Centrify Server Suite 2016

Centrify Identity and Access Management for Hortonworks

March 2016

Centrify Server Suite 2016

Hortonworks 2.3

Legal notice

Contents

. . . . . .

Benefits of integrating with Centrify

Centrify Server Suite is an enterprise-class solution that supports the Hortonworks implementation of Apache Hadoop. Together, Centrify and Hortonworks allow you to use your organization's existing Active Directory infrastructure to deliver access control, privilege management, and user-level auditing.

Centrify Server Suite provides identity, access and privilege management for the Hortonworks Data Platform in the following ways:

- Simplifies Active Directory integration for Hortonworks to run in secure mode.
- Automates service account credential management.
- Simplifies access with AD-based user single sign-on authentication.
- Ensures regulatory compliance with least privilege and auditing.
- Uses developer SDKs for secure client application access to Hadoop.

By installing a Centrify agent on each node, you can provide identity and access management for users who will log on to computers in the cluster with their Active Directory credentials. You can use Centrify to manage user and service principals and corresponding `keytab` files directly on those computer nodes, or centrally from a Windows console on an administrator's workstation.

Preparing for integration with Centrify

The following sections describe how to prepare your Hadoop environment for integration with Centrify. After you have prepared your environment, go to "Integrating Hortonworks and Centrify" on page 12 for details about performing the integration.

Preparing to create unique principal names

The default Hadoop security architecture is based on Kerberos, which is also the core infrastructure for Active Directory authentication services. As a result, all principals are user principals, and there will be an Active Directory account for each service account that requires a Kerberos key table (`keytab`) file. For example, a two-node cluster with six unique distributed services will require 12 Active Directory accounts that will each require a unique Kerberos keytab file.

The key to managing Hadoop clusters in Active Directory is the addition of a cluster prefix to the associated Hortonworks Kerberos principal. The cluster prefix ensures that the user principal name (UPN) and service principal name (SPN) for the account each cluster depends upon are unique across the Active Directory domain.

You should outline a naming convention for all Hadoop service principals that will reside in Active Directory. Ideally, you should be able to identify the service, cluster, and host by the naming convention you establish.

Keep the following things in mind when you establish a naming convention for your environment:

- The `sAMAccountName` attribute has a maximum length of 20 characters and must be unique across the Active Directory domain.
- Because host names in DNS are case-sensitive, you must ensure that the case you use for host names is consistent throughout your environment.

Basic prerequisites

- Active Directory must be installed and at least one domain controller available.
- You must have a Windows workstation joined to the domain where you can run administrative consoles.
- You must have access to the physical or virtual Linux computers that are used as Hadoop nodes. The environment described in this document uses three virtual Linux computers for nodes.
- You must have Centrify Server Suite software installed or available to be installed.

  You can request a free trial of Centrify Server Suite by filling out the http://www.centrify.com/free-trial/server-suite-form/ on the Centrify website and specifying `Hadoop` in the Comments field.
- You should have Centrify Server Suite documentation available for reference.

  You can download documentation from http://www.centrify.com/support/documentation.
- You must have access to Hortonworks Data Platform software.

Planning the organizational units to use

You should use an Active Directory organizational unit (OU) to manage all of your Hadoop clusters, such as `OU=centrify`. Your Active Directory domain administrators might need to delegate administrative rights of this OU to you or your technical lead. The Linux identity, access information, and privilege information are stored within the OU that was created for you (`OU=centrify`).

Each cluster should have its own OU to independently manage its nodes and service accounts. The OU name should reflect the name of the cluster, for example, `OU=hqa1` (the name chosen here represents Hortonworks QA cluster 1). This cluster-level OU is usually created within the OU that was created by the Active Directory administrator and delegated to you so that you can create an OU for each Hadoop cluster and manage the accounts and policies yourself.

See "Creating Active Directory organizational units" on page 7 for more information.

Planning to use Centrify zones for Hadoop clusters

Centrify uses the Zones container to store the access and privilege permissions for the selected Active Directory users that you authorize to access each Hadoop cluster. A typical setup for Hadoop is to create a `global` zone (`OU=zones,OU=global`) containing unique child zones for each Hadoop cluster that you deploy. This arrangement ensures separation of duties and enables delegated administration. Use the child zone name as the same name for the cluster prefix, for example, `hqa1`.

Containers for top-level and child zones are created when you use Access Manager to create and configure zones.See "Creating zones and defining a UNIX user profile" on page 10 for more information.

Creating Active Directory organizational units

You are now ready to create the Active Directory organizational units for the Hadoop cluster. Figure 1 shows Active Directory Users and Computers (ADUC) after you perform this procedure.

1. On the domain controller, open ADUC.
2. Right-click the domain (`centrifybigdata.net` in Figure 1), then select **New > Organizational Unit**.
3. Type the name of the top-level Hadoop OU, then click **OK**.

   For example, you might create a `centrify` OU similar to this:
   `OU=centrify,DC=centrifybigdata,DC=net`
4. Create new OUs for each cluster and for other required and optional objects that support Hadoop:
   a. Select the top-level Hadoop OU (`centrify` in Figure 1).
   b. Right-click.
   c. Select **New > Organizational Unit**.

