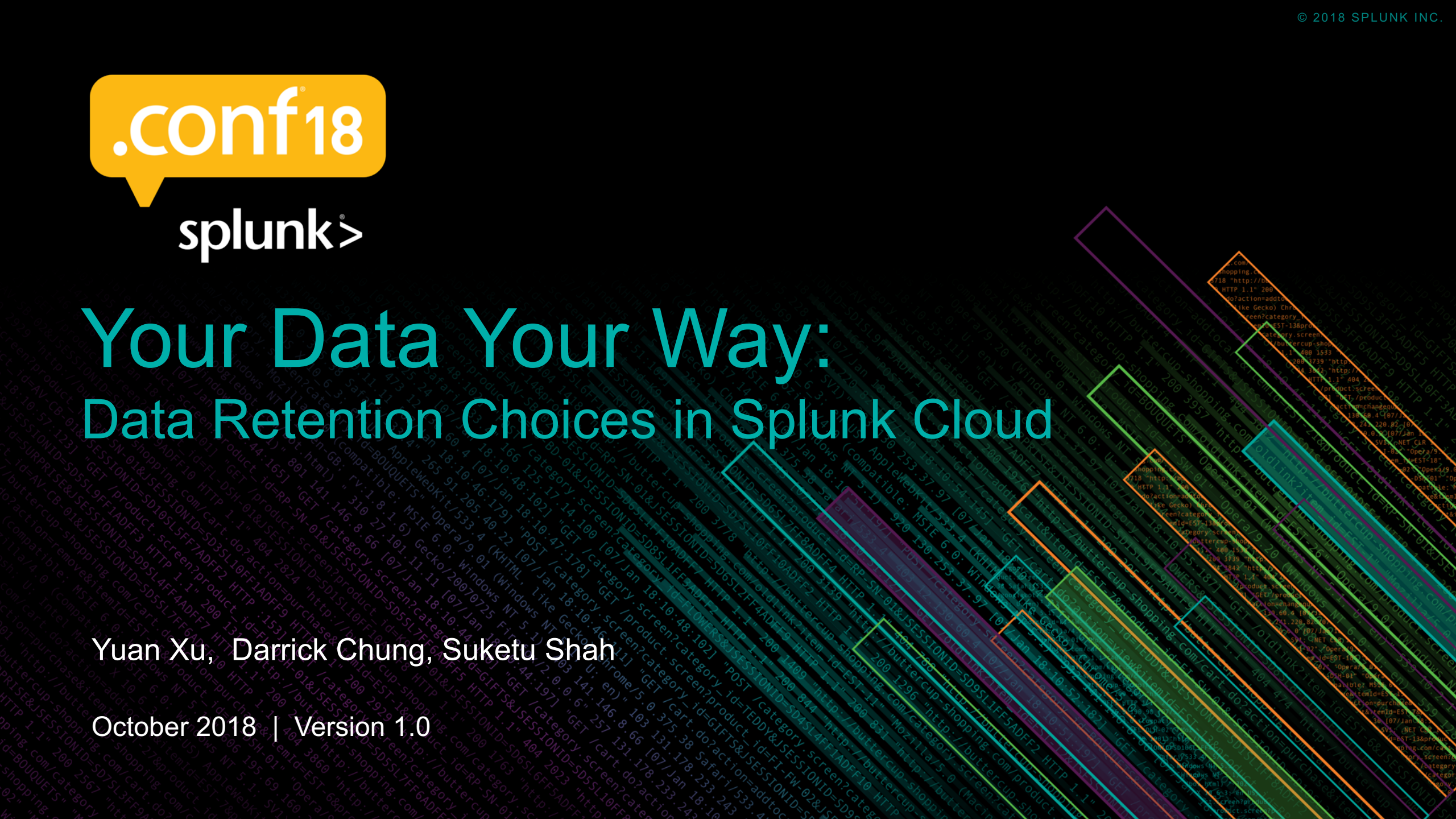


Your Data Your Way:

Data Retention Choices in Splunk Cloud

Yuan Xu, Darrick Chung, Suketu Shah

October 2018 | Version 1.0

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

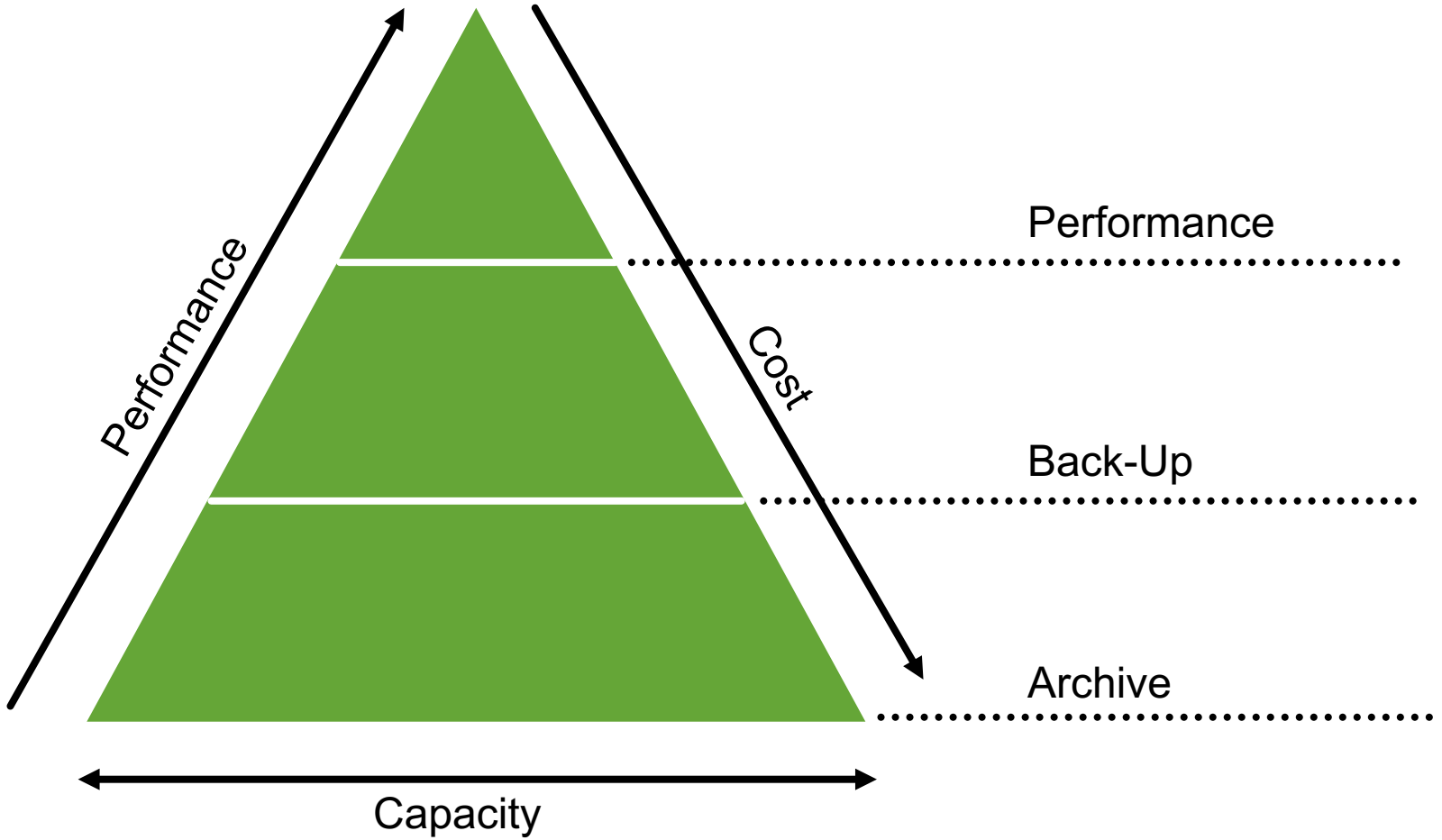
splunk> .conf18

Agenda

This is where the subtitle goes

- Enterprise storage tiers
- Changing customer requirements
- Data retention options
- Design principles
- Technical Architecture
- Demo
- Beta program feedback
- Q&A

Enterprise Storage Tiers



- Actively used data
- Frequent reads, write
- E.g. Analytics use case

- In-frequent data access
- Short-term back-up
- E.g. Disaster recovery use case

- Long term retention
- E.g. Compliance use case

Changing Customer Requirements

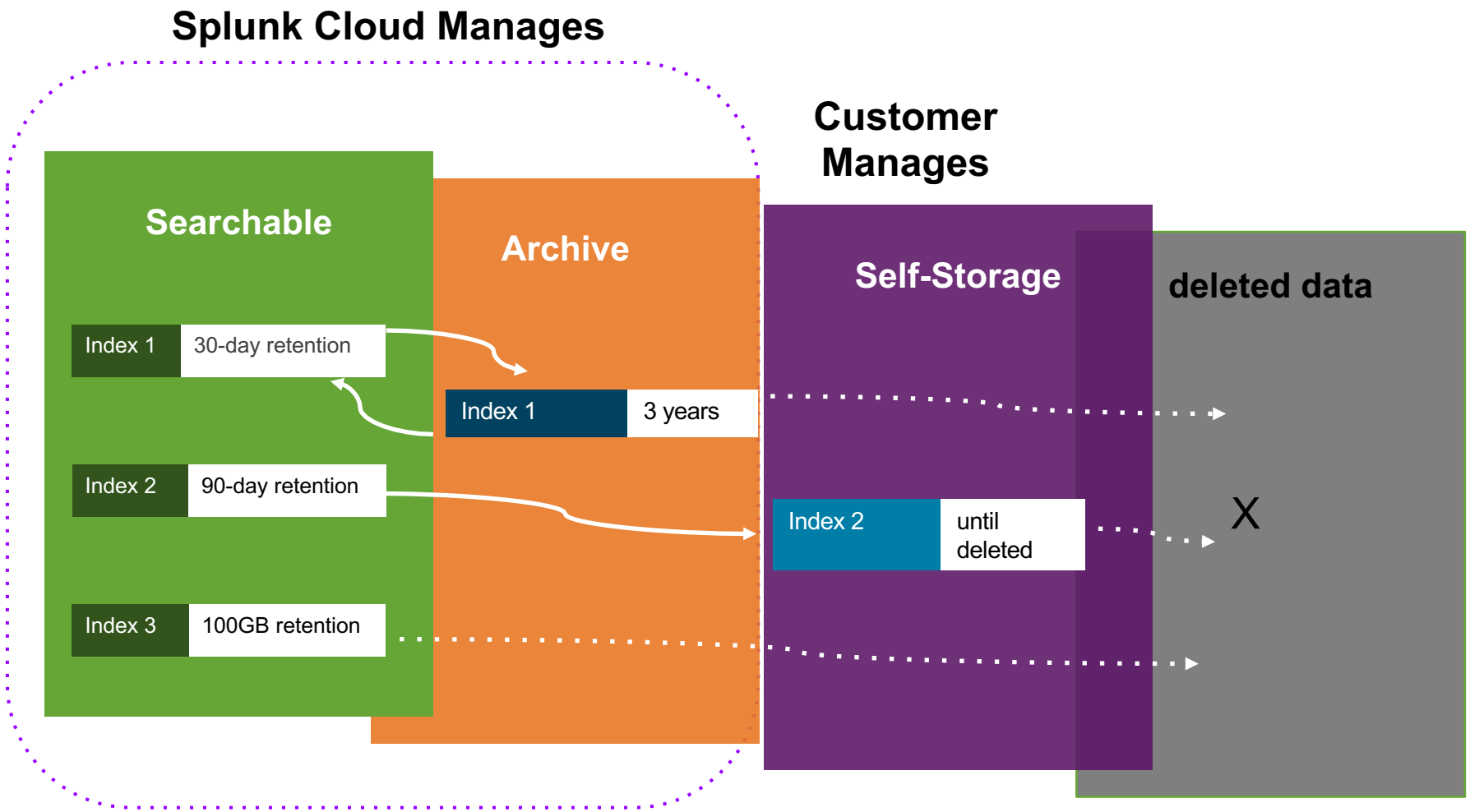
Regulatory compliance

Security Incident investigation

Advanced Analytics

Cost

Data Retention Options (1 of 2)



Data Retention Options (2 of 2)

Splunk Cloud Manages

Searchable

Actively used data

Per Index retention rules (volume or time based)

Available to all Splunk Cloud Customers

90 days of ingest entitlement storage included, purchase additional in 500GB increments

