



Cloudera ODBC Driver for Apache
Hive

Version 2.5.20



Important Notice

© 2010-2016 Cloudera, Inc. All rights reserved.

Cloudera, the Cloudera logo, and any other product or service names or slogans contained in this document, except as otherwise disclaimed, are trademarks of Cloudera and its suppliers or licensors, and may not be copied, imitated or used, in whole or in part, without the prior written permission of Cloudera or the applicable trademark holder.

Hadoop and the Hadoop elephant logo are trademarks of the Apache Software Foundation. All other trademarks, registered trademarks, product names and company names or logos mentioned in this document are the property of their respective owners. Reference to any products, services, processes or other information, by trade name, trademark, manufacturer, supplier or otherwise does not constitute or imply endorsement, sponsorship or recommendation thereof by us.

Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Cloudera.

Cloudera may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Cloudera, the furnishing of this document does not give you any license to these patents, trademarks copyrights, or other intellectual property.

The information in this document is subject to change without notice. Cloudera shall not be liable for any damages resulting from technical errors or omissions which may be present in this document, or from use of this document.

Cloudera, Inc.
1001 Page Mill Road, Building 2
Palo Alto, CA 94304-1008
info@cloudera.com
US: 1-888-789-1488
Intl: 1-650-843-0595
www.cloudera.com

Release Information

Version: 2.5.20

Date: July 12, 2016

Table of Contents

ABOUT THE CLOUDERA ODBC DRIVER FOR APACHE HIVE	5
WINDOWS DRIVER	6
INSTALLING THE DRIVER ON WINDOWS	6
CREATING A DATA SOURCE NAME	6
CONFIGURING A DSN-LESS CONNECTION	9
CONFIGURING AUTHENTICATION	11
CONFIGURING ADVANCED OPTIONS	14
CONFIGURING HTTP OPTIONS	15
CONFIGURING SSL VERIFICATION	16
CONFIGURING THE TEMPORARY TABLE FEATURE	17
CONFIGURING SERVER-SIDE PROPERTIES	18
CONFIGURING LOGGING OPTIONS	19
CONFIGURING KERBEROS AUTHENTICATION FOR WINDOWS	20
VERIFYING THE VERSION NUMBER	24
LINUX DRIVER	26
LINUX SYSTEM REQUIREMENTS	26
INSTALLING THE DRIVER	26
SETTING THE LD_LIBRARY_PATH ENVIRONMENT VARIABLE ON LINUX	27
VERIFYING THE VERSION NUMBER	28
MAC OS X DRIVER	29
INSTALLING THE DRIVER ON MAC OS X	29
VERIFYING THE VERSION NUMBER	29
SETTING THE DYLD_LIBRARY_PATH ENVIRONMENT VARIABLE	30
AIX DRIVER	31
INSTALLING THE DRIVER ON AIX	31
SETTING THE LD_LIBRARY_PATH ENVIRONMENT VARIABLE ON AIX	32
VERIFYING THE VERSION NUMBER	32
DEBIAN DRIVER	33
INSTALLING THE DRIVER ON DEBIAN	33
SETTING THE LD_LIBRARY_PATH ENVIRONMENT VARIABLE ON DEBIAN	34
VERIFYING THE VERSION NUMBER	35
CONFIGURING ODBC CONNECTIONS FOR NON-WINDOWS PLATFORMS	36
CONFIGURATION FILES	36
SAMPLE CONFIGURATION FILES	37
CONFIGURING THE ENVIRONMENT	37

DEFINING DSNS IN ODBC.INI	38
SPECIFYING ODBC DRIVERS IN ODBCINST.INI	39
CONFIGURING DRIVER SETTINGS IN CLOUDERA.HIVEODBC.INI	41
CONFIGURING SERVICE DISCOVERY MODE	41
CONFIGURING AUTHENTICATION	41
CONFIGURING SSL VERIFICATION	43
CONFIGURING SERVER-SIDE PROPERTIES	44
CONFIGURING LOGGING OPTIONS	45
TESTING THE CONNECTION	47
AUTHENTICATION MECHANISMS	50
USING A CONNECTION STRING	52
DSN CONNECTION STRING EXAMPLE	52
DSN-LESS CONNECTION STRING EXAMPLES	52
FEATURES	55
SQL CONNECTOR FOR HIVEQL	55
DATA TYPES	55
CATALOG AND SCHEMA SUPPORT	57
HIVE_SYSTEM TABLE	57
SERVER-SIDE PROPERTIES	57
TEMPORARY TABLES	58
GET TABLES WITH QUERY	59
ACTIVE DIRECTORY	59
WRITE-BACK	59
DYNAMIC SERVICE DISCOVERY USING ZOOKEEPER	59
CONTACT US	61
DRIVER CONFIGURATION OPTIONS	62
CONFIGURATION OPTIONS APPEARING IN THE USER INTERFACE	62
CONFIGURATION OPTIONS HAVING ONLY KEY NAMES	81
ODBC API CONFORMANCE LEVEL	84

About the Cloudera ODBC Driver for Apache Hive

The Cloudera ODBC Driver for Apache Hive is used for direct SQL and HiveQL access to Apache Hadoop / Hive distributions, enabling Business Intelligence (BI), analytics, and reporting on Hadoop / Hive-based data. The driver efficiently transforms an application's SQL query into the equivalent form in HiveQL, which is a subset of SQL-92. If an application is Hive-aware, then the driver is configurable to pass the query through to the database for processing. The driver interrogates Hive to obtain schema information to present to a SQL-based application. Queries, including joins, are translated from SQL to HiveQL. For more information about the differences between HiveQL and SQL, see "SQL Connector for HiveQL" on page 55.

The Cloudera ODBC Driver for Apache Hive complies with the ODBC 3.80 data standard and adds important functionality such as Unicode and 32- and 64-bit support for high-performance computing environments.

ODBC is one of the most established and widely supported APIs for connecting to and working with databases. At the heart of the technology is the ODBC driver, which connects an application to the database. For more information about ODBC, see the *Data Access Standards Glossary*: <http://www.simba.com/resources/data-access-standards-library>. For complete information about the ODBC specification, see the *ODBC API Reference*: [http://msdn.microsoft.com/en-us/library/windows/desktop/ms714562\(v=vs.85\).aspx](http://msdn.microsoft.com/en-us/library/windows/desktop/ms714562(v=vs.85).aspx).

The *Installation and Configuration Guide* is suitable for users who are looking to access data residing within Hive from their desktop environment. Application developers might also find the information helpful. Refer to your application for details on connecting via ODBC.

Windows Driver

Installing the Driver on Windows

On 64-bit Windows operating systems, you can execute both 32- and 64-bit applications. However, 64-bit applications must use 64-bit drivers and 32-bit applications must use 32-bit drivers. Make sure that you use the version of the driver matching the bitness of the client application accessing data in Hadoop / Hive:

- ClouderaHiveODBC32.msi for 32-bit applications
- ClouderaHiveODBC64.msi for 64-bit applications

You can install both versions of the driver on the same machine.

You install the Cloudera ODBC Driver for Apache Hive on client machines that access data stored in a Hadoop cluster with the Hive service installed and running. Each machine that you install the driver on must meet the following minimum system requirements:

- One of the following operating systems:
 - Windows Vista, 7, 8, or 10
 - Windows Server 2008 or later
- 100 MB of available disk space

Important:

To install the driver, you must have Administrator privileges on the machine.

The driver supports Apache Hive versions 0.11 through 1.2.

To install the Cloudera ODBC Driver for Apache Hive:

1. Depending on the bitness of your client application, double-click to run **ClouderaHiveODBC32.msi** or **ClouderaHiveODBC64.msi**.
2. Click **Next**.
3. Select the check box to accept the terms of the License Agreement if you agree, and then click **Next**.
4. To change the installation location, click **Change**, then browse to the desired folder, and then click **OK**. To accept the installation location, click **Next**.
5. Click **Install**.
6. When the installation completes, click **Finish**.

Creating a Data Source Name

Typically, after installing the Cloudera ODBC Driver for Apache Hive, you need to create a Data Source Name (DSN).

Alternatively, for information about DSN-less connections, see "Configuring a DSN-less Connection" on page 9.

To create a Data Source Name:

1. Open the ODBC Administrator:
 - If you are using Windows 7 or earlier, click **Start** , then click **All Programs**, then click the **Cloudera ODBC Driver for Apache Hive 2.5** program group corresponding to the bitness of the client application accessing data in Hadoop / Hive, and then click **ODBC Administrator**.
 - Or, if you are using Windows 8 or later, on the Start screen, type **ODBC administrator**, and then click the **ODBC Administrator** search result corresponding to the bitness of the client application accessing data in Hadoop / Hive.
2. In the ODBC Data Source Administrator, click the **Drivers** tab, and then scroll down as needed to confirm that the Cloudera ODBC Driver for Apache Hive appears in the alphabetical list of ODBC drivers that are installed on your system.
3. Choose one:
 - To create a DSN that only the user currently logged into Windows can use, click the **User DSN** tab.
 - Or, to create a DSN that all users who log into Windows can use, click the **System DSN** tab.

Note:

It is recommended that you create a System DSN instead of a User DSN. Some applications, such as Sisense, load the data using a different user account, and might not be able to detect User DSNs that are created under another user account.

4. Click **Add**.
5. In the Create New Data Source dialog box, select **Cloudera ODBC Driver for Apache Hive** and then click **Finish**.
6. In the **Data Source Name** field, type a name for your DSN.
7. Optionally, in the **Description** field, type relevant details about the DSN.
8. In the **Hive Server Type** drop-down list, select **Hive Server 1** or **Hive Server 2**.

Note:

If you are connecting through Apache ZooKeeper, then **Hive Server 1** is not supported.

9. Choose one:
 - To connect to Hive without using the Apache ZooKeeper service, in the **Service Discovery Mode** drop-down list, select **No Service Discovery**.
 - Or, to enable the driver to discover Hive Server 2 services via the ZooKeeper service, in the **Service Discovery Mode** drop-down list, select **ZooKeeper**.
10. Choose one:
 - If you selected **No Service Discovery** above, then in the **Host(s)** field, type the IP address or host name of the Hive server.

- Or, if you selected **ZooKeeper** above, then in the **Host(s)** field, type a comma-separated list of ZooKeeper servers. Use the following format, where *[ZK_Host]* is the IP address or host name of the ZooKeeper server and *[ZK_Port]* is the number of the TCP port that the ZooKeeper server uses to listen for client connections:

[ZK_Host1]:[ZK_Port1], [ZK_Host2]:[ZK_Port2]

11. In the **Port** field, if you selected **No Service Discovery** above, then type the number of the TCP port that the Hive server uses to listen for client connections. Otherwise, do not type a value in the field.
12. In the **Database** field, type the name of the database schema to use when a schema is not explicitly specified in a query.

Note:

You can still issue queries on other schemas by explicitly specifying the schema in the query. To inspect your databases and determine the appropriate schema to use, type the `show databases` command at the Hive command prompt.

13. In the **ZooKeeper Namespace** field, if you selected **ZooKeeper** above, then type the namespace on ZooKeeper under which Hive Server 2 znodes are added. Otherwise, do not type a value in the field.
14. In the Authentication area, configure authentication as needed. For more information, see "Configuring Authentication" on page 11.

Note:

Hive Server 1 does not support authentication. Most default configurations of Hive Server 2 require User Name authentication. To verify the authentication mechanism that you need to use for your connection, check the configuration of your Hadoop / Hive distribution. For more information, see "Authentication Mechanisms" on page 50.