   For example, Figure 1 shows the creation of the following OUs:

   - `OU=hqa1,OU=centrify,DC=centrifybigdata,DC=net`: This OU is the container for the `hqa1` cluster.

- `OU=licenses,OU=centrify,DC=centrifybigdata,DC=net`: This OU is the container for licenses.
- `OU=role-groups,OU=centrify,DC=centrifybigdata,DC=net`: This OU is for AD groups that you might create later and assign roles to (for example, `unix-global-sysadmin`). Note that this OU is parallel to (rather than under) the `cdev1` OU, allowing you to create groups that span all clusters.
- `OU=unix-groups,OU=centrify,DC=centrifybigdata,DC=net`: This OU is for AD groups that you might create later and map UNIX groups to (for example, `unix-hqa1-unixgroup-supergroup`). Note that this OU is parallel to the `hqa1` OU, allowing you to create groups that span all clusters.
- `OU=zones,OU=centrify,DC=centrifybigdata,DC=net`: This OU is for Centrify zones.
- `OU=staff,OU=centrify,DC=centrifybigdata,DC=net`: This OU is for users that you will create later. (This OU is not shown in Figure 1.)

5 To manage computer nodes in the cluster separately from user accounts and service accounts in the cluster, create child OUs under cluster OUs for computer nodes (`OU=nodes`) and user and service accounts (`OU=accounts`).

a Select the cluster-specific OU (for example, `hqa1`).

b Right-click.

c Select **New > Organizational Unit**.

d Create a `nodes` OU for computer nodes, then repeat these steps to create an `accounts` OU.

For example:

```
OU=nodes, OU=hqa1,OU=centrify,DC=centrifybigdata,DC=net
OU=accounts, OU=hqa1,OU=centrify,DC=centrifybigdata,DC=net
```

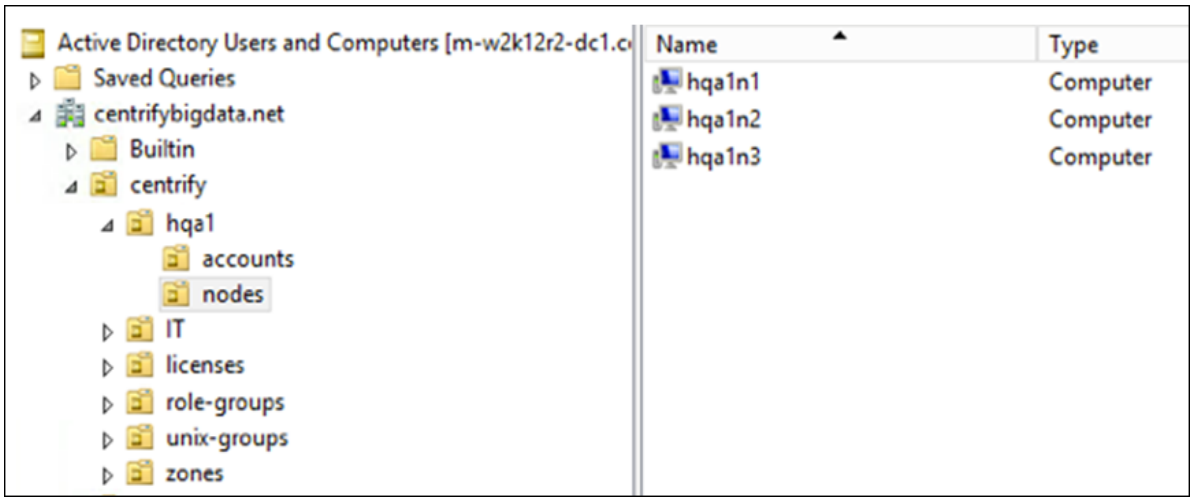

Figure 1 Organizational units and nodes for Centrify/Hortonworks integration

Creating AD groups and users, and delegating AD privileges

After you have created an OU structure similar to the one shown in Figure 1, create groups and users in Active Directory, and delegate Active Directory permissions to them. For example:

1. In the `role-groups` OU that you created in Step 4 on page 7, you might create a group for UNIX administrators (such as `unix-global-sysadmin`).
2. In the `unix-groups` OU that you created earlier, you might create a group for cluster super users (such as `unix-hqa1-unixgroup-supergroup`).
3. In the `staff` OU that you created earlier, create users who will perform various administrative tasks. For example, you might create the following users:
   - An existing employee, who will be the main administrator of the entire Hadoop integration.
   - A `unix.admin` user, who will administer all of the UNIX computers in all clusters.
   - An `hqa1.admin` user, who will be the Hortonworks administrator of the `hqa1` cluster.
   - A `kerberos.admin` user, who will be the Kerberos administrator for the entire integration.
4. Add the users that you created in Step 3 to pre-existing groups, and also to the groups that you created in Step 1 and Step 2.

