Centrify Identity and Access Management for Cloudera

Integration Guide

Abstract

Centrify Server Suite is an enterprise-class solution that secures Cloudera Enterprise Data Hub leveraging an organization's existing Active Directory infrastructure to deliver access control, privilege management and user-level auditing.

Contents

Overview

Centrify Server Suite is an enterprise-class solution that secures even the most complex Hadoop environments leveraging an organization's existing Active Directory infrastructure to deliver access control, privilege management and user-level auditing.

Centrify Server Suite secures the industry's broadest range of mission-critical servers from identity-related insider risks and outsider attacks, making security and regulatory compliance repeatable and sustainable. The solution leverages existing Active Directory infrastructure to centrally manage authentication, access controls, privileged identities, policy enforcement and compliance for on-premises and cloud resources.

Centrify Server Suite provides identity, access and privilege management for Cloudera

- Simplifying AD integration for Cloudera to run in secure mode
- Automating service account credential management
- Simplifying access with AD-based user single sign-on authentication
- Ensuring regulatory compliance with least privilege and auditing
- Developer SDKs for secure client application access to Hadoop

This document will show how to use Cloudera Manager to setup the cluster leveraging Centrify for the management of Kerberos and user access controls to the cluster. If you are not using the Cloudera Manager, you may need to manually create the various service accounts and distribute their keytab files. Centrify will also be used to automate DNS configuration as well as Time sync with the Domain Controllers to simplify the configuration and management of the nodes as well as provide the identity and access management for Active Directory user access.

Planning for Active Directory Integration

Centrify provides a centralized access control and privilege management solution built on top of Active Directory that simply requires the Centrify agent software to be installed on every node within the cluster while administration is performed through Microsoft Management Consoles on an administrator's Windows computer.

Cluster Creation Pre-Requisites

There are several common requirements such as you must have an Active Directory environment running, you will need a Windows workstation joined to the domain where you can run administrative consoles and you will need several Linux systems on which to install Cloudera.

You should request a free trial of Centrify Server Suite if you don't already have access to Centrify software from http://www.centrify.com/lp/server-suite-free-trial/, just specify Hadoop in the Comments field.

You can find the Centrify Documentation online here http://community.centrify.com/t5/custom/page/page-id/Centrify-Documentation after you register for a free trial and setup your Centrify Account here https://www.centrify.com/account/register.asp.

First, You should outline a naming convention for all Hadoop components that will reside in AD. Ideally you will be able to identify the cluster in the names. But keep in mind the limitations of the Active Directory sAMAccountName that has a maximum length of 20 characters and must be unique across the Active Directory environment.

- You will need an Active Directory OU for managing all your Hadoop clusters such as OU=Hadoop. You may have to ask your Active Directory team to create this OU for you. The technical lead or Hadoop admin should have full control of this Hadoop OU. Your Active Directory Domain Admin will need to delegate administrative rights of this OU to your technical lead.
- Each cluster should have it's own OU in order to independently manage it's nodes and service accounts. The OU name should reflect the name of the cluster; e.g. Cloudera1. This is usually created within an OU that was created by the AD staff and delegated to you so that you can create an OU for each Cloudera cluster and manage the accounts and policies yourself.
- Centrify uses Zones as a logical container for storing the access and privilege permissions for the selected Active Directory users who you authorize to access your Cloudera cluster. You will setup a unique Zone for each Cloudera cluster you deploy in order to ensure separation of duties and enable delegated administration. This Linux identity, access and privilege information is stored within the OU that was created for you in the steps above. Use the child zone name as the same name for the cluster prefix, e.g. Cloudera1.

