Centrify Identity and Access Management for MapR

Centrify Server Suite 2016
MapR 5.1

March 2016

Contents

General Information

Hadoop is a distributed computing infrastructure comprised of a Distributed File System (HDFS) and the MapReduce batch-processing engine. You can use the Hadoop framework to process massive data sets with distributed computing systems by using a large number of commodity servers.

While Hadoop is a powerful analytics tool for data sets, it doesn't incorporate user authentication when installed using its predefined defaults, which leaves potentially sensitive information unsecure. To address this potential security breach, you can use Centrify Server Suite, which provides a turnkey NIST 140-2 FIPS certified Kerberos infrastructure, to make use of the Hadoop environment's native Kerberos authentication protocol that requires "tickets" to allow nodes to identify themselves.

Integrating Centrify Server Suite with a Hadoop installation will allow you to use your existing Active Directory environment to enable security for your Hadoop deployment using Kerberos authentication without the need to create a stand-alone Kerberos MIT realm.

Centrify Server Suite for Hadoop Deployments

This document provides the steps and configuration for multiple MapR clusters to be added to Active Directory. The key to multiple clusters in Active Directory is the addition of a cluster prefix to the associated MapR Kerberos principal. Without the cluster prefix, Kerberos principals would have the same User Principal Name (UPN). The account name (UPN) must be unique within the Active Directory domain.

General Benefits

- Native Integration with Active Directory
  This is a proven capability of Centrify across many UNIX/Linux platforms that provides centralized administration.

- System-Level Access Controls and Role-Based Access
  Centrify zones and authorization give you a framework to limit access and privilege, as well as the ability to control exactly who has access to data, while enforcing the separation of duties. This capability is available across both Windows and UNIX/Linux computers.

- Session Capture and Replay
  DirectAudit provides you with the ability to record and replay user sessions.

Hadoop Related Benefits

- Faster Deployment
  Rather than configuring and maintaining a stand-alone Kerberos MIT realm, you can use Centrify to integrate with AD directly.

- Simplicity
  Because you do not have to duplicate capabilities, there is no need to maintain a second, highly available infrastructure that requires you to accommodate moves and changes, or ensure compliance.

- Flexibility
  Centrify provides tools for automation in elastic environments. Examples include:
  a. User Provisioning
     - You can use zone provisioning agent (ZPA), PowerShell, or adedit - as well as existing solutions via AD groups - to manage privileged access.

b. Host Provisioning
   - You can easily automate the provisioning of hosts, keytab creation, and distributions by using native scripts or tools like adedit.

Creating multiple MapR Hadoop clusters with Centrify Server Suite

Introduction

Hadoop's security implementation uses MIT Kerberos. As a result, all principals are user principals and there must be an Active Directory account for each service that requires a keytab. For example, a MapR cluster requires a single principal, so a 3-node cluster will require a single Kerberos keytab.

Cluster Creation Prerequisites

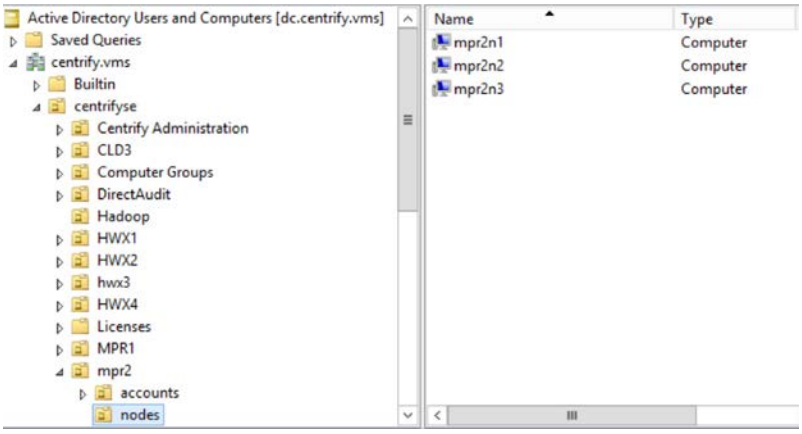
Aside from the typical requirements (Active Directory, a member server for the consoles or audit components such as MS SQL server, and the UNIX/Linux hosts) you should add the following:

- An Active Directory organizational unit (OU) for Hadoop service accounts.
  - Create a sub OU (users) for Hadoop principals.
  - Create a sub OU for Hadoop nodes.
- A technical lead that has full control on the Hadoop OU.
- A **working** 3-node cluster with the latest Centrify agent installed on each node.