   To add a user to a group, right-click the user, select **Properties**, select the **Member Of** tab, and add the user to a group. Users will inherit permissions from the parent group. For example:
   - Add the main Hadoop administrator to these groups: `Domain Admins`, `unix-global-sysadmin`.
   - Add the `unix.admin` user to the `unix-global-sysadmin` group.
   - Add the `kerberos admin` user to the `Domain Admins` group.

Installing Centrify DirectManage Access

You are now ready to install Centrify Server Suite on a Windows administrator's workstation.

Note If DirectManage Access is already installed, go to Creating zones and defining a UNIX user profile and continue from there.

If you downloaded the documentation, you can use the *Centrify Server Suite Quick Start Guide* to guide you through the next steps.

1. Open the Centrify Server Suite ISO or ZIP file for Windows 64-bit on the Windows workstation.

2 Click **Access** on the Getting Started page or run the setup program in the DirectManage folder.

3 Follow the prompts displayed to select the suite edition and components to install.

Creating zones and defining a UNIX user profile

Use Access Manager to set up the Active Directory domain and create the zones for the Hadoop cluster. See Figure 2 for details about what the environment looks like in Access Manager after you perform the procedures described here and in the two sections.

1 Open Access Manager to start the Setup Wizard.

2 Follow the prompts displayed to create the containers for Licenses and Zones.

Set up the containers so that they match the OU structure that you created in Step 4 on page 7 (for example, `centrifybigdata.net/centrify/licenses` and `centrifybigdata.net/centrify/zones`.)

3 In Access Manager, create a top level zone (for example, `global`) that will contain all child zones for the Hadoop integration.

4 Create a child zone for the cluster within the top level (`global`) zone. Name the child zone so that it is easily identified as a cluster zone (for example, `hqa1`, which matches the name of the cluster OU that you created in Step 4 on page 7).

5 Add Active Directory users to the zone. Later, you will grant permission to all of these users so that they can log into the cluster using their Active Directory credentials. All of the Active Directory users that you created in Step 3 on page 9 should be added. To simplify user management, you can add the users to the top level (`global`) zone, not to each cluster zone (`hqa1`).

Select the top level zone (`global`), right-click, then select **Add User** to search for and select an existing Active Directory user.

6 Select **Define user UNIX profile** and deselect **Assign roles**, then click **Next**.

7 Accept the defaults for all fields, click **Next**, then click **Finish**.

Assigning a role to a user in a zone

User profiles are inherited by child zones, so the users that you added to the top level (`global`) zone automatically have a profile in the cluster (`hqa1`) zone. To log on to a computer, however, a user must have both a profile and a role assignment. Access Manager includes a default UNIX Login role that you can assign to enable users to log on.

1 Expand the top level (`global`) zone, **Child Zones**, the cluster (`hqa1`) zone, and **Authorization**.

2 Select **Role Assignments**, right-click, then click **Assign Role**.

3 Select the UNIX Login role from the list of roles and click **OK**.

4 Click **Add AD Account** to search for and select each Active Directory user you added to the `global` zone, then click **OK**.

Zone-enabling Active Directory groups for use with nodes

Use Access Manager to map AD groups to UNIX groups on nodes in the cluster. To set up this mapping, you create a UNIX profile in Access Manager for each AD group that you want to map. After you perform this mapping, the AD groups that you specified are zone enabled. The next time a Linux node computer containing the Centrify agent joins the domain, the UNIX profiles for the zone-enabled groups are installed on the Linux node computer.

To zone-enable AD groups for use on nodes

1 In Access Manager, go to the top level (`global`) zone and create a UNIX group for users who will have global administrator rights. The group, and all of its rights, will be inherited by child zones for each cluster. To create this group:

   a Right click **global > UNIX Data > Groups**, and select **Create UNIX Group**.

   b In the Create UNIX Group dialog name field, type the name one of the groups that you created in Step 1 on page 9 (for example, `unix-global-sysadmin`).

   c Click **OK**.

   d In the Set UNIX Group Profile dialog, type a name of your choice for the UNIX group name.

   e Assign a role to the `unix-global-sysadmin` group that gives the group privileges that are typically given to UNIX system administrators (such as the ability to log into all computers, run all commands as root, and so on).

   This role is not predefined, and must have already been created in Access Manager by you or another user with role creation privileges. For details about creating this role, see the section "Creating a root-equivalent role definition" in the "Creating and assigning custom role definitions" chapter of the *Administrator's Guide for Linux and UNIX*.

2 In Access Manager, go to a cluster zone (such as `hqa1`) and create a UNIX group for users who will have administrative privileges on all nodes in that cluster.

   To create this group, repeat the steps that you performed in Step 1, specifying the name of the cluster-specific UNIX group that you created in Step 2 on page 9 (such as `unix-hqa1-unixgroup-supergroup)` instead of the group name that you specified in Step 1 (`unix-global-sysadmin`).

After you perform these steps:

- The top level (`global`) zone is linked to the AD group `unix-global-sysadmin`, and users in the `unix-global-sysadmin` group have all of the privileges that are defined in

the role that you assigned to `unix-global-sysadmin`. These privileges apply to all zones, and therefore to all clusters.