Additionally, you will need the following:

- At least 3 (preferably 5) Linux systems that are compatible with Cloudera to use for the Hadoop nodes.
- Access to Cloudera software
- Preferably the organization is running their own Hadoop repository/repo (this speeds up any setup)

Preparing Active Directory

Create Active Directory OUs (Organizational Unit is just a container for AD objects). For this task you may need your Active Directory administrator to perform the first step and grant you delegated permission to manage this top level OU for

- Create OUs
  - Create the Hadoop OU; e.g. OU=Hadoop, DC=Company, DC=com
  - Then for each Cluster create another OU under OU=Hadoop; e.g. OU= Cloudera1, OU=Hadoop, DC=Company, DC=Com

        - Next in order to make it easier to manage nodes in the cluster separate from the Service accounts, you may also want to create a set of child OUs with OU=Nodes and OU=Users
    - Create AD groups
        - Create the linux-admins group
        - Create the supergroup group
    - Create AD users
        - Create the linuxadm user
        - Set a password for the linuxadm user
        - Create the srv-cm-krb user
        - Set a password for the srv-cm-krb user
    - Delegate AD privileges
        - Add linuxadm as a member of linux-admins and supergroup
        - Delegate permissions for linuxadm to add computers to the domain
        - Delegate permissions for srv-cm-krb to create and manager accounts in the Hadoop OU

Setup Centrify Zones and setup Roles for Linux login

Start with the Centrify Server Suite Quick Start Guide to install the Management Consoles and to setup your Centrify Zone with the appropriate Roles to grant AD users with login rights to the Linux systems you will join to Active Directory in the next step.

- Run the appropriate setup program from the Management ISO for Windows 32-bit or 64-bit on a Windows administrator's workstation.

  The setup program simply copies the necessary files to the local Windows computer, so there are no special permissions required to run the setup program other than permission to install files on the local computer. Follow the prompts displayed to select the type of suite to install and which components to install.

- Open Access Manager to start the Setup Wizard and create the containers for Licenses and Zones. You can accept the default locations or use create a Centrify organizational unit for the containers.

- In Access Manager, create a new zone with the default options. For example, create a new zone named **Hadoop**.

- In Access Manager, add Active Directory users to the new zone.
    - Select the new **Hadoop** zone.

        - Right-click, then select Add User to search for and select existing Active Directory users.
        - Select Define user UNIX profile and deselect assign roles.
        - Accept the defaults for all fields.
    - Create a child zone.
        - Select the **Hadoop** zone.
        - Right-click, then select Create Child Zone.
        - Type a name for the zone, for example, **Cloudera1** and an optional description, then click Next and Finish to create the new child zone.
    - Assign a role for the users you added to the **Hadoop** zone.
    - User profiles are inherited by child zones, so the users you added to **Hadoop**, automatically have a profile in **Cloudera1**. To login to a machine, a user requires a profile and a role assignment. DirectManage provides a default UNIX Login role that you can assign to enable users to login.
        - Expand **Child Zones**, **Cloudera1**, and **Authorization**.
        - Select **Role Assignments**, right-click, then click **Assign Role**.
        - Select the **UNIX Login** role from the results and click **OK**.
        - Click **Add AD Account**, then search for one of the Active Directory user you added to the **Hadoop** zone. Select this user and click **OK**.
    - Zone enable the AD groups
        - Add the linux-admins group to the Cloudera1 Zone.
        - Add the supergroup group to the Cloudera1 Zone.
    - Zone enable the AD users
        - Add the linuxadm user to the Cloudera1 Zone
    - Delegate Zone privileges
        - Delegate permissions for linuxadm to add computers to the Cloudera1 Zone

Setup Cloudera Cluster with Centrify

Setup the Virtual Machines

- Provision 3 new Centos 6.x virtual machines:
    - cm1, 2 processors, 8gb RAM, 1HD (40gb)

  - node1, 2 processors, 8gb RAM, 1HD (40gb)
  - node2, 2 processors, 8gb RAM, 1HD (40gb)
- For each node, perform the following:
- Ensure hostname is configured
- Ensure DNS is configured to enable access to the Internet
- Install openldap-clients
  - `sudo yum install openldap-clients`
- Add the following to the end of /etc/sudoers
  - `%linux-admins  ALL=(ALL)  NOPASSWD: ALL`
- Disable selinux
  - `sudo vi /etc/selinux/config and set to disabled`
- Install wget via yum if it's not installed already
  - `sudo yum install wget`
- turn off iptables firewall that may prevent access to the cluster
  - `sudo service iptables stop`
  - `sudo chkconfig iptables off`
- Reboot

Install Centrify on each node in the cluster

Install the Centrify Agent and join the nodes to Active Directory.

