

# Dell EMC ECS: Splunk SmartStore Configuration Guide

## Abstract

This document describes how to configure Splunk SmartStore with Dell EMC ECS.

March 2020

## Revisions

Date	Description
June 2019	Initial release
March 2020	Updated document template, component versions, reference links

## Acknowledgements

This paper was produced by the Dell EMC Unstructured Technical Marketing Engineering and Solution Architects team.

Author: [Rich Paulson](#)

The information in this publication is provided “as is.” Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

Copyright © 2020 Dell Inc. or its subsidiaries. All Rights Reserved. Dell Technologies, Dell, EMC, Dell EMC and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be trademarks of their respective owners. [3/9/2020] [Configuration and Deployment] [H17780.2]

# Table of contents

Revisions.....	2
Acknowledgements.....	2
Table of contents .....	3
Executive summary.....	4
Audience .....	4
ECS terminology .....	4
Splunk terminology .....	5
<b>1 Solution overview .....</b>	<b>6</b>
1.1 ECS overview .....	6
1.2 Splunk SmartStore overview .....	6
1.3 Solution architecture .....	7
1.4 Key components .....	8
<b>2 Solution implementation .....</b>	<b>9</b>
2.1 Implementation workflow .....	9
2.1.1 Create the ECS object user, secret key and bucket.....	9
2.1.2 Configure SmartStore indexes with ECS .....	12
2.2 Solution verification .....	13
2.2.1 Verify connectivity from SmartStore to ECS.....	13
2.2.2 Browse the ECS bucket using an S3 client. ....	13
2.2.3 SmartStore activity.....	14
<b>3 SmartStore Sizing Considerations .....</b>	<b>15</b>
<b>4 Best practices .....</b>	<b>16</b>
<b>A Technical support and resources .....</b>	<b>18</b>
A.1 Related resources.....	18

## Executive summary

The explosive growth of unstructured data and cloud-native applications has created demand for scalable on-premises cloud storage infrastructure in the modern data center. Dell EMC™ ECS, the third generation of object store by Dell EMC, is designed from the ground up to take advantage of modern cloud storage APIs and distributed data protection, providing active/active or active/passive availability spanning multiple data centers.

As data volumes continue to increase in both capacity and long-term business analytics value, Splunk® SmartStore indexes and raw data can be stored externally using a cost-effective object store compatible with the Amazon Simple Storage Service (S3). Dell EMC ECS with its distributed multi-site data protection model, erasure coding and a fully compatible S3 API makes it the best choice for customers to scale their indexer storage resources separately.

This white paper is a reference guide for configuring Splunk SmartStore with ECS and the recommended best practices.

## Audience

This document is intended for administrators who deploy and configure Dell EMC ECS with Splunk SmartStore. This guide assumes a high level of technical knowledge for the devices and technologies described.

## ECS terminology

**Replication group:** Replication groups are logical constructs that define where data is protected and accessed. Replication groups can be local or global. Local replication groups protect objects within the same VDC against disk or node failures. Global replication groups span two or more federated VDCs and protect objects against disk, node, and site failures.

The strategy for defining replication groups depends on multiple factors including requirements for data resiliency, the cost of storage, and physical versus logical separation of data. As with storage pools, the minimum number of replication groups required should be implemented. At the core ECS indexing level, each storage pool and replication group pairing is tracked and adds significant overhead. It is best practice to create the absolute minimum number of replication groups required. Generally, there is one replication group for each local VDC, if necessary, and one replication group that contains all sites. Deployments with more than two sites may consider additional replication groups, for example, in scenarios where only a subset of VDCs should participate in data replication, but this decision should not be made lightly.

**Namespace:** Namespaces enable ECS to handle multi-tenant operations. Each tenant is defined by a namespace and a set of users who can store and access objects within that namespace. Namespaces can represent a department within an enterprise, can be created for each unique enterprise or business unit, or can be created for each user. There is no limit to the number of namespaces that can be created from a performance perspective. Time to manage an ECS deployment, on the other hand, or, management overhead, may be a concern in creating and managing many namespaces.

**Bucket:** Buckets are containers for object data. Each bucket is assigned to one replication group. Namespace users with the appropriate privileges can create buckets and objects within buckets for each object protocol using its API. Buckets can be configured to support NFS and HDFS. Within a namespace, it is possible to use buckets as a way of creating subtenants. For performance reasons, it is not recommended to

have more than 1000 buckets per namespace. Generally, a bucket is created per application, workflow, or user

**Object user:** Object users are defined by a username and a secret key that can be used to access the object store. Usernames can be local names or can be domain-style user names that include an @ in the name.