Archive

Long term retention

0 days included in subscription

Lower price/performance alternative to Searchable

Customer Manages

Self-Storage

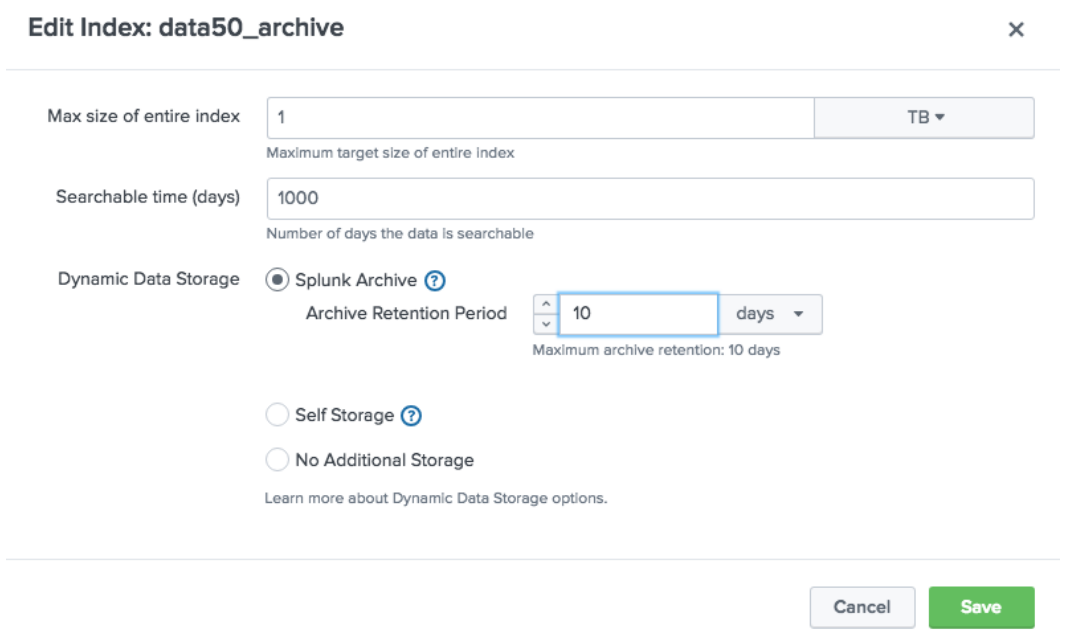
Customers manage their own data

Data resident in customer's AWS S3 account

Included in subscription, customer pays for S3 storage

splunk> .conf18

Dynamic Data Active Archive – Data Aging



Searchable

Per Index retention rule, Aged data is automatically copied to archival storage

On copy completion aged data is deleted from index.

Splunk manages data lifeccyle

Aged Data

Archive

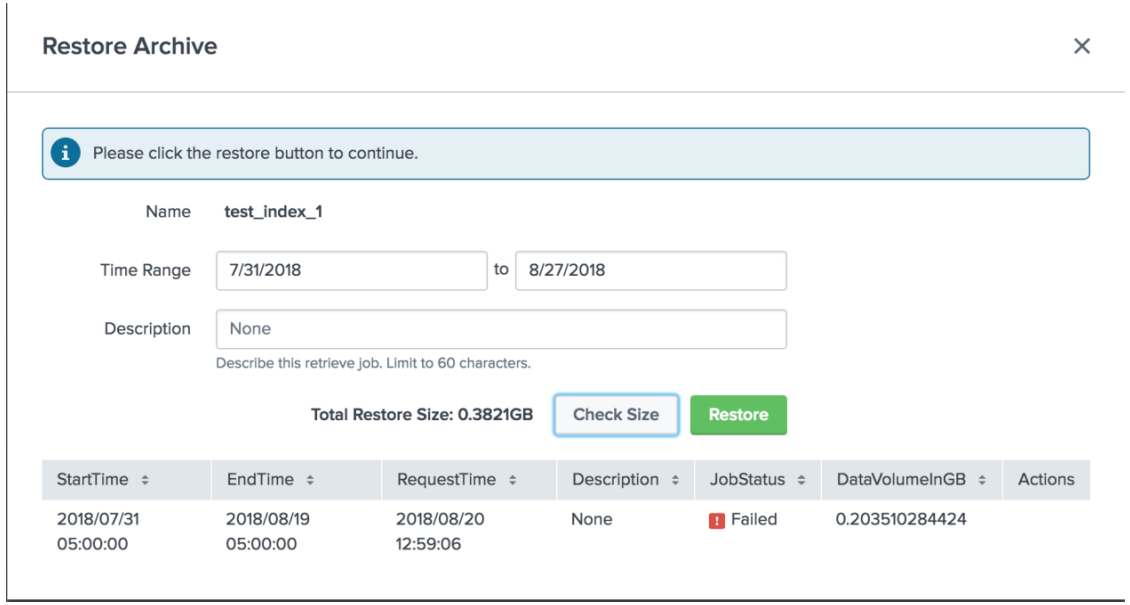
Data will reside in archival storage until the data ages out.

Aged out archived data is deleted.

”Lock” placed on archive to prevent accidental deletes

splunk> .conf18

Dynamic Data Active Archive - Restore of Archive Data



Optionally purchase additional storage

10% entitlement

Searchable (Restored Data)

Customer can optionally purchase more storage

Restored data is searchable for 30 days then deleted

Archive Data

Archive

Restore data to Splunk Cloud instance

No limits for # of restores/year

Up to 10% of the archive storage entitlement can be restored concurrently

Restore takes ~24 hours depending on amount of data.

Data is copied when restored, so Archive is untouched.