15. Optionally, if the operations against Hive are to be done on behalf of a user that is different than the authenticated user for the connection, type the name of the user to be delegated in the **Delegation UID** field.

Note:

This option is applicable only when connecting to a Hive Server 2 instance that supports this feature.

16. In the **Thrift Transport** drop-down list, select the transport protocol to use in the Thrift layer.
17. If the Thrift Transport option is set to HTTP, then to configure HTTP options such as custom headers, click **HTTP Options**. For more information, see "Configuring HTTP Options" on page 15.
18. To configure client-server verification over SSL, click **SSL Options**. For more information, see "Configuring SSL Verification" on page 16.

19. To configure advanced driver options, click **Advanced Options**. For more information, see "Configuring Advanced Options" on page 14.
20. To configure server-side properties, click **Advanced Options** and then click **Server Side Properties**. For more information, see "Configuring Server-Side Properties" on page 18.
21. To configure the Temporary Table feature, click **Advanced Options** and then click **Temporary Table Configuration**. For more information, see "Configuring the Temporary Table Feature" on page 17 and "Temporary Tables" on page 58.

Important:

When connecting to Hive 0.14 or later, the Temporary Tables feature is always enabled and you do not need to configure it in the driver.

22. To configure logging behavior for the driver, click **Logging Options**. For more information, see "Configuring Logging Options" on page 19.
23. To test the connection, click **Test**. Review the results as needed, and then click **OK**.

Note:

If the connection fails, then confirm that the settings in the Cloudera ODBC Driver for Apache Hive DSN Setup dialog box are correct. Contact your Hive server administrator as needed.

24. To save your settings and close the Cloudera ODBC Driver for Apache Hive DSN Setup dialog box, click **OK**.
25. To close the ODBC Data Source Administrator, click **OK**.

Configuring a DSN-less Connection

Some client applications provide support for connecting to a data source using a driver without a Data Source Name (DSN). To configure a DSN-less connection, you can use a connection string or the Cloudera Hive ODBC Driver Configuration tool that is installed with the Cloudera ODBC Driver for Apache Hive. The following section explains how to use the driver configuration tool. For information about using connection strings, see "Using a Connection String" on page 52.

To configure a DSN-less connection using the driver configuration tool:

1. Choose one:
 - If you are using Windows 7 or earlier, click **Start** , then click **All Programs**, and then click the **Cloudera ODBC Driver for Apache Hive 2.5** program group corresponding to the bitness of the client application accessing data in Hadoop / Hive.
 - Or, if you are using Windows 8 or later, click the arrow button at the bottom of the Start screen, and then find the **Cloudera ODBC Driver for Apache Hive 2.5** program group corresponding to the bitness of the client application accessing data in Hadoop / Hive.
2. Click **Driver Configuration**, and then click **OK** if prompted for administrator permission to make modifications to the machine.

Note:

You must have administrator access to the machine to run this application because it makes changes to the registry.

3. In the **Hive Server Type** drop-down list, select **Hive Server 1** or **Hive Server 2**.

Note:

If you are connecting through Apache ZooKeeper, then **Hive Server 1** is not supported.

4. Choose one:
 - To connect to Hive without using the Apache ZooKeeper service, in the **Service Discovery Mode** drop-down list, select **No Service Discovery**.
 - Or, to enable the driver to discover Hive Server 2 services via the ZooKeeper service, in the **Service Discovery Mode** drop-down list, select **ZooKeeper**.
5. In the **ZooKeeper Namespace** field, if you selected **ZooKeeper** above, then type the namespace on ZooKeeper under which Hive Server 2 znodes are added. Otherwise, do not type a value in the field.
6. In the Authentication area, configure authentication as needed. For more information, see "Configuring Authentication" on page 11.

Note:

Hive Server 1 does not support authentication. Most default configurations of Hive Server 2 require User Name authentication. To verify the authentication mechanism that you need to use for your connection, check the configuration of your Hadoop / Hive distribution. For more information, see "Authentication Mechanisms" on page 50.

7. Optionally, if the operations against Hive are to be done on behalf of a user that is different than the authenticated user for the connection, then in the **Delegation UID** field, type the name of the user to be delegated.

Note:

This option is applicable only when connecting to a Hive Server 2 instance that supports this feature.

8. In the **Thrift Transport** drop-down list, select the transport protocol to use in the Thrift layer.
9. If the Thrift Transport option is set to HTTP, then to configure HTTP options such as custom headers, click **HTTP Options**. For more information, see "Configuring HTTP Options" on page 15.
10. To configure client-server verification over SSL, click **SSL Options**. For more information, see "Configuring SSL Verification" on page 16.

11. To configure advanced options, click **Advanced Options**. For more information, see "Configuring Advanced Options" on page 14.
12. To configure server-side properties, click **Advanced Options** and then click **Server Side Properties**. For more information, see "Configuring Server-Side Properties" on page 18.
13. To configure the Temporary Table feature, click **Advanced Options** and then click **Temporary Table Configuration**. For more information, see "Temporary Tables" on page 58 and "Configuring the Temporary Table Feature" on page 17.

Important:

When connecting to Hive 0.14 or later, the Temporary Tables feature is always enabled and you do not need to configure it in the driver.

14. To save your settings and close the Cloudera Hive ODBC Driver Configuration tool, click **OK**.

Configuring Authentication

Some Hive servers are configured to require authentication for access. To connect to a Hive server, you must configure the Cloudera ODBC Driver for Apache Hive to use the authentication mechanism that matches the access requirements of the server and provides the necessary credentials.

For information about how to determine the type of authentication your Hive server requires, see "Authentication Mechanisms" on page 50.

ODBC applications that connect to Hive Server 2 using a DSN can pass in authentication credentials by defining them in the DSN. To configure authentication for a connection that uses a DSN, use the ODBC Data Source Administrator.

Normally, applications that are not Hive Server 2 aware and that connect using a DSN-less connection do not have a facility for passing authentication credentials to the Cloudera ODBC Driver for Apache Hive for a connection. However, the Cloudera Hive ODBC Driver Configuration tool enables you to configure authentication without using a DSN.

Important:

Credentials defined in a DSN take precedence over credentials configured using the driver configuration tool. Credentials configured using the driver configuration tool apply for all connections that are made using a DSN-less connection unless the client application is Hive Server 2 aware and requests credentials from the user.

Using No Authentication

When connecting to a Hive server of type Hive Server 1, you must use No Authentication. When you use No Authentication, Binary is the only Thrift transport protocol that is supported.

To configure a connection without authentication:

1. Choose one:
 - To access authentication options for a DSN, open the ODBC Data Source Administrator where you created the DSN, then select the DSN, and then click **Configure**.
 - Or, to access authentication options for a DSN-less connection, open the Cloudera Hive ODBC Driver Configuration tool.
2. In the **Mechanism** drop-down list, select **No Authentication**.
3. If the Hive server is configured to use SSL, then click **SSL Options** to configure SSL for the connection. For more information, see "Configuring SSL Verification" on page 16.
4. To save your settings and close the dialog box, click **OK**.

Using Kerberos

Kerberos must be installed and configured before you can use this authentication mechanism. For more information, see "Configuring Kerberos Authentication for Windows" on page 20.

This authentication mechanism is available only for Hive Server 2. When you use Kerberos authentication, the Binary transport protocol is not supported.

To configure Kerberos authentication:

1. Choose one:
 - To access authentication options for a DSN, open the ODBC Data Source Administrator where you created the DSN, then select the DSN, and then click **Configure**.
 - Or, to access authentication options for a DSN-less connection, open the Cloudera Hive ODBC Driver Configuration tool.
2. In the **Mechanism** drop-down list, select **Kerberos**.
3. Choose one:
 - To use the default realm defined in your Kerberos setup, leave the **Realm** field empty.
 - Or, if your Kerberos setup does not define a default realm or if the realm of your Hive Server 2 host is not the default, then, in the **Realm** field, type the Kerberos realm of the Hive Server 2.
4. In the **Host FQDN** field, type the fully qualified domain name of the Hive Server 2 host.

Note:

To use the Hive server host name as the fully qualified domain name for Kerberos authentication, in the **Host FQDN** field, type **_HOST**.

5. To allow the driver to pass your credentials directly to the server for use in authentication, set **Delegate Kerberos Credentials**.
6. In the **Service Name** field, type the service name of the Hive server.
7. In the **Thrift Transport** drop-down list, select the transport protocol to use in the Thrift layer.

Important:

When using this authentication mechanism, the Binary transport protocol is not supported.

8. If the Hive server is configured to use SSL, then click **SSL Options** to configure SSL for the connection. For more information, see "Configuring SSL Verification" on page 16.
9. To save your settings and close the dialog box, click **OK**.

Using User Name

This authentication mechanism requires a user name but not a password. The user name labels the session, facilitating database tracking.

This authentication mechanism is available only for Hive Server 2. Most default configurations of Hive Server 2 require User Name authentication. When you use User Name authentication, SSL is not supported and SASL is the only Thrift transport protocol available.

To configure User Name authentication:

1. Choose one:
 - To access authentication options for a DSN, open the ODBC Data Source Administrator where you created the DSN, then select the DSN, and then click **Configure**.
 - Or, to access authentication options for a DSN-less connection, open the Cloudera Hive ODBC Driver Configuration tool.
2. In the **Mechanism** drop-down list, select **User Name**.
3. In the **User Name** field, type an appropriate user name for accessing the Hive server.
4. To save your settings and close the dialog box, click **OK**.

Using User Name And Password

This authentication mechanism requires a user name and a password.

This authentication mechanism is available only for Hive Server 2.

To configure User Name And Password authentication:

1. Choose one:
 - To access authentication options for a DSN, open the ODBC Data Source Administrator where you created the DSN, then select the DSN, and then click **Configure**.
 - Or, to access authentication options for a DSN-less connection, open the Cloudera Hive ODBC Driver Configuration tool.
2. In the **Mechanism** drop-down list, select **User Name And Password**.
3. In the **User Name** field, type an appropriate user name for accessing the Hive server.
4. In the **Password** field, type the password corresponding to the user name you typed above.
5. To save the password, select the **Save Password (Encrypted)** check box.

Important:

The password is obscured, that is, not saved in plain text. However, it is still possible for the encrypted password to be copied and used.

6. In the **Thrift Transport** drop-down list, select the transport protocol to use in the Thrift layer.
7. If the Hive server is configured to use SSL, then click **SSL Options** to configure SSL for the connection. For more information, see "Configuring SSL Verification" on page 16.
8. To save your settings and close the dialog box, click **OK**.

Configuring Advanced Options

You can configure advanced options to modify the behavior of the driver.

To configure advanced options:

1. Choose one:
 - To access advanced options for a DSN, open the ODBC Data Source Administrator where you created the DSN, then select the DSN, then click **Configure**, and then click **Advanced Options**.
 - Or, to access advanced options for a DSN-less connection, open the Cloudera Hive ODBC Driver Configuration tool, and then click **Advanced Options**.
2. To disable the SQL Connector feature, select the **Use Native Query** check box.
3. To defer query execution to SQLExecute, select the **Fast SQLPrepare** check box.
4. To allow driver-wide configurations to take precedence over connection and DSN settings, select the **Driver Config Take Precedence** check box.
5. To use the asynchronous version of the API call against Hive for executing a query, select the **Use Async Exec** check box.

Note:

This option is applicable only when connecting to a Hive cluster running Hive 0.12.0 or later.