## Splunk terminology

**Bucket:** Splunk Enterprise stores indexed data in buckets, which are directories containing both the data and index files into the data. An index typically consists of many buckets, organized by age of the data.

**Indexer:** A Splunk Enterprise instance that indexes data, transforming raw data into events and placing the results into an index. It also searches the indexed data in response to search requests.

**Search:** The primary way users navigate data in Splunk Enterprise. You can write a search to retrieve events from an index, use statistical commands to calculate metrics and generate reports, search for specific conditions within a rolling time range window, identify patterns in your data, predict future trends, and so on.

**SmartStore:** An indexer feature that provides a way to use remote object stores to store indexed data. By reducing reliance on local storage, SmartStore allows you to scale compute and storage resources separately, thus improving the efficiency of resource usage.

**Cache Manager:** The part of an indexer that manages the SmartStore cache. The goal of the cache manager is to optimize the use of local storage. The cache manager handles the transfer of bucket copies between local and remote storage. It also evicts bucket copies from local storage.

# 1 Solution overview

This section provides an overview of the integration of Dell EMC ECS and Splunk SmartStore and the key technologies used.

## 1.1 ECS overview

ECS provides a complete software-defined strongly-consistent, indexed, cloud storage platform that supports the storage, manipulation, and analysis of unstructured data on a massive scale. Client access protocols include S3, with additional Dell EMC extensions to the S3 protocol. Object access for S3 is achieved via REST APIs. Objects are written, retrieved, updated and deleted via HTTP or HTTPS calls using REST verbs such as GET, POST, PUT, DELETE, and HEAD.

ECS was built as a completely distributed system following the principle of cloud applications. In this model, all hardware nodes provide the core storage services. Without dedicated index or metadata master nodes the system has limitless capacity and scalability.

Service communication ports are an integral part of the configuration when configuring a SmartStore index to store index data in ECS. See Table 1 for the associated ports used with the ECS S3 data service protocol.

---

**Note:** An implementation would normally utilize an external IP load balancer to manage traffic flow to the individual ECS nodes.

---

Table 1 ECS S3 protocol and associated ports

Protocol	Transfer Protocol	Port
S3	HTTP	9020
	HTTPS	9021

## 1.2 Splunk SmartStore overview

SmartStore is an indexer capability that provides a way to use remote S3 object stores, such as ECS, to store master copies of warm data.

As a deployment's data volume increases, demand for storage typically outpaces demand for compute resources. SmartStore allows you to manage your indexer storage and compute resources in a cost-effective manner by scaling those resources separately.

SmartStore introduces a remote storage tier and a cache manager. These features allow data to reside either locally on indexers or on ECS. Data movement between the indexer and ECS is managed by the cache manager, which resides on the indexer.

With SmartStore, you can reduce the indexer storage footprint to a minimum and choose I/O optimized compute resources. Most data will reside on ECS, while the indexer maintains a local cache that contains a minimal amount of data: hot buckets, copies of warm buckets participating in active or recent searches, and bucket metadata.

SmartStore offers several advantages to the deployment's indexing tier:

- Reduced storage cost. Your deployment can take advantage of the economy of ECS, instead of relying on costly local storage.
- Access to high availability and data resiliency features available through ECS.
- The ability to scale compute and storage resources separately, thus ensuring that you use resources efficiently.
- Simple and flexible configuration with per-index settings.
- An intelligent cache manager ensures that, for most search use cases, SmartStore provides similar performance to local storage configurations.

## 1.3 Solution architecture

Figure 1 shows the indexing time workflow for Splunk SmartStore.

Data is written to a hot bucket as normal, when a bucket rolls to warm, it is uploaded to remote storage. The local copy will remain as a cached bucket. The Cache Manager monitors the cache to evict any buckets not participating in recent searches in order to free cache space. Upon eviction, the bucket directory will be emptied but not deleted.

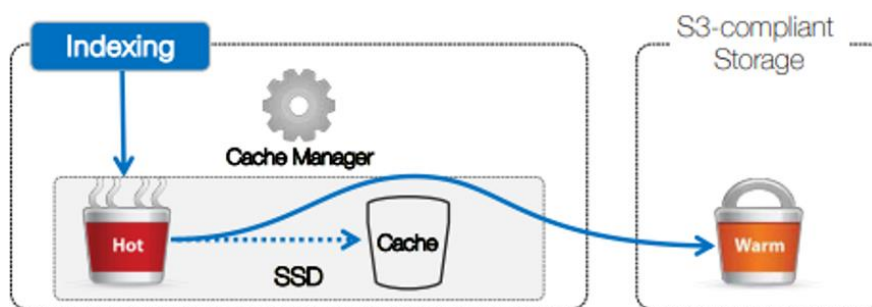


Figure 1 SmartStore indexing workflow

Figure 2 shows the workflow as event data is ingested, indexed and uploaded/downloaded to an ECS multi-site configuration. When the hot bucket rolls to warm, the source indexer copies the warm bucket to ECS while leaving the existing copy in its cache. ECS ensures high availability without the need to maintain multiple local copies so the target indexers will delete their copies. The master copy of the bucket resides in ECS.

