



# INFORME DE CONEXIÓN

---

## **HIVE – SQOOP/BEELINE POR MEDIO DE DRIVER JDBC**

Jorge Andrés Salazar

10/09/2020

# CONTENIDO

INTRODUCCIÓN.....	3
DESCRIPCIÓN DEL PROBLEMA.....	3
OBJETIVO .....	3
DESCRIPCIÓN DE LA SOLUCIÓN.....	3
1. Estructura de conexión .....	3
2. Procedimiento .....	3

# INTRODUCCIÓN

A continuación, en este documento se detalla el procedimiento, así como estructura de conexión al cluster Kerberizado de UIAF, tanto vía Beeline como Sqoop. En adición, se muestra la estructura de las cadenas a utilizar junto a los parámetros que conforman estas, teniendo en cuenta los requerimientos para una correcta autenticación sobre la base de datos de Hive.

## DESCRIPCIÓN DEL PROBLEMA

Se presenta un problema de conexión al cluster Kerberizado UIAF, impidiendo tanto la obtención como carga de datos de la base de datos de Hive asociada a este cluster.

## OBJETIVO

Generar una cadena válida para conexión al cluster Kerberizado de UIAF tanto por vía Beeline como Sqoop.

## DESCRIPCIÓN DE LA SOLUCIÓN

El desarrollo de esta solución se divide en dos (2) posibles cadenas de conexión, las cuales son: Beeline y Sqoop.

### 1. ESTRUCTURA DE CONEXIÓN

La estructura de cadena de conexión vía Beeline, se detalla a continuación:

```
!connect
jdbc:hive2://<host>:<port>/<database>;principal=<principal>;ssl=<true/false>;sslTrustStore=<ssltruststore>;sslTrustStorePassword=<ssltruststorepassword>
```

La estructura de cadena de conexión vía Sqoop, se detalla a continuación:

```
sqoop                                eval                                --connect
'jdbc:hive2://<host>:<port>/<database>;principal=<principal>;ssl=<true/false>;sslTrustStore=<ssltruststore>;sslTrustStorePassword=<ssltruststorepassword>' --query 'show databases' --driver org.apache.hive.jdbc.HiveDriver
```

### 2. PROCEDIMIENTO

A continuación, se presenta el paso a paso tanto para la obtención de los parámetros de las cadenas, como para la asignación de estos.

1. Obtener los archivos de configuración de Hive llamados: “hive-site.xml” y “ssl-client.xml”

Nota: Estos archivos se encuentran en la plataforma de Hive en Cloudera Manager al hacer clic en la lista desplegable llamada "Actions" y en la opción "Download Client Configuration".

- Una vez obtenidos estos archivos, podrán ser abiertos con la ayuda de cualquier editor de texto para obtener dentro de ellos los parámetros detallados en la Tabla 1.

Tabla 1. Parámetros y descripción para cadenas de conexión.

Parámetro	Descripción	Ejemplo
host	host o ip de conexión a la base de datos de Hive	node02.uaifdev.local
port	Puerto de conexión a la base de datos de hive	10000
database	Nombre de la Base de datos	default
principal	hive.server2.authentication.kerberos.principal, propiedad del archivo hive-site.xml	_HOST@REALM
SSL	true or false, si hive.server2.use.SSL se encuentra en true en el archivo hive-site.xml se deben agregar las líneas de conexión de ssl	true
sslTrustStore	ssl.client.truststore.location, propiedad del archivo ssl-client.xml. Se debe buscar la dirección completa de la ubicación de aquel archivo .jks	/var/lib/cloudera-scm-agent/agent-cert/cm-auto-global_truststore.jks
sslTrustStorePassword	ssl.client.truststore.password, propiedad del archivo ssl-client.xml. Contraseña.	dnfjqwbnxadjos34243fdsafdf32d
--query	Consulta a realizarse en la base de datos	Show databases
--driver	Driver de conexión de jdbc con Hive	org.apache.hive.jdbc.HiveDriver

- Tras la obtención de estos valores a continuación serán reemplazados en las cadenas proporcionadas en la Sección 1, como se muestra a continuación para el caso de conexión Sqoop de acuerdo con los valores ejemplo de la Tabla 1.

Sqoop eval --connect

```
'jdbc:hive2://node02.uaifdev.local:10000/default;principal=_HOST@REALM;ssl=true;sslTrustStore=/var/lib/cloudera-scm-agent/agent-cert/cm-auto-global_truststore.jks;sslTrustStorePassword=dnfjqwbnxadjos34243fdsafdf32d' --query 'show databases' --driver org.apache.hive.jdbc.HiveDriver
```