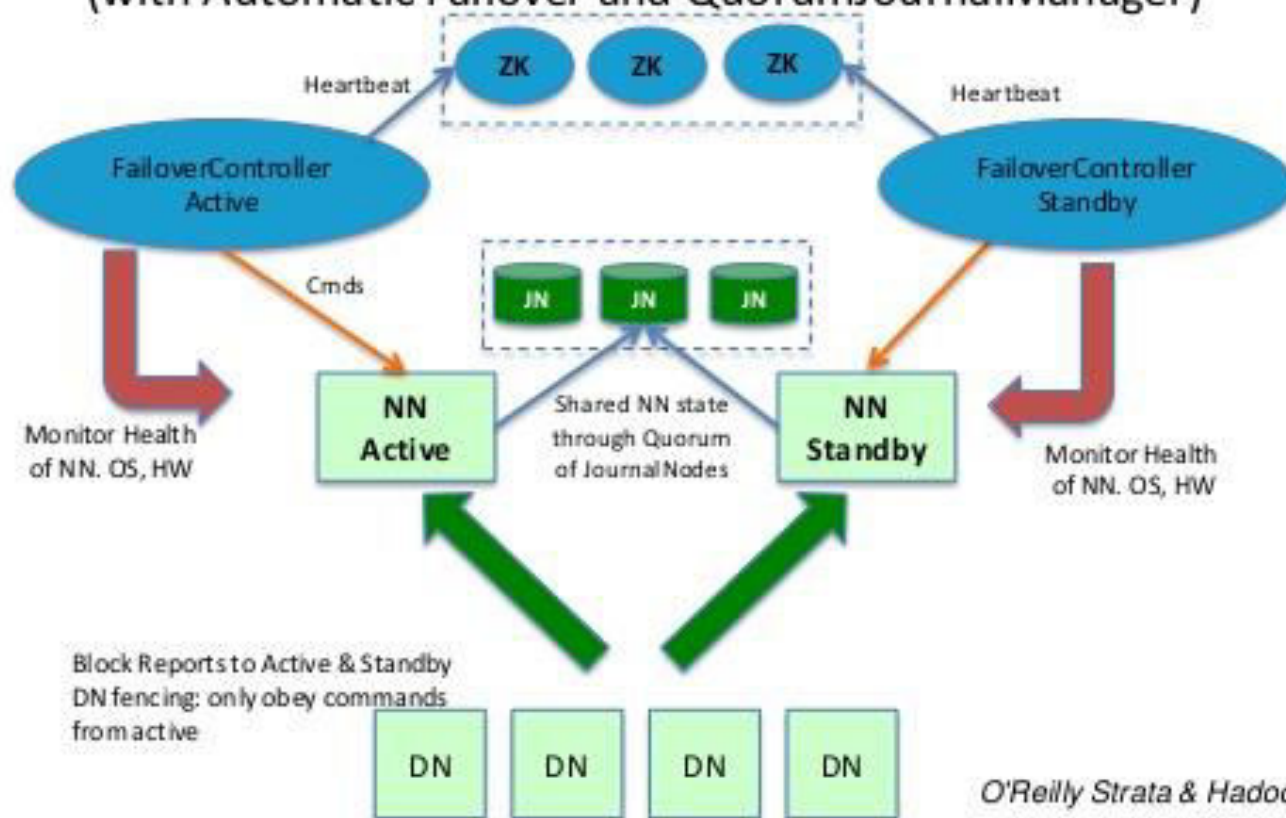


# Namenode HA

## HDFS HA Architecture (with Automatic Failover and QuorumJournalManager)



O'Reilly Strata & Hadoop World

# Enabling HA

Enable Na

Support

service	HDFS <b>Enable</b> High Availability
	Oozie <b>Enable</b> High Availability
	HDFS Add <b>Nameservice</b>
	YARN (MR2 Included) <b>Enable</b> High Availability
config	Cloudera <b>Management</b> Service: <b>Enable</b> Expression Input
	Cloudera <b>Management</b> Service: <b>Enable</b> Audit Collection
	Cloudera <b>Management</b> Service: <b>Navigator</b> Audit Server Database...
	HDFS: <b>NameNode</b> Safemode Health Test
	HDFS: <b>NameNode</b> Process Health Test
	Cloudera <b>Management</b> Service: <b>Navigator</b> Metadata Server Data...
	HDFS: Bind <b>NameNode</b> to Wildcard Address
	Cloudera <b>Management</b> Service: <b>Navigator</b> Audit Server Database...
	Cloudera <b>Management</b> Service: <b>Navigator</b> Metadata Server Data...
	Cloudera <b>Management</b> Service: AWS Credential
	Hive: <b>Enable</b> Lineage Collection
role	HDFS: <b>NameNode</b> (ip-172-31-15-57)

Recent Commands

which uses an en

Cha

Clus

percent

2h 1d 7d

:45

/s

# Enabling HA



cloudera MANAGER

Support ▾ admin

## Enable High Availability for HDFS

### Getting Started

This wizard leads you through adding a standby NameNode, restarting this HDFS service and any dependent services, and then re-deploying client configurations.

Nameservice Name

Enabling High Availability creates a new nameservice. Accept the default name **nameservice1** or provide another name in **Nameservice Name**.

Back

1 2 3 4 5

Continue

# Enabling HA



cloudera MANAGER

Support ▾ admin

## Enable High Availability for HDFS

### Assign Roles

NameNode Hosts

ip-172-31-15-57.us-west-2.compute.internal (Current)

Select a host

JournalNode Hosts

Select hosts

We recommend that JournalNodes be hosted on machines of similar hardware specifications as the NameNodes. The hosts of NameNodes and the ResourceManager are generally good options. You must have a minimum of three and an odd number of JournalNodes.

Back

1 2 3 4 5

Continue

:2-52-37-178-78.us-west-2.compute.amazonaws.com:7180/cmf/services/16/enableHA/index#

# Enabling HA

2 Hosts Selected