  Note: Though it isn't necessary, it is suggested that you set up your own Hadoop repository to speed up the set up process.

Configure Organizational Units and Centrify Zones

- Outline a naming convention for all Hadoop components that will reside in AD. Ideally you will be able to identify the cluster in the name. The sAMAccountName must have a maximum of 20 characters.
- In Centrify Access Manager, create a child zone for each cluster (for example, mapr2). Use the child zone name as the name for the cluster prefix (mapr2). Typical node names within the cluster might be mpr2n1 (representing MapR Cluster 2/ Node 1), mpr2n2 (MapR Cluster 2/ Node 2), and mapr2n3 (MapR Cluster 2/ Node 3).



- In ADUC, create a sub OU for each cluster and use the cluster name for the sub OU. For example, mpr2.
- Within each sub OU create two additional sub OUs (nodes & accounts). The nodes follow the same naming convention, so that mapr2n1 would represent cluster two, node one.

Enabling Hadoop Security using MapR

Provision Node VMs

- Provision 3 identical Centos 6.x virtual machines that will be used for cluster nodes. For example:
  - mpr2n1.centrify.vms, 2 processors, 8GB RAM, 2 HD (100/ 150 GB)
  - mpr2n2.centrify.vms, 2 processors, 8GB RAM, 2 HD (100/ 150 GB)
  - mpr2n3.centrify.vms, 2 processors, 8GB RAM, 2 HD (100/ 150 GB)
- Create the corresponding A records in the centrify.vms DNS zone.
- Create and name the Hadoop Sub OU under UNIX; e.g. mpr2
- Follow the instructions to setup the MapR cluster
  - MapR Installation Located: https://www.mapr.com/products/hadoop-download?cid=newhomedownload

Configure Node VMs

Perform the following steps on each node VM.

In AWS environments:

1. Modify `/etc/cloud/cloud.cfg` by commenting out the lines containing `–set_hostname` and `-update_hostname`.
2. Modify `/etc/hostname` by replacing the AWS generated hostname with the node shortname (for example, `mpr2n1`).
3. Execute the `hostname` command to set the hostname to the node shortname (for example, `hostname mpr2n1`)
4. Reboot the computer.

In all environments:

Execute the following commands to install required software, or verify that the software described here is already installed:

```
yum install –y perl

yum install –y wget

yum install –y openldap-clients

yum install –y krb5-workstation

yum install –y ntp
```

After NTP installed, execute the following commands to enable and start it:

```
systemctl enable ntpd

systemctl start ntpd
```

Install the Centrify Agent on Node VMs

Perform the following steps on each node to install the Centrify agent:

1. Obtain the Centrify agent installation package for your node operating system and make the packace available on each node.
2. Untar the package file (`tar –xvf package_file_name`).
3. Run the agent installation script (`./install.sh`). **Do not join the node computer to the domain during the agent installation process.**

   On the head node (for example, mpr2n1), install DirectControl, Centrify OpenSSH, LDAP Proxy, and Auditing. . Installing the Centrify LDAP proxy is necessary because MapR requires the ability to communicate over secured LDAP (LDAPS) to Active Directory. The procedure for installing and setting the secured LDAP proxy is described in Centrify's *UNIX Administrator's Guide* (PDF) beginning on page 216.

   On other nodes, install DirectControl, Centrify Open SSH, and Auditing.

4. In the Centrify configuration file `/etc/centrifydc/centrifydc.conf`, make the following changes:
   a. Change the default setting of the `adclient.krb5.service.principals` parameter to `ftp cifs`. When you are done, the line should look like this:
      `adclient.krb5.service.principals ftp cifs`
   b. On a new line, add the parameter `client.dynamic.dns.enabled: true`
   c. On a new line, add the parameter `client.dynamic.dns.refresh.interval: 3600`
5. Join the node computer to the domain:

   `adjoin –u admin_user_name –-zone zone_name –-container OU_pathname, Domain_Name`

   For example:

   ```
   adjoin –u administrator –-zone mpr2 –-container
   ou=nodes,ou=mpr2,ou=centrifyse,dc=centrify,dc=vms
   ```

6. Update DNS (`addns –U`).
7. If you did not configure LDAPS when you installed the Centrify LDAP proxy on the head node earlier, do so now by following the instructions in the Centrify *UNIX Administrator's Guide* (PDF) beginning on page 216.

