Centrify Server Suite 2016

Centrify Identity and Access Management for Cloudera

March 2016

Centrify Server Suite 2016

Cloudera Enterprise 5.5.0

Legal notice

Contents

Benefits of integrating with Centrify 5

Preparing for integration with Centrify 5

- Preparing to create unique principal names 5
- Basic prerequisites 6
- Planning the organizational units to use 6
- Planning to use Centrify zones for Hadoop clusters 7
- Creating Active Directory organizational units 7
- Creating AD groups and users, and delegating AD privileges 9
- Installing Centrify DirectManage Access 10
- Creating zones and defining a user UNIX profile 11
- Assigning a role to a user in a zone 11
- Zone-enabling AD groups for use with nodes 12

Integrating Cloudera and Centrify 15

- Create a Cloudera cluster that uses Centrify 15
- Create home directories and authorize cluster users in HDFS 20
- Testing the cluster 21
- Enabling security for the cluster 21
- Validating cluster security 22

Maintaining your Centrify Hadoop environment 24

- Keeping the Hadoop service account keytab up to date 24
- Configuring Active Directory user accounts not to expire 25
- Configuring Kerberos credentials not to expire 25

• • • • • •

Benefits of integrating with Centrify

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

By installing a Centrify agent on each node in the Hadoop cluster, you can provide identity and access management for the users who will log on to computers in the cluster with their Active Directory credentials.

In addition, by installing the agent on the control node in the cluster, you can centrally create, secure, and distribute the service accounts and Kerberos key table (`keytab`) files that your cluster requires for distributed computing. The service accounts are stored securely in Active Directory with the domain controller acting as the Kerberos key distribution center (KDC).

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 Cloudera and Centrify" on page 15 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 or computer principals, and there will be an Active Directory account for each service account that requires a Kerberos key table (`keytab`) file.

The key to managing Hadoop clusters in Active Directory is the addition of a cluster prefix to the associated 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.

After you install the Centrify agent on each node, you can use Centrify to manage user and service principals and corresponding `keytab` files on those computer nodes or centrally from a Windows console on an administrator's workstation.

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 should have a Windows workstation joined to the domain where you can run administrative consoles.
- You should have access to at least three physical or virtual Linux computers to use as Hadoop nodes.
- You should 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 https://www.centrify.com/support/documentation.

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=cdev1` (the name chosen here represents Cloudera development 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, `cdev1`.

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 user UNIX profile" on page 11 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 (`devbigdata.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 the OU in a location similar to this:
   `OU=centrify,DC=devbigdata,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=cdev1,OU=centrify,DC=devbigdata,DC=net`: This OU is the container for the `cdev1` cluster.
   - `OU=licenses,OU=centrify,DC=devbigdata,DC=net`: This OU is the container for licenses.
   - `OU=role-groups,OU=centrify,DC=devbigdata,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=devbigdata,DC=net`: This OU is for AD groups that you will create later and map UNIX groups to (for example, `unix-cdev1-unixgroup-supergroup` as described in "Create AD groups that will be linked to supergroups on cluster nodes" on page 9). Note that this OU is parallel to the `cdev1` OU, allowing you to create groups that span all clusters.
- `OU=zones,OU=centrify,DC=devbigdata,DC=net`: This OU is for Centrify zones.
- `OU=staff,OU=centrify,DC=devbigdata,DC=net`: This OU is for users that you will create later.

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

a Select the cluster-specific OU (for example, `cdev1`).