Select hosts for a new or existing role. The host list is filtered to remove hosts that are not valid candidates; these include hosts that are unhealthy, members of other clusters, or have an incompatible version of CDH installed on them.

Enter hostnames: host01, host[01-10], IP addresses or rack.

Search

	Hostname	IP Address	Rack	Cores	Physical Memory	Existing Roles	Added Roles
<input checked="" type="checkbox"/>	ip-172-31-11-93.us-west-2.compute.internal	172.31.11.93	/default	2	7.7 GiB	M B SNN G HMS HS2 HS ICS ISS LHBI AP ES HM RM SM OS SS G HS S	NN
<input type="checkbox"/>	ip-172-31-12-112.us-west-2.compute.internal	172.31.12.112	/default	2	7.7 GiB	RS DN G ID G NM S	
<input type="checkbox"/>	ip-172-31-14-9.us-west-2.compute.internal	172.31.14.9	/default	2	7.7 GiB	RS DN G ID G NM S	
<input checked="" type="checkbox"/>	ip-172-31-15-57.us-west-2.compute.internal	172.31.15.57	/default	2	7.7 GiB	NN G G JHS RM S	
<input type="checkbox"/>	ip-172-31-8-159.us-west-2.compute.internal	172.31.8.159	/default	2	7.7 GiB	RS DN G ID G NM S	

OK

Cancel

# Enabling HA

3 Hosts Selected

Select hosts for a new or existing role. The host list is filtered to remove hosts that are not valid candidates; these include hosts that are unhealthy, members of other clusters, or have an incompatible version of CDH installed on them.

Enter hostnames: host01, host[01-10], IP addresses or rack.

Search

<input type="checkbox"/>	Hostname	IP Address	Rack	Cores	Physical Memory	Existing Roles	Added Roles
<input checked="" type="checkbox"/>	ip-172-31-11-93.us-west-2.compute.internal	172.31.11.93	/default	2	7.7 GiB	<div>MBSNNGHMSHS2HSICSISSLHBIAPESHMRRMSOSSSGHS</div>	<div>NNJN</div>
<input type="checkbox"/>	ip-172-31-12-112.us-west-2.compute.internal	172.31.12.112	/default	2	7.7 GiB	<div>RSNDGIDGNMSS</div>	
<input checked="" type="checkbox"/>	ip-172-31-14-9.us-west-2.compute.internal	172.31.14.9	/default	2	7.7 GiB	<div>RSNDGIDGNMSS</div>	<div>JN</div>
<input checked="" type="checkbox"/>	ip-172-31-15-57.us-west-2.compute.internal	172.31.15.57	/default	2	7.7 GiB	<div>NNGGJHSRMS</div>	<div>JN</div>
<input type="checkbox"/>	ip-172-31-8-159.us-west-2.compute.internal	172.31.8.159	/default	2	7.7 GiB	<div>RSNDGIDGNMSS</div>	