6. To retrieve table names from the database by using the SHOW TABLES query, select the **Get Tables With Query** check box.

Note:

This option is applicable only when connecting to Hive Server 2.

7. To enable the driver to return SQL_WVARCHAR instead of SQL_VARCHAR for STRING and VARCHAR columns, and SQL_WCHAR instead of SQL_CHAR for CHAR columns, select the **Unicode SQL Character Types** check box.

8. To enable the driver to return the `hive_system` table for catalog function calls such as `SQLTables` and `SQLColumns`, select the **Show System Table** check box.
9. To handle Kerberos authentication using the SSPI plugin instead of MIT Kerberos by default, select the **Use Only SSPI** check box.
10. To enable the driver to automatically open a new session when the existing session is no longer valid, select the **Invalid Session Auto Recover** check box.

Note:

This option is applicable only when connecting to Hive Server 2.

11. In the **Rows Fetched Per Block** field, type the number of rows to be fetched per block.
12. In the **Default String Column Length** field, type the maximum data length for STRING columns.
13. In the **Binary Column Length** field, type the maximum data length for BINARY columns.
14. In the **Decimal Column Scale** field, type the maximum number of digits to the right of the decimal point for numeric data types.
15. In the **Socket Timeout** field, type the number of seconds that an operation can remain idle before it is closed.

Note:

This option is applicable only when asynchronous query execution is being used against Hive Server 2 instances.

16. To save your settings and close the Advanced Options dialog box, click **OK**.

Configuring HTTP Options

You can configure options such as custom headers when using the HTTP transport protocol in the Thrift layer.

To configure HTTP options:

1. Choose one:
 - If you are configuring HTTP for a DSN, open the ODBC Data Source Administrator where you created the DSN, then select the DSN, then click **Configure**, and then make sure that the Thrift Transport option is set to **HTTP**.
 - Or, if you are configuring HTTP for a DSN-less connection, open the Cloudera Hive ODBC Driver Configuration tool and then make sure that the Thrift Transport option is set to **HTTP**.
2. To access HTTP options, click **HTTP Options**.

Note:

The HTTP options are available only when the Thrift Transport option is set to **HTTP**.

3. In the **HTTP Path** field, type the partial URL corresponding to the Hive server.
4. To create a custom HTTP header, click **Add**, then type appropriate values in the **Key** and **Value** fields, and then click **OK**.
5. To edit a custom HTTP header, select the header from the list, then click **Edit**, then update the **Key** and **Value** fields as needed, and then click **OK**.
6. To delete a custom HTTP header, select the header from the list, and then click **Remove**. In the confirmation dialog box, click **Yes**.
7. To save your settings and close the HTTP Options dialog box, click **OK**.

Configuring SSL Verification

If you are connecting to a Hive server that has Secure Sockets Layer (SSL) enabled, you can configure verification between the client and the Hive server over SSL.

To configure SSL verification:

1. Choose one:
 - To access SSL options for a DSN, open the ODBC Data Source Administrator where you created the DSN, then select the DSN, then click **Configure**, and then click **SSL Options**.
 - Or, to access advanced options for a DSN-less connection, open the Cloudera Hive ODBC Driver Configuration tool, and then click **SSL Options**.
2. Select the **Enable SSL** check box.
3. To allow self-signed certificates from the server, select the **Allow Self-signed Server Certificate** check box.
4. To allow the common name of a CA-issued SSL certificate to not match the host name of the Hive server, select the **Allow Common Name Host Name Mismatch** check box.
5. Choose one:
 - To configure the driver to load SSL certificates from a specific PEM file when verifying the server, specify the full path to the file in the **Trusted Certificates** field.
 - Or, to use the trusted CA certificates PEM file that is installed with the driver, leave the **Trusted Certificates** field empty.
6. To configure two-way SSL verification, select the **Two Way SSL** check box and then do the following:
 - a. In the **Client Certificate File** field, specify the full path of the PEM file containing the client's certificate.
 - b. In the **Client Private Key File** field, specify the full path of the file containing the client's private key.
 - c. If the private key file is protected with a password, type the password in the **Client Private Key Password** field. To save the password, select the **Save Password (Encrypted)** check box.

Important:

The password is obscured, that is, not saved in plain text. However, it is still possible for the encrypted password to be copied and used.

7. To save your settings and close the SSL Options dialog box, click **OK**.

Configuring the Temporary Table Feature

You can configure the driver to create temporary tables. For more information about this feature, including details about the statement syntax used for temporary tables, see "Temporary Tables" on page 58.

Important:

When connecting to Hive 0.14 or later, the Temporary Tables feature is always enabled and you do not need to configure it in the driver.

To configure the Temporary Table feature:

1. Choose one:
 - To configure the temporary table feature for a DSN, open the ODBC Data Source Administrator where you created the DSN, then select the DSN and click **Configure**, then click **Advanced Options**, and then click **Temporary Table Configuration**.
 - Or, to configure the temporary table feature for a DSN-less connection, open the Cloudera Hive ODBC Driver Configuration tool, then click **Advanced Options**, and then click **Temporary Table Configuration**.
2. To enable the Temporary Table feature, select the **Enable Temporary Table** check box.
3. In the **Web HDFS Host** field, type the host name or IP address of the machine hosting both the namenode of your Hadoop cluster and the WebHDFS service. If this field is left blank, then the host name of the Hive server is used.
4. In the **Web HDFS Port** field, type the WebHDFS port for the namenode.
5. In the **HDFS User** field, type the name of the HDFS user that the driver uses to create the necessary files for supporting the Temporary Table feature.
6. In the **Data File HDFS Dir** field, type the HDFS directory that the driver uses to store the necessary files for supporting the Temporary Table feature.

Note:

Due to a known issue in Hive (see <https://issues.apache.org/jira/browse/HIVE-4554>), HDFS paths with space characters do not work with versions of Hive prior to 0.12.0.

7. In the **Temp Table TTL** field, type the number of minutes that a temporary table is guaranteed to exist in Hive after it is created.
8. To save your settings and close the Temporary Table Configuration dialog box, click **OK**.

Configuring Server-Side Properties

You can use the driver to apply configuration properties to the Hive server.

To configure server-side properties:

1. Choose one:
 - To configure server-side properties for a DSN, open the ODBC Data Source Administrator where you created the DSN, then select the DSN, then click **Configure**, then click **Advanced Options**, and then click **Server Side Properties**.
 - Or, to configure server-side properties for a DSN-less connection, open the Cloudera Hive ODBC Driver Configuration tool, then click **Advanced Options**, and then click **Server Side Properties**.
2. To create a server-side property, click **Add**, then type appropriate values in the **Key** and **Value** fields, and then click **OK**. For example, to set the value of the `mapreduce.job.queuename` property to `myQueue`, type `mapreduce.job.queuename` in the **Key** field and then type `myQueue` in the **Value** field.

Note:

For a list of all Hadoop and Hive server-side properties that your implementation supports, type `set -v` at the Hive CLI command line or Beeline. You can also execute the `set -v` query after connecting using the driver.

3. To edit a server-side property, select the property from the list, then click **Edit**, then update the **Key** and **Value** fields as needed, and then click **OK**.
4. To delete a server-side property, select the property from the list, and then click **Remove**. In the confirmation dialog box, click **Yes**.
5. To change the method that the driver uses to apply server-side properties, do one of the following:
 - To configure the driver to apply each server-side property by executing a query when opening a session to the Hive server, select the **Apply Server Side Properties With Queries** check box.
 - Or, to configure the driver to use a more efficient method for applying server-side properties that does not involve additional network round-tripping, clear the **Apply Server Side Properties With Queries** check box.

Note:

The more efficient method is not available for Hive Server 1, and it might not be compatible with some Hive Server 2 builds. If the server-side properties do not take effect when the check box is clear, then select the check box.

6. To configure the driver to convert server-side property key names to all lower-case characters, select the **Convert Key Name To Lower Case** check box.
7. To save your settings and close the Server Side Properties dialog box, click **OK**.

Configuring Logging Options

To help troubleshoot issues, you can enable logging. In addition to functionality provided in the Cloudera ODBC Driver for Apache Hive, the ODBC Data Source Administrator provides tracing functionality.

Important:

Only enable logging or tracing long enough to capture an issue. Logging or tracing decreases performance and can consume a large quantity of disk space.

The driver allows you to set the amount of detail included in log files. The following table lists the logging levels provided by the Cloudera ODBC Driver for Apache Hive, in order from least verbose to most verbose.

Logging Level	Description
OFF	Disables all logging.
FATAL	Logs severe error events that lead the driver to abort.
ERROR	Logs error events that might allow the driver to continue running.
WARNING	Logs potentially harmful situations.
INFO	Logs general information that describes the progress of the driver.
DEBUG	Logs detailed information that is useful for debugging the driver.
TRACE	Logs all driver activity.

To enable driver logging:

1. To access logging options, open the ODBC Data Source Administrator where you created the DSN, then select the DSN, then click **Configure**, and then click **Logging Options**.
2. From the **Log Level** drop-down list, select the desired level of information to include in log files.
3. In the **Log Path** field, specify the full path to the folder where you want to save log files.
4. If requested by Technical Support, type the name of the component for which to log messages in the **Log Namespace** field. Otherwise, do not type a value in the field.
5. In the **Max Number Files** field, type the maximum number of log files to keep.

Note:

After the maximum number of log files is reached, each time an additional file is created, the driver deletes the oldest log file.

6. In the **Max File Size** field, type the maximum size of each log file in megabytes (MB).

Note:

After the maximum file size is reached, the driver creates a new file and continues logging.

7. Click **OK**.
8. Restart your ODBC application to make sure that the new settings take effect.

The Cloudera ODBC Driver for Apache Hive produces a log file named `HiveODBC_driver.log` at the location that you specify in the Log Path field.

To disable driver logging:

1. Open the ODBC Data Source Administrator where you created the DSN, then select the DSN, then click **Configure**, and then click **Logging Options**.
2. From the **Log Level** drop-down list, select **LOG_OFF**.
3. Click **OK**.

To start tracing using the ODBC Data Source Administrator:

1. In the ODBC Data Source Administrator, click the **Tracing** tab.
2. In the **Log File Path** area, click **Browse**. In the Select ODBC Log File dialog box, browse to the location where you want to save the log file, then type a descriptive file name in the **File Name** field, and then click **Save**.
3. On the Tracing tab, click **Start Tracing Now**.

To stop ODBC Data Source Administrator tracing:

- On the Tracing tab in the ODBC Data Source Administrator, click **Stop Tracing Now**.

For more information about tracing using the ODBC Data Source Administrator, see "How to Generate an ODBC Trace with ODBC Data Source Administrator" on the Microsoft Support website: <http://support.microsoft.com/kb/274551>.

Configuring Kerberos Authentication for Windows

Active Directory

The Cloudera ODBC Driver for Apache Hive supports Active Directory Kerberos on Windows. There are two prerequisites for using Active Directory Kerberos on Windows:

- MIT Kerberos is not installed on the client Windows machine.
- The MIT Kerberos Hadoop realm has been configured to trust the Active Directory realm so that users in the Active Directory realm can access services in the MIT Kerberos Hadoop realm.

MIT Kerberos

Downloading and Installing MIT Kerberos for Windows 4.0.1

For information about Kerberos and download links for the installer, see the MIT Kerberos website: <http://web.mit.edu/kerberos/>.

