Lab: Securing a Hadoop Cluster

|  |  |
| --- | --- |
| **Objective:** | To understand how to configure security for HDP. |
| **Successful Outcome:** | The **hadoop** user will be authenticated and authorized to view the contents of the **/user** folder in HDFS. |
| **Before You Begin:** | SSH into **hadoop-master**. |

1. Create a New User
   1. On **hadoop-slave1** as **root**, create a new user named **hadoop**:

# useradd -g hadoop hadoop

* 1. Switch to the **hdfs** user and create a new directory in HDFS named **/user/hadoop**.
  2. Change ownership of **/user/hadoop** in HDFS to the **hadoop** user.
  3. Exit out from the **hdfs** user, and switch to the **hadoop** user.
  4. Check whether you can do a listing of the **/user** directory successfully:

$ hadoop fs -ls /user

* 1. Exit out from the **hadoop** user.

**NOTE**: The current cluster is not a secure cluster so you can easily do a listing of the **/user** directory in HDFS successfully.

1. Install Kerberos
   1. As the root user on **hadoop-slave1**, install the following packages:

# yum -y install krb5-server krb5-libs krb5-auth-dialog krb5-workstation

* 1. Login to **hadoop-slave2**, **hadoop-slave3** and **hadoop-slave4** and install the Kerberos client only:

# yum -y install krb5-workstation

1. Configure Kerberos
   1. On **hadoop-slave1**, edit the configuration file **/etc/krb5.conf** and modify the value for **kdc** and **admin\_server** to **hadoop-slave1**:

# vi /etc/krb5.conf

[realms]

EXAMPLE.COM = {

kdc = hadoop-slave1

admin\_server = hadoop-slave1

}

* 1. Copy the **/etc/krb5.conf** file to all the other nodes:

# ~/scripts/distFile.sh /etc/krb5.conf /etc/krb5.conf

* 1. Enter the following command to create a Kerberos database using the **kdb5\_util** utility:

# kdb5\_util create -s

During this step it will ask you to define a master key. Enter **1234** as the key.

1. Start Kerberos
   1. Start the KDC server by executing following commands:

# /etc/rc.d/init.d/krb5kdc start

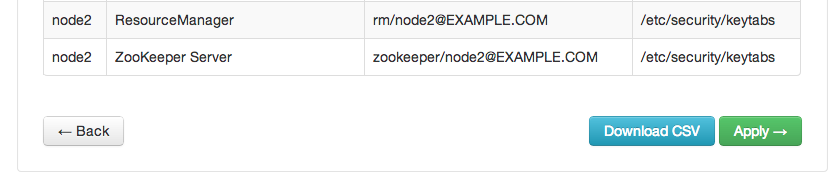
# /etc/rc.d/init.d/kadmin start

1. Run the Enable Security Wizard
   1. In Ambari, go the **Admin** page and click on the **Security** link:

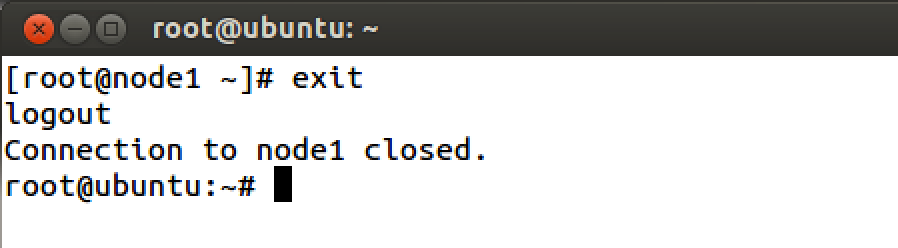


* 1. Click the **Enable Security** button.
  2. Notice on the **Get Started** step of the **Enable Security Wizard** there are four manual steps that must be completed first. We have already executed the first 2 steps. For the remaining 2 steps, click the **Next** button.
  3. The default settings on the **Configure Services** step of the wizard are fine, so click **Next** to continue.

