

# SPLUNK ENTERPRISE AND ECS TECHNICAL SOLUTION GUIDE

## Splunk Frozen and Archive Buckets on ECS

### **ABSTRACT**

This technical solution guide describes a solution for archiving Splunk frozen buckets to ECS. It also describes how Hadoop Data Roll can be used with ECS to search archived data that is no longer available in Splunk.

April, 2017

The information in this publication is provided “as is.” EMC Corporation 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 EMC software described in this publication requires an applicable software license.

EMC<sup>2</sup>, EMC, the EMC logo, ECS are registered trademarks or trademarks of EMC Corporation in the United States and other countries. All other trademarks used herein are the property of their respective owners. © Copyright 2017 EMC Corporation. All rights reserved. Published in the USA. 4/17, White Paper, H15955

EMC believes the information in this document is accurate as of its publication date. The information is subject to change without notice.

EMC is now part of the Dell group of companies.

# TABLE OF CONTENTS

<b>ABSTRACT</b> .....	<b>1</b>
<b>INTRODUCTION</b> .....	<b>5</b>
<b>SOLUTION PURPOSE</b> .....	<b>5</b>
<b>SOLUTION ARCHITECTURE</b> .....	<b>5</b>
SOLUTION ARCHITECTURE OVERVIEW.....	5
SPLUNK ARCHIVE METHODS .....	6
HADOOP DATA ROLL.....	6
NFS AND S3.....	6
SOLUTION COMPONENTS .....	7
DELL EMC ELASTIC CLOUD STORAGE (ECS) .....	7
SPLUNK ENTERPRISE.....	7
HORTONWORKS DATA PLATFORM (HDP) HADOOP CLUSTER.....	7
<b>PRODUCTS AND VERSIONS</b> .....	<b>8</b>
<b>TESTING ENVIRONMENT</b> .....	<b>8</b>
<b>INSTALLATION AND CONFIGURATION</b> .....	<b>8</b>
SPLUNK.....	8
ECS.....	9
HADOOP DATA ROLL.....	9
HORTONWORKS DATA PLATFORM (HDP) .....	9
ADD SPLUNK USER .....	10
CREATE VIRTUAL INDEX PROVIDER FOR HADOOP .....	10
CREATE ARCHIVED INDEX .....	12
FROZEN PATH TO NFS.....	13
ECS.....	13
MOUNT THE NFS EXPORT ON SPLUNK HOSTS .....	15
SPLUNK .....	15
COLD TO FROZEN S3 .....	15
ECS.....	15
ECS-SYNC .....	16
COLD TO FROZEN ECS S3 SCRIPT .....	16
SPLUNK .....	16

**USAGE**..... 16

    HADOOP DATA ROLL..... 16

    FROZEN PATH TO NFS..... 17

    COLD TO FROZEN S3..... 17

**CONCLUSION** ..... 17

**REFERENCES**..... 17

**DOCUMENT VERSION HISTORY** ..... 17

## INTRODUCTION

Splunk Enterprise makes it simple to collect, analyze and act upon the untapped value of the big data generated by your technology infrastructure, security systems and business applications—giving you the insights to drive operational performance and business results.

By monitoring and analyzing everything from customer clickstreams and transactions to security events and network activity, Splunk Enterprise helps you gain valuable Operational Intelligence from your machine-generated data. And with a full range of powerful search, visualization and pre-packaged content for use cases, any user can quickly discover and share insights. Just point your raw data at Splunk Enterprise and start analyzing your world.

- Collects and indexes log and machine data from any source
- Powerful search, analysis and visualization capabilities empower users of all types
- Provides solutions for security, IT ops, business analysis and more
- Enables visibility across on-premises, cloud and hybrid environments
- Delivers the scale, security and availability to suit any organization

Splunk implements a form of storage tiering called hot/warm and cold buckets of data to optimize performance for newly indexed data and to provide an option to keep older data for longer periods on higher capacity storage. As buckets age, they "roll" from one stage to the next eventually rolling from cold to frozen. By default Splunk deletes frozen buckets, however it can be configured to copy frozen buckets to another location, or in this case, copied to an ECS cluster for long term archiving. Frozen data is no longer searchable within Splunk without invoking a manual process to "thaw" the data.

Alternatively, Splunk has a new feature called Hadoop Data Roll that can copy warm and cold buckets to a Hadoop file system such as EMC Elastic Cloud Storage (ECS) for long term storage without losing the ability to search the archived data. This allows Splunk to utilize a Hadoop cluster's MapReduce framework to search through the archived buckets just as a Splunk user would search through the hot/warm and cold buckets, although with reduced performance.