If a search includes data that is stored in ECS but does not reside in the local warm cache, then that bucket data is downloaded to local storage and presented in search results. In this example, there are two federated ECS sites. Data is replicated between sites so in the event of a site outage, data can still be accessed with the use of Global server load balancing.

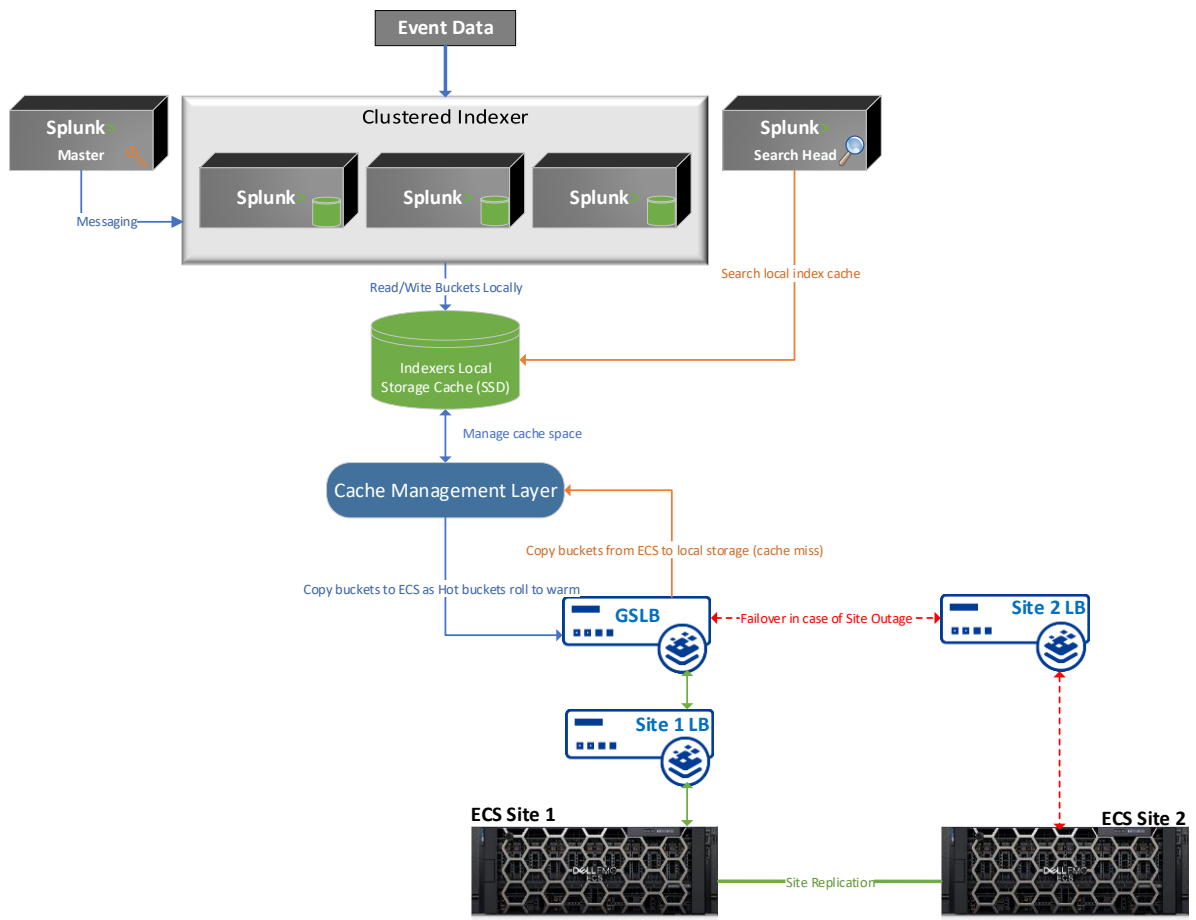


Figure 2 Architectural workflow

## 1.4 Key components

This section describes the Dell EMC ECS integration components and versions.

Table 2 Del EMC components

Component	Description
Dell EMC ECS	Release 3.2.2 and above.