- The zone and cluster `hqa1` are linked to the AD group `unix-hqa1-unixgroup-supergroup`. When an agent-managed node computer joins the domain, the users in the AD group `unix-hqa1-unixgroup-supergroup` have all of the privileges on the `hqa1` cluster that are defined in the role that you assigned to `unix-hqa1-unixgroup-supergroup`.

Integrating Hortonworks and Centrify

The following sections describe how to configure a demonstration environment that uses three Linux virtual machines with the Centrify agent.

Create a Hortonworks cluster that uses Centrify

The instructions that follow describe how to prepare three CentOS virtual machines for use as a cluster. The first VM is for the name node. The two other VMs are for additional nodes. You will install the Centrify agent on each VM, and install Hortonworks on the name node.

Note If you already have a Hortonworks cluster consisting of three or more physical or virtual machines in an isolated environment for testing, you can install the Centrify agent as described in "Install the Centrify agent on each node" on page 13, and then continue on to "Enabling security for the cluster" on page 16.

Prepare virtual machines

Perform the following steps to prepare the virtual environment for testing with three CentOS computers:

1 Provision three new CentOS 6.x 64-bit virtual machines using the following settings:
   - For node 1 (the name node: `hqa1n1.centrifybigdata.net`):

     2 processors, 8GB RAM, 1 HD (120gb)
   - For node 2 (`hqa1n2.centrifybigdata.net`):

     2 processors, 8GB RAM, 1 HD (120 GB)
   - For node 3 (`hqa1n3.centrifybigdata.net`):

     2 processors, 8GB RAM, 1 HD (120 GB)

2 Ensure that DNS is configured correctly:
   a Create the corresponding DNS address (A) records in the appropriate DNS zone.
   b Create the proper reverse DNS entries.

3 If Perl is not installed on each node, install it now:

```
yum install -y perl
```

4 If Wget is not installed on each node, install it now:

```
yum install -y wget
```

5 If NTP is not installed and running on each node, install and start it now:

```
yum install -y ntp
systemctl enable ntpd
systemctl start ntpd
```

6 Depending on the OS installed on each node, the version of Snappy provided with the OS might be incompatible with some software in the cluster environment. (For example, in some situations the version of Snappy is too new.) In this situation, replace the installed version of Snappy with a supported version:

```
yum remove -y snappy
yum install -y snappy-devel
```

Note For other possible node requirements, see http://hortonworks.com/hdp/downloads/.

Install the Centrify agent on each node

You can now install the Centrify agent on each node computer in the cluster and join the nodes to an Active Directory domain. Perform the following steps on each node.

1 Download the appropriate tarred and zipped Centrify agent. Copy the agent `.tgz` file to each node computer in the cluster.

2 Unzip and extract the agent package.

For example:

```
gunzip centrify-suite-2016-centos-x86_64.tgz
tar -xvf centrify-suite-2016-centos-x86_64.tar
```

3 Run the `install.sh` installation script interactively.

For example:

```
./install.sh
```

4 Follow the prompts displayed to install Standard Edition (S).

You can press ENTER to accept the default for any prompt. The following instructions are provided for only the prompts in which you should provide a non-default response.

Note During installation, do not join the computer to Active Directory. After installation, you must edit configuration files as described in Step 5 before joining the computer to Active Directory.

In the following prompt, type the user name and password for an Active Directory user with permissions to update the top-level Hadoop organizational unit as described in "Creating Active Directory organizational units" on page 7.

```
Enter the Active Directory authorized user [administrator]:
Enter the password for the Active Directory user:
```

In the following prompts, it is not necessary to specify the information for the organizational unit structure that you created in "Creating Active Directory organizational units" on page 7. Instead, you will specify the information when you join the computer to Active Directory in Step 6.

```
Enter the container DN [Computers]:
Enter the name of the zone:
```

5 Open `/etc/centrifydc/centrifydc.conf` for editing and make the following changes.

Note If you do not perform this step, the Kerberos configuration wizard will not complete, and the cluster will fail to start.

a Uncomment the `adclient.krb5.service.principals` line.

b Remove `http` from the `adclient.krb5.service.principals` line so that the line appears as follows:

```
adclient.krb5.service.principals: ftp difs nfs
```

c If you selected NFS Gateway during cluster configuration, remove `nfs` from the `adclient.krb5.service.principals` line so that the line appears as follows:

```
adclient.krb5.service.principals: ftp difs
```

6 Join the computer to Active Directory. In the command shown in this step:

- The user specified in `--user` must have domain administration rights. For the environment described in this guide, you can specify either main administrator or the AD administrator `kerberos.admin` (both were created earlier in Step 3 on page 9).
- In the --container parameter, specify the container that contains node computers (in the environment described in this guide, the container is `OU=nodes,OU=hqa1` under `OU=centrify`).

```
adjoin --force --user kerberos.admin --name hqa1n1 --zone hqa1 --container
“ou=nodes,ou=hqa1,ou=centrify,dc=centrifybigdata,dc=net” centrifybigdata.net
```

7 Optional: Install the Centrify Audit agent and enable auditing.

```
rpm -Uhv centrifyda-release.version
```

8 Reboot the VM.

You should see confirmation that the computer has successfully joined Active Directory. After the computer restarts, you can log on using the Active Directory user name and password for the user you previously assigned the UNIX Login role.