EMC ECS is a software-defined, cloud-scale, object storage platform that combines the cost advantage of commodity infrastructure with the reliability, availability and serviceability of traditional arrays. With ECS, any organization can deliver scalable and simple public cloud services with the reliability and control of a private-cloud infrastructure.

ECS provides comprehensive protocol support for unstructured object and file workloads on a single, cloud-scale storage platform. With ECS, you can easily manage your globally distributed storage infrastructure under a single global namespace with access to content anywhere. ECS features a flexible software-defined architecture that is layered to promote limitless scalability. Each layer is completely abstracted and independently scalable with high availability and no single point of failure.

## SOLUTION PURPOSE

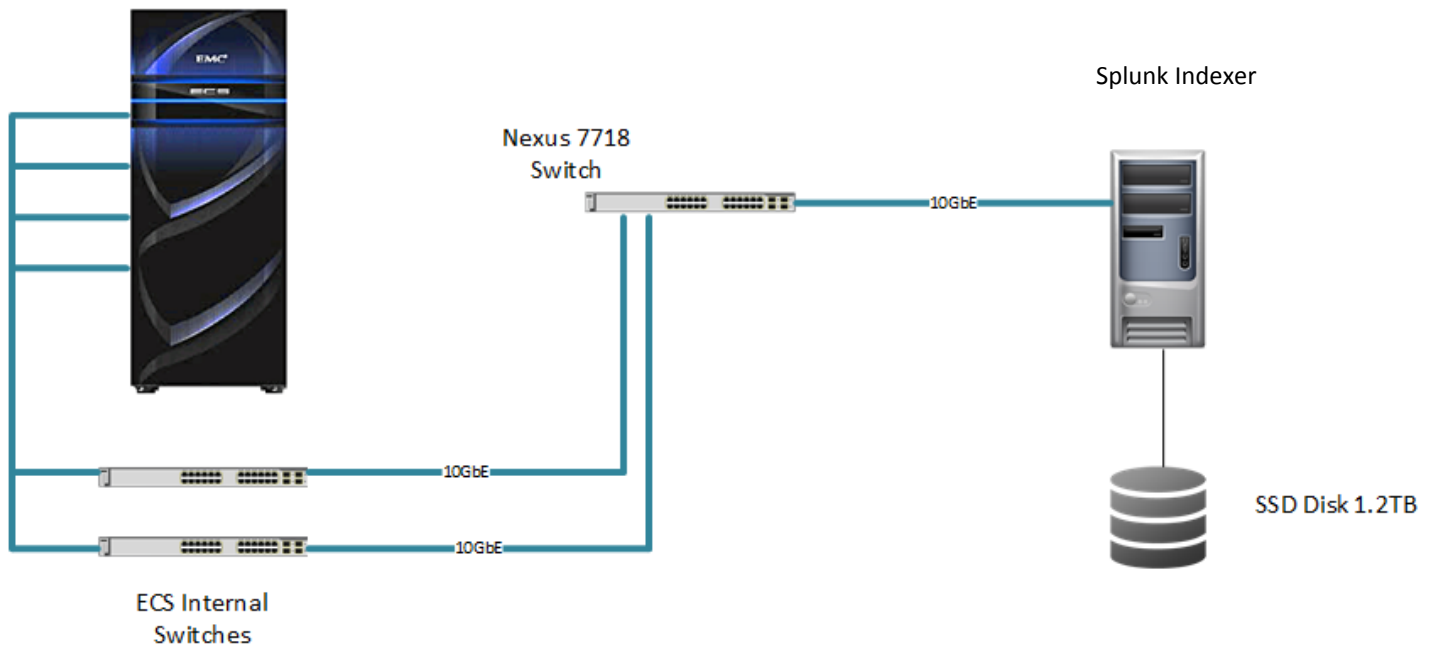
This document describes several methods for using ECS to store archived or frozen Splunk buckets. In the case of Hadoop Data Roll, these archived buckets are searchable but with significantly reduced performance compared to Splunk's hot/warm and cold buckets. This document does **not** provide a complete Splunk solution. See the References section for complete Splunk solutions offered by Dell EMC.

## SOLUTION ARCHITECTURE

### SOLUTION ARCHITECTURE OVERVIEW

**Figure 1. Solution Architecture Overview**

ECS U1500



## SPLUNK ARCHIVE METHODS

This document describes three different methods of archiving Splunk buckets to ECS.

