

LDAP Authentication for Hive in Big Data Management[®]

Abstract

You can configure Hive to use LDAP authentication on Cloudera CDH and Hortonworks HDP clusters. This article discusses how Big Data Management® integrates with the authentication mechanisms of the Hadoop cluster and Hive.

Supported Versions

- Big Data Management 10.2.2

Table of Contents

Overview.	2
Authentication for Hive in a Cloudera CDH Cluster.	2
Cluster Does Not Use Kerberos Authentication.	3
Cluster Uses Kerberos authentication.	3
Authentication for Hive in a Hortonworks HDP Cluster.	3
Cluster Does Not Use Kerberos Authentication.	3
Summary.	3

Overview

Big Data Management supports Kerberos authentication for Hadoop clusters. Based on the Hadoop distribution, you can configure Hive to use LDAP or Kerberos authentication. When you configure Hive for authentication, the Hive metastore, HiveServer2, and the Hive clients use the authentication mechanism.

Cloudera CDH and Hortonworks HDP support LDAP authentication for HiveServer2. The Data Integration Service depends on HiveServer2 for the following tasks:

- Run mappings in the native environment
- Preview data
- Access Hive metadata at run time
- Access Hive tables enabled for SQL authorization
- Access bucketed or ACID Hive tables

Based on whether the Hadoop cluster uses Kerberos authentication, Big Data Management has certain guidelines to use LDAP authentication for Hive. To access Hive sources and targets, you can use LDAP authentication. However, to run mappings on a cluster that uses Kerberos authentication, the Data Integration Service always uses Kerberos authentication. It does not use LDAP authentication because the mappings can contain sources and targets other than Hive. For example, a mapping can read from a Hive source and write to an HDFS target.

Authentication for Hive in a Cloudera CDH Cluster

Cloudera CDH supports both Kerberos and LDAP to authenticate HiveServer2. When the cluster uses Kerberos authentication, Cloudera recommends using LDAP or Kerberos authentication for HiveServer2. Review the Big Data Management rules and guidelines to use LDAP authentication for Hive.

Cluster Does Not Use Kerberos Authentication

When Hive uses LDAP authentication in a cluster that does not use Kerberos authentication, consider the following guidelines:

- To access Hive metadata and preview Hive data, you can configure Big Data Management to use LDAP credentials. In the Hive connection, set the user name and password properties for the LDAP user.
- If the Data Integration Service uses operating system profiles, set the user name of the operating system profile as the Hive user.

Cluster Uses Kerberos authentication

In a Cloudera CDH cluster that uses Kerberos authentication, you can configure Big Data Management to access Hive metadata using LDAP authorization. Consider the following guidelines:

- To use LDAP authorization for metadata access, set the Hive connection user name as the LDAP user name.
- Data preview and mappings that depend on HiveServer2 use Kerberos authentication. To run mappings that depend on HiveServer2, specify the Hadoop Kerberos SPN as the LDAP user in the Data Integration Service properties.
- If the Data Integration Service uses operating system profiles, set the impersonation user as the LDAP user.

Authentication for Hive in a Hortonworks HDP Cluster

Hortonworks HDP supports LDAP to authenticate Hive. You can connect to Hive only with LDAP credentials. You cannot connect to Hive with Kerberos credentials. If the cluster uses Kerberos authentication, Informatica does not support LDAP authentication for Hive. Review the Big Data Management rules and guidelines to use LDAP authentication for Hive.

Cluster Does Not Use Kerberos Authentication

When Hive uses LDAP authentication in a cluster that does not use Kerberos authentication, consider the following guidelines:

- To access Hive metadata and preview Hive data, you can configure Big Data Management to use LDAP credentials. In the Hive connection, set the user name and password properties for the LDAP user.
- If the Data Integration Service uses operating system profiles, set the user name of the operating system profile as the Hive user.

Summary

Big Data Management supports LDAP authentication for Hive as follows:

Cluster does not use Kerberos authentication.

Big Data Management can use LDAP authentication to access Hive metadata and preview Hive data. To run mappings, Informatica recommends setting the impersonation user or the operating system profile user name as the Hive connection user name.

Cluster uses Kerberos authentication.

Big Data Management can use LDAP authentication to access Hive metadata. However, Kerberos authentication takes precedence when you run Big Data Management mappings because the mappings can contain Hive, HDFS, or any other sources and targets.

In a Hortonworks HDP environment that uses Kerberos authentication, Informatica does not recommend LDAP to authenticate Hive.

Author

Indra Sivakumar

Acknowledgements

The author would like to acknowledge Sumeet Kumar Agrawal, Mathew Kovoov, Naveen Garla Babu, and Asif M for their contributions.