In Access Manager, you can verify that the CentOS VMs are running and joined to Active Directory:

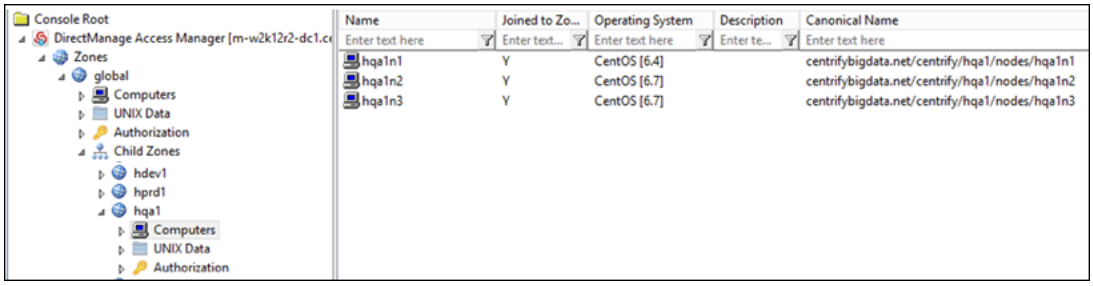

Figure 2 Cluster VMs are joined to Active Directory and can be viewed in Access Manager

Install the Centrify LDAP proxy on the name node

Install the Centrify LDAP proxy on the name node (`hqa1n1`) in the cluster. Installing the Centrify LDAP proxy is necessary because Hortonworks 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.

Install Hortonworks on the name node

You can now install Hortonworks on the name node (`hqa1n1`). For detailed instructions about installing Hortonworks, http://hortonworks.com/hdp/downloads/.

After you install Hortonworks, the Ambari Dashboard launches. At this time, the cluster is still not secured because security has not been enabled.

Test the cluster and authorize a cluster user

Before enabling security, log into the name node (`hqa1n1`) to test the cluster and authorize a cluster user (the user `ed` in the following example). If your cluster test uses installed example files, include the full path to the example files in the command, or change to the example file directory. Example files are typically located in:
`/usr/hdp/version/hadoop-mapreduce`

```
Using Kerberos authentication
Using principal ed@CENTRIFYBIGDATA.NET
Got host ticket host/hqa1n1.centrifybigdata.net@CENTRIFYBIGDATA.NET
login as edCENTRIFYBIGDATA.NET
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.centrifybigdata.net
[ed@hqa1n1 ~]$ dzdo su hdfs
[dzdo] password for ed:
[hdfs@hqa1n1 ed.]$ hadoop fs -ls /user
Found 7 items
drwx------   - accumulo  hdfs          0 2015-12-01 09:59 /user/accumulo
drwxrwx---   - ambari-qa hdfs          0 2015-12-01 10:05 /user/ambari-qa
drwxr-xr-x   - hcat      hdfs          0 2015-12-01 10:00 /user/hcat
drwx------   - hdfs      hdfs          0 2015-12-01 11:59 /user/hdfs
drwxr-xr-x   - hive      hdfs          0 2015-12-01 10:01 /user/hive
drwxrwxr-x   - oozie     hdfs          0 2015-12-01 10:01 /user/oozie
drwxrwxr-x   - spark     hdfs          0 2015-12-01 09:58 /user/spark
[hdfs@hqa1n1 ed]$ hadoop fs -mkdir /user/ed

[hdfs@hqa1n1 ed]$ hadoop fs -chown ed:ed /user/ed
```

Enabling security for the cluster

Configure the cluster to operate in secure mode by using the Kerberos infrastructure that was enabled by the Centrify agent on each node.

To enable cluster security

1. In Ambari, go to **Admin > Kerberos** and click **Enable Security**.
2. On the Get Started page, select **Existing Active Directory**, and then select all prerequisites.
3. Click **Next**.
4. In the Configure Kerberos page, complete the KDC and Kadmin pages. The required fields are:
   - **KDC host**: The FQDN of your domain controller. For example, `m-w2k12r2-dc1.centrifybigdata.net`.
   - **REALM name**: The domain name (must be in all uppercase characters). For example, `CENTRIFYBIGDATA.NET`.
   - **LDAP url**: The domain controller (must be in secured format, `ldaps://`). For example, `ldaps://m-w2k12r2-dc1.centrifybigdata.net`.

- **Container DN**: location the service principals (accounts) will be created in. Use domain component format. For example, `OU=accounts, OU=hqa1,DC=centrifybigdata,DC=net`.
- **Kadmin Host**: The domain controller for the Admin principals to authenticate to. For example, `s-w2k12r2-m1.centrify.bigdata.net`.
- **Admin principal**: An Active Directory account that has permission to create service principals. For example, `ed@centrifybigdata.net`.
- **Admin password**: The Admin principal's password.

5 In the Advanced krb5-conf page, deselect **Manage Kerberos client krb5.conf**. The Centrify agent will update the local Kerberos files.