**Table 1. Comparison of archive methods**

METHOD	APPLIES TO	MOVE OR COPY	SEARCHABLE
Hadoop Data Roll	Buckets older than a specific age	Copy	Yes. Unified Search also available.
Frozen Path to NFS	Oldest cold buckets that cause the index to exceed the maximum size	Move	No
Cold to Frozen S3	Oldest cold buckets that cause the index to exceed the maximum size	Move	No

## HADOOP DATA ROLL

Hadoop Data Roll provides the unique ability to allow searching through the archived buckets by utilizing the Hadoop cluster's MapReduce framework. However, archiving performance is relatively poor compared to archiving using NFS or S3. Further, the search performance for buckets archived using Hadoop Data Roll is much lower than the performance when searching hot/warm or cold buckets. This means that searching through the archived buckets should generally be a rare activity or at least span a very narrow time range to limit the amount of data that must be processed.

When archiving using Hadoop Data Roll, the source buckets are copied and the source buckets will remain as warm or cold buckets. Independent of Hadoop Data Roll, the warm or cold buckets will be tiered to frozen according to the index's policies such as the maximum index size.

When Hadoop Data Roll archives data to ECS, the Splunk server uses a locally installed Hadoop client (including the ECS HDFS client) to copy files directly to ECS. A Hadoop cluster is not needed nor are Hadoop YARN Node Managers, Resources Managers or any other Hadoop component involved in the archive process. The only requirement is to have the Hadoop client installed and properly configured which is best done with Ambari and a full Hortonworks HDP cluster. When searching archived data, the Hadoop cluster's MapReduce framework is used which does utilize the YARN Node Managers and Resource Managers.

## NFS AND S3

When archiving using either the NFS or S3 methods, archived cold buckets are moved. This means that the source bucket is copied to ECS using NFS or S3 and then it is deleted from the cold storage.

There is virtually no difference in performance between the NFS and the S3 methods. The NFS method takes more work to configure the NFS export and security on ECS but configuring Splunk to use it is extremely easy as only the path needs to be entered into the Splunk UI.

In contrast, the S3 method is very easy to configure on ECS but requires the installation of `ecs-sync` and the custom cold to frozen ECS S3 script. Due to its reliance on this custom script, it may be not thoroughly tested or officially supported by either Dell EMC or Splunk.

Buckets archived using either the NFS or S3 methods can be "thawed" by manually copying them back to a Splunk server in a special *thawedb* directory. This will cause Splunk to recreate the TSIDX files and allow very quick searching against the bucket. However, recreation of the TSIDX files is an expensive operation.

## SOLUTION COMPONENTS

### DELL EMC ELASTIC CLOUD STORAGE (ECS)

The third generation object platform from Dell EMC, Elastic Cloud Storage (ECS) is designed for traditional and next-generation applications with unmatched storage efficiency, resiliency and simplicity. ECS has been true software-defined storage that is available as a turnkey appliance, as ECS Software that could be deployed on supported industry-standard hardware, or as a fully EMC-managed ECS-as-a-Service. ECS supports standard object protocols such as S3 as well as file system protocols such as NFS and it provides a Hadoop Compatible File System (HCFS) interface.

ECS provides the following key features relevant to this solution.

- ECS supports a variety of storage protocols, including S3, CAS, and NFS. For access from Hadoop, a Hadoop Compatible File System (HCFS) client-side Java library is provided.
- ECS supports 1 to 8 sites in a federation. Each site can be in a different data center anywhere in the world.
- Replication occurs asynchronously which provides low latency updates even with a world-wide federation.
- Replication is configured per bucket. A bucket on ECS is the root of a particular file system. An ECS cluster can have multiple buckets and each bucket can be configured with different replication properties. For instance, one bucket can be configured to replicate between three specific sites while another bucket can be configured to only exist at a single site.
- Data Protection
  - Data is protected *across* sites (VDCs) using either XOR erasure coding or mirrors. When a bucket is replicated across three sites, for example, 1/3rd of the stored bytes are XOR parity. With four sites, it becomes 1/4th, and so on. All data is fully recoverable from the complete failure of any single site.
  - Data is protected *within* a site (VDC) using Reed Solomon erasure coding. For each 12 blocks of data, 4 additional blocks of parity are calculated and stored. Each of these 16 blocks is distributed on different disks within the site. All data is fully recoverable from the failure of any four drives or a single node within the site.
  - All data is fully recoverable from the complete failure of a single site plus four drives in each site.
  - All data is fully recoverable from the complete failure of a single site plus one node in each site.
  - The two layers of data protection (across sites and within a site) work together to provide an efficiency starting at 38% for two sites and increasing up to 66% for eight sites. Storage efficiency is the effective percentage of raw disk bytes that are usable by your data.