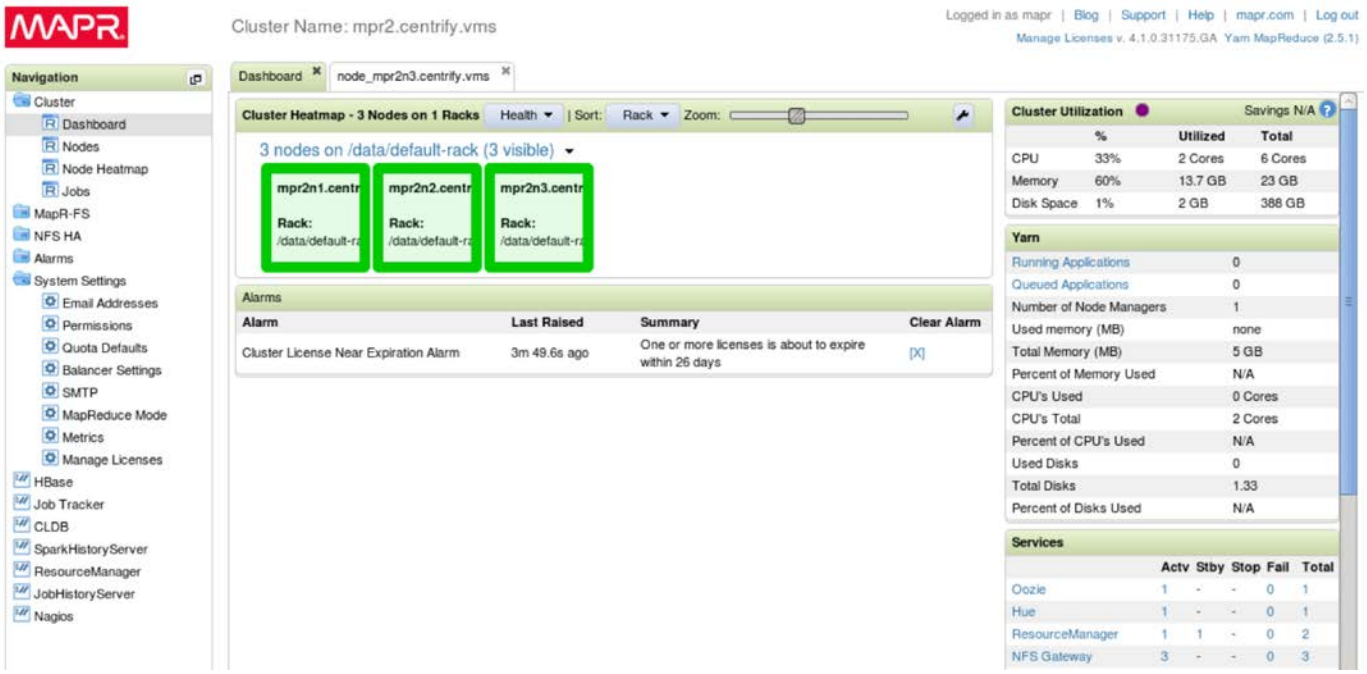
Create home directories and authorize cluster users in HDFS

On Node 1 (mpr2n1), for each Active Directory user that will submit a job, create a home directory and give the user ownership of files in the cluster. The following example shows these steps being taken by the Active Directory administrator `ed` for himself and the Active Directory user `wade`. After these steps are complete, `ed` and `wade` can run jobs in the cluster.

```
Using Kerberos authentication
Using principal ed@CENTRIFY.VMS
Got host ticket host/mpr2n1.centrify.vms@ CENTRIFY.VMS
login as edCENTRIFY.VMS
Successful Kerberos connection
CentOS release 6.7 (Final)
Kernel 2.6.32-358.el6.x86_64 on an x86_64
Password will expire in 11 days
Last login: Tue Nov 10 14:24:02 2015 from m-w2k12r2-dc1.centrify.vms
[ed@mpr2n1 ~]$ dzdo su hdfs
[dzdo] password for ed:
[hdfs@mpr2n1 ed]$ hadoop fs -mkdir /user/ed
[hdfs@mpr2n1 ed]$ hadoop fs -chown ed:ed /user/ed
[hdfs@mpr2n1 ed]$ hadoop fs -mkdir /user/wade
[hdfs@mpr2n1 ed]$ hadoop fs -chown wade:wade /user/wade
```