1. Create the Principals and Keytabs
   1. On the **Create Principals and Keytabs** step of the wizard, all the required default settings for Kerberos are shown in a tabular format. Click the **Download CSV** button at the bottom of the page and save the file:



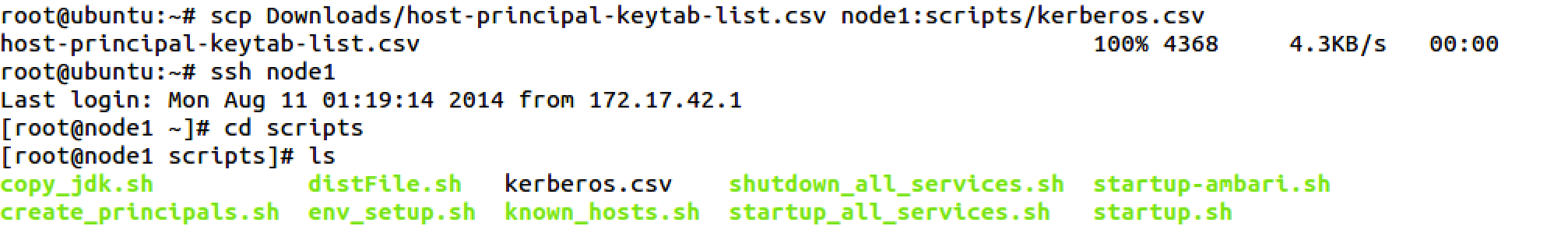
* 1. Exit from **hadoop-slave1**, so that you will be no more on any node. The command prompt should show root@ubutnu:



* 1. Execute following command to copy CSV file to **hadoop-slave1** (type this command in one line):

# scp Downloads/host-principal-keytab-list.csv hadoop-slave1:scripts/kerberos.csv

* 1. Now login back to hadoop-slave1 and navigate to /root/scripts folder. You must see **kerberos.csv** file there.