To download and install MIT Kerberos for Windows 4.0.1:

1. Download the appropriate Kerberos installer:
 - For a 64-bit machine, use the following download link from the MIT Kerberos website: <http://web.mit.edu/kerberos/dist/kfw/4.0/kfw-4.0.1-amd64.msi>.
 - For a 32-bit machine, use the following download link from the MIT Kerberos website: <http://web.mit.edu/kerberos/dist/kfw/4.0/kfw-4.0.1-i386.msi>.

Note:

The 64-bit installer includes both 32-bit and 64-bit libraries. The 32-bit installer includes 32-bit libraries only.

2. To run the installer, double-click the .msi file that you downloaded above.
3. Follow the instructions in the installer to complete the installation process.
4. When the installation completes, click **Finish**.

Setting Up the Kerberos Configuration File

Settings for Kerberos are specified through a configuration file. You can set up the configuration file as an .ini file in the default location, which is the C:\ProgramData\MIT\Kerberos5 directory, or as a .conf file in a custom location.

Normally, the C:\ProgramData\MIT\Kerberos5 directory is hidden. For information about viewing and using this hidden directory, refer to Microsoft Windows documentation.

Note:

For more information on configuring Kerberos, refer to the MIT Kerberos documentation.

To set up the Kerberos configuration file in the default location:

1. Obtain a krb5.conf configuration file. You can obtain this file from your Kerberos administrator, or from the /etc/krb5.conf folder on the machine that is hosting the Hive Server 2 instance.
2. Rename the configuration file from krb5.conf to krb5.ini.
3. Copy the krb5.ini file to the C:\ProgramData\MIT\Kerberos5 directory and overwrite the empty sample file.

To set up the Kerberos configuration file in a custom location:

1. Obtain a `krb5.conf` configuration file. You can obtain this file from your Kerberos administrator, or from the `/etc/krb5.conf` folder on the machine that is hosting the Hive Server 2 instance.
2. Place the `krb5.conf` file in an accessible directory and make note of the full path name.
3. Open the System window:
 - If you are using Windows 7 or earlier, click **Start** , then right-click **Computer**, and then click **Properties**.
 - Or, if you are using Windows 8 or later, right-click **This PC** on the Start screen, and then click **Properties**.
4. Click **Advanced System Settings**.
5. In the System Properties dialog box, click the **Advanced** tab and then click **Environment Variables**.
6. In the Environment Variables dialog box, under the System Variables list, click **New**.
7. In the New System Variable dialog box, in the **Variable Name** field, type **KRB5_CONFIG**.
8. In the **Variable Value** field, type the full path to the `krb5.conf` file.
9. Click **OK** to save the new variable.
10. Make sure that the variable is listed in the System Variables list.
11. Click **OK** to close the Environment Variables dialog box, and then click **OK** to close the System Properties dialog box.

Setting Up the Kerberos Credential Cache File

Kerberos uses a credential cache to store and manage credentials.

To set up the Kerberos credential cache file:

1. Create a directory where you want to save the Kerberos credential cache file. For example, create a directory named `C:\temp`.
2. Open the System window:
 - If you are using Windows 7 or earlier, click **Start** , then right-click **Computer**, and then click **Properties**.
 - Or, if you are using Windows 8 or later, right-click **This PC** on the Start screen, and then click **Properties**.
3. Click **Advanced System Settings**.
4. In the System Properties dialog box, click the **Advanced** tab and then click **Environment Variables**.
5. In the Environment Variables dialog box, under the System Variables list, click **New**.
6. In the New System Variable dialog box, in the **Variable Name** field, type **KRB5CCNAME**.

7. In the **Variable Value** field, type the path to the folder you created above, and then append the file name krb5cache. For example, if you created the folder C:\temp, then type C:\temp\krb5cache.

Note:

krb5cache is a file (not a directory) that is managed by the Kerberos software, and it should not be created by the user. If you receive a permission error when you first use Kerberos, make sure that the krb5cache file does not already exist as a file or a directory.

8. Click **OK** to save the new variable.
9. Make sure that the variable appears in the System Variables list.
10. Click **OK** to close the Environment Variables dialog box, and then click **OK** to close the System Properties dialog box.
11. To make sure that Kerberos uses the new settings, restart your machine.

Obtaining a Ticket for a Kerberos Principal

A principal refers to a user or service that can authenticate to Kerberos. To authenticate to Kerberos, a principal must obtain a ticket by using a password or a keytab file. You can specify a keytab file to use, or use the default keytab file of your Kerberos configuration.

To obtain a ticket for a Kerberos principal using a password:

1. Open MIT Kerberos Ticket Manager.
2. In MIT Kerberos Ticket Manager, click **Get Ticket**.
3. In the Get Ticket dialog box, type your principal name and password, and then click **OK**.

If the authentication succeeds, then your ticket information appears in MIT Kerberos Ticket Manager.

To obtain a ticket for a Kerberos principal using a keytab file:

1. Open a command prompt:
 - If you are using Windows 7 or earlier, click **Start** , then click **All Programs**, then click **Accessories**, and then click **Command Prompt**.
 - If you are using Windows 8 or later, click the arrow button at the bottom of the Start screen, then find the Windows System program group, and then click **Command Prompt**.
2. In the Command Prompt, type a command using the following syntax:

```
kinit -k -t [KeytabPath] [Principal]
```

[KeytabPath] is the full path to the keytab file. For example:
C:\mykeytabs\myUser.keytab.

[Principal] is the Kerberos user principal to use for authentication. For example:
myUser@EXAMPLE.COM.

3. If the cache location KRB5CCNAME is not set or used, then use the **-c** option of the kinit command to specify the location of the credential cache. In the command, the **-c** argument must appear last. For example:

```
kinit -k -t C:\mykeytabs\myUser.keytab myUser@EXAMPLE.COM  
-c C:\ProgramData\MIT\krbcache
```

Krbcache is the Kerberos cache file, not a directory.

To obtain a ticket for a Kerberos principal using the default keytab file:

Note:

For information about configuring a default keytab file for your Kerberos configuration, refer to the MIT Kerberos documentation.

1. Open a command prompt:
 - If you are using Windows 7 or earlier, click **Start** , then click **All Programs**, then click **Accessories**, and then click **Command Prompt**.
 - If you are using Windows 8 or later, click the arrow button at the bottom of the Start screen, then find the Windows System program group, and then click **Command Prompt**.
2. In the Command Prompt, type a command using the following syntax:

```
kinit -k [principal]
```

[principal] is the Kerberos user principal to use for authentication. For example:
MyUser@EXAMPLE.COM.

3. If the cache location KRB5CCNAME is not set or used, then use the **-c** option of the kinit command to specify the location of the credential cache. In the command, the **-c** argument must appear last. For example:

```
kinit -k -t C:\mykeytabs\myUser.keytab myUser@EXAMPLE.COM  
-c C:\ProgramData\MIT\krbcache
```

Krbcache is the Kerberos cache file, not a directory.

Verifying the Version Number

If you need to verify the version of the Cloudera ODBC Driver for Apache Hive that is installed on your Windows machine, you can find the version number in the ODBC Data Source Administrator.

To verify the version number:

1. Open the ODBC Administrator:
 - If you are using Windows 7 or earlier, click **Start** , then click **All Programs**, then click the **Cloudera ODBC Driver for Apache Hive 2.5** program group corresponding to the bitness of the client application accessing data in Hive, and then click **ODBC Administrator**.
 - Or, if you are using Windows 8 or later, on the Start screen, type **ODBC administrator**, and then click the **ODBC Administrator** search result corresponding to the bitness of the client application accessing data in Hive.
2. Click the **Drivers** tab and then find the Cloudera ODBC Driver for Apache Hive in the list of ODBC drivers that are installed on your system. The version number is displayed in the **Version** column.

Linux Driver

Linux System Requirements

You install the Cloudera ODBC Driver for Apache Hive on client machines that access data stored in a Hadoop cluster with the Hive service installed and running. Each machine that you install the driver on must meet the following minimum system requirements:

- One of the following distributions:
 - Red Hat® Enterprise Linux® (RHEL) 5, 6, or 7
 - CentOS 5, 6, or 7
 - SUSE Linux Enterprise Server (SLES) 11 or 12
- 150 MB of available disk space
- One of the following ODBC driver managers installed:
 - iODBC 3.52.7 or later
 - unixODBC 2.2.14 or later

The driver supports Apache Hive versions 0.11 through 1.2.

Installing the Driver

There are two versions of the driver for Linux:

- ClouderaHiveODBC-32bit-[Version]-[Release].[LinuxDistro].i686.rpm for the 32-bit driver
- ClouderaHiveODBC-[Version]-[Release].[LinuxDistro].x86_64.rpm for the 64-bit driver

[Version] is the version number of the driver, and [Release] is the release number for this version of the driver.

The bitness of the driver that you select should match the bitness of the client application accessing your Hadoop / Hive-based data. For example, if the client application is 64-bit, then you should install the 64-bit driver. Note that 64-bit editions of Linux support both 32- and 64-bit applications. Verify the bitness of your intended application and install the appropriate version of the driver.

Important:

Make sure that you install the driver using the RPM corresponding to your Linux distribution.

The Cloudera ODBC Driver for Apache Hive files are installed in the following directories:

- /opt/cloudera/hiveodbc contains release notes, the *Cloudera ODBC Driver for Apache Hive Installation and Configuration Guide* in PDF format, and a Readme.txt file that provides plain text installation and configuration instructions.

- /opt/cloudera/hiveodbc/ErrorMessages contains error message files required by the driver.
- /opt/cloudera/hiveodbc/Setup contains sample configuration files named odbc-c.ini and odbcinst.ini.
- /opt/cloudera/hiveodbc/lib/32 contains the 32-bit shared libraries and the cloudera.hiveodbc.ini configuration file.
- /opt/cloudera/hiveodbc/lib/64 contains the 64-bit shared libraries and the cloudera.hiveodbc.ini configuration file.

To install the Cloudera ODBC Driver for Apache Hive:

1. Choose one:

- In Red Hat Enterprise Linux or CentOS, log in as the root user, then navigate to the folder containing the driver RPM packages to install, and then type the following at the command line, where [RPMFileName] is the file name of the RPM package containing the version of the driver that you want to install:

```
yum --nogpgcheck localinstall [RPMFileName]
```

- Or, in SUSE Linux Enterprise Server, log in as the root user, then navigate to the folder containing the driver RPM packages to install, and then type the following at the command line, where [RPMFileName] is the file name of the RPM package containing the version of the driver that you want to install:

```
zypper install [RPMFileName]
```

The Cloudera ODBC Driver for Apache Hive depends on the following resources:

- cyrus-sasl-2.1.22-7 or later
- cyrus-sasl-gssapi-2.1.22-7 or later
- cyrus-sasl-plain-2.1.22-7 or later

If the package manager in your Linux distribution cannot resolve the dependencies automatically when installing the driver, then download and manually install the packages required by the version of the driver that you want to install.

Setting the LD_LIBRARY_PATH Environment Variable on Linux

The LD_LIBRARY_PATH environment variable must include the paths to the installed ODBC driver manager libraries.

For example, if ODBC driver manager libraries are installed in /usr/local/lib, then set LD_LIBRARY_PATH as follows:

```
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/usr/local/lib
```

For information about how to set environment variables permanently, refer to your Linux shell documentation.

For information about creating ODBC connections using the Cloudera ODBC Driver for Apache Hive, see "Configuring ODBC Connections for Non-Windows Platforms" on page 36.