Summary

After the setup is complete, a dashboard view of the cluster is available from the browser user interface at https://hostname:8443:



Test the Cluster

Before securing, verify the cluster is operating normally before listing out the HDFS and provisioning a user.

```
[root@mpr2n1 hadoop-0.20.2]# su mapr

[mapr@mpr2n1 hadoop-0.20.2]$ hadoop fs -ls /user

Found 4 items

drwxr-xr-x   - dwirth    dwirth          0 2015-07-10 11:01 /user/dwirth

drwxr-xr-x   - mapr      root            2 2015-07-06 08:21 /user/mapr

drwxr-xr-x   - ned_atlas ned_atlas       0 2015-07-10 11:01 /user/ned_atlas

drwxr-xr-x   - root      root            0 2015-07-12 04:59 /user/root

[mapr@mpr2n1 hadoop-0.20.2]$ cd /opt/mapr/hadoop/hadoop-0.20.2/

[mapr@mpr2n1 hadoop-0.20.2]$ hadoop jar hadoop-0.20.2-dev-examples.jar pi 5 10

Number of Maps  = 5

Samples per Map = 10

Job Finished in 19.27 seconds

Estimated value of Pi is 3.28000000000000000000

[mapr@mpr2n1 hadoop-0.20.2]$
```

Enable Security

1. Install Centrify's Direct Control agent on all nodes and join to Active Directory.
2. Shutdown Cluster (stop services on all nodes): stop Warden first
    a. service mapr-warden stop
    b. service mapr-zookeeper stop
3. Start with your install node (node the web interface)
4. From the command line, create the MapR keytab
    a. Command example
    adkeytab - -new - -upn mapr/Mapr-Cluster-Name@REALM –P mapr/Mapr-Cluster-Name@REALM - -keytab /opt/mapr/conf/mapr.keytab -c - -ou-AD-path - -ignore - -V -M AD-account-name
        i. - -new: create a principal
        ii. - -upn: user principal to be created
        iii. - P: service prinicap to be created
        iv. - - keytab: Linux fs location of key table file
        v. - c: location to create the account in AD
        vi. - -ignore: Ignore directory security permissions
        vii. - -V: Verbose mode
        viii. –M: create the account as a machine (computer)
    b. [root@mpr2n1 hadoop-0.20.2]# adkeytab --new --upn mapr/mpr2.centrify.vms@CENTRIFY.VMS -P mapr/mpr2.centrify.vms@CENTRIFY.VMS --

keytab /opt/mapr/conf/mapr.keytab -c ou=accounts,ou=mpr2,ou=centrifyse -V -M mpr2-mapr
ADKeyTab version: CentrifyDC 5.2.2-154
Options
----
use machine ccache: no
domain: centrify.vms
server: null
gc: null
user: null
container: ou=accounts,ou=mpr2,ou=centrifyse
account: mpr2-mapr
trust: no
des: no
Administrator@CENTRIFY.VMS's password:
Attempting bind to centrify.vms site:Demo-Network server:dc.centrify.vms: ccache:MEMORY:0x56f2e0
Bind successful to server dc.centrify.vms
Attempting bind to GC domain:centrify.vms site:Demo-Network gcserver:dc.centrify.vms ccache:MEMORY:0x56f2e0
Bound to GC server:dc.centrify.vms domain:CENTRIFY.VMS
Building Container DN from ou=accounts,ou=mpr2,ou=centrifyse
Searching for AD Object: filter = (samAccountName=mpr2-mapr$), root = DC=centrify,DC=vms
AD Object not found.
Account with samAccountName 'mpr2-mapr$' does not exist
Search for account in GC: filter = (samAccountName=mpr2-mapr$), root = DC=CENTRIFY,DC=VMS
SAM name 'mpr2-mapr$' not found in GC
Searching for AD Object: filter = (samAccountName=mpr2-mapr$), root = DC=centrify,DC=vms
AD Object found: CN=mpr2-mapr,OU=accounts,OU=mpr2,OU=centrifyse,DC=centrify,DC=vms
Key Version = 1
Adding managed account keys to configuration file: mpr2-mapr
Changing account 'mpr2-mapr' password with user 'Administrator@CENTRIFY.VMS' credentials.
Searching for AD Object: filter = (samAccountName=mpr2-mapr$), root = DC=centrify,DC=vms
AD Object found: CN=mpr2-mapr,OU=accounts,OU=mpr2,OU=centrifyse,DC=centrify,DC=vms
Key Version = 2
Updated properties to config file /etc/centrifydc/centrifydc.conf.
Success: New Account: mpr2-mapr Change Ownership to mapr

