Configure for Kerberos Integration

Hiemdal

Contents:

This document describes how to set up a Trifacta® user in Kerberos.

- Kerberos provides authentication services across a wide variety of platforms. See *http://www.kerberos.org/*.

Pre-requisites for Kerberos integration

Before you begin, please verify the following:

1. The `[hadoop.user` (default=`trifacta)]` user is created and enabled on each node in the Hadoop cluster.

   > **NOTE:** If LDAP is enabled, the `trifacta` user should be created in the same realm as the cluster.

2. On the Trifacta host, the directory `/opt/trifacta` is owned by the `[hadoop.user]` user .
3. The `[hadoop.user]` user exists on each node in the Hadoop cluster.

   > **NOTE:** The `[hadoop.user]` must have the same user ID and group ID on each node in the cluster. Depending on your cluster's configuration, this requirement may require an LDAP command. Configuring LDAP is beyond the scope of this document.

4. The `[hadoop.user]` user must be a member of any special group that is permitted to access HDFS or to run Hadoop jobs.

Configure the KDC

Steps:

1. On your KDC node, configure a Kerberos principal for the Trifacta platform:
   a. The principal's identifier has two parts: its **name** and its **realm**. For example, the principal `trifacta @HADOOPVAL.MSSVC.LOCAL` has the name `trifacta` and the realm `HADOOPVAL.MSSVC.LOCAL`.
   b. Retain the name and principal for later configuration.
2. Create a keytab file for the Trifacta principal. Command:

   ```
   kadmin xst -k trifacta.keytab <full principal identifier>
   ```

   where:
   `<full_principal_identifier>` is the principal identifier in Kerberos.

On the KDC, you may have to run `kadmin.local` instead of `kadmin`. The rest of the arguments should remain the same.

NOTE: If you're creating a keytab file in an AD environment, alternative instructions may need to be applied. See below.

3. Verify that the keytab is working. Command:

```
klist -e -k -t trifacta.keytab
```

4. Copy the keytab to the Trifacta node in the following directory:
   `/opt/trifacta/conf/trifacta.keytab`
5. Configure the keytab file so that it is owned by the `[hadoop.user]` user. It should only be readable by that user.

NOTE: Verify that all user principals that use the platform are also members of the group of the keytab user.

Create keytab in Active Directory environments

Some additional instructions are provided for the following environments.

For MIT Kerberos

See *https://kb.iu.edu/d/aumh* :

```
> ktutil
  ktutil:  addent -password -p username@EXAMPLE.COM -k 1 -e rc4-hmac
  Password for username@EXAMPLE.COM: [enter your password]
  ktutil:  addent -password -p username@EXAMPLE.COM -k 1 -e aes256-cts
  Password for username@EXAMPLE.COM: [enter your password]
  ktutil:  wkt username.keytab
  ktutil:  quit
```

For Heimdal Kerberos

```
  > ktutil -k username.keytab add -p username@EXAMPLE.COM -e arcfour-hmac-md5 -V 1
```

If the keytab created in Heimdal does not work, you may need an `aes256-cts` entry. In this case, locate a machine with MIT Kerberos, and use the MIT Kerberos method instead.

Configure the Trifacta platform for Kerberos

You can apply this change through the *Admin Settings Page* (recommended) or

`trifacta-conf.json`

. For more information, see *Platform Configuration Methods*.

Locate the `kerberos` section, which controls Kerberos authentication.

Example configuration:

Substitute your own values in place of the example values as appropriate.

```
"kerberos.enabled": true,
"kerberos.principal": "trifacta",
"kerberos.kdc": "kdc.mssvc.local",
"kerberos.realm": "HADOOPVAL.MSSVC.LOCAL",
"kerberos.keytab": "/opt/trifacta/conf/trifacta.keytab"
"kerberos.principals.hive": "<UNUSED>",
"kerberos.principals.namenode": "nn/_HOST@EXAMPLE.COM"
"kerberos.principals.resourcemanager": "<YOUR_VALUE_HERE>",
```

Parameter	Description
enabled	To enable Kerberos authentication, set this value to `true`.
principal	The name part of the principal you created in the KDC
kdc	The host of the KDC
realm	Realm of the KDC
keytab	Directory in the Trifacta deployment where the Kerberos keytab file is stored
principals	List of jobtrackers and namenodes that are governed by Kerberos **NOTE:** `kerberos.principals.hive` is unused. This value must be inserted into the Hive connection definition. See *Create Hive Connections*. **NOTE:** If you don't know the values to use here, see *Set principal values* below. **NOTE:** If you don't specify principal names in the `principals` definition section, the default names are used: `mapred/<jobtracker host>@<realm>`. You should specify the principals explicitly.

At this point, you should be able to load files from HDFS and run jobs against the kerberized Hadoop cluster.

Set principal values for YARN

Check the following Hadoop config properties in `yarn-site.xml` :

```
principals.jobtracker = yarn.resourcemanager.principal
principals.namenode = dfs.namenode.kerberos.principal
```

Configure Kerberos-delegated relational connections

When Kerberos has been enabled in the platform, you can apply the global keytab to be used for SSO connections to relational sources of data. For more information, see *Enable SSO for Relational Connections*.