### SPLUNK ENTERPRISE

As this document is limited in scope to the archival of Splunk buckets, the architecture of the Splunk Enterprise servers is out of scope.

### HORTONWORKS DATA PLATFORM (HDP) HADOOP CLUSTER

To allow searching of Splunk buckets archived with Hadoop Data Roll, a Hadoop cluster must be installed and properly configured. Dell EMC recommends Hortonworks Data Platform (HDP) as this is supported with ECS.

## PRODUCTS AND VERSIONS

Below is the list of products and versions that this guide is based on.

**Table 1. Products and Versions**

PRODUCT / COMPONENT	VERSION	DETAILS
ECS OS	3.0.0.0-1422.d46985b.663	ecs-os-setup-target.x86_64-3.0.0.0-1422.d46985b.663.install.iso
ECS Fabric and Object	3.0.0.0.85807.98632a9	ecs-3.0.0.0-2398.9f9f451.582-production.tgz
ECS Client	3.0.0.0.85807.98632a9	hdfsclient-3.0.0.0.85807.98632a9.zip
Splunk Enterprise	6.5.2	
Hortonworks Hadoop Data Platform (HDP)	2.4.2	Includes: Hadoop 2.7.1.2.4
Apache Ambari	2.2.2.0	
CentOS Linux	7.2.1511 x64	Used for all hosts except ECS nodes
ecs-sync	3.1.2	

## TESTING ENVIRONMENT

**Table 2. List of resources and technologies used to support the testing process**

DESCRIPTION	VERSION	DETAILS & SPECIFICATIONS
(1) ECS U1500 rack	3.0.0.0	(4) nodes per rack Each node has: (60) 6 TB, 7200 RPM drives (1) Intel® Xeon® CPU E5-2609 v2 @ 2.50GHz 64 GB RAM (2) 10 GbE Arista 7124 switch (4) 10 Gbps uplinks to Node Managers 1440 TB raw capacity
(1) Splunk Enterprise Node		VMware Virtual Machine 64 GiB Memory 16 CPU Cores 1000 GB virtual hard disk stored on a VNX 7600 FlashVP tier
(6) Node Managers		2x Intel Xeon E5-2650 v3 Processor (2.3 GHz, 10 Cores, 10x 256KB L2 and 25MB L3 Cache) 256 GB RAM 2x 10GbE (Intel X520-DA2), 4x 1GbE (AF1 Slot), 4x 1GbE (AF2 Slot) [Only a single 10 GbE NIC was used for data] 1x 1Gb OOB (Motherboard) 1x 32 GB SSD [used for OS and application binaries] 12x 600GB, SAS, 10K, 2.5 inch [all 12 used for shuffle/intermediate data] CentOS 7.2
(3) Hadoop Master Nodes		VMware Virtual Machine 32 GB Memory 4 CPU Cores 128 GB hard disk CentOS 7.2
(1) Cisco Nexus 7718 switch		The ECS rack connects to this core switch via its Arista 7124 uplink switches (four 10 Gbps link each). All other hosts connect directly to this core switch (single 10 Gbps link for each Node Manager).

## INSTALLATION AND CONFIGURATION

### SPLUNK



The steps below describe a basic and minimal installation of Splunk that is sufficient to perform the archive processes. It should not be relied upon to perform a production installation of Splunk.

For complete details, refer to <http://docs.splunk.com/Documentation/Splunk/6.5.2/Installation/InstallonLinux>.

We'll create a user called *splunk* to run the Splunk processes.

```
[root@splunk-01 ~]#
adduser splunk
mkdir /opt/splunk
chown splunk /opt/splunk
sudo -u splunk -i

[splunk@splunk-01 ~]$
wget -O splunk-6.5.2-67571ef4b87d-Linux-x86_64.tgz
'https://www.splunk.com/bin/splunk/DownloadActivityServlet?architecture=x86_64&platform=linux&version=6.5.2&product=splunk&filename=splunk-6.5.2-67571ef4b87d-Linux-x86_64.tgz&wget=true'

tar -xvzf splunk-6.5.2-67571ef4b87d-Linux-x86_64.tgz -C /opt
SPLUNK_HOME=/opt/splunk
$SPLUNK_HOME/bin/splunk start
```

Open your browser to <http://splunk-01.example.com:8000>.