splunk> .conf18

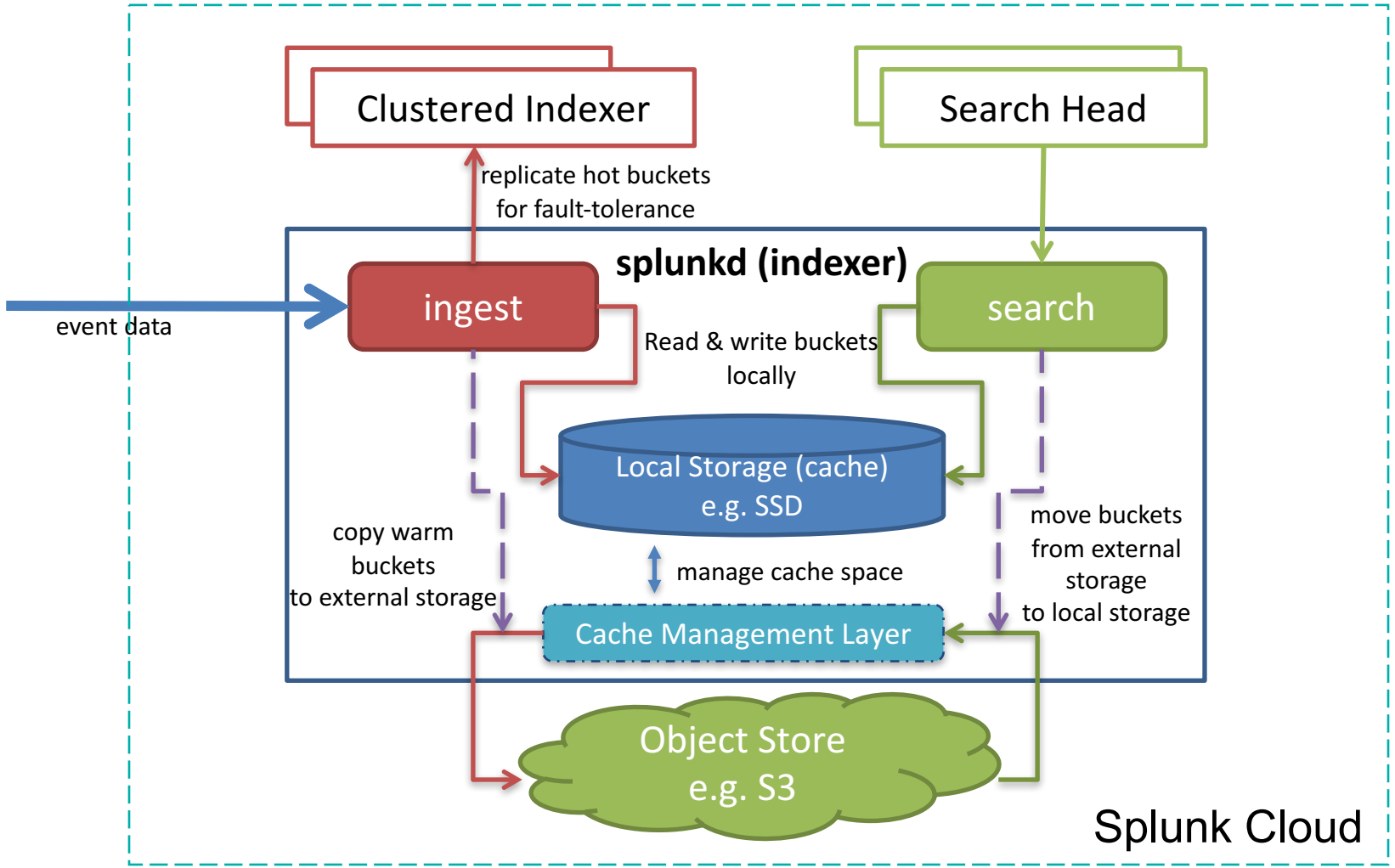
Restore Entitlement Example

A customer with 100GB daily ingestion subscription starting Jan 1 2019. They also purchased 365 days of archive that matches their ingestion subscription.

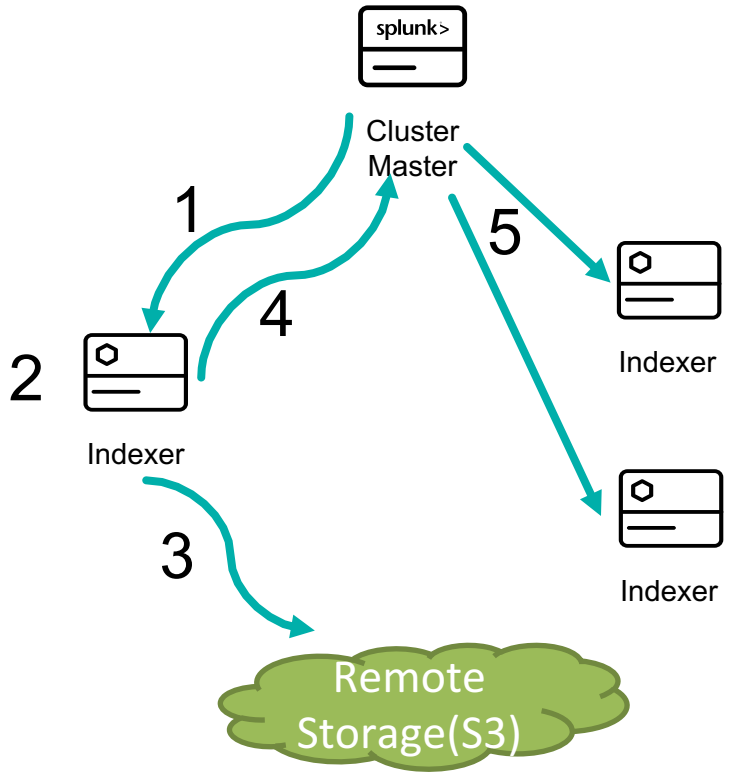
Restore Date	Restore Entitlement	Restored Amount from Archive	Within Entitlement?
3/1/19	3.65 TB	500 GB	Yes
6/1/19	3.65 TB	2 TB	Yes
9/15/19	3.65 TB	4 TB	No
11/1/19	3.65 TB	3 TB	Yes
11/3/19	3.65 TB	3 TB (+ 3 TB from 11/1) = 6 TB	No
12/15/19	3.65 TB	3 TB	Yes

