find answers in their massive amounts of machine data to gain operational intelligence that supports better customer service. In Splunk, events are stored as raw data in a compressed format and add indexes that point to the raw data. Together, these files constitute the Splunk Enterprise Index, also known as the indexer, with data now transformed into searchable events waiting to respond to new search requests.

The Splunk Enterprise index typically consists of many buckets, organized by age of the data. Some directories contain newly indexed data; others contain previously indexed data, and the number of directories can become large, depending on how much data is being indexed. As data ages, Splunk's indexer gracefully handles and moves data through the following five stages in a transparent, policy-driven mode—where policies are set as attributes in the index configuration—until by default at a preset time, it removes old data from the system:

Bucket stages	Description	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
Cold	Data rolled from warm. There are many cold buckets.	Yes
Frozen	Data rolled from cold. 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

Yet 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 specific long-term data retention and compliance mandates require enterprises to 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

### Key Benefits

- Cohesity and Splunk together reduce overall data protection costs with a space and cost-efficient secondary data and apps storage tier
- Web-scale platform with unlimited scalability, non-disruptive upgrades, pay-as-you-grow scalability
- NFS, SMB, and S3 compatible object storage with multiprotocol access supporting NFSv3, CIFS, SMB2.x, SMB 3.0, and S3 APIs
- Global deduplication and compression
- Global Google-like search on all file and object metadata
- User and file system quotas with audit logs
- Public cloud integration—Amazon Web Services, Microsoft Azure, and Google Cloud Platform—for archival, tiering, and replication



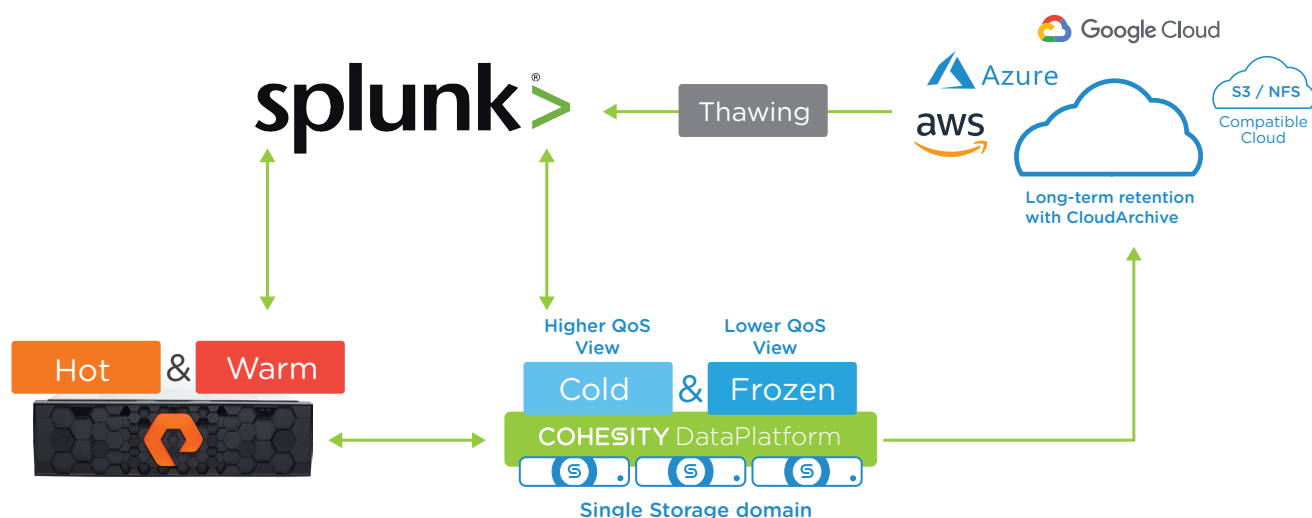
only on Splunk for data storage with a new challenge: where to store data for varying lengths of time and how to quickly retrieve it, if and when needed.

## Storing Cold and Frozen Splunk Buckets in Cohesity

Cohesity overcomes variable-length data storage obstacles by providing enterprises with a hyperconverged platform designed to converge secondary data silos, including backup and recovery infrastructure and target storage. Because the paths for the hot/warm, cold, and thawed directories are configurable, enterprises can store buckets in one stage in a separate location from buckets in other stages for optimal access and streamlined compliance.

Cohesity is the ideal solution to store cold and frozen buckets of Splunk, while providing complete visibility to the Splunk administrator via the Splunk interface. Concurrently, hot and warm data can be stored in primary storage arrays, such as Pure Storage, and accessed via Splunk. Then as data becomes cold, it will automatically move to a Cohesity View with high quality of service (QoS) settings that can be accessed via NFS endpoints.

Cohesity simplifies cold and frozen data storage by providing enterprises with globally distributed NFS, SMB, and S3 object storage on Cohesity DataPlatform, a revolutionary web-scale secondary data and application platform. NFS, SMB, and S3 volumes are provisioned as Views on Cohesity and leverage all of the DataPlatform capabilities including scale-out storage, global variable-length deduplication and compression, and unlimited snapshots and clones.



Cohesity capabilities for storing Splunk buckets include the following:

Capacity optimization	Globally, variable-length deduplicated data is distributed across all the 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 cold 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 NFS endpoints. However, data in a frozen bucket cannot be searched and requires manual intervention to be either deleted or archived. Enterprises choosing the archival option can take advantage of Cohesity long-term retention and archival with CloudArchive, reducing reliance on tape and lowering TCO while also gaining an easy way to retrieve data back on-premises or recover data to a different site.

By combining Cohesity and Splunk, enterprises can uncover more value from all their data while achieving data retention goals. They also can modify policies as needed quickly to meet ever-changing industry and regulatory requirements. If your enterprise is already using Splunk, Cohesity is the ideal complementary secondary storage solution.

To learn more, visit [www.cohesity.com](http://www.cohesity.com).