## ECS

Follow the procedures in the *ECS Administrator's Guide* to perform the steps below.

1. Configure storage pools.
2. Configure Virtual Data Centers (VDCs).
3. Configure an appropriate replication group.
4. Configure a namespace. This document will use the namespace *ns1* but any valid name can be used.
5. Configure users. At a minimum, create a user named *splunk*.

## HADOOP DATA ROLL

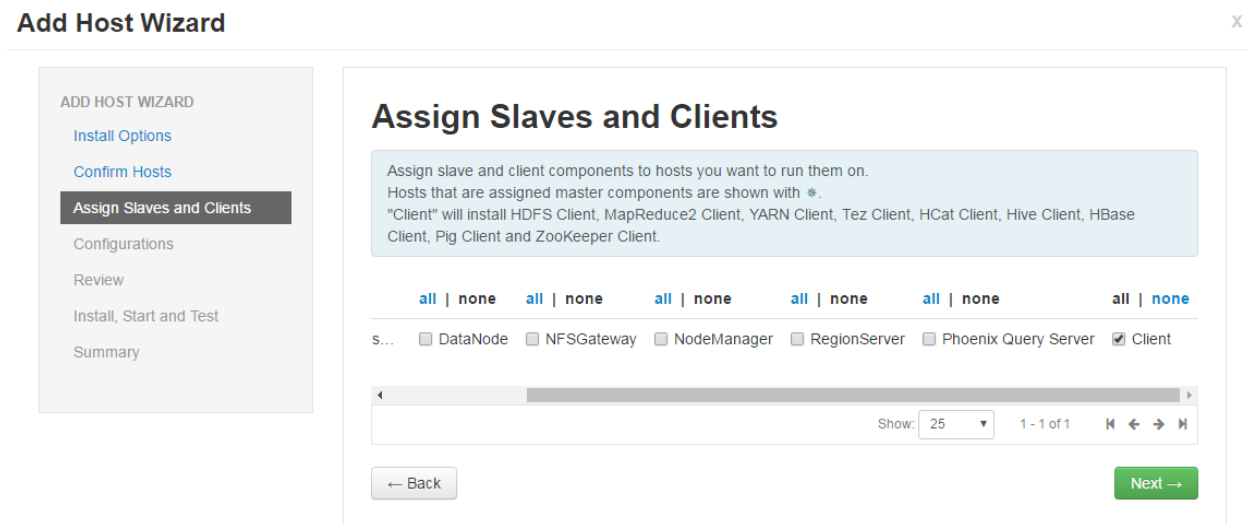
This section describes the additional steps that are specific to Hadoop Data Roll. If you will only use the NFS or S3 method, you may skip this entire section.

### HORTONWORKS DATA PLATFORM (HDP)

To use Hadoop Data Roll, you must install and configure an HDP cluster for ECS. Refer to the *ECS Data Access Guide* for details. You may choose to secure the HDP cluster with Kerberos if desired.

Once HDP is installed, you must install the Hadoop client on all Splunk nodes. Use Ambari's Add Host Wizard to install the Client components.

Figure 2. Add Host Wizard



### Adding the Hadoop client to a Splunk node

Next, you must deploy the ECS HDFS client to all Splunk nodes. Follow the same procedure used to deploy the ECS HDFS client to the HDP nodes.

#### ADD SPLUNK USER

```
[hdfs@hdp-01 ~]$  
hadoop fs -mkdir -p /user/splunk  
hadoop fs -chown splunk /user/splunk
```

If using Kerberos, you must also create the user on your KDC, create a keytab, distribute the keytab, and update the security metadata in the secure ECS bucket. Refer to the *ECS Data Access Guide* for details.

#### CREATE VIRTUAL INDEX PROVIDER FOR HADOOP

Next, create the Virtual Index Provider using the Splunk UI. Go to Settings -> Data -> Virtual Indexes -> New Provider.