- After downloading Centrify agents disk image, just copy the appropriate tgz file from the ISO to the Nodes, un pack the file and run the install.sh
- Install.sh will ask several questions if you run it interactively which is suggested this first time, however the installation can be automated with a custom config file for silent installation. Just install Standard Edition of Centrify Suite and do not join Active Directory, we will need to do that after making a few changes to the configuration files.
- Edit the /etc/centrifydc/centrifydc.conf file and uncomment he adclient.krb5.service.principals line and remove the http principal.

  Note: this step is required or the cluster will not start. Centrify should not create servicePrincipalName for the http service since Cloudera will need to do this later.

- Add the following values to /etc/cenridydc/centrifydc.conf.
  - `client.dynamic.dns.enabled: true`
  - `client.dynamic.dns.refresh.interval: 3600`

    Note: 86400 seconds is the default value for Windows clients, configure appropriately for the environment, using 1 hour in test environment
- Join the node to the domain
  - `sudo adjoin --force -u linuxadm --name NODENAME --zone cloudera1 --container ou=Cloudera,ou=Hadoop,dc=company,dc=com company.com`
  - Enter password for linuxadm when prompted
- Optional: Install the Centrify Audit agent and enable audit (rpm –Uvh centrifyda-<version>)
- To force a dynamic DNS update, run the following command as root.
  - `addns -Um`
- The computer should now be joined to AD and reboot. At this point, you should be able to login with an AD userid and password for the user you granted login rights to previously.

Install Cloudera on each node in the cluster

You will need to install Cloudera Manager on the first node in the cluster, CM1.

- Logon via ssh to CM1
- Obtain the Cloudera Manager installer
  - `wget http://archive.cloudera.com/cm5/installer/latest/cloudera-manager-installer.bin`
  - `chmod 755 cloudera-manager-installer.bin`
- Launch the Cloudera Manager installer
  - `sudo ./cloudera-manager-installer.bin`
- Accept the license agreements and complete installation

Install and Configure Cluster

- Log into the Cloudera Manager web console
- Click Continue until prompted to specify hosts for the clusters
- Enter cm1, node[1-2] and click search

- Continue until the Oracle Binary Code License Agreement is displayed
- Select Install Oracle Java SE Development kit (JDK) and Install Java Unlimited Strength Encryption Policy Files
- Do not select Single User Mode
- Click Continue until prompted for SSH login credentials
    - Select Another User
    - Enter the linuxadm password configured in Active Directory
- Click Continue, selecting Hadoop Core and all the defaults
- Once the installation is complete, disable Host Clock Offset Thresholds monitoring
- Click on the Hosts tab
- Under Hosts, click on the Configuration tab
- Select the Monitoring Category on the left side of the screen
- Locate the Host Clock Offset Thresholds property
    - Click on the Value field
    - Change both values from Specify to Never
    - Click Save Changes

Test Cluster

- Ensure all health indicators on the Cloudera Manager homepage are green
- Logon to the CM1 node and run a test job
    - `sudo -u hdfs hadoop jar /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar pi 10 10`

You now have a working environment with Centrify controlling access to Linux hosts and coexisting with the Cloudera cluster.

Enable Security

The next step is to configure the cluster to operate in secure mode leveraging the Kerberos that was enabled by the Centrify agent on each of the nodes.