Verifying the Version Number

If you need to verify the version of the Cloudera ODBC Driver for Apache Hive that is installed on your Linux machine, you can query the version number through the command-line interface if the driver was installed using an RPM file.

To verify the version number:

- Depending on your package manager, at the command prompt, run one of the following commands:
 - `yum list | grep ClouderaHiveODBC`
 - `rpm -qa | grep ClouderaHiveODBC`

The command returns information about the Cloudera ODBC Driver for Apache Hive that is installed on your machine, including the version number.

Mac OS X Driver

Installing the Driver on Mac OS X

The Cloudera ODBC Driver for Apache Hive supports both 32- and 64-bit client applications.

You install the Cloudera ODBC Driver for Apache Hive on client machines that access data stored in a Hadoop cluster with the Hive service installed and running. Each machine that you install the driver on must meet the following minimum system requirements:

- Mac OS X version 10.9, 10.10, or 10.11
- 100 MB of available disk space
- iODBC 3.52.7 or later

The driver supports Apache Hive versions 0.11 through 1.2.

The Cloudera ODBC Driver for Apache Hive files are installed in the following directories:

- /opt/cloudera/hiveodbc contains release notes and the *Cloudera ODBC Driver for Apache Hive Installation and Configuration Guide* in PDF format.
- /opt/cloudera/hiveodbc/ErrorMessages contains error message files required by the driver.
- /opt/cloudera/hiveodbc/Setup contains sample configuration files named odbc.ini and odbcinst.ini.
- /opt/cloudera/hiveodbc/lib contains the driver binaries and the cloudera.hiveodbc.ini configuration file.

To install the Cloudera ODBC Driver for Apache Hive:

1. Double-click **ClouderaHiveODBC.dmg** to mount the disk image.
2. Double-click **ClouderaHiveODBC.pkg** to run the installer.
3. In the installer, click **Continue**.
4. On the Software License Agreement screen, click **Continue**, and when the prompt appears, click **Agree** if you agree to the terms of the License Agreement.
5. Optionally, to change the installation location, click **Change Install Location**, then select the desired location, and then click **Continue**.
6. To accept the installation location and begin the installation, click **Install**.
7. When the installation completes, click **Close**.

Verifying the Version Number

If you need to verify the version of the Cloudera ODBC Driver for Apache Hive that is installed on your Mac OS X machine, you can query the version number through the Terminal.

To verify the version number:

- At the Terminal, run the following command:

```
pkgutil --info com.cloudera.hiveodbc
```

The command returns information about the Cloudera ODBC Driver for Apache Hive that is installed on your machine, including the version number.

Setting the DYLD_LIBRARY_PATH Environment Variable

The DYLD_LIBRARY_PATH environment variable must include the paths to the installed ODBC driver manager libraries.

For example, if ODBC driver manager libraries are installed in /usr/local/lib, then set DYLD_LIBRARY_PATH as follows:

```
export DYLD_LIBRARY_PATH=$DYLD_LIBRARY_PATH:/usr/local/lib
```

For information about how to set environment variables permanently, refer to your Mac OS X shell documentation.

For information about creating ODBC connections using the Cloudera ODBC Driver for Apache Hive, see "Configuring ODBC Connections for Non-Windows Platforms" on page 36.

AIX Driver

Installing the Driver on AIX

There are two versions of the driver for AIX:

- ClouderaHiveODBC-32bit-[Version]-[Release].ppc.rpm for the 32-bit driver
- ClouderaHiveODBC-[Version]-[Release].ppc.rpm for the 64-bit driver

[Version] is the version number of the driver, and [Release] is the release number for this version of the driver.

The bitness of the driver that you select should match the bitness of the client application accessing your Hadoop / Hive-based data. For example, if the client application is 64-bit, then you should install the 64-bit driver. Note that 64-bit editions of AIX support both 32- and 64-bit applications. Verify the bitness of your intended application and install the appropriate version of the driver.

You install the Cloudera ODBC Driver for Apache Hive on client machines that access data stored in a Hadoop cluster with the Hive service installed and running. Each machine that you install the driver on must meet the following minimum system requirements:

- IBM AIX 5.3, 6.1, or 7.1 (32- and 64-bit editions are supported)
- 150 MB of available disk space
- One of the following ODBC driver managers installed:
 - iODBC 3.52.7 or later
 - unixODBC 2.3.0 or later

The driver supports Apache Hive versions 0.11 through 1.2.

The Cloudera ODBC Driver for Apache Hive driver files are installed in the following directories:

- /opt/cloudera/hiveodbc contains release notes, the *Cloudera ODBC Driver for Apache Hive Installation and Configuration Guide* in PDF format, and a Readme.txt file that provides plain text installation and configuration instructions.
- /opt/cloudera/hiveodbc/ErrorMessages contains error message files required by the driver.
- /opt/cloudera/hiveodbc/Setup contains sample configuration files named odbc.ini and odbcinst.ini.
- /opt/cloudera/hiveodbc/lib/32 contains the 32-bit driver and the cloudera.hiveodbc.ini configuration file.
- /opt/cloudera/hiveodbc/lib/64 contains the 64-bit driver and the cloudera.hiveodbc.ini configuration file.

To install the Cloudera ODBC Driver for Apache Hive:

1. Log in as the root user, then navigate to the folder containing the driver RPM packages to install, and then type the following at the command line, where *[RPMFileName]* is the file name of the RPM package containing the version of the driver that you want to install:

```
rpm --install [RPMFileName]
```

Setting the LD_LIBRARY_PATH Environment Variable on AIX

The LD_LIBRARY_PATH environment variable must include the path to the installed ODBC driver manager libraries.

For example, if ODBC driver manager libraries are installed in /usr/local/lib, then set LD_LIBRARY_PATH as follows:

```
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/usr/local/lib
```

For information about how to set environment variables permanently, refer to your AIX shell documentation.

For information about creating ODBC connections using the Cloudera ODBC Driver for Apache Hive, see "Configuring ODBC Connections for Non-Windows Platforms" on page 36.

Verifying the Version Number

If you need to verify the version of the Cloudera ODBC Driver for Apache Hive that is installed on your AIX machine, you can query the version number through the command-line interface.

To verify the version number:

- At the command prompt, run the following command:

```
rpm -qa | grep ClouderaHiveODBC
```

The command returns information about the Cloudera ODBC Driver for Apache Hive that is installed on your machine, including the version number.

Debian Driver

Installing the Driver on Debian

You install the Cloudera ODBC Driver for Apache Hive on client machines that access data stored in a Hadoop cluster with the Hive service installed and running. Each machine that you install the driver on must meet the following minimum system requirements:

- Debian 6 or 7 (Ubuntu 12.04 LTS and Ubuntu 14.04 LTS)
- 150 MB of available disk space
- One of the following ODBC driver managers installed:
 - iODBC 3.52.7 or later
 - unixODBC 2.2.14 or later

The driver supports Apache Hive versions 0.11 through 1.2. It supports both 32- and 64-bit client applications.

There are two versions of the driver for Debian:

- `ClouderaHiveODBC-32bit-[Version]-[Release]_i386.deb` for the 32-bit driver
- `ClouderaHiveODBC-[Version]-[Release]_amd64.deb` for the 64-bit driver

`[Version]` is the version number of the driver, and `[Release]` is the release number for this version of the driver.

The bitness of the driver that you select should match the bitness of the client application accessing your Hadoop / Hive-based data. For example, if the client application is 64-bit, then you should install the 64-bit driver. Note that 64-bit editions of Debian support both 32- and 64-bit applications. Verify the bitness of your intended application and install the appropriate version of the driver.

The Cloudera ODBC Driver for Apache Hive driver files are installed in the following directories:

- `/opt/cloudera/hiveodbc` contains release notes, the *Cloudera ODBC Driver for Apache Hive Installation and Configuration Guide* in PDF format, and a `Readme.txt` file that provides plain text installation and configuration instructions.
- `/opt/cloudera/hiveodbc/ErrorMessages` contains error message files required by the driver.
- `/opt/cloudera/hiveodbc/Setup` contains sample configuration files named `odb-c.ini` and `odbcinst.ini`.
- `/opt/cloudera/hiveodbc/lib/32` contains the 32-bit shared libraries and the `cloudera.hiveodbc.ini` configuration file.
- `/opt/cloudera/hiveodbc/lib/64` contains the 64-bit shared libraries and the `cloudera.hiveodbc.ini` configuration file.

To install the Cloudera ODBC Driver for Apache Hive:

1. In Ubuntu, log in as the root user, then navigate to the folder containing the driver Debian packages to install, and double-click **ClouderaHiveODBC-32bit-Version-Release_i386.deb** or **ClouderaHiveODBC-Version-Release_amd64.deb**.
2. Follow the instructions in the installer to complete the installation process.

The Cloudera ODBC Driver for Apache Hive depends on the following resources:

- cyrus-sasl-2.1.22-7 or above
- cyrus-sasl-gssapi-2.1.22-7 or above
- cyrus-sasl-plain-2.1.22-7 or above

If the package manager in your Ubuntu distribution cannot resolve the dependencies automatically when installing the driver, then download and manually install the packages required by the version of the driver that you want to install.

Setting the LD_LIBRARY_PATH Environment Variable on Debian

The LD_LIBRARY_PATH environment variable must include the path to the installed ODBC driver manager libraries.

For example, if ODBC driver manager libraries are installed in /usr/local/lib, then set LD_LIBRARY_PATH as follows:

```
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/usr/local/lib
```

For information about how to set environment variables permanently, refer to your Ubuntu shell documentation.

For information about creating ODBC connections using the Cloudera ODBC Driver for Apache Hive, see "Configuring ODBC Connections for Non-Windows Platforms" on page 36.

Verifying the Version Number

If you need to verify the version of the Cloudera ODBC Driver for Apache Hive that is installed on your Debian machine, you can query the version number through the command-line interface.

To verify the version number:

- At the command prompt, run the following command:

```
dpkg -l | grep ClouderaHiveODBC
```

The command returns information about the Cloudera ODBC Driver for Apache Hive that is installed on your machine, including the version number.

Configuring ODBC Connections for Non-Windows Platforms

The following sections describe how to configure ODBC connections when using the Cloudera ODBC Driver for Apache Hive with non-Windows platforms:

- "Configuration Files" on page 36
- "Sample Configuration Files" on page 37
- "Configuring the Environment" on page 37
- "Defining DSNs in odbc.ini" on page 38
- "Specifying ODBC Drivers in odbcinst.ini" on page 39
- "Configuring Driver Settings in cloudera.hiveodbc.ini" on page 41
- "Configuring Service Discovery Mode" on page 41
- "Configuring Authentication" on page 41
- "Configuring SSL Verification" on page 43
- "Configuring Server-Side Properties" on page 44
- "Configuring Logging Options" on page 45
- "Testing the Connection" on page 47

Configuration Files

ODBC driver managers use configuration files to define and configure ODBC data sources and drivers. By default, the following configuration files are used:

- `.odbc.ini` is used to define ODBC data sources, and it is required for DSNs.
- `.odbcinst.ini` is used to define ODBC drivers, and it is optional.

These files are located in the user's home directory.

Also, by default the Cloudera ODBC Driver for Apache Hive is configured using the `cloudera.hiveodbc.ini` file. This file is located in one of the following directories depending on the version of the driver that you are using:

- `/opt/cloudera/hiveodbc/lib/32` for the 32-bit driver on Linux/AIX/Debian.
- `/opt/cloudera/hiveodbc/lib/64` for the 64-bit driver on Linux/AIX/Debian.
- `/opt/cloudera/hiveodbc/lib` for the driver on Mac OS X.