6 In the KDC page, click **Test KDC Connection**. The expected response is **Connection OK**.

7 Select the Install and Test Kerberos Client page, review the information, and select **Next**.

8 Continue to click **Next** until the Confirm Configuration screen displays. Confirm that the displayed information matches your input from earlier steps, and click **Next**. The Kerberize Cluster page displays.

9 In the Kerberize Cluster page, click Next. Continue clicking Next in subsequent pages until the Enable Kerberos wizard finishes, and all services are restarted.

The console should appear similar to the following after services are restarted:

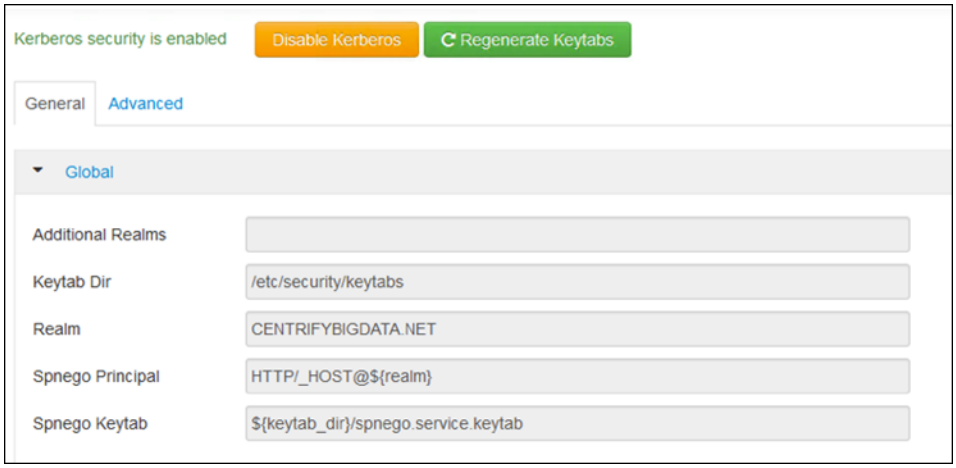

Figure 3. Console after security is enabled and services are restarted

Verifying Active Directory managed service accounts

In ADUC, browse to the `centrify/hqa1/accounts` OU. You should see service accounts in AD similar to those shown here (not all accounts are shown in this example):

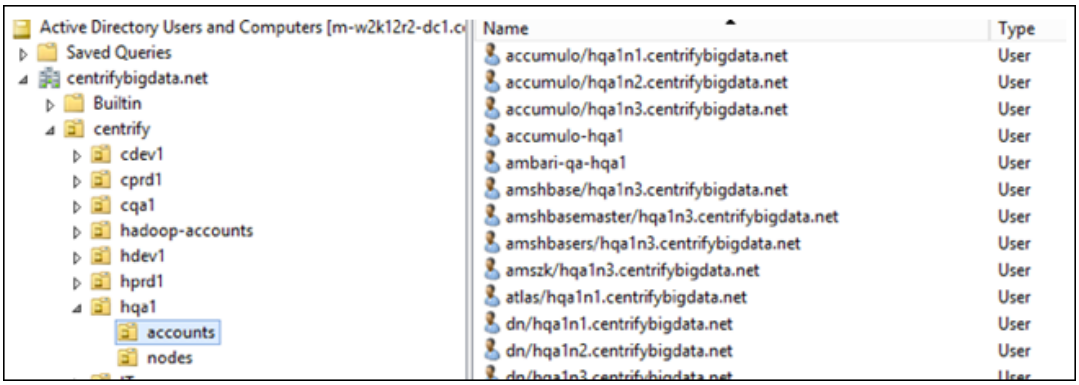

Figure 4. Service accounts are viewable in ADUC after security is enabled

On each cluster node, log in and verify that keytabs have the appropriate permissions:

```
-r--r----- 1 accumulo  hadoop 373 Dec  7 15:36 accumulo.headless.keytab
-r-------- 1 accumulo  hadoop 488 Dec  7 15:36 accumulo.service.keytab
-r-------- 1 accumulo  hadoop 363 Dec  7 15:36 accumulo-tracer.headless.keytab
-r-------- 1 atlas     hadoop 473 Dec  7 15:36 atlas.service.keytab
-r-------- 1 hdfs      hadoop 458 Dec  7 15:36 dn.service.keytab
-r--r----- 1 hbase     hadoop 358 Dec  7 15:36 hbase.headless.keytab
-r-------- 1 hbase     hadoop 473 Dec  7 15:36 hbase.service.keytab
-r--r----- 1 hdfs      hadoop 353 Dec  7 15:36 hdfs.headless.keytab
-r-------- 1 kafka     hadoop 473 Dec  7 15:36 kafka.service.keytab
-r-------- 1 knox      hadoop 468 Dec  7 15:36 knox.service.keytab
-r-------- 1 hdfs      hadoop 463 Dec  7 15:36 nfs.service.keytab
-r-------- 1 yarn      hadoop 458 Dec  7 15:36 nm.service.keytab
-r-------- 1 hdfs      hadoop 458 Dec  7 15:36 nn.service.keytab
-r--r----- 1 ambari-qa hadoop 378 Dec  7 15:36 smokeuser.headless.keytab
-r-------- 1 spark     hadoop 358 Dec  7 15:36 spark.headless.keytab
-r--r----- 1 root      hadoop 468 Dec  7 15:36 spnego.service.keytab
-r-------- 1 storm     hadoop 358 Dec  7 15:36 storm.headless.keytab
-r-------- 1 zookeeper hadoop 493 Dec  7 15:36 zk.service.keytab
```

Verifying cluster security

Now that the Hortonworks cluster is using Centrify for Active Directory based authentication, a user can log in using Active Directory credentials directly at the console prompt, or could use a Kerberized SSH client such as the Centrify version of PuTTY to get single sign-on access to the cluster.