Tip: Click the first checkbox, hold down the Shift key and click the last checkbox to select a range.

OK

Cancel

© Copyright, IntelliPaat Software Solutions Pvt. Ltd. All rights reserved.

# Enabling HA



cloudera MANAGER

Support ▾ admin

## Enable High Availability for HDFS

### Review Changes

Set the following configuration values for your new role(s). Required values are marked with \*.

Parameter	Group ⓘ	Value	Description
<b>Service HDFS</b>			
NameNode Data Directories* dfs.namenode.name.dir	ip-172-31-11-93	/dfs/nn Inherited from: NameNode Default Group	Determines where on the local file system the NameNode should store the name table (fsimage). For redundancy, enter a comma-delimited list of directories to replicate the name table in all of the directories. Typical values are /data/N/dfs/nn where N=1..3.
	ip-172-31-15-57	/dfs/nn Inherited from: NameNode Default Group	
JournalNode Edits Directory* dfs.journalnode.edits.dir	ip-172-31-11-93	Inherited value is empty. Click to edit.	Directory on the local file system where NameNode edits are written.
	ip-172-31-14-9	Inherited value is empty. Click to edit.	
	ip-172-31-15-57	Inherited value is empty. Click to edit.	

### Extra Options

Back

1 2 3 4 5

Continue



# Enabling HA

## Enable High Availability for HDFS

### Review Changes

Set the following configuration values for your new role(s). Required values are marked with \*.

Parameter	Group ?	Value	Description
<b>Service HDFS</b>			
NameNode Data Directories* dfs.namenode.name.dir	ip-172-31-11-93	/dfs/nn Inherited from: NameNode Default Group	Determines where on the local file system the NameNode should store the name table (fsimage). For redundancy, enter a comma-delimited list of directories to replicate the name table in all of the directories. Typical values are /data/N/dfs/nn where N=1..3.
	ip-172-31-15-57	/dfs/nn Inherited from: NameNode Default Group	
JournalNode Edits Directory* dfs.journalnode.edits.dir	ip-172-31-11-93	<input type="text" value="/jn"/> <a href="#">Reset to empty default value ↺</a>	Directory on the local file system where NameNode edits are written.
	ip-172-31-14-9	<input type="text" value="/jn"/> <a href="#">Reset to empty default value ↺</a>	
	ip-172-31-15-57	<input type="text" value="/jn"/> <a href="#">Reset to empty default value ↺</a>	

Back

1 2 3 4 5

Continue

# Enabling HA



cloudera MANAGER

Support admin

## Enable High Availability for HDFS

✓ Enable High Availability Command

Status: **Finished** Context: [HDFS](#) Start Time: Mar 15, 9:50:28 PM Duration: 6m

Successfully enabled High Availability and Automatic Failover

### Details

Completed 20 of 20 step(s).

☒ All ☐ Failed Only ☐ Running Only

Step	Context	Start Time	Duration	Actions
✓ Check that name directories for the new Standby NameNode either do not exist or are writable and empty. Can optionally clear directories. Process host-validate-writable-empty-dirs (id=170) on host ip-172-31-11-93.us-west-2.compute.internal (id=3) exited with 0 and expected 0	<a href="#">ip-172-31-11-93.us-west-2.compute.internal</a>	Mar 15, 9:50:28 PM	628ms	
➤ ✓ Check that edits directories for the nameservice either do not exist or are writable and empty. Can optionally clear directories. Successfully completed 3 steps.		Mar 15, 9:50:28 PM	869ms	
➤ ✓ Stop hdfs and its dependent services All services successfully stopped.	<a href="#">Bias Cluster</a>	Mar 15, 9:50:29 PM	28.29s	

Back

1 2 3 4 5

Continue

**cloudera** MANAGER

## Enable High Availability for HDFS

Congratulations!