   c. Change keytab ownership and permissions
      i. >chmod 600 /opt/mapr/conf/mapr.keytab
      ii. >chown mapr:mapr /opt/mapr/conf/mapr.keytab
   d. Copy (scp) the keytab to other nodes in the cluster
      i. Change ownership to mapr:mapr
5. Verify the following are not in the /opt/mapr/conf directory; if they are, delete them:
   a. cldb.key
   b. maprserviceticket
   c. ssl_keystore
   d. ssl_truststore
6. Verify that Mapr security setting is set to false.

a. More contents of /opt/mapr/conf/mapr-clusters.conf
    i. For example:
    [root@mpr2n1 conf]# more /opt/mapr/conf/mapr-clusters.conf

    mpr2.centrify.vms secure=false mpr2n1.centrify.vms:7222 mpr2n2.centrify.vms:7222 mpr2n3.centrify.vms:7222

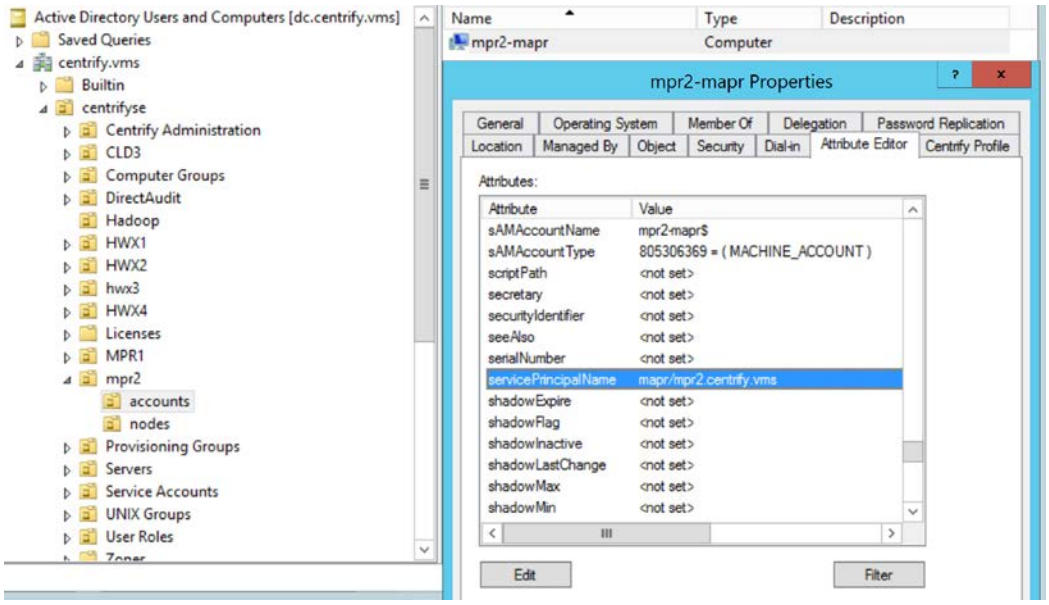
7. Enable MapR security with configure.sh on the install "control" node
    a. /opt/mapr/server/configure.sh –secure –genkeys –C CLDB-Node –Z zookeeper-Node –K –P Principal-Name –N cluster-name
    b. Use Example:
    [root@mpr2n1 conf]# /opt/mapr/server/configure.sh -secure -genkeys -C mpr2n1.centrify.vms,mpr2n2.centrify.vms,mpr2n3.centrify.vms -Z mpr2n1.centrify.vms,mpr2n2.centrify.vms,mpr2n3.centrify.vms -K -P mapr/mpr2.centrify.vms@CENTRIFY.VMS -N mpr2.centrify.vms
    Configuring Hadoop-2.5.1 at /opt/mapr/hadoop/hadoop-2.5.1
    Done configuring Hadoop
    CLDB node list:
    mpr2n1.centrify.vms:7222,mpr2n2.centrify.vms:7222,mpr2n3.centrify.vms:7222
    Zookeeper node list:
    mpr2n1.centrify.vms:5181,mpr2n2.centrify.vms:5181,mpr2n3.centrify.vms:5181
    Node setup configuration: cldb fileserver hbinternal nfs nodemanager resourcemanager webserver zookeeper
    Log can be found at: /opt/mapr/logs/configure.log
    Creating 10 year self signed certificate with subjectDN='CN=*.centrify.vms'
    sed: -e expression #1, char 49: unknown option to `s'
    Warden is not running. Starting mapr-warden. Warden will then start all other configured services on this node
    Starting WARDEN, logging to /opt/mapr/logs/warden.log.
    .
    For diagnostics look at /opt/mapr/logs/ for createsystemvolumes.log, warden.log and configured services log files
    Warden respawn not found in inittab. Adding entry.
    Creating warden respawn in inittab with ID of "a1"
    ... Starting cldb
    ... Starting fileserver
    ... Starting hbinternal
    ... Starting nfs
    ... Starting nodemanager
    ... Starting resourcemanager
    ... Starting webserver
    To further manage the system, use "maprcli", or connect browser to https://mpr2n1.centrify.vms:8443/
    To stop and start this node, use "service mapr-warden stop/start"
    [root@mpr2n1 conf]#
8. Enable security on other nodes in the cluster.
    a. Copy key files from the secured control node to all other nodes in the cluster. To the/opt/mapr/conf directory
        i. cldb.key with permissions 600
        ii. maprserviceticket with permissions 600
        iii. ssl_keystore with permissions 400
        iv. ssl_truststore with permissions 666
    b. Enable security via the configure.sh, on all other odes in the cluster.