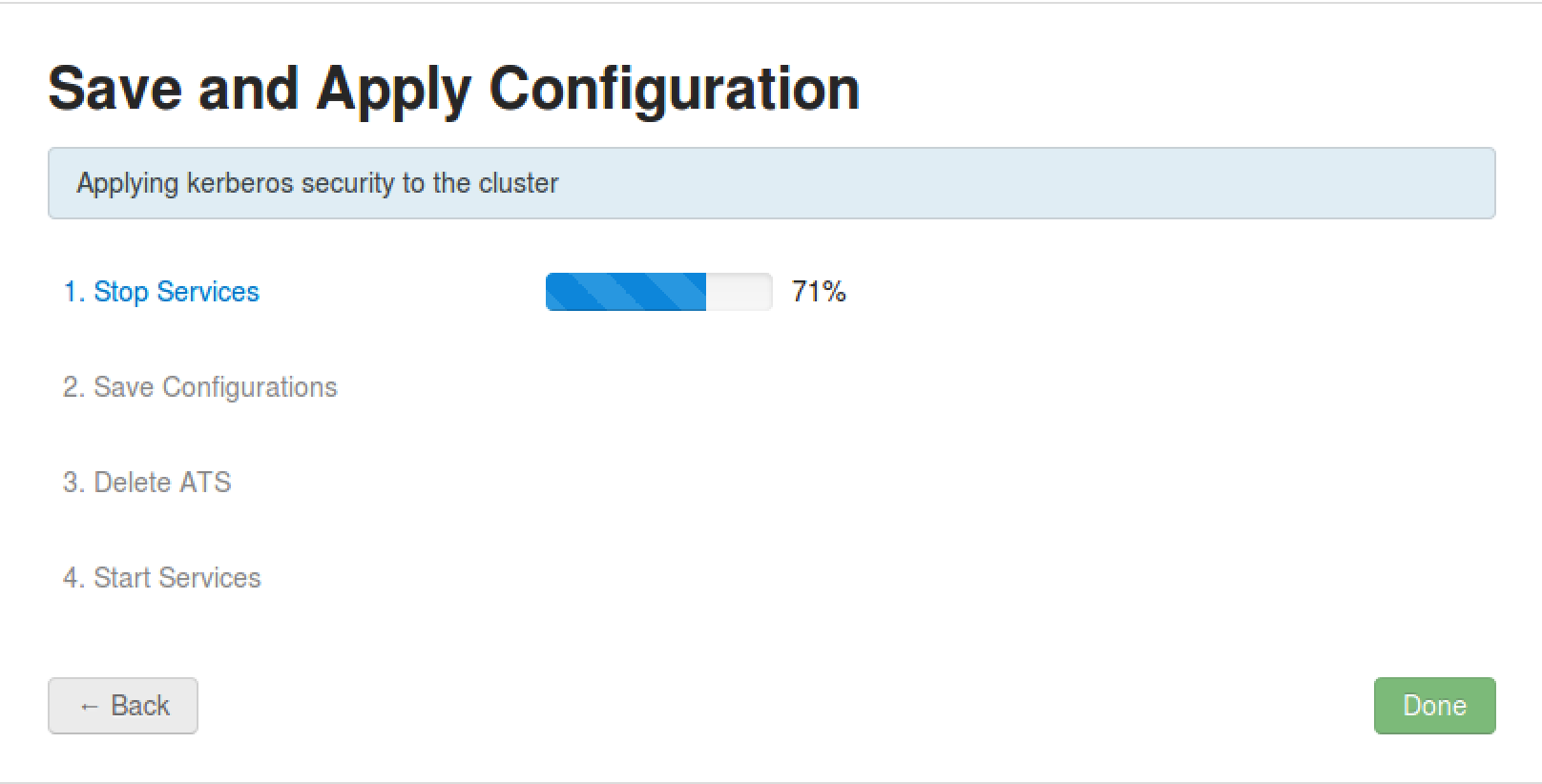


* 1. Run the pre-written script to create all required principals and keytabs. It will ask for the location of the CSV file you created. Provide the full path **(/root/scripts/Kerberos.csv**) to the file:

# /root/scripts/create\_principals.sh

**NOTE**: This step will create all required principals and keytab files on all the nodes. Once you are done with the step, go back to Ambari UI.

1. Finish the Enable Security Wizard
   1. We have completed all 4 required steps. Now it is time to enable security through Ambari. Click the **Apply** button.
   2. The **Save and Apply Configuration** step can take 5-10 minutes. When the task is complete, click the **Done** button:



1. Verify Security is Enabled
   1. On **hadoop-slave1**, switch to the **hadoop** user and try to list the contents of **/user** in HDFS:

# su - hadoop

$ hadoop fs -ls /user

The command should throw an error.

1. Configure User Permissions
   1. Logout from ‘hadoop’ user.
   2. As the **root** user, add **hadoop** to the Kerberos database. Start the **kadmin** service by typing following:

# kadmin.local

* 1. Type following command to create a **hadoop** principal:

**kadmin.local:** addprinc -randkey hadoop@EXAMPLE.COM

* 1. Now create a keytab file in the **/etc/security/keytabs** directory using the following command:

**kadmin.local:** xst -norandkey -k /etc/security/keytabs/hadoop.headless.keytab hadoop@EXAMPLE.COM

**kadmin.local:** exit

* 1. Set appropriate permissions for the keytab file for the **hadoop** user:

# chown hadoop:hadoop /etc/security/keytabs/hadoop.headless.keytab

# chmod 440 /etc/security/keytabs/hadoop.headless.keytab

* 1. Switch to the **hadoop** user and initialize the keytab file:

# su - hadoop

$ kinit -kt /etc/security/keytabs/hadoop.headless.keytab hadoop@EXAMPLE.COM

* 1. Now try to list the contents of **/user** in HDFS again. This time you should be able to view the folder’s contents!

**RESULT**: You have enabled Kerberos security for your HDP cluster.