- In Cloudera Manager, click on Select Enable Kerberos from the cluster options

- Check all the boxes on the first screen of the wizard and click Continue
- On the KDC Information screen, select Active Directory for the KDC Type
- Enter the following on the KDC Information screen
  - KDC Server Host: dom1.example.com
  - Kerberos Security Realm: DOM1.EXAMPLE.COM
  - Kerberos Encryption Types: rc4-hmac
  - Active Directory Account Prefix: srv-cdh-
  - Active Directory Suffix: ou=hadoop,DC=dom1,DC=example,DC=com
- Click Continue
- On the KRB5 Configuration screen, make sure "Manage krb5.conf through Cloudera Manager" is NOT selected and click Continue
- On the KDC Account Manager Credentials screen, enter the credentials for the srv-cm-krb user
- Click Continue until the wizard is complete, selecting "Yes, I am ready to restart the cluster now." when prompted.

Verify Proper Operation

Now that the Cloudera cluster is using Centrify for Active Directory based authentication, Active Directory users who are members of the Centrify Zone for the specified cluster can now login using her Active Directory credentials directly at the console prompt or could use a Kerberized SSH client such as Centrify's version of PuTTY to get Single Sign-on to the Cluster. Once logged in, the user will have Kerberos credentials and will be able to run a Hadoop job such as the example used below that computes the value of Pi. Since the cluster is now running in secure mode, users without Kerberos will not be able to successfully submit a job to the cluster.

- Ensure all health indicators on the Cloudera Manager homepage are green
- Logon to the CM1 node and run the original test
  - `sudo -u hdfs hadoop jar /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar p`
  - This test should fail
- Login from a Windows computer using Centrify PuTTY after selecting Kerberos as the authentication method for SSH.
- Ensure the current user is a member of the supergroup
  - run the id command

- Ensure the current user has Kerberos credentials
  - run the klist command
  - If no valid credentials are listed, run kinit to populate the credential cache
- Run the test as the current user
  - `hadoop jar /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar pi 10 10`
  - This test should succeed

Conclusion

Centrify Server Suite — the industry's most widely deployed solution for securing identity on Linux- and Windows-based servers and applications — provides several benefits for Hadoop and Big Data environments including:

- **Simple and secure access to Hadoop environments.** Centrify makes it simple to run Hadoop in secure mode by leveraging existing identity management infrastructure—Active Directory—without the hassle of introducing alternative solutions that do not scale and are not enterprise ready. Centrify Server Suite also saves money by letting organizations leverage existing skill sets within the enterprise.
- **Single sign-on for IT administrators and big data users.** By extending the power of Active Directory's Kerberos and LDAP capabilities to Hadoop clusters, Centrify Server Suite lets organizations leverage existing Active Directory-based authentication for Hadoop administrators and end users. New SSO functionality in Big Data environments makes users more productive and secure by allowing them to login in as themselves, rather than sharing privileged accounts.
- **Secure machine-to-machine communications.** Centrify Server Suite automates Hadoop service account management within Active Directory. By automating machine-to-machine credential management, Centrify not only secures user identity but also system and service account identity.
- **Reduced identity-related risks and greater regulatory compliance.** The reality is that Hadoop environments store most if not all of an organization's most important data. Centrify Server Suite tracks user activity back to an individual in Active Directory, thereby making data more secure. Centrify also reports on who did what across Hadoop clusters, nodes and services. And, by enforcing access controls and least-privilege security across Hadoop, Centrify delivers cost-effective compliance through combined access and activity reporting.
- **Certified solution for superior compatibility and support.** Centrify has worked closely with Cloudera and has received product certification. This ensures product compatibility and technical support collaboration between customers, Cloudera and Centrify.

How to Contact Centrify

North America
(And All Locations Outside EMEA)

Centrify Corporation
3393 Octavius Dr, Suite 100
Santa Clara, CA 95054
United States

Sales: +1 (669) 444-5200
Online: www.centrify.com/contact

Europe, Middle East, Africa
(EMEA)

Centrify EMEA
Lilly Hill House
Lilly Hill Road
Bracknell, Berkshire RG12 2SJ
United Kingdom

Sales: +44 (0) 1344 317950