Table 3 Splunk components

Component	Description
Splunk Enterprise	Version 7.3.x, 8.0.x



## 2 Solution implementation

This section describes the steps required to configure a Splunk SmartStore index with Dell EMC ECS.

### 2.1 Implementation workflow

The following are the minimal steps required to implement the solution.

- Step 1: Create an ECS Object User and Secret Key**  
Authentication details to access ECS
- Step 2: Create an ECS Bucket**  
Container for the warm Splunk index data
- Step 3: Configure a SmartStore index to use ECS**  
Use ECS for remote storage
- Step 4: Review SmartStore index options**  
Versioning, etc..

Figure 3 Configuration steps

#### 2.1.1 Create the ECS object user, secret key and bucket

An object user, S3 secret key, and bucket will need to be created for SmartStore to access ECS to store index data. This can be done from the ECS Web Portal or Management API.

1. Create a new object user and generate an S3 secret key.
  - a. From the ECS Web Portal, navigate to **Manage > Users** and click the **New Object User** button.
  - b. Enter the name of the user and a namespace.

Figure 4 Create an ECS object user

- c. Click the **Next to Add Passwords** button and click the **Generate & Add Secret Key** button under the **Object Access** section.

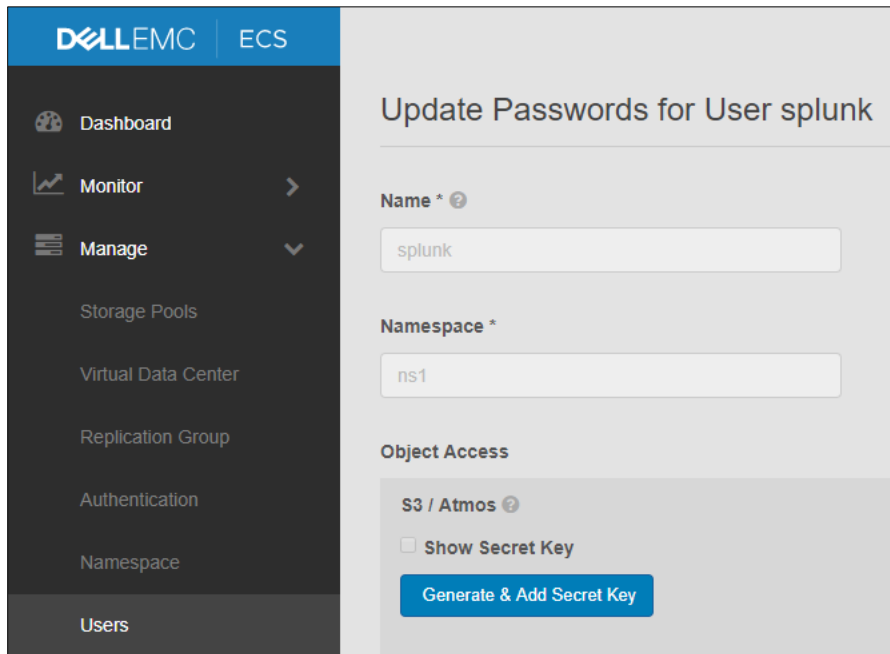


Figure 5 Generate an S3 secret key

- d. Click the **Show Secret Key** box to display the secret key. Click the **Close** button at the bottom of the page.

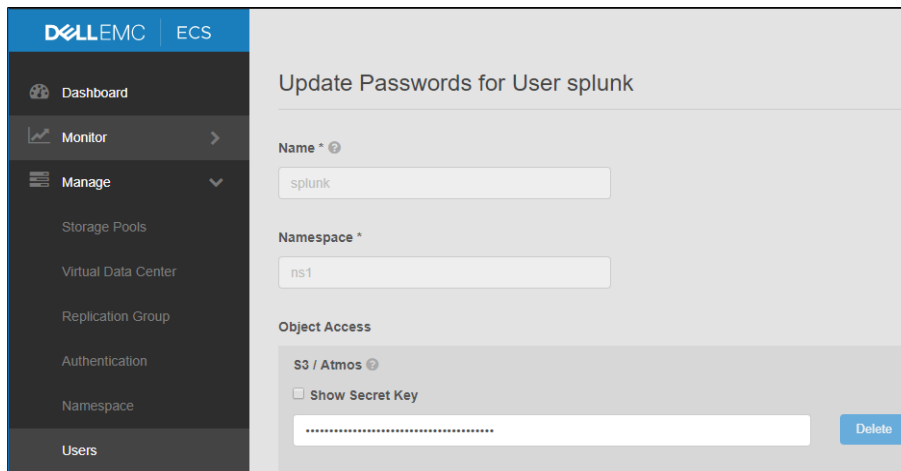


Figure 6 ECS Object User with an S3 secret key

## 2. Create an ECS bucket.

- a. Navigate to **Manage > Buckets** and select the namespace that you chose when creating the Object User (in our example that would be ns1). Click the **New Bucket** button.