The settings in the screen shots below were successfully tested but you will need to edit them for your specific environment. If you are unsure of the specific paths in your environment, run "hadoop envvars".

```
[splunk@hdp-01 ~]$ hadoop envvars  
JAVA_HOME='/usr/jdk64/jdk1.8.0_60'  
HADOOP_COMMON_HOME='/usr/hdp/2.4.2.0-258/hadoop'  
HADOOP_COMMON_DIR='./'  
HADOOP_COMMON_LIB_JARS_DIR='lib'  
HADOOP_COMMON_LIB_NATIVE_DIR='lib/native'  
HADOOP_CONF_DIR='/usr/hdp/2.4.2.0-258/hadoop/conf'  
HADOOP_TOOLS_PATH='/usr/hdp/2.4.2.0-258/hadoop/share/hadoop/tools/lib/*'
```

**Figure 3. Create a Virtual Index Provider**

Edit provider  
Virtual indexes > Edit provider

Name  
hdp9

Description  
optional

Provider Family  
hadoop ▾

### Environment Variables

#### Java Home

/usr/jdk64/jdk1.8.0\_60

Example: /usr/jdk

#### Hadoop Home

/usr/hdp/2.4.2.0-258/hadoop

Example: /usr/hadoop

### Hadoop Cluster Information

#### Hadoop Version

Hadoop 2.x, (Yarn) ▾

#### File System

viprfs://hdp5root.ns1.federation1

Example: hdfs://namenode.example.com:8020

Enable Pass Through Authentication

#### Resource Manager Address

hop-claudio-hdp5-master-0.solarch.lab.emc.com:8050

#### Resource Scheduler Address

hop-claudio-hdp5-master-0.solarch.lab.emc.com:8030

Figure 4. Create a Virtual Index Provider (continued)

### Splunk Settings

HDFS Working Directory ?

/user/splunk/hop-claudio-splunk1-01

Example: /user/splunk/hunk-01.example.com/

Job Queue ?

default

Example: highPriorityQ

Add Secure Cluster

### Security Settings

Mode

kerberos

Kerberos Server Configuration:

Configuration File Path ▾

Kerberos Configuration File Path

/etc/krb5.conf

Hadoop Kerberos Credentials:

Kerberos Principal Name

splunk@KR.SOLARCH.LAB.EMC.COM

Kerberos Keytab Path

/mnt/home/faheyc/splunk/splunk.keytab

HDFS Principal

hdfs-hdp9@KR.SOLARCH.LAB.EMC.COM

Resource Manager Principal

rm/\_HOST@KR.SOLARCH.LAB.EMC.COM

Node Manager Principal

nm/\_HOST@KR.SOLARCH.LAB.EMC.COM

### CREATE ARCHIVED INDEX

Create an Archived Index in the Splunk UI. Go to Virtual Indexes -> Archived Indexes -> New Archived Index.

**Figure 5. Create an Archived Index**

Splunk Index(es) ?

× main

Archived Index Name Suffix ?

\_archive

Destination Provider

hdp5 ▾

Destination Path in HDFS ?

/user/splunk/archived

Examples: /home/data/archive/ or hdfs://namenode:8020/user/john/data/

Older Than ?

1 minute(s) ▾

Enable Unified Search  
Unified Search may increase search runtime.

Cutoff Time

second(s) ▾

Query against virtual index for events older than this value. [Learn More](#)

## FROZEN PATH TO NFS

This section describes the additional steps that are specific to using archiving Splunk buckets to ECS using an NFS mount.

### ECS

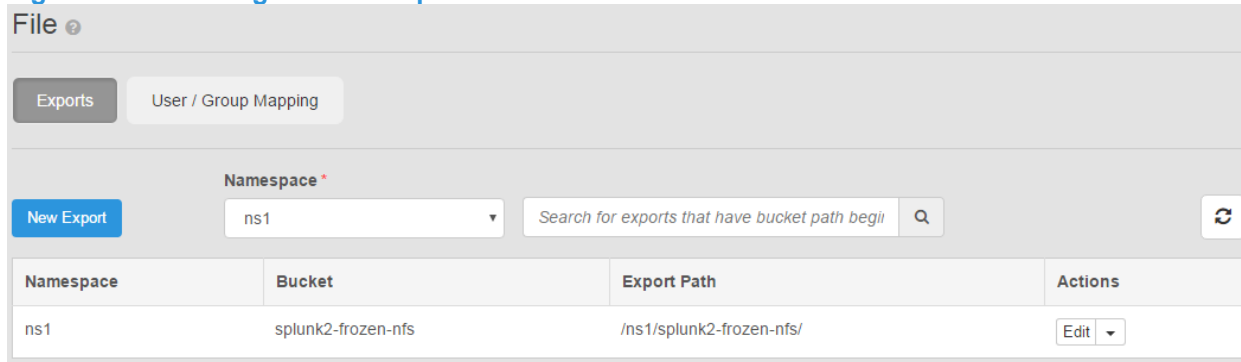
1. Create an ECS bucket
  - a. Create a bucket named *splunk-frozen-nfs*. A different name can be used if desired.
  - b. Bucket Owner: *splunk*
  - c. File System: Enabled
  - d. Default Bucket Group: *splunk*
  - e. Group File Permissions: Read, Write checked
  - f. Group Directory Permissions: Read, Write, Execute checked
2. Edit bucket ACL
  - a. User ACLs
    - i. splunk: Full Control
  - b. Custom Group ACLs:
    - i. splunk: Full Control