i. /opt/mapr/server/configure.sh –secure –C CLDB-node-list –Z zookeeper-node-list –K –P Principal-Name –N cluster-name
ii. Use example, /opt/mapr/server/configure.sh -secure -C mpr2n1.centrify.vms,mpr2n2.centrify.vms,mpr2n3.centrify.vms -Z mpr2n1.centrify.vms,mpr2n2.centrify.vms,mpr2n3.centrify.vms -K -P mapr/mpr2.centrify.vms@CENTRIFY.VMS -N mpr2.centrify.vms

9. Verify cluster is secured.
   a. [root@mpr2n1 conf]# more mapr-clusters.conf
   mpr2.centrify.vms secure=true kerberosEnable=true cldbPrincipal=mapr/mpr2.centrify.vms@CENTRIFY.VMS mpr2n1.centrify.vms:7222 mpr2n2.centrify.vms:7222 mpr2n3.centrify.vms:7222
10. If needed restart Cluster (per node).
   a. Shutdown: stop Warden first
      i. service mapr-warden stop
      ii. service mapr-zookeeper stop
   b. Start: start Zookeeper first
      i. service mapr-zookeeper start
      ii. service mapr-warden start

Verify AD and the Keytabs

In ADUC, browse to the UNIX/Hadoop/ Mapr OU. You should see the following AD service account:



Maintaining your Centrify Hadoop environment

This section describes the actions you should take to ensure that your integrated Centrify Hadoop environment continues to operate correctly.

Hadoop creates Kerberos principals for service accounts. Those principals are governed by the same Active Directory polices that govern user accounts and computer accounts. That arrangement differs from MIT Kerberos implementations, and requires the following maintenance procedures after your environment is set up.

Keeping the Hadoop service account keytab up to date

Centrify Server Suite automatically maintains the keytab entries for computer accounts when the Centrify agent updates keytab entries every 28 days (or at a different interval if you specify a value other than the default of 28 days). However, other keytab entries, such as those created for user accounts and that reside on each node, are not automatically refreshed. If you created the Hadoop service account as a user account, you must ensure that keytab entries for Hadoop-specific user principals are automatically updated.

Note: You do not need to perform this procedure if you created the Hadoop service account as a computer account.

You can perform this configuration by writing a script that issues the adkeytab -C command, so that the keytab entry for the specified user account is updated. When the Centrify agent updates the user account, it obtains a new key version number (KVNO). The script must update every keytab on every node in the cluster.

Also, you must ensure that Hadoop service accounts are zone enabled.

Configuring Active Directory user accounts not to expire

Active Directory user accounts (user principals) are governed by Active Directory group policy objects for users. Organizations typically change user passwords every 30 to 60 days, or automatically expire accounts.