The `cloudera.hiveodbc.ini` file is required.

Note:

The `cloudera.hiveodbc.ini` file provides default settings for most configuration options available in the Cloudera ODBC Driver for Apache Hive.

You can set driver configuration options in your `odbc.ini` and `cloudera.hiveodbc.ini` files. Configuration options set in a `cloudera.hiveodbc.ini` file apply to all connections, whereas configuration options set in an `odbc.ini` file are specific to a connection. Configuration

options set in `odbc.ini` take precedence over configuration options set in `cloudera.hiveodbc.ini`. For information about the configuration options available for controlling the behavior of DSNs that are using the Cloudera ODBC Driver for Apache Hive, see "Driver Configuration Options" on page 62.

Sample Configuration Files

The driver installation contains the following sample configuration files in the `Setup` directory:

- `odbc.ini`
- `odbcinst.ini`

These sample configuration files provide preset values for settings related to the Cloudera ODBC Driver for Apache Hive.

The names of the sample configuration files do not begin with a period (.) so that they appear in directory listings by default. A file name beginning with a period (.) is hidden. For `odbc.ini` and `odbcinst.ini`, if the default location is used, then the file names must begin with a period (.).

If the configuration files do not exist in the home directory, then you can copy the sample configuration files to the home directory, and then rename the files. If the configuration files already exist in the home directory, then use the sample configuration files as a guide to modify the existing configuration files.

Configuring the Environment

Optionally, you can use three environment variables, `ODBCINI`, `ODBCSYSINI`, and `CLOUDERAHIVEODBCINI`, to specify different locations for the `odbc.ini`, `odbcinst.ini`, and `cloudera.hiveodbc.ini` configuration files by doing the following:

- Set `ODBCINI` to point to your `odbc.ini` file.
- Set `ODBCSYSINI` to point to the directory containing the `odbcinst.ini` file.
- Set `CLOUDERAHIVEODBCINI` to point to your `cloudera.hiveodbc.ini` file.

For example, if your `odbc.ini` and `cloudera.hiveodbc.ini` files are located in `/etc` and your `odbcinst.ini` file is located in `/usr/local/odbc`, then set the environment variables as follows:

```
export ODBCINI=/etc/odbc.ini
export ODBCSYSINI=/usr/local/odbc
export CLOUDERAHIVEODBCINI=/etc/cloudera.hiveodbc.ini
```

The following search order is used to locate the `cloudera.hiveodbc.ini` file:

1. If the `CLOUDERAHIVEODBCINI` environment variable is defined, then the driver searches for the file specified by the environment variable.

Note:

CLOUDERAHIVEODBCINI must specify the full path, including the file name.

2. The directory containing the driver's binary is searched for a file named `cloudera-hiveodbc.ini` (not beginning with a period).
3. The current working directory of the application is searched for a file named `cloudera-hiveodbc.ini` (not beginning with a period).
4. The directory `~/`, that is, `$HOME`, is searched for a hidden file named `cloudera-hiveodbc.ini` (beginning with a period).
5. The directory `/etc` is searched for a file named `cloudera.hiveodbc.ini` (not beginning with a period).

Defining DSNs in `odbc.ini`

Note:

If you are using a DSN-less connection, then you do not need to configure the `odbc.ini` file. For information about configuring a DSN-less connection, see "Using a Connection String" on page 52.

ODBC Data Source Names (DSNs) are defined in the `odbc.ini` configuration file. This file is divided into several sections:

- `[ODBC]` is optional. This section is used to control global ODBC configuration, such as ODBC tracing.
- `[ODBC Data Sources]` is required. This section lists the DSNs and associates them with a driver.
- A section having the same name as the data source specified in the `[ODBC Data Sources]` section is required to configure the data source.

The following is an example of an `odbc.ini` configuration file for Linux/AIX/Debian:

```
[ODBC Data Sources]
Cloudera Hive DSN 32=Cloudera Hive ODBC Driver 32-bit

[Cloudera Hive DSN 32]
Driver=/opt/cloudera/hiveodbc/lib/32/libclouderahiveodbc32.so
HOST=[MyHiveServer]
PORT=10000
```

`[MyHiveServer]` is the IP address or host name of the Hive server.

The following is an example of an `odbc.ini` configuration file for Mac OS X:

[ODBC Data Sources]

Cloudera Hive DSN=Cloudera Hive ODBC Driver

[Cloudera Hive DSN]

Driver=/opt/cloudera/hiveodbc/lib/universal/libclouderahiveodbc.dylib

HOST=[MyHiveServer]

PORt=10000

[MyHiveServer] is the IP address or host name of the Hive server.

To create a Data Source Name:

1. In a text editor, open the `odbc.ini` configuration file.
2. In the `[ODBC Data Sources]` section, add a new entry by typing the Data Source Name (DSN), then an equal sign (=), and then the driver name.
3. Add a new section to the file, with a section name that matches the DSN you specified above, and then add configuration options to the section. Specify the configuration options as key-value pairs.

Note:

Hive Server 1 does not support authentication. Most default configurations of Hive Server 2 require User Name authentication, which you configure by setting the `AuthMech` key to 2. To verify the authentication mechanism that you need to use for your connection, check the configuration of your Hadoop / Hive distribution. For more information, see "Authentication Mechanisms" on page 50.

4. Save the `odbc.ini` configuration file.

For information about the configuration options available for controlling the behavior of DSNs that are using the Cloudera ODBC Driver for Apache Hive, see "Driver Configuration Options" on page 62.

Specifying ODBC Drivers in `odbcinst.ini`

ODBC drivers are defined in the `odbcinst.ini` configuration file. This configuration file is optional because drivers can be specified directly in the `odbc.ini` configuration file, as described in "Defining DSNs in `odbc.ini`" on page 38.

The `odbcinst.ini` file is divided into the following sections:

- `[ODBC Drivers]` lists the names of all the installed ODBC drivers.
- For each driver, a section having the same name as the driver name specified in the `[ODBC Drivers]` section lists the driver attributes and values.

The following is an example of an `odbcinst.ini` configuration file for Linux/AIX/Debian:

```
[ODBC Drivers]
Cloudera Hive ODBC Driver 32-bit=Installed
Cloudera Hive ODBC Driver 64-bit=Installed

[Cloudera Hive ODBC Driver 32-bit]
Description=Cloudera ODBC Driver for Apache Hive (32-bit)
Driver=/opt/cloudera/hiveodbc/lib/32/libclouderahiveodbc32.so

[Cloudera Hive ODBC Driver 64-bit]
Description=Cloudera ODBC Driver for Apache Hive (64-bit)
Driver=/opt/cloudera/hiveodbc/lib/64/libclouderahiveodbc64.so
```

The following is an example of an `odbcinst.ini` configuration file for Mac OS X:

```
[ODBC Drivers]
Cloudera Hive ODBC Driver=Installed

[Cloudera Hive ODBC Driver]
Description=Cloudera ODBC Driver for Apache Hive
Driver=/opt/cloudera/hiveodbc/lib/universal/libclouderahiveodbc.dylib
```

To define a driver:

1. In a text editor, open the `odbcinst.ini` configuration file.
2. In the `[ODBC Drivers]` section, add a new entry by typing the driver name and then typing `=Installed`.

Note:

Give the driver a symbolic name that you want to use to refer to the driver in connection strings or DSNs.

3. Add a new section to the file with a name that matches the driver name you typed above, and then add configuration options to the section based on the sample `odbcinst.ini` file provided in the Setup directory. Specify the configuration options as key-value pairs.
4. Save the `odbcinst.ini` configuration file.

Configuring Driver Settings in `cloudera.hiveodbc.ini`

The `cloudera.hiveodbc.ini` file contains configuration settings for the Cloudera ODBC Driver for Apache Hive. Settings that you define in this file apply to all connections that use the driver.

You do not need to modify the settings in the `cloudera.hiveodbc.ini` file to use the driver and connect to your data source.

However, to help troubleshoot issues, you can configure the `cloudera.hiveodbc.ini` file to enable logging in the driver. For information about configuring logging, see "Configuring Logging Options" on page 45.

Configuring Service Discovery Mode

You can configure the Cloudera ODBC Driver for Apache Hive to discover Hive Server 2 services via ZooKeeper.

To enable service discovery via ZooKeeper:

1. Open the `odbc.ini` configuration file in a text editor.
2. Set the `ServiceDiscoveryMode` connection attribute to 1.
3. Set the `ZKNamespace` connection attribute to specify the namespace on ZooKeeper under which Hive Server 2 znodes are added.
4. Set the `Host` connection attribute to specify the ZooKeeper ensemble as a comma-separated list of ZooKeeper servers. For example, type the following, where `[ZK_Host]` is the IP address or host name of the ZooKeeper server and `[ZK_Port]` is the number of the TCP port that the ZooKeeper server uses to listen for client connections:

```
[ZK_Host1]:[ZK_Port1], [ZK_Host2]:[ZK_Port2]
```

Important:

When `ServiceDiscoveryMode` is set to 1, connections to Hive Server 1 are not supported and the `Port` connection attribute is not applicable.

5. Save the `odbc.ini` configuration file.

Depending on whether service discovery mode is enabled or disabled, you might need to provide different connection attributes or values in your connection string or DSN. For more information about connection attributes, see "Driver Configuration Options" on page 62.

Configuring Authentication

Some Hive servers are configured to require authentication for access. To connect to a Hive server, you must configure the Cloudera ODBC Driver for Apache Hive to use the authentication mechanism that matches the access requirements of the server and provides the necessary credentials.

For information about how to determine the type of authentication your Hive server requires, see "Authentication Mechanisms" on page 50.

You can select the type of authentication to use for a connection by defining the `AuthMech` connection attribute in a connection string or in a DSN (in the `odbc.ini` file). Depending on the authentication mechanism you use, there might be additional connection attributes that you must define. For more information about the attributes involved in configuring authentication, see "Driver Configuration Options" on page 62.

Using No Authentication

When connecting to a Hive server of type Hive Server 1, you must use No Authentication. When you use No Authentication, Binary is the only Thrift transport protocol that is supported.

To configure a connection without authentication:

1. Set the `AuthMech` connection attribute to 0.
2. If the Hive server is configured to use SSL, then configure SSL for the connection. For more information, see "Configuring SSL Verification" on page 43.

Using Kerberos

Kerberos must be installed and configured before you can use this authentication mechanism. For more information, refer to the MIT Kerberos Documentation: <http://web.mit.edu/kerberos/krb5-latest/doc/>.

This authentication mechanism is available only for Hive Server 2. When you use Kerberos authentication, the Binary transport protocol is not supported.

To configure Kerberos authentication:

1. Set the `AuthMech` connection attribute to 1.
2. Choose one:
 - To use the default realm defined in your Kerberos setup, do not set the `KrbRealm` attribute.
 - Or, if your Kerberos setup does not define a default realm or if the realm of your Hive server is not the default, then set the appropriate realm using the `KrbRealm` attribute.
3. Set the `KrbHostFQDN` attribute to the fully qualified domain name of the Hive Server 2 host.

Note:

To use the Hive server host name as the fully qualified domain name for Kerberos authentication, set `KrbHostFQDN` to `_HOST`.

4. Set the `KrbServiceName` attribute to the service name of the Hive server.
5. To allow the driver to pass your credentials directly to the server for use in authentication, set `DelegateKrbCreds` to 1.