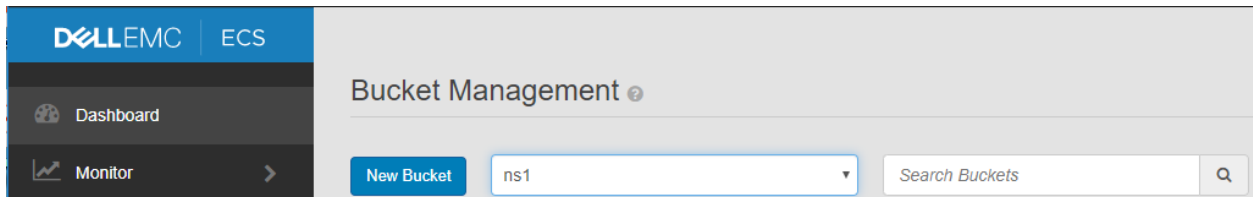


Figure 7 Create a new ECS bucket

- b. Enter a name and bucket owner. The bucket owner is the ECS Object user which was created in step 1.
- c. Click **Next** at the bottom of the page

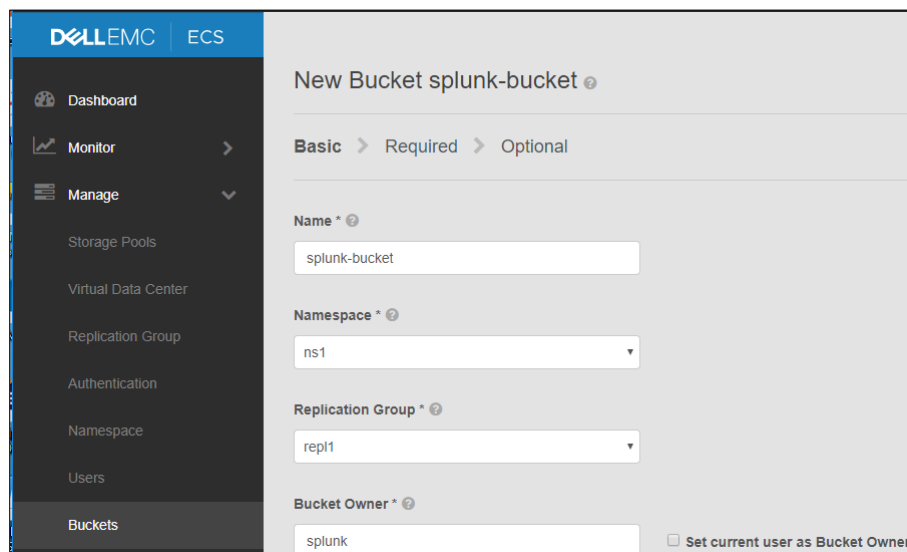


Figure 8 ECS Bucket name and owner

---

**Note:** Do not enable File System unless you've reviewed the best practices section of this document. The Required and Optional configuration settings can all be left as the defaults however the ECS documentation should be reviewed to better understand each option. Access During Outage should be enabled if the replication group chosen for the bucket spans multiple ECS sites. Reference the Temporary Site Outage section in the [ECS Overview and Architecture](#) document for details of this feature.

---

- d. Click the **Save** button at the bottom of the final page to create the bucket.

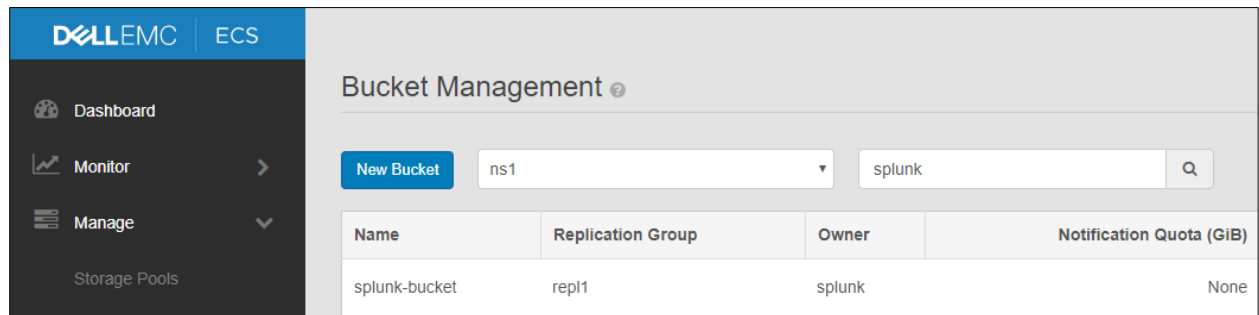


Figure 9 ECS bucket to store Splunk SmartStore warm index data

## 2.1.2 Configure SmartStore indexes with ECS

The SmartStore settings in `indexes.conf` enable and control SmartStore indexes. You can enable SmartStore for all indexes, or you can enable it on an index-by-index basis, allowing a mix of SmartStore and non-SmartStore indexes on the same indexer.