If you created the Hadoop service account as a user account, you must ensure that passwords for Hadoop-specific user principals are set to never expire.

Note: You do not need to perform this procedure if you created the Hadoop service account as a computer account.

To perform this configuration in ADUC:

1. Go to the Users organizational unit.
2. Right-click the user account that you want to have never expire.
3. Select Properties.
4. Select the Account tab.
5. Select the Password never expires option.
6. Configuring Kerberos credentials not to expire

Configuring Kerberos credentials not to expire

For your Centrify Hadoop environment to operate correctly in the long term, you must ensure that Kerberos tickets that are linked to user principals do not expire. Starting with Server Suite 2015.1, you can perform this configuration in one of these ways:

- To specify that all user credentials are automatically reissued when they expire, enable the **Renew credentials automatically** group policy or set the `krb5.cache.infinite.renewal` configuration parameter to `true`.

  To enable the **Renew credentials automatically** group policy, open Group Policy Management Editor on the domain controller and go to **Computer Configuration > Policies > Centrify Settings > DirectControl Settings > Kerberos Settings**.

To set the `krb5.cache.infinite.renewal` parameter to `true`, edit `/etc/centrifydc/centrifydc.conf` on each node computer.

- To specify that credentials for only certain users are automatically reissued when they expire, enable the **Specify users to infinitely renew Kerberos credentials** group policy or set the `krb5.cache.infinite.renewal.batch.users` configuration parameter to `true`.
- To specify that credentials for only certain groups are automatically reissued when they expire, enable the **Specify groups to infinitely renew Kerberos credentials** group policy or set the `krb5.cache.infinite.renewal.batch.groups` configuration parameter to true.

See the *Configuration and Tuning Reference Guide* for more information about setting configuration parameters. See the *Group Policy Guide* for more information about setting group policies.

Testing your cluster's security

Login as dwirth then attempt to access HDFS without a Kerberos ticket. A Kerberos ticket is now required to access the cluster with a kinit and maprlogin kerberos. The initial "Hadoop" command should fail, as dwirth does not have a valid ticket.

You should see the following error message:

```
Using Kerberos authentication
Using principal dwirth@CENTRIFY.VMS
Got host ticket host/mpr2n1.centrify.vms@CENTRIFY.VMS
login as dwirth
Successful Kerberos connection
-------------
WARNING
-------------
THIS IS A PRIVATE COMPUTER SYSTEM.
All computer systems may be monitored for all lawful purposes. This is to ensure that their use is authorized.
During monitoring, information may be examined, recorded, copied and used for authorized purposes. All
information including personal information, placed on or sent over this system may be monitored. Uses of this
system, authorized or unauthorized, constitutes consent to monitoring of this system. Use of this system
constitutes consent to monitoring for these purposes.
-----------
Last login: Sun Jul 12 07:52:57 2015 from dc.centrify.vms
[dwirth@mpr2n1 ~]$ klist
klist: No credentials cache found (ticket cache FILE:/tmp/krb5cc_1040188499)
[dwirth@mpr2n1 ~]$ maprlogin authtest
Attempting to pick up default credentials for cluster mpr2.centrify.vms
Failure during kerberos authentication. null
[dwirth@mpr2n1 ~]$ hadoop fs -ls /user
ls: failure to login: Unable to obtain MapR credentials
```