### 3. File

#### a. Exports

- i. Create a new export.
- ii. Namespace: ns1
- iii. Bucket: splunk-frozen-nfs
- iv. Export Path: /ns1/splunk-frozen-nfs
- v. Export Host Options:
  1. Host: IP address(es) for your Splunk hosts
  2. Permissions: Read/Write
  3. Write Transfer Policy: Async
  4. Authentication: Sys
  5. Mounting Directories Allowed: Yes

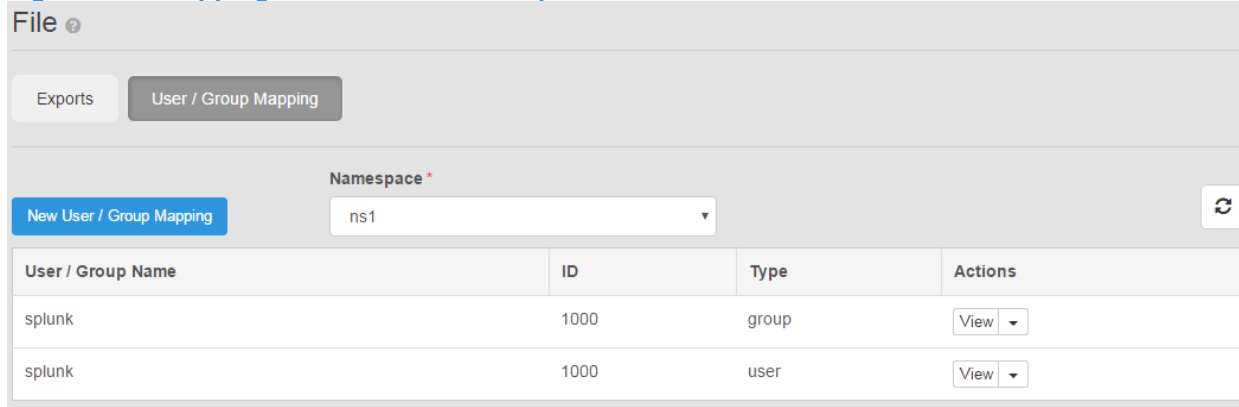
**Figure 6. Creating the NFS Export**



#### b. User / Group Mapping

- i. Create a mapping for the *splunk* user. The ID must match the uid for the *splunk* user on all of the Splunk hosts. Run "id" to see your uid.
- ii. Create a mapping for the *splunk* group. The ID must match the gid for the *splunk* group on all of the Splunk hosts. Run "id" to see your gid.

**Figure 7. Mapping NFS Users and Groups**



### MOUNT THE NFS EXPORT ON SPLUNK HOSTS

On each Splunk host, add the following line to the file `/etc/fstab`. It must all be on one line.

```
ecs1.example.com:/ns1/splunk-frozen-nfs /mnt/splunk-frozen-nfs nfs
async,vers=3,nolock,rsize=524288,wsiz=524288 0 0
```

Then run the following:

```
[root@splunk-01 ~]#
mkdir -p /mnt/splunk-frozen-nfs
mount -a
sudo -u splunk -i
[splunk@splunk-01 ~]$
ls /mnt/splunk-frozen-nfs
echo test > /mnt/splunk-frozen-nfs/testfile
rm /mnt/splunk-frozen-nfs/testfile
```

### SPLUNK

Lastly, for each index that you want to archive to NFS, use the Splunk UI to edit the index and set the *Frozen Path* to:

```
/mnt/splunk-frozen-nfs/your_index_name
```

Additionally, set *Max Size of Entire Index* to an appropriate value. When the size of the index exceeds this amount, Splunk will begin moving the oldest buckets to the Frozen Path.

You may need to restart Splunk for the changes to take effect.

### COLD TO FROZEN S3

This section describes the additional steps that are specific to using archiving Splunk buckets to ECS using S3.