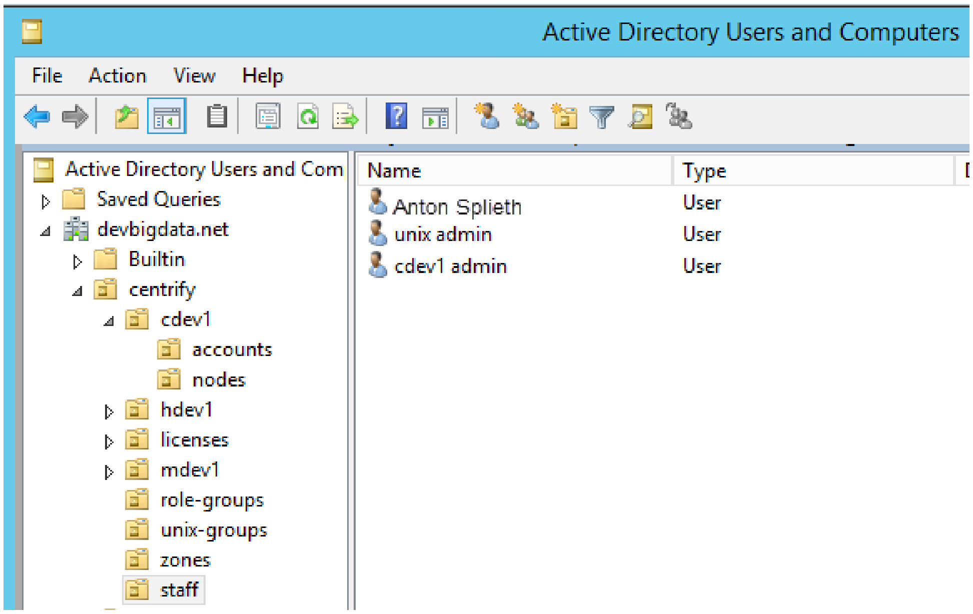
b Right-click.