Successfully enabled High Availability.

# Enabling HA



```
[root@ip-172-31-12-112 ~]# head -15 /etc/hadoop/conf/core-site.xml
<?xml version="1.0" encoding="UTF-8"?>

<!--Autogenerated by Cloudera Manager-->
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://nameservice1</value>
  </property>
  <property>
    <name>fs.trash.interval</name>
    <value>1</value>
  </property>
  <property>
    <name>io.compression.codecs</name>
    <value>org.apache.hadoop.io.compress.DefaultCodec,org.apache.hadoop.io.compress.GzipCodec,org.apache.hadoop.io.compress.BZip2C
odec,org.apache.hadoop.io.compress.DeflateCodec,org.apache.hadoop.io.compress.SnappyCodec,org.apache.hadoop.io.compress.Lz4Codec</
value>
```

# Enabling HA



```
<property>
  <name>dfs.ha.namenodes.nameservice1</name>
  <value>namenode95,namenode57</value>
</property>
<property>
  <name>dfs.namenode.rpc-address.nameservice1.namenode95</name>
  <value>ip-172-31-11-93.us-west-2.compute.internal:8020</value>
</property>
<property>
  <name>dfs.namenode.servicerpc-address.nameservice1.namenode95</name>
  <value>ip-172-31-11-93.us-west-2.compute.internal:8022</value>
</property>
<property>
  <name>dfs.namenode.http-address.nameservice1.namenode95</name>
  <value>ip-172-31-11-93.us-west-2.compute.internal:50070</value>
</property>
<property>
  <name>dfs.namenode.https-address.nameservice1.namenode95</name>
  <value>ip-172-31-11-93.us-west-2.compute.internal:50470</value>
</property>
<property>
  <name>dfs.namenode.rpc-address.nameservice1.namenode57</name>
  <value>ip-172-31-15-57.us-west-2.compute.internal:8020</value>
</property>
<property>
  <name>dfs.namenode.servicerpc-address.nameservice1.namenode57</name>
  <value>ip-172-31-15-57.us-west-2.compute.internal:8022</value>
</property>
<property>
```

# Access using logical name









```
[root@ip-172-31-12-112 ~]# hadoop fs -ls hdfs://nameservice1:8020/
```

```
Found 6 items
```

drwxr-xr-x	-	hbase	hbase	0	2017-03-15	21:54	hdfs://nameservice1:8020/hbase
drwxr-xr-x	-	hdfs	supergroup	0	2017-03-15	07:17	hdfs://nameservice1:8020/hello
drwxrwxr-x	-	solr	solr	0	2017-03-15	06:14	hdfs://nameservice1:8020/solr
drwxr-xr-x	-	hdfs	supergroup	0	2017-03-15	07:42	hdfs://nameservice1:8020/system
drwxrwxrwt	-	hdfs	supergroup	0	2017-03-15	06:39	hdfs://nameservice1:8020/tmp
drwxr-xr-x	-	hdfs	supergroup	0	2017-03-15	07:18	hdfs://nameservice1:8020/user

```
[root@ip-172-31-12-112 ~]# █
```

# Failover

Demo CDM	i-01bf264dcca3e6dbb	t2.large	us-west-2c	 running
Demo Namenode	i-03eb66e0a08a089ac	t2.large	us-west-2c	 stopped
Demo SNameode	i-057575c3ce81de46b	t2.large	us-west-2c	 running
Demo DataNode1	i-0711013525c2586aa	t2.large	us-west-2c	 running
Demo DatanNode2	i-085f5a4f098fb9f35	t2.large	us-west-2c	 running
Demo DataNode3	i-0fd070bf5c899b169	t2.large	us-west-2c	 running