**Note:** When you configure these settings on an indexer cluster's peer nodes, you must deploy the settings through the configuration bundle method. As with all settings in `indexes.conf`, SmartStore settings must be the same across all peer nodes.

This example configures SmartStore for an indexer cluster. On the master node, `cd` to `/$SPLUNK_HOME/etc/master/_cluster/local` and create a file named **`indexes.conf`**.

### Sample `indexes.conf` file:

```
[default]
remotePath = volume:remote_store/$_index_name
repFactor = auto
maxDataSize = auto

[volume:remote_store]
storageType = remote

path = s3://splunk-bucket/indexes
remote.s3.access_key = splunk
remote.s3.secret_key = <ECS Object Users S3 Secret Key>
remote.s3.endpoint = <Endpoint to access ECS nodes>

[index_name]
homePath = $SPLUNK_DB/index_name/db
thawedPath = $SPLUNK_DB/index-name/thaweddb
# Note: SmartStore-enabled indexes do not use coldPath,
# but you must still specify it here.
coldPath = $SPLUNK_DB/index-name/colddb
```

The highlighted parameters above are the values that need to be modified to store indexed data in ECS.

Note that the `StorageType` and `RemotePatch` entries under the stanza are what determines this is a SmartStore index.

Table 4 Indexes.conf configuration attributes

Attribute Name	Description
path	The above example is configured to use our bucket example in section 2.1.1, Step 2. A prefix/path named "indexes" is also being used. The bucket must be created on ECS prior to deploying the configuration bundle but any prefixes will automatically be created by SmartStore.
remote.s3.access_key	The ECS Object User name that was created in section 2.1.1
remote.s3.secret_key	The ECS Object Users S3 secret key generated in section 2.1.1
remote.s3.endpoint	The HTTP HTTPS endpoint to access the ECS nodes. This would typically be the IP load balancer in front of the ECS cluster
maxDataSize	Recommended setting for SmartStore bucket size is to use the value "auto" (750MB). This is to help avoid timeouts when downloading large sized buckets from the remote store back to the local cache.

Push the configuration bundle to the peer node(s) from the master using the CLI or GUI. This action may cause the peer node(s) to restart.

## 2.2 Solution verification

There are several tools available which can be used to verify that the SmartStore indexes are being uploaded to ECS. In addition to the below methods, the Splunk logs contain insight into SmartStore operations.

Reference the SmartStore [troubleshooting](#) document on Splunk.com for detailed troubleshooting details.

### 2.2.1 Verify connectivity from SmartStore to ECS

SmartStore contains a CLI utility which can be used to verify connectivity to the remote store.

The syntax for running the command is:

```
./splunk cmd splunkd rfs -- ls --starts-with volume:remote_store
```

The output of this command will list the contents of the ECS bucket.

### 2.2.2 Browse the ECS bucket using an S3 client.

S3 Browser is a freeware Windows client for S3 which can be used to browse ECS buckets.

To add a new account, select **Add new account** from the Accounts menu item. Provide an account name to identify the connection, select **S3 Compatible Storage** and enter the endpoint to the ECS cluster. ECS supports both V2 and V4 Signatures. Lastly, enter you're ECS Object User ID and S3 password, ensure that **Use secure transfer (SSL/TLS)** is checked if your using HTTPS.