#### ECS

1. If you do not already have one, create a user named *splunk*.
2. Edit the *splunk* user. Click the button to *Generate & Add Password*. Record the S3 password.
3. Create an ECS bucket
  - a. Create a bucket named *splunk-frozen-s3*. A different name can be used if desired.
  - b. Bucket Owner: *splunk*

- c. File System: Disabled
- 4. Edit bucket ACL
  - a. User ACLs
    - i. splunk: Full Control

## ECS-SYNC

On each Splunk host, ecs-sync must be installed. This application is used to copy files to ECS using the S3 protocol. Execute the steps below.

```
[root@splunk-01 ~]#  
mkdir -p /opt/ecs-sync  
chown splunk /opt/ecs-sync  
wget https://github.com/EMCECS/ecs-sync/releases/download/v3.1.2/ecs-sync-3.1.2.zip  
unzip ecs-sync-3.1.2.zip  
yum -y install java-1.7.0-openjdk
```

## COLD TO FROZEN ECS S3 SCRIPT

This is a Python script based on Splunk's coldToFrozenExample.py that copies a cold Splunk bucket to ECS using ecs-sync.

```
[splunk@splunk-01 ~]$  
cd /opt/splunk/bin  
wget -O coldToFrozenECSS3.py \  
https://gist.githubusercontent.com/claudiofahey/adeaf399dd62225526b8903cf9134d9c/raw/ddf0ca4286ca510da2b4da3f0500a2b70cb44812/coldToFrozenECSS3.py  
vi coldToFrozenECSS3.py
```

Edit the configuration parameters at the top of the script to match your environment.

## SPLUNK

Edit the file `$$SPLUNK_HOME/etc/system/local/indexes.conf` and set the `coldToFrozenScript` parameter.

```
coldToFrozenScript = "$SPLUNK_HOME/bin/python" "$SPLUNK_HOME/bin/coldToFrozenECSS3.py"
```

Additionally, set *Max Size of Entire Index* to an appropriate value. When the size of the index exceeds this amount, Splunk will begin moving the oldest buckets by running the cold to frozen script.

You will need to restart Splunk for the changes to take effect.

## USAGE

### HADOOP DATA ROLL

It is assumed that some events have already been loaded into Splunk. Hadoop Data Roll will not archive hot buckets. If you only have newly indexed events, force the hot bucket to become warm by restarting Splunk.

When enabled, Hadoop Data Roll will automatically copy old warm and cold buckets to the Hadoop file system (ECS) every 1 hour. To force the archive process to execute immediately, open the Splunk search app and type the following:

```
| archivebuckets forcerun=1
```



If any errors occur, review the log file `/opt/splunk/var/log/splunk/splunk_archiver.log`.

To perform a search on the archived index using the Hadoop cluster's MapReduce framework, simply search against the `index_archive` index. For example:

```
index=main_archive productID=1000
```

If any errors occur, review the job logs in the YARN Resource Manager. You can also review the log files `/opt/splunk/var/run/splunk/dispatch/*/search.log`.

## FROZEN PATH TO NFS

When the size of the index exceeds *Max Size of Entire Index*, Splunk will begin moving the oldest buckets to the Frozen Path.

You may review the log file `/opt/splunk/var/log/splunk/splunkd.log` and the target directory `/mnt/splunk-frozen/main`.

## COLD TO FROZEN S3

When the size of the index exceeds *Max Size of Entire Index*, Splunk will begin moving the oldest buckets by running the cold to frozen script.

You may review the log file `/opt/splunk/var/log/splunk/splunkd.log` and the target bucket using a tool such as S3 Browser.

## CONCLUSION

This document shows how ECS can be used effectively as an archive target for Splunk frozen buckets using three different methods.

## REFERENCES

**Table 3. References**

DESCRIPTION	DETAIL / LINKS
ECS Product Page	<a href="http://www.emc.com/ecs">http://www.emc.com/ecs</a>
ECS 3.0 Product Documentation	<a href="https://community.emc.com/docs/DOC-53956">https://community.emc.com/docs/DOC-53956</a>
Hortonworks Data Platform	<a href="http://hortonworks.com/">http://hortonworks.com/</a>
Splunk Documentation	<a href="http://docs.splunk.com/Documentation/Splunk/6.5.2">http://docs.splunk.com/Documentation/Splunk/6.5.2</a>
ecs-sync	<a href="https://github.com/EMCECS/ecs-sync">https://github.com/EMCECS/ecs-sync</a>
Dell EMC Reference Architecture for Splunk	<a href="http://www.emc.com">http://www.emc.com</a>

## DOCUMENT VERSION HISTORY

**Table 4. Document Version History**

VERSION #	REVISION DATE	REVISION AUTHOR	REVISION DETAIL
1.0	28 April 2017	Claudio Fahey	Initial document