Dynamic Data - Foundation

SmartStore Architecture

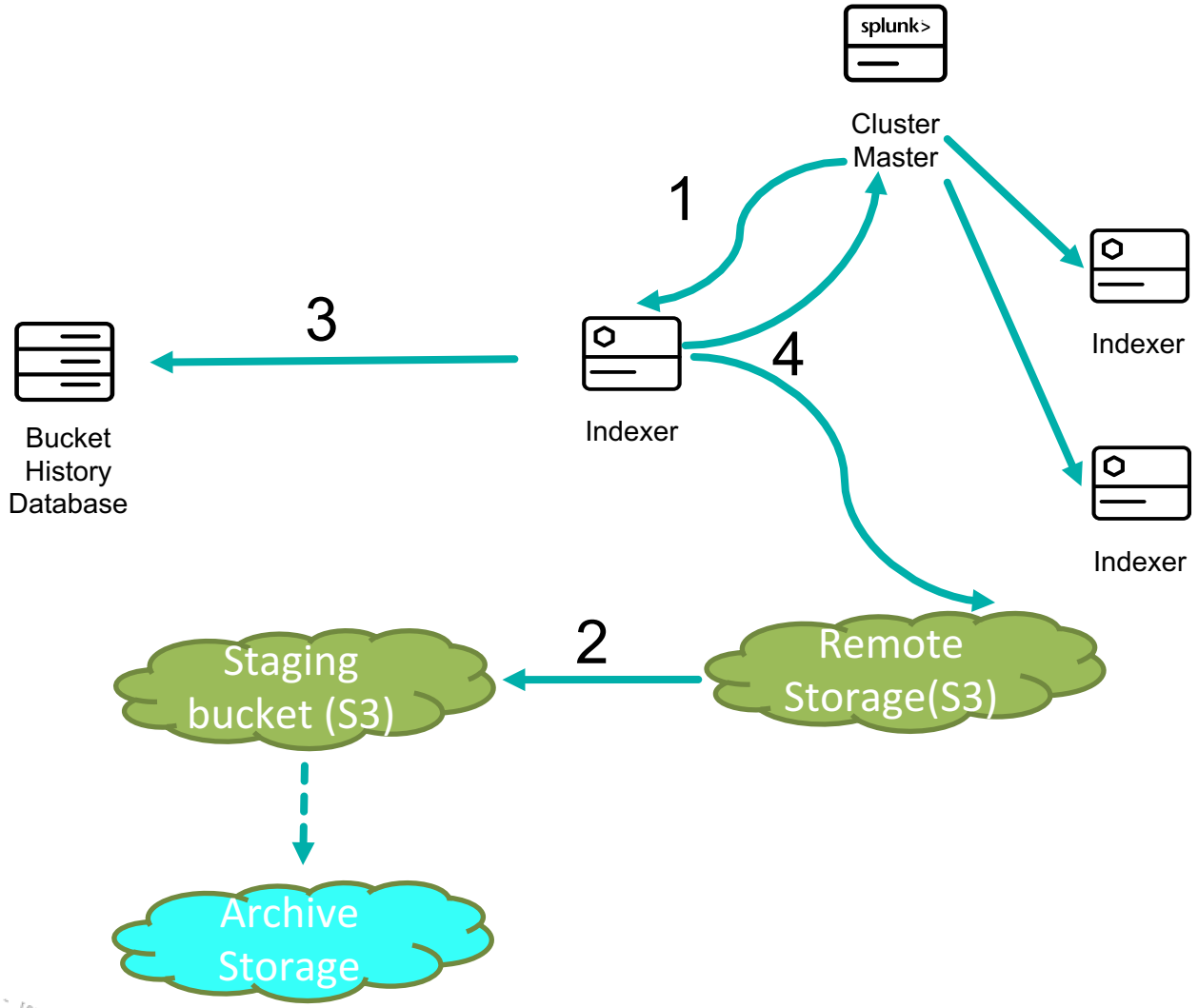


SmartStore Data Pruning



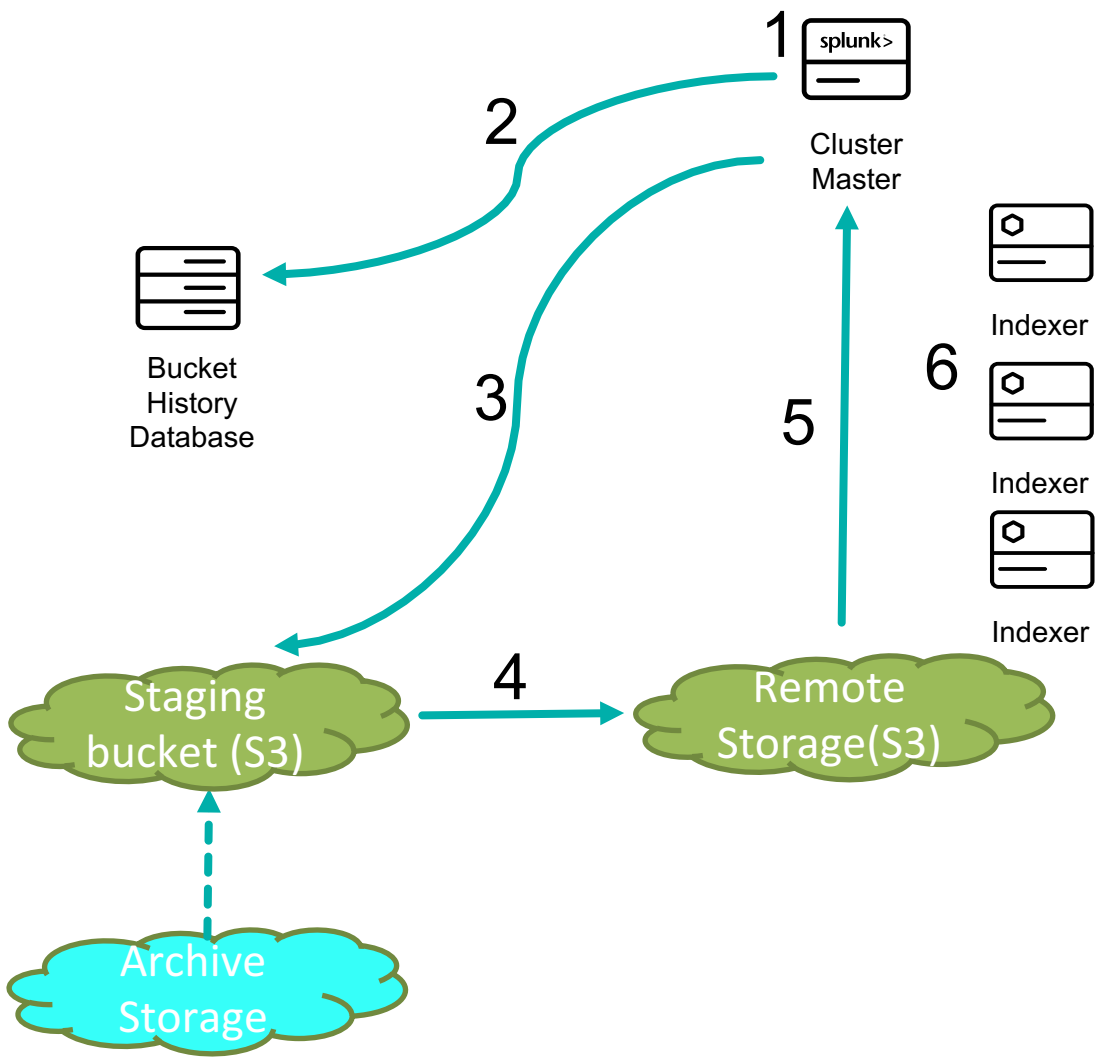
- Every 15 minutes, CM checks for expired buckets based on:
  - Time limit: Searchable time
  - Size limit: MaxGlobalDataSizeMB
- Bucket freezing workflow:
  - 1. CM picks an indexer to send the freezing request
  - 2. Indexer removes the bucket from local disk
  - 3. Indexer removes the bucket from S3
  - 4. Indexer notifies CM
  - 5. CM requests other indexers to remove the same bucket

Archive Workflow



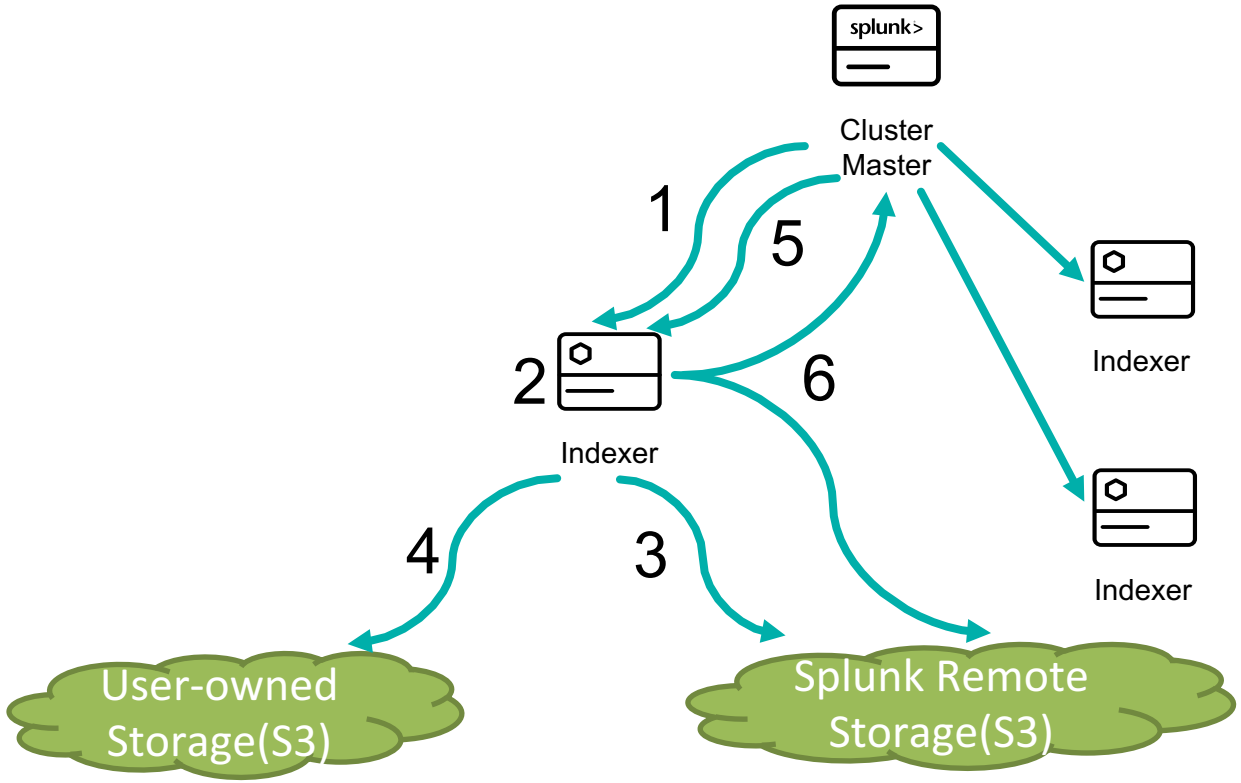
1. CM picks an indexer to send freezing request
2. Indexer copies the data on remote storage to a staging bucket configured with life cycle rule. Within 48 hours, the data will be moved to archive storage.
3. Indexer writes the bucket metadata to bucket history table
4. Indexer goes to normal freezing workflow, removing the bucket from local disk, remote storage and other indexers.