After the user dwirth obtains Kerberos credentials, the credentials are used to obtain a MapR ticket. Now, when dwirth tries to run a Hadoop job, the attempt succeeds:

```
[dwirth@mpr2n1 ~]$ kinit
Password for dwirth@CENTRIFY.VMS:
```

```
[dwirth@mpr2n1 ~]$ maprlogin kerberos
MapR credentials of user 'dwirth' for cluster 'mpr2.centrify.vms' are written to '/tmp/maprticket_1040188499'
[dwirth@mpr2n1 ~]$ hadoop fs -ls /user
Found 4 items
drwxr-xr-x   - dwirth    dwirth          0 2015-07-12 07:02 /user/dwirth
drwxr-xr-x   - mapr      root            2 2015-07-12 05:01 /user/mapr
drwxr-xr-x   - ned_atlas ned_atlas       0 2015-07-10 11:01 /user/ned_atlas
drwxr-xr-x   - root      root            0 2015-07-12 04:59 /user/root
[dwirth@mpr2n1 ~]$ cd /opt/mapr/hadoop/hadoop-0.20.2/
[dwirth@mpr2n1 hadoop-0.20.2]$ hadoop jar hadoop-0.20.2-dev-examples.jar pi 5 10
Number of Maps  = 5
Samples per Map = 10
Wrote input for Map #0
Wrote input for Map #1
Wrote input for Map #2
Wrote input for Map #3
Wrote input for Map #4
Starting Job
15/07/12 07:54:05 INFO client.MapRZKBasedRMFailoverProxyProvider: Updated RM address to
mpr2n1.centrify.vms/192.168.1.120:8032
15/07/12 07:54:05 INFO input.FileInputFormat: Total input paths to process : 5
15/07/12 07:54:05 INFO mapreduce.JobSubmitter: number of splits:5
15/07/12 07:54:05 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1436701210336_0003
15/07/12 07:54:05 INFO security.ExternalTokenManagerFactory: Initialized external token manager class -
com.mapr.hadoop.yarn.security.MapRTicketManager
15/07/12 07:54:05 INFO impl.YarnClientImpl: Submitted application application_1436701210336_0003
15/07/12 07:54:05 INFO mapreduce.Job: The url to track the job:
https://mpr2n1.centrify.vms:8090/proxy/application_1436701210336_0003/
15/07/12 07:54:05 INFO mapreduce.Job: Running job: job_1436701210336_0003
15/07/12 07:54:11 INFO mapreduce.Job: Job job_1436701210336_0003 running in uber mode : false
15/07/12 07:54:11 INFO mapreduce.Job:  map 0% reduce 0%
15/07/12 07:54:16 INFO mapreduce.Job:  map 20% reduce 0%
15/07/12 07:54:19 INFO mapreduce.Job:  map 80% reduce 0%
15/07/12 07:54:21 INFO mapreduce.Job:  map 100% reduce 0%
15/07/12 07:54:25 INFO mapreduce.Job:  map 100% reduce 100%
15/07/12 07:54:25 INFO mapreduce.Job: Job job_1436701210336_0003 completed successfully
15/07/12 07:54:25 INFO mapreduce.Job: Counters: 46
    File System Counters
        FILE: Number of bytes read=0
        FILE: Number of bytes written=491977
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        MAPRFS: Number of bytes read=1597
        MAPRFS: Number of bytes written=567
        MAPRFS: Number of read operations=195
        MAPRFS: Number of large read operations=0
        MAPRFS: Number of write operations=143
    Job Counters
        Launched map tasks=5
```

```
        Launched reduce tasks=1
        Data-local map tasks=5
        Total time spent by all maps in occupied slots (ms)=27583
        Total time spent by all reduces in occupied slots (ms)=5595
        Total time spent by all map tasks (ms)=27583
        Total time spent by all reduce tasks (ms)=1865
        Total vcore-seconds taken by all map tasks=27583
        Total vcore-seconds taken by all reduce tasks=1865
        Total megabyte-seconds taken by all map tasks=28244992
        Total megabyte-seconds taken by all reduce tasks=5729280
        DISK_MILLIS_MAPS=13793
        DISK_MILLIS_REDUCES=2480
    Map-Reduce Framework
        Map input records=5
        Map output records=10
        Map output bytes=90
        Map output materialized bytes=0
        Input split bytes=675
        Combine input records=0
        Combine output records=0
        Reduce input groups=2
        Reduce shuffle bytes=120
        Reduce input records=10
        Reduce output records=0
        Spilled Records=20
        Shuffled Maps =5
        Failed Shuffles=0
        Merged Map outputs=6
        GC time elapsed (ms)=644
        CPU time spent (ms)=4730
        Physical memory (bytes) snapshot=1538961408
        Virtual memory (bytes) snapshot=12815409152
        Total committed heap usage (bytes)=1339555840
    Shuffle Errors
        IO_ERROR=0
    File Input Format Counters
        Bytes Read=590
    File Output Format Counters
        Bytes Written=97
Job Finished in 20.231 seconds
Estimated value of Pi is 3.28000000000000000000
```

Conclusion

Centrify Server Suite provides these benefits to organizations with Active Directory and a Linux-based Hadoop cluster:

- Faster implementation
- Infrastructure simplicity

- Process reuse
- Role-based access controls
- A tool set designed for automation
- Best compatibility with Microsoft's implementation of Kerberos