6. Set the `ThriftTransport` connection attribute to the transport protocol to use in the Thrift layer.

Important:

When using this authentication mechanism, Binary (`ThriftTransport=0`) is not supported.

7. If the Hive server is configured to use SSL, then configure SSL for the connection. For more information, see "Configuring SSL Verification" on page 43.

Using User Name

This authentication mechanism requires a user name but does not require a password. The user name labels the session, facilitating database tracking.

This authentication mechanism is available only for Hive Server 2. Most default configurations of Hive Server 2 require User Name authentication. When you use User Name authentication, SSL is not supported and SASL is the only Thrift transport protocol available.

To configure User Name authentication:

1. Set the `AuthMech` connection attribute to 2.
2. Set the `UID` attribute to an appropriate user name for accessing the Hive server.

Using User Name And Password

This authentication mechanism requires a user name and a password.

This authentication mechanism is available only for Hive Server 2.

To configure User Name And Password authentication:

1. Set the `AuthMech` connection attribute to 3.
2. Set the `UID` attribute to an appropriate user name for accessing the Hive server.
3. Set the `PWD` attribute to the password corresponding to the user name you provided above.
4. Set the `ThriftTransport` connection attribute to the transport protocol to use in the Thrift layer.
5. If the Hive server is configured to use SSL, then configure SSL for the connection. For more information, see "Configuring SSL Verification" on page 43.

Configuring SSL Verification

If you are connecting to a Hive server that has Secure Sockets Layer (SSL) enabled, you can configure the driver to connect to an SSL-enabled socket.

To configure SSL verification:

1. Open the `odbc.ini` configuration file in a text editor.
2. To enable SSL connections, set the `SSL` attribute to `1`.
3. To allow self-signed certificates from the server, set the `AllowSelfSignedServerCert` attribute to `1`.
4. To allow the common name of a CA-issued SSL certificate to not match the host name of the Hive server, set the `CAIssuedCertNamesMismatch` attribute to `1`.
5. Choose one:
 - To configure the driver to load SSL certificates from a specific PEM file when verifying the server, set the `TrustedCerts` attribute to the full path of the PEM file.
 - Or, to use the trusted CA certificates PEM file that is installed with the driver, do not specify a value for the `TrustedCerts` attribute.
6. To configure two-way SSL verification, set the `TwoWaySSL` attribute to `1` and then do the following:
 - a. Set the `ClientCert` attribute to the full path of the PEM file containing the client's certificate.
 - b. Set the `ClientPrivateKey` attribute to the full path of the file containing the client's private key.
 - c. If the private key file is protected with a password, set the `ClientPrivateKeyPassword` attribute to the password.
7. Save the `odbc.ini` configuration file.

Configuring Server-Side Properties

You can use the driver to apply configuration properties to the Hive server. You can set these server-side properties in a DSN (in the `odbc.ini` file) or in a connection string.

Important:

If server-side properties are set in both the `odbc.ini` file and the connection string, the ones set in the connection string take precedence.

To configure server-side properties:

1. To set a server-side property, use the syntax `SSP_[SSPKey]=[SSPValue]`, where `[SSPKey]` is the name of the server-side property and `[SSPValue]` is the value to specify for that property. For example, to set the `mapreduce.job.queuename` property to `myQueue`, type the following:

```
SSP_mapreduce.job.queuename=myQueue
```

Note:

- When setting a server-side property in a connection string, it is recommended that you enclose the value in braces ({}) to make sure that special characters can be properly escaped.
- For a list of all Hadoop and Hive server-side properties that your implementation supports, type `set -v` at the Hive CLI command line or Beeline. You can also execute the `set -v` query after connecting using the driver.

2. To change the method that the driver uses to apply server-side properties, do one of the following:

- To configure the driver to apply each server-side property by executing a query when opening a session to the Hive server, set the `ApplySSPWithQueries` property to 1.
- Or, to configure the driver to use a more efficient method for applying server-side properties that does not involve additional network round-tripping, set the `ApplySSPWithQueries` property to 0.

Note:

The more efficient method is not available for Hive Server 1, and it might not be compatible with some Hive Server 2 builds. If the server-side properties do not take effect when the `ApplySSPWithQueries` property is set to 0, then set it to 1.

3. To disable the driver's default behavior of converting server-side property key names to all lower-case characters, set the `LCaseSspKeyName` property to 0.
4. Save the `odbc.ini` configuration file.

Configuring Logging Options

To help troubleshoot issues, you can enable logging in the driver.

Important:

Only enable logging long enough to capture an issue. Logging decreases performance and can consume a large quantity of disk space.

Use the `LogLevel` key to set the amount of detail included in log files. The following table lists the logging levels provided by the Cloudera ODBC Driver for Apache Hive, in order from least verbose to most verbose.

LogLevel Value	Description
0	Disables all logging.
1	Logs severe error events that lead the driver to abort.

LogLevel Value	Description
2	Logs error events that might allow the driver to continue running.
3	Logs potentially harmful situations.
4	Logs general information that describes the progress of the driver.
5	Logs detailed information that is useful for debugging the driver.
6	Logs all driver activity.

To enable logging:

1. Open the `cloudera.hiveodbc.ini` configuration file in a text editor.
2. Set the `LogLevel` key to the desired level of information to include in log files. For example:

`LogLevel=2`

3. Set the `LogPath` key to the full path to the folder where you want to save log files. For example:

`LogPath=/localhome/employee/Documents`

4. Set the `LogFileCount` key to the maximum number of log files to keep.

Note:

After the maximum number of log files is reached, each time an additional file is created, the driver deletes the oldest log file.

5. Set the `LogFileSize` key to the maximum size of each log file in megabytes (MB).

Note:

After the maximum file size is reached, the driver creates a new file and continues logging.

6. Save the `cloudera.hiveodbc.ini` configuration file.
7. Restart your ODBC application to make sure that the new settings take effect.

The Cloudera ODBC Driver for Apache Hive produces a log file named `HiveODBC_driver.log` at the location you specify using the `LogPath` key.

To disable logging:

1. Open the `cloudera.hiveodbc.ini` configuration file in a text editor.
2. Set the `LogLevel` key to 0.

3. Save the `cloudera.hiveodbc.ini` configuration file.

Testing the Connection

To test the connection, you can use an ODBC-enabled client application. For a basic connection test, you can also use the test utilities that are packaged with your driver manager installation. For example, the iODBC driver manager includes simple utilities called `iodbctest` and `iodbctestw`. Similarly, the unixODBC driver manager includes simple utilities called `isql` and `iusql`.

Using the iODBC Driver Manager

You can use the `iodbctest` and `iodbctestw` utilities to establish a test connection with your driver. Use `iodbctest` to test how your driver works with an ANSI application, or use `iodbctestw` to test how your driver works with a Unicode application.

Note:

There are 32-bit and 64-bit installations of the iODBC driver manager available. If you have only one or the other installed, then the appropriate version of `iodbctest` (or `iodbctestw`) is available. However, if you have both 32- and 64-bit versions installed, then you need to make sure that you are running the version from the correct installation directory.

For more information about using the iODBC driver manager, see <http://www.iodbc.org>.

To test your connection using the iODBC driver manager:

1. Run `iodbctest` or `iodbctestw`.
2. Optionally, if you do not remember the DSN, then type a question mark (?) to see a list of available DSNs.
3. Type an ODBC connection string using the following format, specifying additional connection attributes as needed:

`DSN=[DataSourceName];[Key]=[Value]`

`[DataSourceName]` is the DSN that you are using for the connection. `[Key]` is any connection attribute that is not already specified as a configuration key in the DSN, and `[Value]` is the value for the attribute. Add key-value pairs to the connection string as needed, separating each pair with a semicolon (;).

Or, If you are using a DSN-less connection, then type an ODBC connection string using the following format, specifying additional connection attributes as needed:

`Driver=[DriverNameOrFile];HOST=[MyHiveServer];
PORT=[PortNumber];Schema=[DefaultSchema];
HiveServerType=[ServerType]`

The placeholders in the connection string are defined as follows:

- *[DriverNameOrFile]* is either the symbolic name of the installed driver defined in the `odbcinst.ini` file or the absolute path of the shared object file for the driver. If you use the symbolic name, then you must make sure that the `odbcinst.ini` file is configured to point the symbolic name to the shared object file. For more information, see "Specifying ODBC Drivers in `odbcinst.ini`" on page 39.
- *[MyHiveServer]* is the IP address or hostname of the Hive Server.
- *[PortNumber]* is the number of the TCP port that the Hive server uses to listen for client connections.
- *[DefaultSchema]* is the database schema to use when a schema is not explicitly specified in a query.
- *[ServerType]* is either 1 (for Hive Server 1) or 2 (for Hive Server 2).

If the connection is successful, then the `SQL>` prompt appears.

Note:

For information about the connection attributes that are available, see "Driver Configuration Options" on page 62.

Using the unixODBC Driver Manager

You can use the `isql` and `iusql` utilities to establish a test connection with your driver and your DSN. `isql` and `iusql` can only be used to test connections that use a DSN. Use `isql` to test how your driver works with an ANSI application, or use `iusql` to test how your driver works with a Unicode application.

Note:

There are 32-bit and 64-bit installations of the unixODBC driver manager available. If you have only one or the other installed, then the appropriate version of `isql` (or `iusql`) is available. However, if you have both 32- and 64-bit versions installed, then you need to make sure that you are running the version from the correct installation directory.

For more information about using the unixODBC driver manager, see <http://www.unixodbc.org>.

To test your connection using the unixODBC driver manager:

- Run `isql` or `iusql` by using the corresponding syntax:
- `isql [DataSourceName]`
 - `iusql [DataSourceName]`

[DataSourceName] is the DSN that you are using for the connection.

If the connection is successful, then the `SQL>` prompt appears.

Note:

For information about the available options, run `isql` or `iusql` without providing a DSN.

Authentication Mechanisms

To connect to a Hive server, you must configure the Cloudera ODBC Driver for Apache Hive to use the authentication mechanism that matches the access requirements of the server and provides the necessary credentials. To determine the authentication settings that your Hive server requires, check the server configuration and then refer to the corresponding section below.

Hive Server 1

You must use No Authentication as the authentication mechanism. Hive Server 1 instances do not support authentication.

Hive Server 2

Note:

Most default configurations of Hive Server 2 require User Name authentication.

Configuring authentication for a connection to a Hive Server 2 instance involves setting the authentication mechanism, the Thrift transport protocol, and SSL support. To determine the settings that you need to use, check the following three properties in the `hive-site.xml` file in the Hive server that you are connecting to:

- `hive.server2.authentication`
- `hive.server2.transport.mode`
- `hive.server2.use.SSL`

Use the following table to determine the authentication mechanism that you need to configure, based on the `hive.server2.authentication` value in the `hive-site.xml` file:

<code>hive.server2.authentication</code>	Authentication Mechanism
NOSASL	No Authentication
KERBEROS	Kerberos
NONE	User Name
LDAP	User Name and Password

Use the following table to determine the Thrift transport protocol that you need to configure, based on the `hive.server2.authentication` and `hive.server2.transport.mode` values in the `hive-site.xml` file:

<code>hive.server2.authentication</code>	<code>hive.server2.transport.mode</code>	Thrift Transport Protocol
NOSASL	binary	Binary

hive.server2.authentication	hive.server2.transport.mode	Thrift Transport Protocol
KERBEROS	binary or http	SASL or HTTP
NONE	binary or http	SASL or HTTP
LDAP	binary or http	SASL or HTTP

To determine whether SSL should be enabled or disabled for your connection, check the `hive.server2.use.SSL` value in the `hive-site.xml` file. If the value is true, then you must enable and configure SSL in your connection. If the value is false, then you must disable SSL in your connection.