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=cdev1,OU=centrify,DC=devbigdata,DC=net
OU=accounts, OU=cdev1,OU=centrify,DC=devbigdata,DC=net
```



Name	Type
Anton Splieth	User
unix admin	User
cdev1 admin	User

Figure 1 Organizational units and users for Centrify/Cloudera 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.

Create AD groups that will be linked to supergroups on cluster nodes

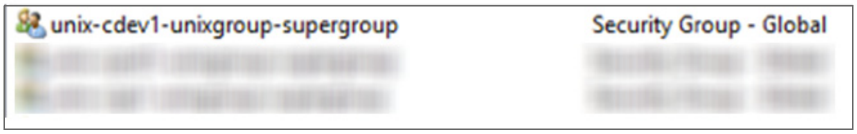
Cloudera uses Linux supergroups to manage the Hadoop Distributed File System (HDFS). The members of supergroups are able to provision and de-provision users, set file system permissions, and so on. Supergroups are configured in the `core-site.xml` file on each node in the cluster.

In environments that do not use Centrify, a Linux administrator typically creates a local supergroup on each node and manually adjusts the membership in each supergroup. After installing Centrify, an administrator can use Centrify's zone-enabled Active Directory group feature to simplify the setup and membership of supergroups. Additionally, different clusters (for example, production and development) can have their own AD groups linked to supergroups.

The first step in configuring supergroups is to create an AD group that contains AD users who will be super users for the cluster. Later (as described in "Zone-enabling AD groups for use with nodes" on page 12), you will link (or *zone-enable*) this AD group to a UNIX group for a Hadoop-recognized supergroup on each node. After you perform those steps, each AD user that you assign to the AD group automatically becomes a super user for the cluster.

To create an AD group for the cluster

1 In the `unix-groups` OU that you created earlier, use ADUC to create an AD group (such as `unix-cdev1-unixgroup-supergroup`) for super users. Later, you will use Access Manager to link this group to a UNIX group for the Hadoop-recognized supergroup. When you are done, you should see the following in ADUC:



unix-cdev1-unixgroup-supergroup Security Group - Global

Figure 2. In ADUC, verify details for the AD group

2 In the group that you created, use ADUC to add users to whom you want to grant cluster super user privileges.

To add a user to an AD 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 these users to the supergroup:

- The main Hadoop administrator (`Anton Splieth` in Figure 1).
- The `cdev1-admin` user.
- Other users to whom you will grant cluster super user privileges.

Create additional users and groups

The following procedure describes additional users and groups to create. All steps are performed in ADUC.

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 `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 administrator of the entire Hadoop integration (`Anton Splieth` in Figure 1).
   - A `unix.admin` user, who will administer all of the UNIX computers in all clusters.
   - A `cdev1.admin` user, who will be the Cloudera administrator of Cluster 1 (`cdev1`).
   - A `kerberos admin` user (not shown in Figure 1), who will be the Kerberos administrator for the entire integration.

3. Add the users that you created in Step 2 to pre-existing groups, and also to the group that you created in Step 1. 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 (`Anton Splieth` in Figure 1) to these groups: `Domain Admins`, `unix-global-sysadmin`.

     Note that you should have already added this user to the AD group that will be used with supergroups (`unix-cdev1-unixgroup-supergroup`) as described in "Create AD groups that will be linked to supergroups on cluster nodes" on page 9.

   - 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 user UNIX 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 user UNIX profile

Use Access Manager to set up the Active Directory domain and create the zones for the Hadoop cluster. See Figure 4 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, `devbigdata.net/centrify/licenses` and `devbigdata.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, `cdev1`, which matches the name of the cluster OU that you created in Step 4 on page 7).

5 Add AD users to the zone. All of the AD users that you created for inclusion in the node supergroup (Step 2 on page 9), as well as additional users (Step 2 on page 10), should be added. To simplify user management, add users to the top level (`global`) zone, not to each cluster zone (`cdev1`).

   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 (`cdev1`) 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 (`cdev1`) 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 AD 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. By zone-enabling groups that will be included in a supergroup on each node, you eliminate the need to provision each supergroup member individually on each node in the cluster.

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 10 (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 `cdev1`) and create a UNIX group for users who will have administrative privileges on all nodes in that cluster. To create this group:

  a Right click **cdev1 > UNIX Data > Groups** and select **Create UNIX Group**.

  b In the Create UNIX Group dialog **Name** field, type the name of the AD group that you created in Step 1 on page 9 (`unix-cdev1-unixgroup-supergroup`).

  c Click **OK**.

  d In the Set UNIX Group Profile dialog, type the name of the supergroup that is defined in the `core-site.xml` file on each node in the `cdev1` cluster, and click **OK**.

  For example, if the `core-site.xml` file on each node in the `cdev1` cluster defines a supergroup named `supergroup`, type `supergroup` in the Set UNIX Group Profile dialog.

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.

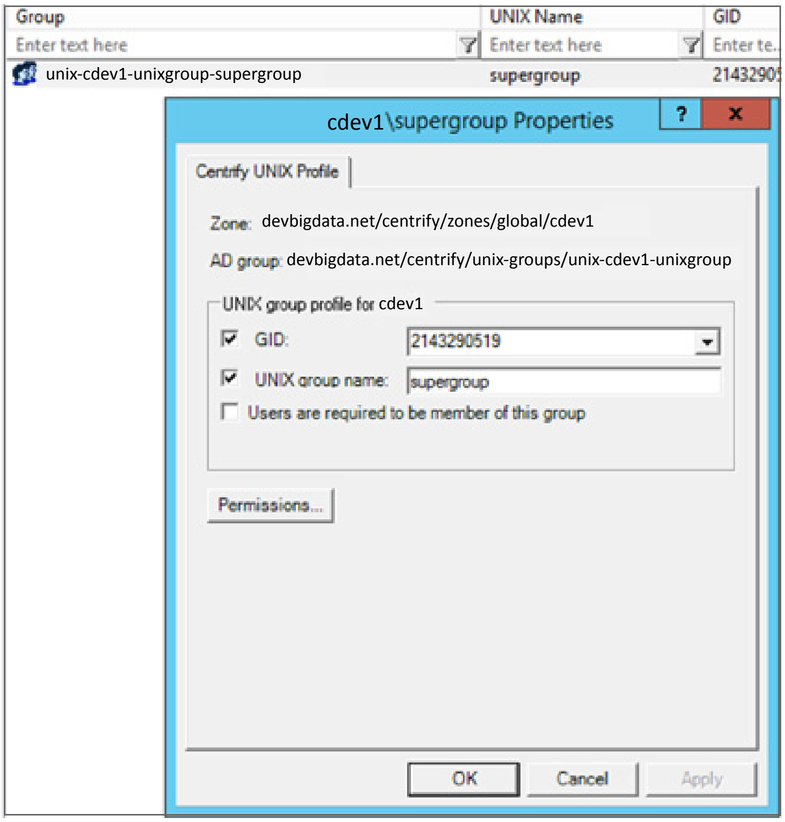
- The zone and cluster `cdev1` are linked to the AD group `unix-cdev1-unixgroup-supergroup.` When an agent-managed node computer joins the domain, the members of the AD group `unix-cdev1-unixgroup-supergroup` automatically become super users of the `cdev1` cluster.

Verifying supergroup configuration

Perform the following steps to verify that AD users and groups, and UNIX users and groups, are configured correctly for use with supergroups on nodes in the cluster:

1 In Access Manager, review the properties of the `cdev1-unixgroup-supergroup` group as shown in Figure 3.

When you view properties of `cdev1-unixgroup-supergroup` in Access Manager, you can see that `unix-cdev1-unixgroup-supergroup` is mapped to the `supergroup` UNIX group, and has a typical GID (2143290519 in this example):



Group	UNIX Name	GID
Enter text here	Enter text here	Enter te..
unix-cdev1-unixgroup-supergroup	supergroup	2143290!

cdev1\supergroup Properties

Centrify UNIX Profile

Zone: devbigdata.net/centrify/zones/global/cdev1

AD group: devbigdata.net/centrify/unix-groups/unix-cdev1-unixgroup

UNIX group profile for cdev1

☑ GID: 2143290519

☑ UNIX group name: supergroup

□ Users are required to be member of this group

Figure 3. In Access Manager, review UNIX group properties for the supergroup

2 Execute the following command on the first node in the `cdev1` cluster:

```
adquery group supergroup
```

The output should display the names of the supergroup members.

3 After Kerberos security is enabled for each cluster (as described in "Enabling security for the cluster" on page 21), verify that a member of the supergroup can create a home directory for a user only after obtaining a Kerberos ticket by executing `kinit`.

For example, if the user `ed` is a member of the supergroup and attempts to create a home directory for the user `gary` before obtaining a Kerberos ticket by executing `kinit`, the attempt should fail:

```
[ed@cdev1n1~]$ hadoop fs -mkdir /user/gary

PriviledgedActionException as:ed (auth:KERBEROS)
cause:javax.security.sasl.SaslException: GSS initiate failed [Caused by
GSSException: No valid credentials provided
...
```

After the user `ed` executes `kinit`, he is able to create a home directory for the user `gary`:

```
[ed@cdev1n1~]$ kinit
Password for ed@DEVBIGDATA.NET:
[ed@cdev1n1 ~]$ hadoop fs -mkdir /user/gary
[ed@cdev1n1 ~]$ hadoop fs -chown gary:gary /user/gary
```

Integrating Cloudera and Centrify

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

Create a Cloudera cluster that uses Centrify

The instructions that follow describe how to prepare three CentOS virtual machines (one VM for the control node to host Cloudera Manager, and two VMs for additional nodes), install the Centrify agent, and install Cloudera to create a three-node cluster. If you have already installed Cloudera on 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 page 17, then continue on to "Enabling security for the cluster" on page 21.

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 (named `cdev1n1` in this document), which hosts Cloudera Manager:

  2 processors, 16GB RAM, 1 HD (40 GB)

- For node 2 (named `cdev1n2`):

  2 processors, 8GB RAM, 1 HD (40 GB)

- For node 3 (named `cdev1n3`):

  2 processors, 8GB RAM, 1 HD (40 GB)

2 Ensure that hostname is configured correctly by reviewing host configuration files. If necessary, execute the `hostname` command to set the hostname to the computer where Access Manager is installed.

Depending on the version of UNIX or Linux in your environment, you should review one or more of these files:

- `/etc/hosts`
- `/etc/resolv.conf`
- `/etc/hostname`
- `/etc/sysconfig/network-scripts/ifgcfg-eth0`

3 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.

4 Optional: To minimize the likelihood of cluster communication issues, it is recommended that you disable and stop the `iptables` service. It is possible to continue without turning off and stopping the `iptables` service, but if you experience cluster communication issues you might have to perform these steps later.

   To disable the iptables service:

```
sudo service iptables stop
sudo chkconfig iptables off
```

5 Install openldap-clients and krb5-workstation.

```
sudo yum install openldap-clients
sudo yum install krb5-workstation
```

6 Disable SELinux.

```
sestatus
sudo setenforce 0
sudo vi /etc/selinux/config (and set SELINUX=disabled)
```

7 If `wget` is not installed, install it with `yum`.

```
sudo yum install wget
```

8 Ensure that root umask is set to `022`.

9 If transparent huge page compaction is enabled, disable it now by performing the following procedure.

   a Manually disable transparent huge page compaction:

```
echo never > /sys/kernel/mm/transparent_hugepage/defrag
```

   b Add the same command to an init script (such as `/etc/rc.local`) so that compaction is disabled every time the system is rebooted. After you edit `/etc/rc.local`, it should contain this line:

```
echo never > /sys/kernel/mm/transparent_hugepage/defrag
```

   Note If a Cloudera installation message describes a different procedure than the one described here, follow the instructions in the installation message.

10 If VM swappiness is set to a value greater than 10, set it to 10 now by performing the following procedure.

   a Manually set VM swappiness to 10 in `/etc/sysctl.conf`:

```
sysctl -w vm.swappiness=10
```

   b Verify that the setting is implemented:

```
cat /proc/sys/vm/swappiness
```

c Add the command from Step A to the /etc/sysctl.conf file so that VM swappiness is set to 10 every time the system is rebooted. After you edit /etc/sysctl.conf, it should contain this line:

```
sysctl -w vm.swappiness=10
```

Note If a Cloudera installation message describes a different procedure than the one described here, follow the instructions in the installation message.

11 Ensure that /tmp is not a mount point.

12 Enable the ntpd service.

```
chkconfig ntpd on
```

13 Enable the EPEL repository on Red Hat:

```
wget http://download.fedoraproject.org/pub/epel/6/x86_64/epel-release-6-
8.noarch.rpm
rpm -ivh epel-release-6-8.noarch.rpm
yum install sshpass
```

14 Reboot.

Install the Centrify agent

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

1 Download the appropriate tarred and zipped Centrify agent for the operating system of the virtual machines. 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) or Enterprise Edition (E).

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 The cluster will fail to start if you do not perform this step.

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

b Remove http from the adclient.krb5.service.principals line.

c Add the following lines:

```
client.dynamic.dns.enabled: true
client.dynamic.dns.refresh.interval: 3600
```

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 the AD administrator kerberos.admin (as shown in this example), or the overall administrator anton.splieth (not shown in this example).
- In the --container parameter, specify the container that contains node computers (in the environment described in this guide, the container is OU=nodes,OU=cdev1 under OU=centrify).