Restore Workflow



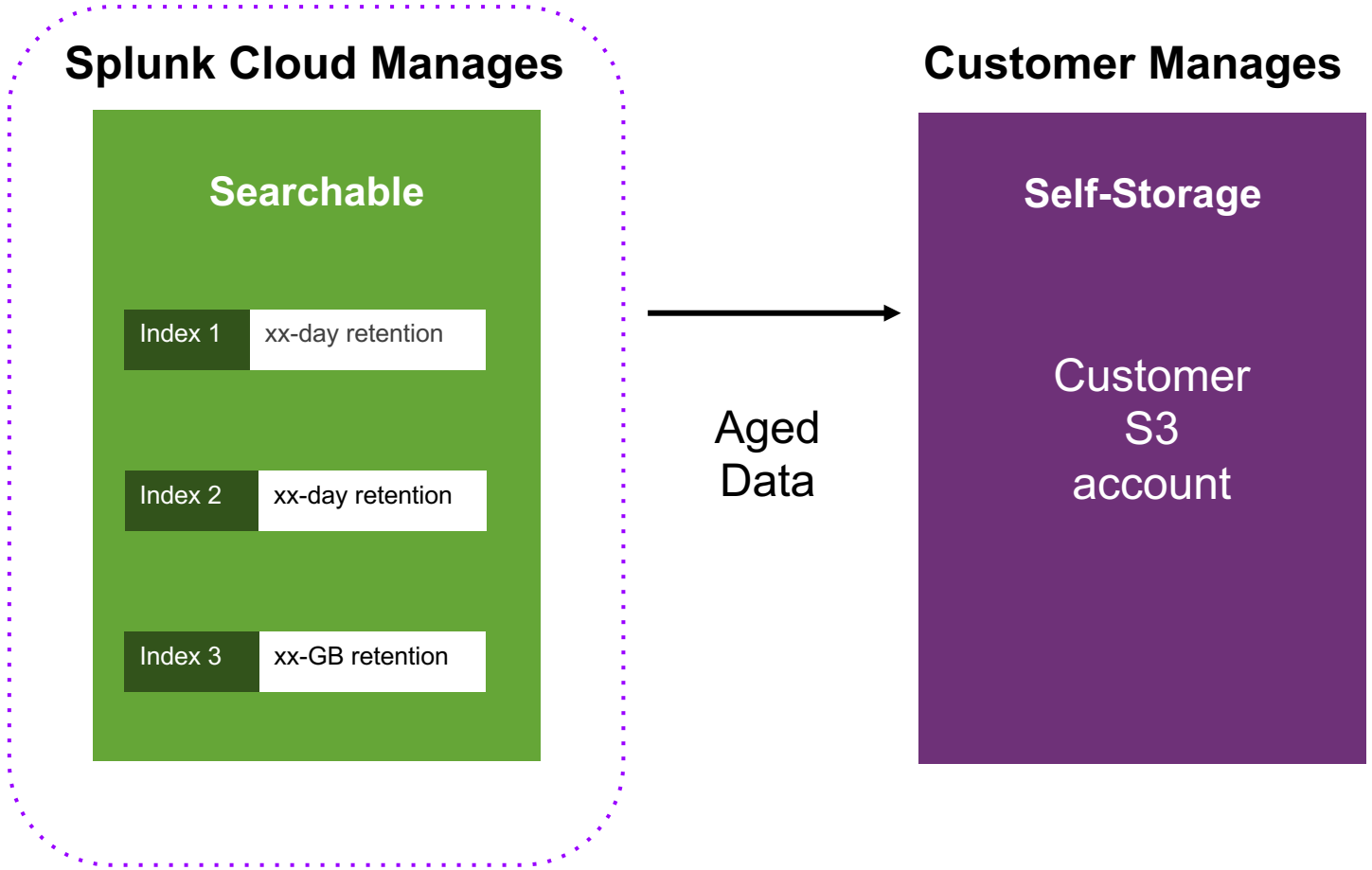
1. CM starts a long-running process for serving restoration requests
2. Restore process queries the Bucket History Table to find the buckets whose time ranges overlap with the restore time range
3. Restore process calls S3 restore APIs for each of the required buckets, then wait for data to be restored to S3 (6 hours)
4. Restore process copies the restored buckets back to remote storage
5. Restore process notifies CM about the restored buckets
6. Restore buckets are searchable!

Self-Storage Workflow



1. CM Picks an indexer to send freezing request
2. Indexer maintains a self-storage thread pool. It puts a new self-storage task into its task queue then short-circuits the freezing request
3. A worker thread picks up the task, downloads the (decrypted) rawdata folder from Splunk remote storage
4. The work thread copies the rawdata folder to the customer-owned S3 bucket
5. After 15 minutes, CM sends the freezing request to the same indexer
6. Indexer checks the progress of self-storage. If the rawdata has been transferred to user-owned storage, go to normal freezing workflow