Running a test job

Because the cluster is now running in secure mode, a user without a Kerberos ticket will not be able to successfully submit a job to the cluster.

If a user (such as `ed` in the following example) does not have a Kerberos ticket and tries to run a Hadoop job after logging into a cluster node (`hqa1n1`), the attempt fails as shown in the following example.

If your cluster test uses installed example files, include the full path to the example files in the command, or change to the example file directory. Example files are typically located in:

`/usr/hdp/`*`version`*`/hadoop-mapreduce`

```
[ed@hqa1n1 hadoop-mapreduce]$ yarn jar hadoop-mapreduce-examples.jar pi 5 10
Number of Maps  = 5
Samples per Map = 10

15/12/07 18:04:14 WARN ipc.Client: Exception encountered while connecting to the
server : javax.security.sasl.SaslException: GSS initiate failed [Caused by
GSSException: No valid credentials provided (Mechanism level: Failed to find any
Kerberos tgt)]
```

After the user `ed` obtains a Kerberos ticket and tries to run the same Hadoop job, the attempt succeeds (not all output is shown here):

```
[ed@hqa1n1 hadoop-mapreduce]$ kinit
Password for ed@CENTRIFYBIGDATA.NET:
Warning: Your password will expire in 5 days on Sun Dec 13 04:03:13 2015
[ed@hqa1n1 hadoop-mapreduce]$ yarn jar hadoop-mapreduce-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

Job Finished in 29.291 seconds

Estimated value of Pi is 3.28000000000000000000
```

Maintaining your Centrify Hortonworks 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.

Configuring Active Directory user accounts not to expire

Note You do not need to perform this procedure if you created the Hadoop service account as a computer account, or if you used the Hortonworks Enable Kerberos Wizard to create the Hadoop service account (the wizard automatically configures 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 manually created the Hadoop service account as a user account (that is, you did not use the Hortonworks Enable Kerberos Wizard to create the account), you must ensure that passwords for Hadoop-specific user principals are set to never expire.

To configure Hadoop user accounts to never expire

1. In ADUC, 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.

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. 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`.

Note 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.

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 as described in the next section.

Zone-enabling Hadoop accounts

The Ambari installer automatically creates a number of local accounts on each node in the cluster. To simplify the management of those accounts, you can create Active Directory users and groups for them, and then link the Active Directory users and groups to the local accounts so that the accounts are zone-enabled. After you perform those steps, you can remove the local accounts from each node, because the accounts are created automatically on each node computer when the node computer joins the domain.

To zone-enable local Hadoop accounts

1 In ADUC, create AD groups for the local Hadoop groups that exist on each cluster.

For example, if a local node group named `hadoop` was created by Ambari, you might create a `unix-hqa1-unixgroup-hadoop-service-accounts` AD group that will be linked to the local `hadoop` group.

2 In Access Manager, go to the cluster zone (for example, `hqa1`), expand **UNIX Data > Groups**, and select **Create UNIX Group**. Use the resulting dialog to zone-enable the AD group that you created in Step 1. For details about zone-enabling a group, see "Zone-

enabling Active Directory groups for use with nodes" on page 11, or the *Administrator's Guide for Linux and UNIX*.

When you are done, you see that the AD group `unix-hqa1-unixgroup-hadoop-service-accounts` in the `hqa1` zone is linked to the `hadoop` local node group:

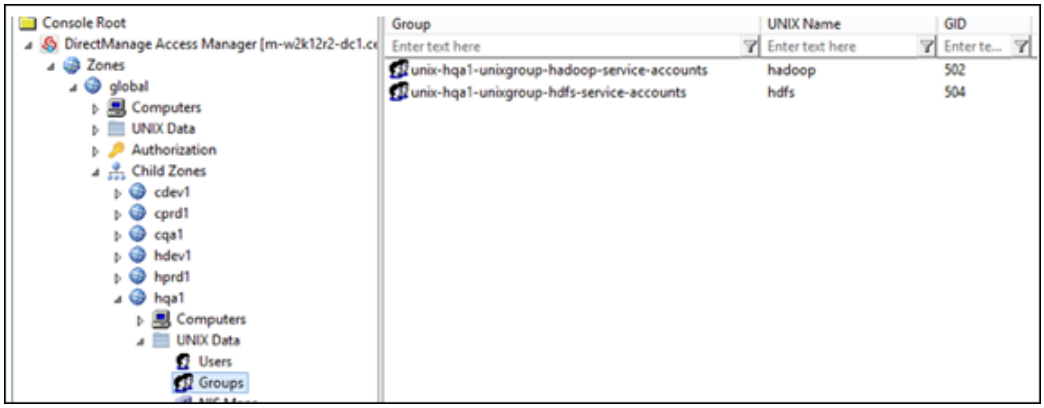

Figure 5. Zone-enabled group accounts for local cluster groups

3 You can perform similar steps to zone-enable local user accounts. For example, the following figure shows the local user accounts `hdfs` and `yarn` that are linked to the AD user accounts `hdfs-hqa1` and `yarn/hqa1n2.centrifybigdata.net`, except for the headless accounts. The headless accounts are created during keytab creation with a specific UPN, and are cluster- wide. However, the headless accounts still must be zone enabled.

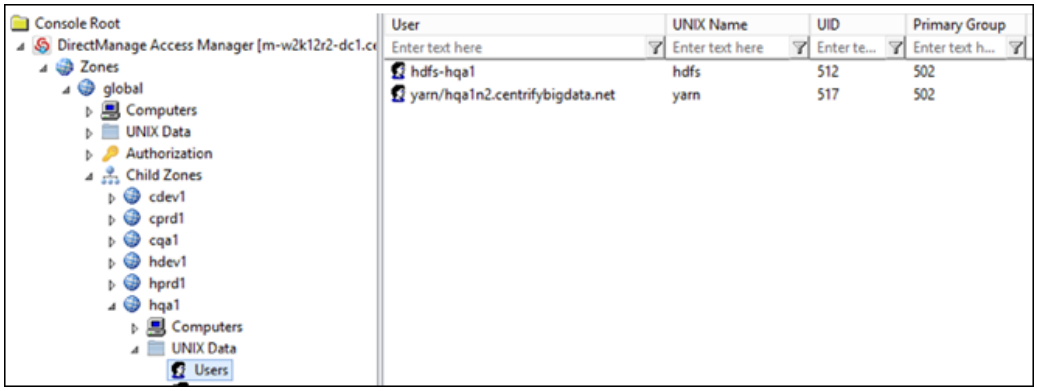

Figure 6. Zone-enabled user accounts for local cluster users

4 Remove the local accounts from each node in the cluster.

Using a zone-enabled account to manage HDFS

After you perform the steps in the preceding sections, you can use a zone-enabled account (such as `hdfs`) to manage HDFS in a cluster (`hqa1`, in this case).

The following example shows how to:

- Verify that the `hdfs` account is zone enabled and has a non-expiring Kerberos ticket.
- Use `dzdo` to access the `hdfs` account in a secured cluster and create a home directory for the user `wade`.

```
Using Kerberos authentication
Using principal ed@CENTRIFYBIGDATA.NET
Got host ticket host/hqa1n1.centrifybigdata.net@CENTRIFYBIGDATA.NET
login as ed@CENTRIFYBIGDATA.NET
Successful Kerberos connection
CentOS release 6.7 (Final)
Kernel 2.6.32-358.el6.x86_64 on an x86_64

Last login: Wed Dec 16 04:40:03 2015 from m-w2k12r2-dc1.centrifybigdata.net
[ed@hqa1n1 ~]$ dzdo su - hdfs
[hdfs@hqa1n1 ~]$ klist

Ticket cache: FILE:/tmp/krb5cc_515
Default principal: hdfs-hqa1@CENTRIFYBIGDATA.NET
Valid starting     Expires            Service principal
12/16/15 04:39:33  12/16/15 14:39:33  krbtgt/
CENTRIFYBIGDATA.NET@CENTRIFYBIGDATA.NET

        renew until 12/17/15 04:39:33

12/16/15 04:39:33  12/16/15 14:39:33  HTTP/
hqa1n1.centrifybigdata.net@CENTRIFYBIGDATA.NET

        renew until 12/17/15 04:39:33

[hdfs@hqa1n1 ~]$ hadoop fs -mkdir /user/wade
[hdfs@hqa1n1 ~]$ hadoop fs -chown wade:wade /user/wade
[hdfs@hqa1n1 ~]$ hadoop fs -ls /user
Found 9 items

drwx------   - accumulo    hdfs0 2015-12-01 09:59 /user/accumulo
drwxrwx---   - ambari-qa   hdfs0 2015-12-07 15:50 /user/ambari-qa
drwxr-xr-x   - ed          ed  0 2015-12-16 03:13 /user/ed
drwxr-xr-x   - hcat        hdfs0 2015-12-01 10:00 /user/hcat
drwx------   - hdfs        hdfs0 2015-12-01 11:59 /user/hdfs
drwxr-xr-x   - hive        hdfs0 2015-12-01 10:01 /user/hive
drwxrwxr-x   - oozie       hdfs0 2015-12-01 10:01 /user/oozie
drwxrwxr-x   - spark       hdfs0 2015-12-01 09:58 /user/spark
drwxr-xr-x   - wade    wade0 2015-12-16 04:42 /user/wade
```