```
adjoin --force --user kerberos.admin --name cdevn1 --zone cdev1 --container
"OU=nodes,OU=cdev1,OU=centrify,DC=devbigdata,DC=net" devbigdata.net
```

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

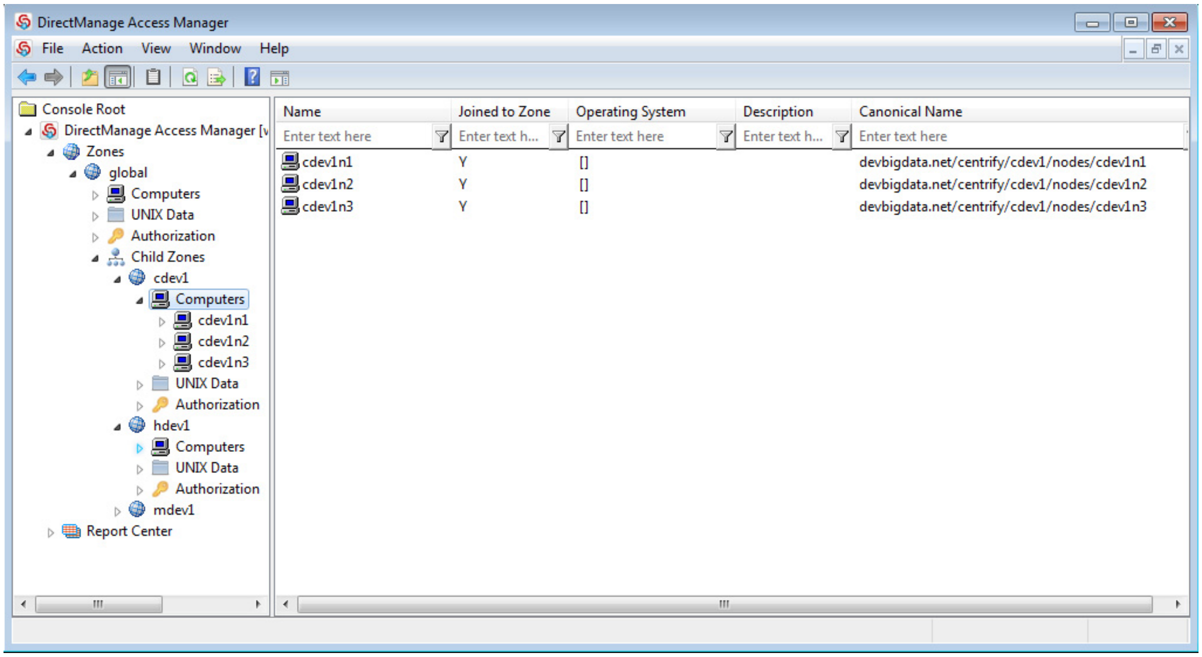
```
rpm -Uhv centrifyda-release.version
```

8 Perform a forced dynamic DNS update.

```
addns -Um
```

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:



Name	Joined to Zone	Operating System	Description	Canonical Name
cdev1n1	Y	[]		devbigdata.net/centrify/cdev1/nodes/cdev1n1
cdev1n2	Y	[]		devbigdata.net/centrify/cdev1/nodes/cdev1n2
cdev1n3	Y	[]		devbigdata.net/centrify/cdev1/nodes/cdev1n3

Figure 4 Cluster VMs are joined to Active Directory and can be viewed in Access Manager

Install the Centrify LDAP proxy on Node 1

Install the Centrify LDAP proxy on Node 1 (`cdev1n1`) in the cluster. Installing the Centrify LDAP proxy is necessary because Cloudera 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 Cloudera on each node in the cluster

You can now install Cloudera on each computer in the cluster as described in http://www.cloudera.com/content/cloudera/en/documentation/core/latest/PDF/cloudera-installation.pdf.

Install Cloudera Manager on Node 1 (cdev1n1) First, install Cloudera Manager on the CentOS workstation VM (`cdev1n1`). Installing Cloudera Manager first makes it available to use when installing Cloudera on the other nodes. When prompted to select a version of Cloudera Manager, select either **Cloudera Enterprise** or **Cloudera Enterprise Data Hub Edition Trial**.

Install Cloudera on other nodes (cdev1n2 and cdev1n3) and configure the cluster Perform the following steps to install and configure Cloudera on the remaining nodes in the cluster.

1 Log on to the Cloudera Manager web console. The URL is shown in the last screen of the Cloudera Manager installation wizard.

2 Click **Continue** until you are prompted to specify hosts for the clusters.

3 At the cluster host prompt, enter the following for the cluster in this example:
`cdev1n[1-3].devbigdata.net`

4 Click **Continue** until you reach the JDK Installation Options page. Select **Install Oracle Java SE Development Kit (JDK)** and **Install Java Unlimited Strength Encryption Policy Files**.

5 Click **Continue** until you reach the Enable Single User Mode page. Ensure that **Single User Mode** is not selected.

6 Click **Continue** until you are prompted for SSH login credentials. Log in as root, or as the unix.admin user. Select **All hosts accept same password**.

7 Click **Continue** until you are prompted to choose the CDH5 services to install on your cluster. Select **Core with HBase**.

8 Click **Continue** until the installation finishes.

9 Optional: After the installation finishes, disable Host Clock Offset Thresholds Monitoring.

Create home directories and authorize cluster users in HDFS

On Node 1 (`cdev1n1`), 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@DEVBIGDATA.NET
Got host ticket host/cdev1n1.devbigdata.net@DEVBIGDATA.NET
login as edDEVBIGDATA.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.devbigdata.net
[ed@cdev1n1 ~]$ dzdo su hdfs
[dzdo] password for ed:
[hdfs@cdev1n1 ed]$ hadoop fs -mkdir /user/ed
[hdfs@cdev1n1 ed]$ hadoop fs -chown ed:ed /user/ed
[hdfs@cdev1n1 ed]$ hadoop fs -mkdir /user/wade
[hdfs@cdev1n1 ed]$ hadoop fs -chown wade:wade /user/wade
```