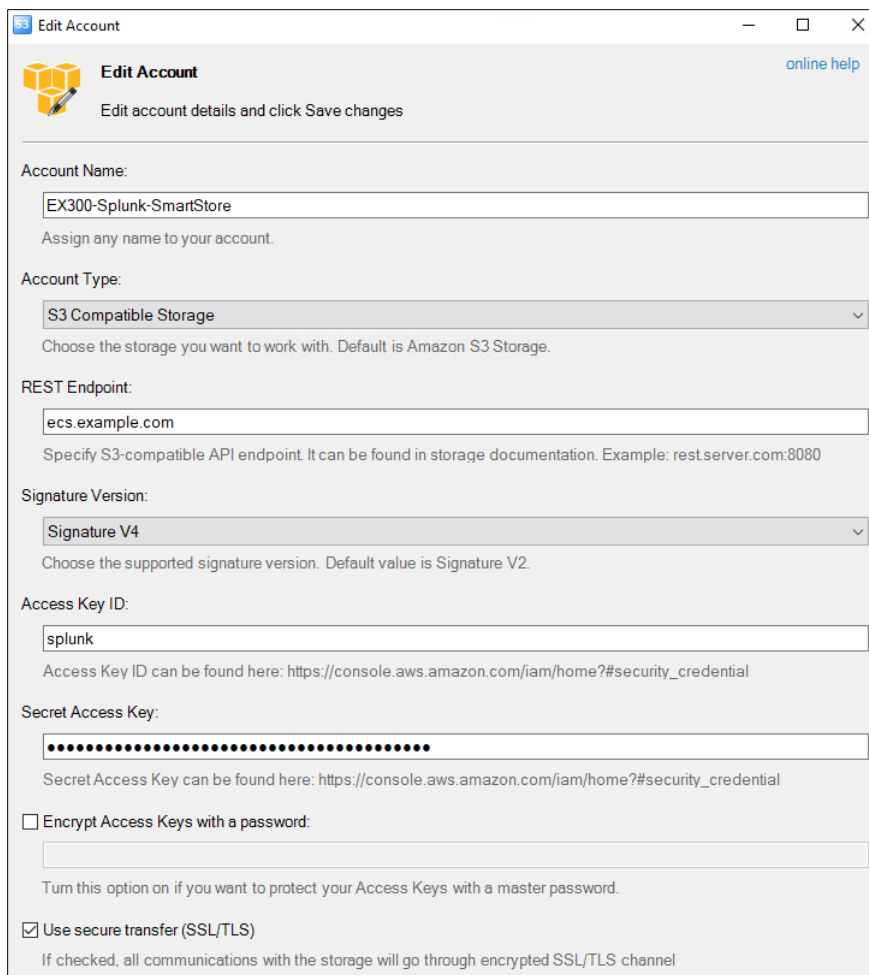


Figure 10 S3 Browser configuration example

### 2.2.3 SmartStore activity

The Splunk internal logs (`_internal` index) stores details related to the SmartStore CacheManager component. This means CacheManager related activity can be searched.

For example, the below search looks for any uploads the CacheManager initiated.

```
index=_internal component=CacheManager action=upload
```

The Monitoring Console also contains SmartStore specific dashboards which report several metrics.

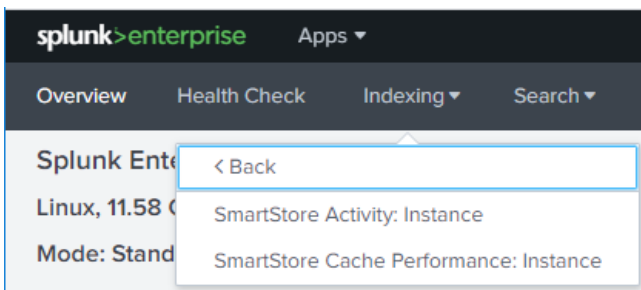


Figure 11 SmartStore Activity Dashboards

### 3 SmartStore Sizing Considerations

- The cached days (Hot/Cache) should be sized as
  - 90 Days if using Enterprise Security
  - 30 Days if only using Splunk Enterprise
- Size the Hot/Cache appropriately.
  - $\text{<Daily Ingest> * .5 = Daily size on disk * <days of retention> = Total storage needed}$
- Use SSD or NVMe for Hot/Cache
- Not for frequent rare searches. ~95% of searches should hit the Hot/Cache
- Replicating factor (RF) and Search Factor (SF) must be equal.
- SmartStore assumes the remote storage manages data availability (HA).
- Splunk's compression ratio factor is generally 50% for most data (15% from raw data, 35% from index files)

## 4 Best practices

The following best practices are recommended when configuring SmartStore indexes with Dell EMC ECS.