# Failover



```
Caused by: java.net.NoRouteToHostException: No route to host
    at sun.nio.ch.SocketChannelImpl.checkConnect(Native Method)
    at sun.nio.ch.SocketChannelImpl.finishConnect(SocketChannelImpl.java:739)
    at org.apache.hadoop.net.SocketIOWithTimeout.connect(SocketIOWithTimeout.java:206)
    at org.apache.hadoop.net.NetUtils.connect(NetUtils.java:530)
    at org.apache.hadoop.net.NetUtils.connect(NetUtils.java:494)
    at org.apache.hadoop.ipc.Client$Connection.setupConnection(Client.java:615)
    at org.apache.hadoop.ipc.Client$Connection.setupIOstreams(Client.java:714)
    at org.apache.hadoop.ipc.Client$Connection.access$2900(Client.java:376)
    at org.apache.hadoop.ipc.Client.getConnection(Client.java:1525)
    at org.apache.hadoop.ipc.Client.call(Client.java:1448)
    ... 29 more
```

Found 6 items

drwxr-xr-x	-	hbase	hbase	0	2017-03-15	21:54	hdfs://nameservice1:8020/hbase
drwxr-xr-x	-	hdfs	supergroup	0	2017-03-15	07:17	hdfs://nameservice1:8020/hello
drwxrwxr-x	-	solr	solr	0	2017-03-15	06:14	hdfs://nameservice1:8020/solr
drwxr-xr-x	-	hdfs	supergroup	0	2017-03-15	07:42	hdfs://nameservice1:8020/system
drwxrwxrwt	-	hdfs	supergroup	0	2017-03-15	06:39	hdfs://nameservice1:8020/tmp
drwxr-xr-x	-	hdfs	supergroup	0	2017-03-15	07:18	hdfs://nameservice1:8020/user



# Failover



## Log Details

Host [ip-172-31-11-93.us-west-2.compute.internal](#) [Change](#) Role Failover Controller - [Change](#)

File `/var/log/hadoop-hdfs/hadoop-cmf-hdfs-FAILOVERCONTROLLER-ip-172-31-11-93.us-west-2.compute.internal.log.out` [Download Full Log](#)

March 15, 2017 9:53 PM - March 15, 2017 10:05 PM

```
java.nio.channels.SocketChannel[connection-pending remote=ip-172-31-15-57.us-west-
  at org.apache.hadoop.net.NetUtils.connect(NetUtils.java:533)
  at org.apache.hadoop.net.NetUtils.connect(NetUtils.java:494)
  at org.apache.hadoop.ipc.Client$Connection.setupConnection(Client.java:611)
  at org.apache.hadoop.ipc.Client$Connection.setupIOstreams(Client.java:714)
  at org.apache.hadoop.ipc.Client$Connection.access$2900(Client.java:376)
  at org.apache.hadoop.ipc.Client.getConnection(Client.java:1525)
  at org.apache.hadoop.ipc.Client.call(Client.java:1448)
  ... 14 more
```

Mar 15, 10:05:38.096 PM	INFO	org.apache.hadoop.ha.NodeFencer	===== Beginning Service Fencing Process... =====
Mar 15, 10:05:38.096 PM	INFO	org.apache.hadoop.ha.NodeFencer	Trying method 1/1: org.apache.hadoop.ha.ShellCommandFencer(true)
Mar 15, 10:05:38.110 PM	INFO	org.apache.hadoop.ha.ShellCommandFencer	Launched fencing command 'true' with pid 1767
Mar 15, 10:05:38.114 PM	INFO	org.apache.hadoop.ha.NodeFencer	===== Fencing successful by method org.apache.hadoop.ha.ShellCommandFencer(true)
Mar 15, 10:05:38.114 PM	INFO	org.apache.hadoop.ha.ActiveStandbyElector	Writing znode /hadoop-ha/nameservice1/ActiveBreadCrumb to indicate that the local
Mar 15, 10:05:38.121 PM	INFO	org.apache.hadoop.ha.ZKFailoverController	Trying to make NameNode at ip-172-31-11-93.us-west-2.compute.internal/172.31.11.93
Mar 15, 10:05:38.909 PM	INFO	org.apache.hadoop.ha.ZKFailoverController	Successfully transitioned NameNode at ip-172-31-11-93.us-west-2.compute.internal/

# Thank You