Testing the cluster

Perform the following steps to ensure that the cluster is fully functional, and that Centrify controls access to cluster nodes.

1 On the Cloudera Manager home page, verify that all indicators are green. If some indicators are not green, perform the following steps:

   a Restart all VMs (`cdev1n1`, `cdev1n2`, `cdev1n3`).

   b Restart the Cloudera Management service.

   c Restart the cluster (`cdev1`).

2 Log on to the workstation node VM (`cdev1n1`) as the user `hdfs` and verify that the following test job completes successfully:

```
hadoop jar /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce/hadoop-mapreduce-
examples.jar pi 10 10
```

Enabling security for the cluster

The procedures in this section assume that you will use Cloudera Manager to configure, manage, and enable security for clusters.

Key tasks for enabling security

Configuring Kerberos authentication for a Cloudera cluster involves both manual and automated tasks. As a preview, you should plan to perform the following tasks:

- Use Cloudera Manager to enable Kerberos security.
- Use Cloudera Manager to generate service principals, accounts, and `keytab` files.
- Restart all Hadoop and Cloudera management service.

Enabling Kerberos security in Cloudera Manager

Use Cloudera Manager to enable Kerberos security for the cluster. For details about using Cloudera Manager to enable Kerberos security, see http://www.cloudera.com/content/cloudera/en/documentation/core/latest/topics/cm_sg_intro_kerb.html.

1 On the Cloudera Manager home page, open the drop-down menu for the cluster (`cdev1`) and select **Enable Kerberos**.

2 Select all of the check boxes on the first screen of the wizard and click **Continue**.

3 On the KDC Information page, select **Active Directory** for the KDC Type and enter information about the KDC server, Kerberos configuration, and Active Directory configuration.

You would provide the following information if you set up the environment described throughout this document:

KDC Type: `Active Directory`

KDC Server Host: `wdev1-dc1.devbigdata.net`

Kerberos Security Realm: `DEVBIGDATA.NET`

Kerberos Encryption Types: `rc4-hmac`

Active Directory Suffix:
`OU=accounts,OU=cdev1,OU=centrify,DC=devbigdata,DC=net`

Active Directory Account Prefix: `cdev1-`

Active Directory Domain Controller Override: `wdev1-dc1.devbigdata.net`

Note As an alternative to specifying the Active Directory Domain Controller Override, in the Kerberos Configuration area you can select the option to use DNS SRV records to find domain controllers. That option is not selected by default.