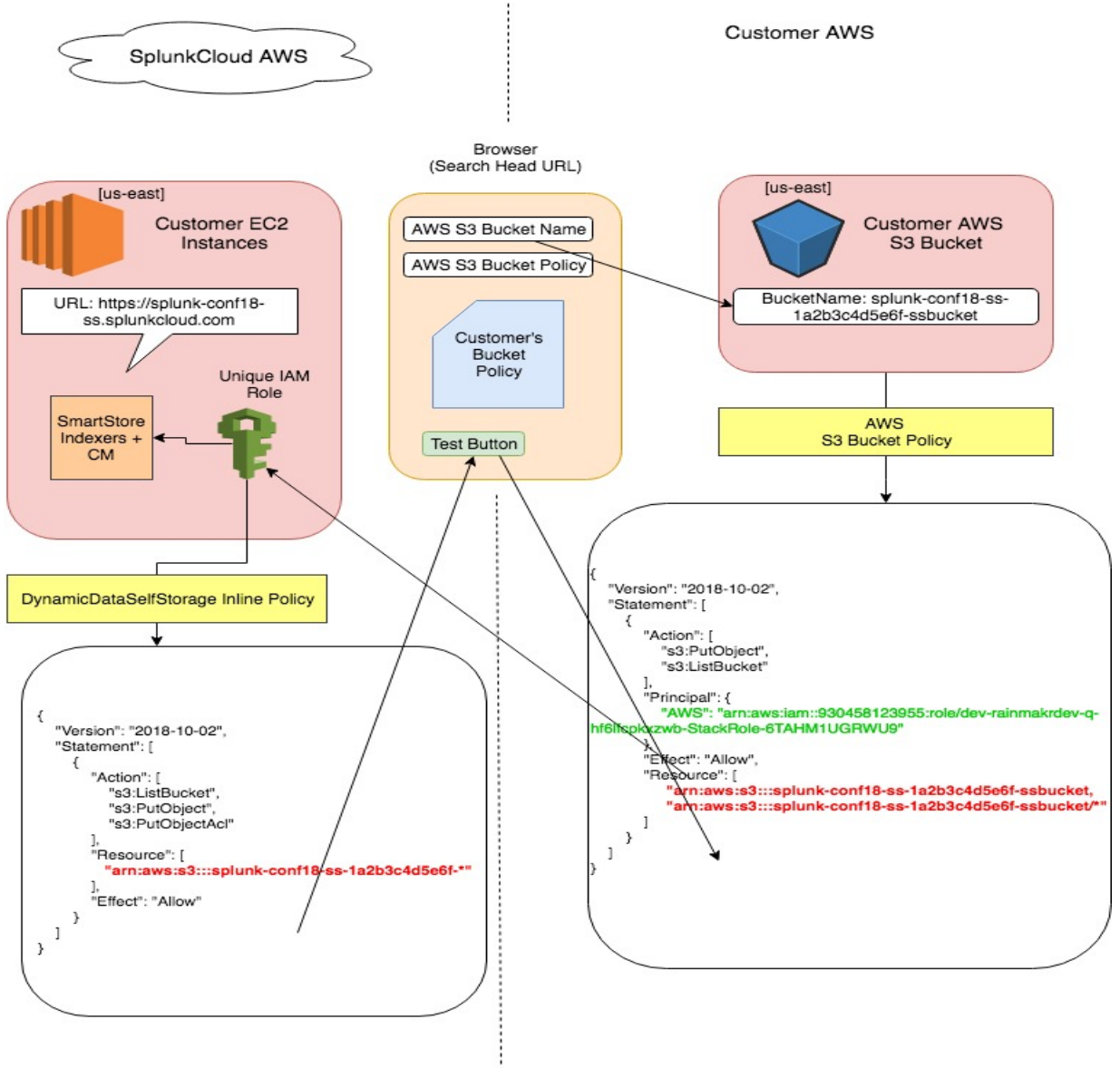
Dynamic Data Self Storage



1. **Create Amazon S3 bucket to store aged data**
   a. Amazon S3 bucket must be in the same region
2. **Enable Self-Storage for index(es)**
   a. Export the raw data only
   b. Tied to the freezing logic.
   c. Data exported in tar.gz format
3. **Verify Self-Storage is working**
   a. Customer is responsible for monitoring
   b. Review in splunkd.log
4. **To search aged data**
   a. Cannot be in Splunk Cloud
   b. BYOL on on-prem, no additional license cost to index and search the data.
   c. Restore this data by moving the exported data into a thawed directory on a Splunk Enterprise instance, such as `$SPLUNK_HOME/var/lib/splunk/defaultdb/thaweddb`.

https://www.splunk.com/blog/2018/04/24/dynamic-data-self-storage-compliance-cloud-and-data-lifecycle.html

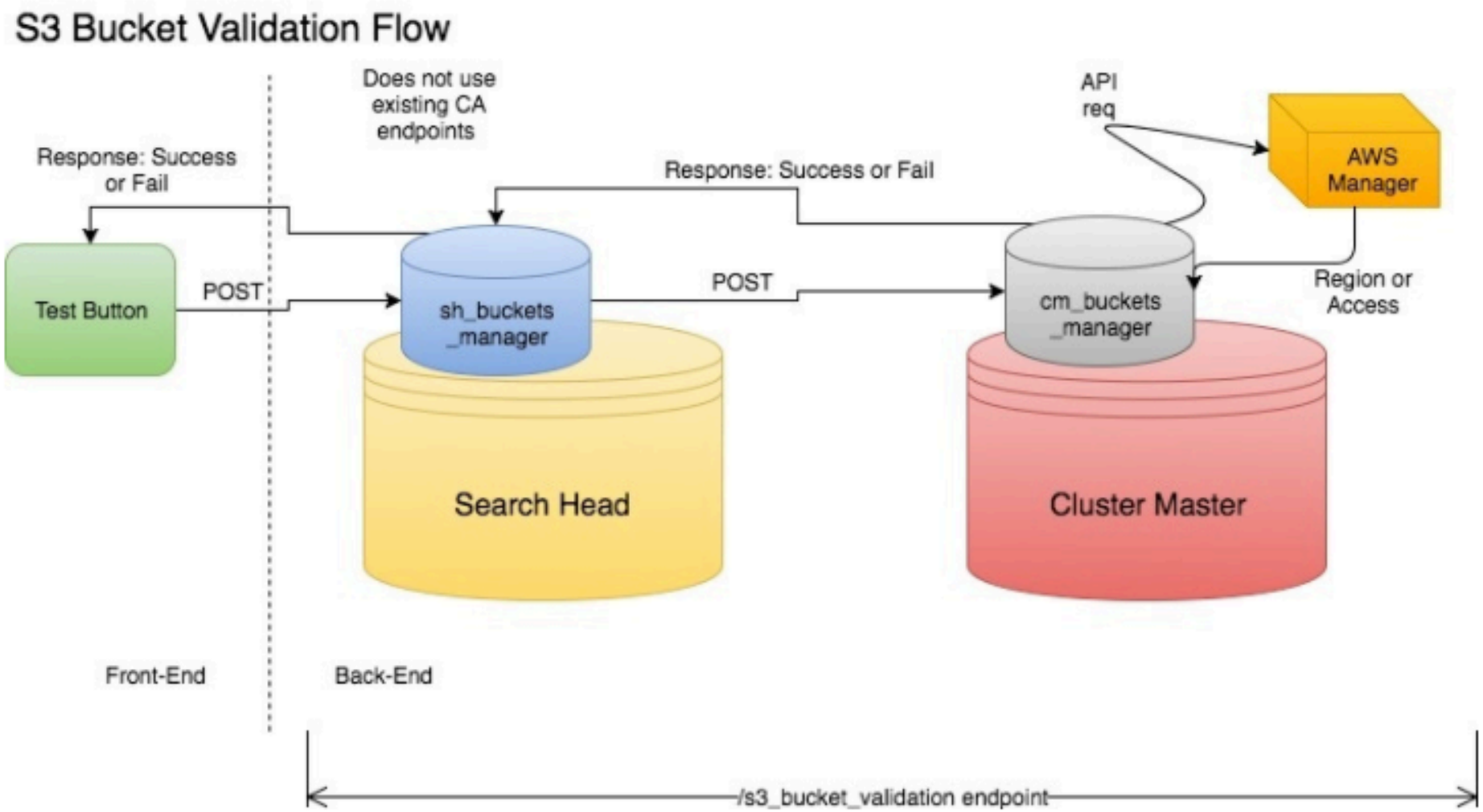
Dynamic Data Self Storage – Security Architecture