Recommendation	Details
Managing traffic flow to ECS	We recommend utilizing an external IP traffic load balancer with ECS to manage the health and distribute traffic to all ECS nodes in the cluster.
remote.s3.use_delimiter to improve listing performance	<p>ECS Release 3.2.2 is the first release that supports delimiter “strings” to support the default value for the SmartStore parameter <b>remote.s3.use_delimiter</b>.</p> <p>This means that ECS Release 3.2.2 or greater cannot be used with file system enabled buckets when this parameter is enabled, as file system semantics for ECS native NFS and Hadoop support require a standard “/” delimiter in file paths.</p> <p>Disabling this parameter for use with legacy versions of ECS or filesystem enabled buckets has neither been tested nor is it a best practice for ECS performance purposes.</p>
Local storage requirements	<p>Splunk recommends to provision enough local storage to accommodate the equivalent of 30 days' worth of indexed data. At a minimum, provision enough storage to keep at least 7-10 days of data in cache, as searches typically occur on data indexed within the last 7 - 10 days.</p> <p>When using Splunk Enterprise Security, provision enough local storage to accommodate 90 days' worth of indexed data, rather than the otherwise recommended 30 days.</p>
Cold Buckets	<p>SmartStore index buckets generally roll to frozen directly from warm.</p> <p>Cold buckets can exist in a SmartStore-enabled index, but only under limited circumstances. Specifically, if you migrate an index from non-SmartStore to SmartStore, any migrated cold buckets use the existing cold path as their cache location, post-migration.</p>
Data Retention	Data retention policy for SmartStore indexes on indexer clusters is configured using settings like those for non-SmartStore indexes. However, with SmartStore indexes, data retention is managed cluster-wide, rather than on a per-indexer basis.
Multi-part upload/download part size	The default settings for multipart download and upload part size is 128MB and should not be modified unless that value has proven to improve throughput.
Multi-part threads	<p><b>remote.s3.multipart_max_connections</b> Specifies the maximum number of HTTP connections to have in progress for either multipart download or upload.</p> <p>The default is 8</p>
Versioning	When <b>remote.s3.supports_versioning = true</b> , SmartStore will iterate over all versions of the object and remove them. Otherwise, it does a simple remove on the object. This means that if set to true, all versions will be removed, and the object contents are irretrievable.



Recommendation	Details
	<p>If set to <b>false</b>, the behavior is as follows:</p> <ol style="list-style-type: none"><li>1) if versioning is disabled on the bucket, the object is deleted.</li><li>2) if versioning is enabled on the bucket, the "remove object" operation simply puts a delete marker on top.</li></ol> <p><b>Note:</b> If versioning is enabled on the ECS bucket storing index data and the above setting is set to <b>false</b>, then it is recommended to set a lifecycle policy on that bucket to eventually remove the versioned objects to free capacity.</p>

## A Technical support and resources

[Dell.com/support](https://www.dell.com/support) is focused on meeting customer needs with proven services and support.

[Storage technical documents and videos](#) provide expertise that helps to ensure customer success on Dell EMC storage platforms.

### A.1 Related resources

#### ECS product documentation

- Dell EMC ECS product documentation
  - <https://www.dell.com/support/article/us/en/19/sln319451/dell-emc-ecs-3-4-x-product-documentation-index?lang=en>
- Dell EMC ECS Architecture and Overview
  - <https://www.dell.com/whitepapers/products/storage-1/h14071-ecs-architectural-guide-wp.pdf>
- Dell EMC ECS Networking and Best Practices
  - <https://www.dell.com/whitepapers/products/software/h15718-ecs-networking-bp-wp.pdf>
- Dell EMC ECS Best Practices
  - <https://www.dell.com/whitepapers/products/storage/h16016-ecs-best-practices-guide-wp.pdf>

#### Splunk SmartStore documentation

- Splunk SmartStore Deployment and Configuration Guide
  - <https://docs.splunk.com/Documentation/Splunk/8.0.2/Indexer/AboutSmartStore>
- SmartStore Architectural Overview
  - <https://docs.splunk.com/Documentation/Splunk/8.0.2/Indexer/SmartStorearchitecture>
- Restrictions when using SmartStore indexes
  - [https://docs.splunk.com/Documentation/Splunk/8.0.2/Indexer/ConfigureSmartStore#Settings\\_in\\_indexes.conf\\_that\\_are\\_incompatible\\_with\\_SmartStore\\_or\\_otherwise\\_restricted](https://docs.splunk.com/Documentation/Splunk/8.0.2/Indexer/ConfigureSmartStore#Settings_in_indexes.conf_that_are_incompatible_with_SmartStore_or_otherwise_restricted)
- SmartStore Troubleshooting Guide
  - <https://docs.splunk.com/Documentation/Splunk/8.0.2/Indexer/TroubleshootSmartStore>
- Configuring Data Retention for SmartStore indexes
  - <https://docs.splunk.com/Documentation/Splunk/8.0.2/Indexer/SmartStoredataretention>
- SmartStore “indexes.conf” settings (see S3 Settings section)
  - <https://docs.splunk.com/Documentation/Splunk/8.0.2/Admin/Indexesconf#indexes.conf.spec>