4 On the KRB5 Configuration screen, ensure that **Manage krb5.conf through Cloudera Manager** is *not* selected so that Centrify, and not Cloudera Manager, will manage `krb5.conf`. Click **Continue**.

5 On the KDC Account Manager Credentials screen, enter `kerberos.admin@DEVBIGDATA.NET`, and provide the password for that user.

6 Click **Continue** (and accept all default values) until you reach the Configure Ports screen. Keep the default ports, select **Yes, I am ready to restart the cluster now**, and click **Continue**.

7 Click **Continue** until the wizard finishes.

Restarting Hadoop and Cloudera services

Log on to the computer where Cloudera Manager is running. In Cloudera Manager, start all Hadoop services and Cloudera management services.

Validating cluster security

Now that the Cloudera 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

After logging into a cluster node (for example `cdev1n1`), the user (`cdev1.admin@cdev1n1` in this example) has Kerberos credentials and will be able to run Hadoop jobs, such as the example shown here that computes the value of Pi. You can verify that a user has Kerberos credentials by executing a command similar to the following. This example includes the command and the resulting output.

```
[cdev1.admin@cdev1n1 ~]$ klist
Ticket cache: FILE:/tmp/krb5cc_1094714474
Default principal: cdev1.admin@DEVBIGDATA.NET
Valid starting      Expires            Service principal
07/14/15 21:59:02   07/15/15 07:59:03  krbtgt/DEVBIGDATA.NET@DEVBIGDATA.NET
renew until 07/21/15 21:59:02
```

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

If a user (such as `hdfs` in the following example) does not have Kerberos credentials and tries to run a Hadoop job after logging into a cluster node (for example `cdev1n1`), the attempt fails:

```
[root@cdev1n1 ~]# su - hdfs
[hdfs@cdev1n1 ~]$ klist
klist: No credentials cache found (ticket cache FILE:/tmp/krb5cc_494)
[hdfs@cdev1n1 ~]$ hadoop fs -ls /user
15/07/14 21:49:06 WARN security.UserGroupInformation:
PriviledgedActionException as:hdfs (auth:KERBEROS)
cause:javax.security.sasl.SaslException: GSS initiate failed [Caused by
GSSException: No valid credentials provided (Mechanism level: Failed to find
any Kerberos tgt)]
...

ls: Failed on local exception: java.io.IOException:
javax.security.sasl.SaslException: GSS initiate failed [Caused by
GSSException: No valid credentials provided (Mechanism level: Failed to find
any Kerberos tgt)]; Host Details : local host is: "cdev1n1.devbigdata.net/
172.27.10.191"; destination host is: "cdev1n1.devbigdata.net":8020;
```

If a user (such as `cdev1.admin@cdev1n1` in the following example) has Kerberos credentials and tries to run a Hadoop job, the attempt succeeds:

```
[cdev1.admin@cdev1n1 ~]$ hadoop jar /opt/cloudera/parcels/CDH/lib/hadoop-
mapreduce/hadoop-mapreduce-examples.jar pi 10 10
...
Starting Job
15/07/14 22:03:31 INFO client.RMProxy: Connecting to ResourceManager at
cdev1n1.devbigdata.net/172.27.10.191
...
15/07/14 22:04:09 INFO mapreduce.Job: Job job_1436934856496_0001 completed
successfully
...
Job Finished in 100.085 seconds
```

Additional tests

Optionally perform the following tests to verify correct operation:

1. As an AD user who is a member of the zone for the cluster (`cdev1`), log into Node 1 (`cdev1n1`) from a Windows computer using Centrify PuTTY after selecting Kerberos as the authentication method for SSH.
2. Run the `id` command. The UID should be greater than 1000, and the user should be a member of the `supergroup` group if the user is an HDFS admin who manages user directories in HDFS.

For additional information about validating cluster security, see http://www.cloudera.com/content/cloudera/en/documentation/core/latest/topics/cm_sg_s8_verify_kerb.html.

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.

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.