Note: ***splunk-conf18-ss*** *is the stack-id from the customer URL, which we validate in customer's s3 bucket.*

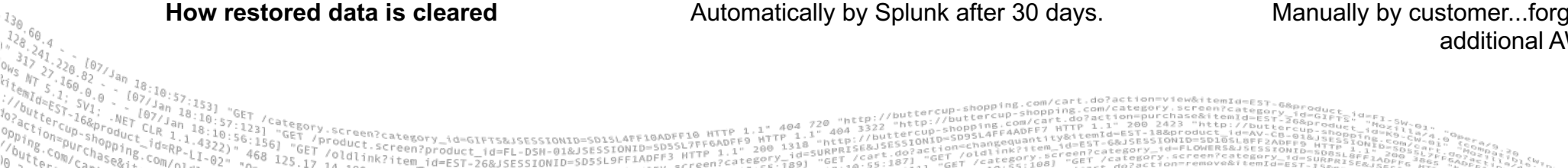
splunk> .conf18

Dynamic Data Self Storage – Bucket validation



Archive vs Self-Storage

	Archive	Self-Storage
Automated Aged Data Movement	Yes	Yes
Setup experience	Unified in Splunk Web: Enable Archive and customer sets the archive duration.	Distributed across Splunk Web, AWS Console and customer manually managing the lifecycle of exported data.
Storage Cost	Included in subscription.	Customer pays AWS separately for S3 storage consumed by exported data.
Monitoring	Provided by Splunk and customer monitors via Cloud Monitoring Console dashboard.	Customer monitors by searching splunkd.log
Restore archive data to Splunk Cloud	Yes	No, must restore to BYOL or on-prem Splunk.
Restore experience	Automated: In Splunk Web, select index(es), date range of data to be restored and restore data.	Manual: Build Splunk infrastructure, install AWS CLI, configure AWS credentials, use recursive copy command (or create a script) to download data.
When restored data is searchable	Within 24 hours.	TBD when data is searchable since customer may have to build the BYOL environment before they can restore.
Restore entitlement	10% of archive entitlement included in subscription. No limit on numbers of restores during subscription period.	Customer pays AWS separately for any BYOL infrastructure required when restoring.
How restored data is cleared	Automatically by Splunk after 30 days.	Manually by customer...forgetting to delete will incur additional AWS cost.



splunk> .conf18

Splunk Demo

Presented by Buttercup Splunker

splunk> .conf18

Beta Program Customer Feedback

Simple, straightforward and very easy to use capability

Active Archive will enable us to meet our compliance and regulatory requirements in a cost effective manner. We can't wait for this capability to be rolled out!

Restoring archive data is a key capability that will enable us to fulfill our auditing requirements, and perform security incident investigations in a time efficient manner

It is exciting to see Splunk continue to innovate in Splunk Cloud.

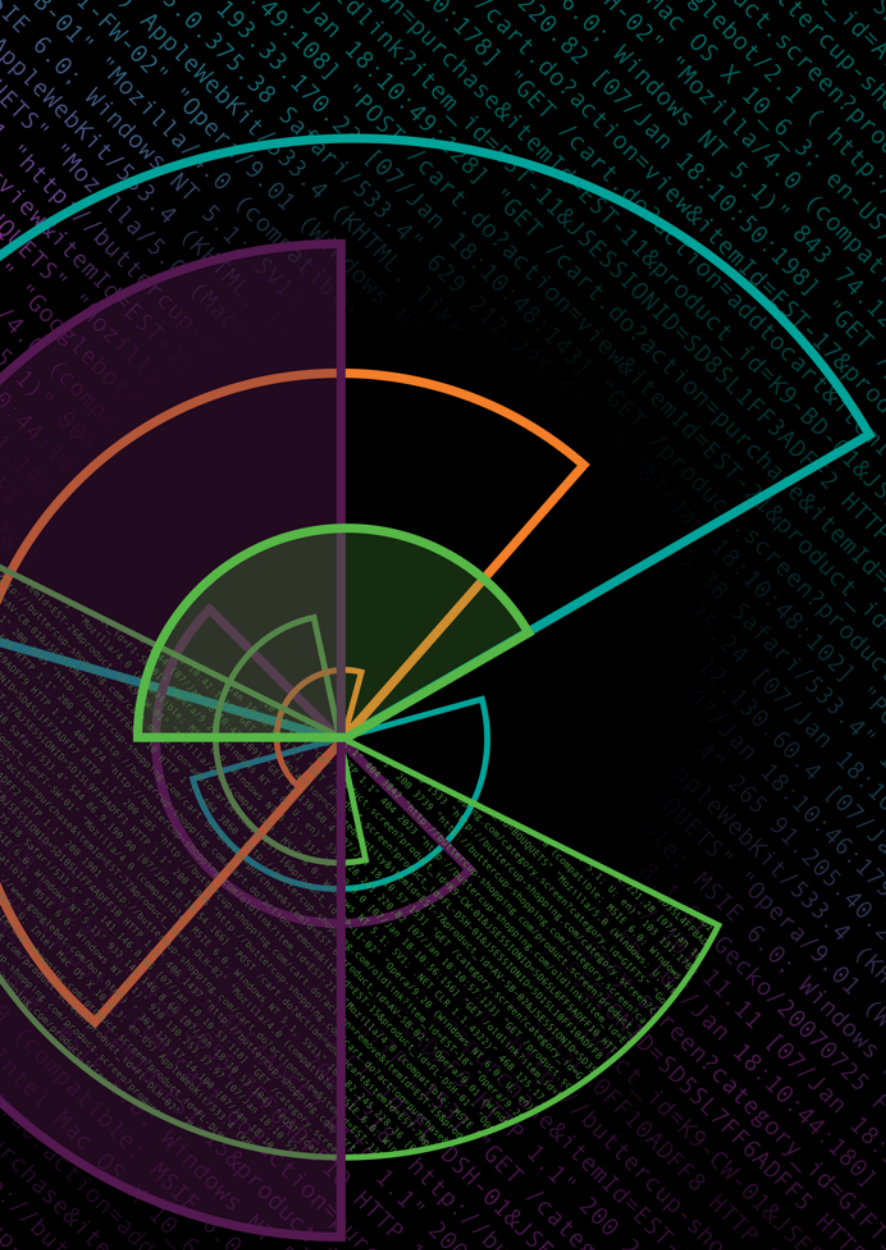
Key Takeaways

This is where the subtitle goes

1. Dynamic Data gives customers flexible data retention options in Splunk Cloud

2. Self Storage -> customer managed, available to all Splunk Cloud customers, manual data restore

3. Active Archive -> Splunk managed, optional/additional subscription service, automatic data restore



Q&A

splunk> .conf18

Thank You

Don't forget to rate this session in the .conf18 mobile app