For detailed instructions on how to configure authentication when using the Windows driver, see "Configuring Authentication" on page 11.

For detailed instructions on how to configure authentication when using a non-Windows driver, see "Configuring Authentication" on page 41.

Using a Connection String

For some applications, you might need to use a connection string to connect to your data source. For detailed information about how to use a connection string in an ODBC application, refer to the documentation for the application that you are using.

The connection strings in the following sections are examples showing the minimum set of connection attributes that you must specify to successfully connect to the data source. Depending on the configuration of the data source and the type of connection you are working with, you might need to specify additional connection attributes. For detailed information about all the attributes that you can use in the connection string, see "Driver Configuration Options" on page 62.

DSN Connection String Example

The following is an example of a connection string for a connection that uses a DSN:

```
DSN=[DataSourceName];
```

[DataSourceName] is the DSN that you are using for the connection.

You can set additional configuration options by appending key-value pairs to the connection string. Configuration options that are passed in using a connection string take precedence over configuration options that are set in the DSN.

DSN-less Connection String Examples

Some applications provide support for connecting to a data source using a driver without a DSN. To connect to a data source without using a DSN, use a connection string instead.

The placeholders in the examples are defined as follows, in alphabetical order:

- *[DomainName]* is the fully qualified domain name of the Hive server host.
- *[Namespace]* is the namespace on ZooKeeper under which Hive Server 2 znodes are added.
- *[PortNumber]* is the number of the TCP port that the Hive server uses to listen for client connections.
- *[Realm]* is the Kerberos realm of the Hive server host.
- *[Server]* is the IP address or host name of the Hive server to which you are connecting.
- *[ServiceName]* is the Kerberos service principal name of the Hive server.
- *[YourPassword]* is the password corresponding to your user name.
- *[YourUserName]* is the user name that you use to access the Hive server.

Connecting to a Hive Server 1 Instance

The following is the format of a DSN-less connection string that connects to a Hive Server 1 instance:

```
Driver=Cloudera Hive ODBC Driver;HiveServerType=1;
```

```
Host=[Server];Port=[PortNumber];
```

For example:

```
Driver=Cloudera Hive ODBC Driver;HiveServerType=1;
Host=192.168.222.160;Port=10000;
```

Connecting to a Standard Hive Server 2 Instance

The following is the format of a DSN-less connection string for a standard connection to a Hive Server 2 instance. By default, the driver is configured to connect to a Hive Server 2 instance that requires User Name authentication, and the driver uses **anonymous** as the user name.

```
Driver=Cloudera Hive ODBC Driver;Host=[Server];
Port=[PortNumber];
```

For example:

```
Driver=Cloudera Hive ODBC Driver;Host=192.168.222.160;
Port=10000;
```

Connecting using Dynamic Service Discovery

The following is the format of a DSN-less connection string that discovers Hive Server 2 services via the ZooKeeper service.

```
Driver=Cloudera Hive ODBC Driver;ServiceDiscoveryMode=1;
Host=[Server1]:[PortNumber1], [Server2]:[PortNumber2], [Server3]:
[PortNumber3];ZKNamespace=[Namespace];
```

For example:

```
Driver=Cloudera Hive ODBC Driver;ServiceDiscoveryMode=1;
Host=192.168.222.160:10000, 192.168.222.165:10000,
192.168.222.231:10000;ZKNamespace=hiveserver;
```

Connecting to a Hive Server 2 Instance Without Authentication

The following is the format of a DSN-less connection string that for a Hive Server 2 instance that does not require authentication.

```
Driver=Cloudera Hive ODBC Driver;Host=[Server];
Port=[PortNumber];AuthMech=0;
```

For example:

```
Driver=Cloudera Hive ODBC Driver;Host=192.168.222.160;
Port=10000;AuthMech=0;
```

Using a Connection String

Connecting to a Hive Server that Requires Kerberos Authentication

The following is the format of a DSN-less connection string that connects to a Hive Server 2 instance requiring Kerberos authentication:

```
Driver=Cloudera Hive ODBC Driver;Host=[Server];  
Port=[PortNumber];AuthMech=1;KrbRealm=[Realm];  
KrbHostFQDN=[DomainName];KrbServiceName=[ServiceName];
```

For example:

```
Driver=Cloudera Hive ODBC Driver;Host=192.168.222.160;  
Port=10000;AuthMech=1;KrbRealm=CLOUDERA;  
KrbHostFQDN=localhost.localdomain;KrbServiceName=hive;
```

Connecting to a Hive Server that Requires User Name And Password Authentication

The following is the format of a DSN-less connection string that connects to a Hive Server 2 instance requiring User Name And Password authentication:

```
Driver=Cloudera Hive ODBC Driver;Host=[Server];  
Port=[PortNumber];AuthMech=3;UID=[YourUserName];  
PWD=[YourPassword];
```

For example:

```
Driver=Cloudera Hive ODBC Driver;Host=192.168.222.160;  
Port=10000;AuthMech=3;UID=cloudera;PWD=cloudera;
```

Features

For more information on the features of the Cloudera ODBC Driver for Apache Hive, see the following:

- "SQL Connector for HiveQL" on page 55
- "Data Types" on page 55
- "Catalog and Schema Support" on page 57
- "hive_system Table" on page 57
- "Server-Side Properties" on page 57
- "Get Tables With Query" on page 59
- "Active Directory" on page 59
- "Write-back" on page 59
- "Dynamic Service Discovery using ZooKeeper" on page 59

SQL Connector for HiveQL

The native query language supported by Hive is HiveQL. For simple queries, HiveQL is a subset of SQL-92. However, the syntax is different enough that most applications do not work with native HiveQL.

To bridge the difference between SQL and HiveQL, the SQL Connector feature translates standard SQL-92 queries into equivalent HiveQL queries. The SQL Connector performs syntactical translations and structural transformations. For example:

- **Quoted Identifiers:** The double quotes ("") that SQL uses to quote identifiers are translated into back quotes (`) to match HiveQL syntax. The SQL Connector needs to handle this translation because even when a driver reports the back quote as the quote character, some applications still generate double-quoted identifiers.
- **Table Aliases:** Support is provided for the AS keyword between a table reference and its alias, which HiveQL normally does not support.
- **JOIN, INNER JOIN, and CROSS JOIN:** SQL JOIN, INNER JOIN, and CROSS JOIN syntax is translated to HiveQL JOIN syntax.
- **TOP N/LIMIT:** SQL TOP N queries are transformed to HiveQL LIMIT queries.

Data Types

The Cloudera ODBC Driver for Apache Hive supports many common data formats, converting between Hive data types and SQL data types.

The following table lists the supported data type mappings.

Hive Type	SQL Type
BIGINT	SQL_BIGINT

Hive Type	SQL Type
BINARY	SQL_VARBINARY
BOOLEAN	SQL_BIT
CHAR(n)	SQL_CHAR <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> Note: SQL_WCHAR is returned instead if the Unicode SQL Character Types configuration option (the UseUnicodeSqlCharacterTypes key) is enabled. </div>
DATE	SQL_TYPE_DATE
DECIMAL	SQL_DECIMAL
DECIMAL(p,s)	SQL_DECIMAL
DOUBLE	SQL_DOUBLE
FLOAT	SQL_REAL
INT	SQL_INTEGER
SMALLINT	SQL_SMALLINT
STRING	SQL_VARCHAR <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> Note: SQL_WVARCHAR is returned instead if the Unicode SQL Character Types configuration option (the UseUnicodeSqlCharacterTypes key) is enabled. </div>
TIMESTAMP	SQL_TYPE_TIMESTAMP
TINYINT	SQL_TINYINT
VARCHAR(n)	SQL_VARCHAR

Note:

The aggregate types (ARRAY, MAP, and STRUCT) are not supported. Columns of aggregate types are treated as STRING columns.

The interval types (YEAR TO MONTH and DAY TIME) are supported only in query expressions and predicates. Interval types are not supported as column data types in tables.

Catalog and Schema Support

The Cloudera ODBC Driver for Apache Hive supports both catalogs and schemas to make it easy for the driver to work with various ODBC applications. Since Hive only organizes tables into schemas/databases, the driver provides a synthetic catalog named HIVE under which all of the schemas/databases are organized. The driver also maps the ODBC schema to the Hive schema/database.

hive_system Table

A pseudo-table called `hive_system` can be used to query for Hive cluster system environment information. The pseudo-table is under the pseudo-schema called `hive_system`. The table has two STRING type columns, `envkey` and `envvalue`. Standard SQL can be executed against the `hive_system` table. For example:

```
SELECT * FROM HIVE.hive_system.hive_system WHERE envkey LIKE
'%hive%'
```

The above query returns all of the Hive system environment entries whose key contains the word "hive". A special query, `set -v`, is executed to fetch system environment information. Some versions of Hive do not support this query. For versions of Hive that do not support querying system environment information, the driver returns an empty result set.

Server-Side Properties

The Cloudera ODBC Driver for Apache Hive allows you to set server-side properties via a DSN. Server-side properties specified in a DSN affect only the connection that is established using the DSN.

You can also specify server-side properties for connections that do not use a DSN. To do this, use the Cloudera Hive ODBC Driver Configuration tool that is installed with the Windows version of the driver, or set the appropriate configuration options in your connection string or the `cloudera.hiveodbc.ini` file. Properties specified in the driver configuration tool or the `cloudera.hiveodbc.ini` file apply to all connections that use the Cloudera ODBC Driver for Apache Hive.

For more information about setting server-side properties when using the Windows driver, see "Configuring Server-Side Properties" on page 18. For information about setting server-side properties when using the driver on a non-Windows platform, see "Driver Configuration Options" on page 62.

Temporary Tables

The driver supports the creation of temporary tables and lets you insert literal values into temporary tables. Temporary tables are only accessible by the ODBC connection that created them, and are dropped when the connection is closed.

CREATE TABLE Statement for Temporary Tables

The driver supports the following DDL syntax for creating temporary tables:

```

<create table statement> := CREATE TABLE <temporary table name>
<left paren><column definition list><right paren>
<column definition list> := <column definition>[, <column
definition>]*
<column definition> := <column name> <data type>
<temporary table name> := <double quote><number sign><table
name><double quote>
<left paren> := (
<right paren> := )
<double quote> :=
<number sign> :=

```

The following is an example of a SQL statement for creating a temporary table:

```
CREATE TABLE "#TEMPTABLE1" (C1 DATATYPE_1, C2 DATATYPE_2, ...,
Cn DATATYPE_n)
```

The temporary table name in a SQL query must be surrounded by double quotes ("), and the name must begin with a number sign (#).

Note:

You can only use data types that are supported by Hive.

INSERT Statement for Temporary Tables

The driver supports the following DDL syntax for inserting data into temporary tables:

```

<insert statement> := INSERT INTO <temporary table name> <left
paren><column name list><right paren> VALUES <left paren><literal
value list><right paren>
<column name list> := <column name>[, <column name>]*
<literal value list> := <literal value>[, <literal value>]*
<temporary table name> := <double quote><number sign><table
name><double quote>
<left paren> := (

```

```

<right paren> := )
<double quote> := "
<number sign> := #

```

The following is an example of a SQL statement for inserting data into temporary tables:

```
INSERT INTO "#TEMPTABLE1" values (VAL(C1), VAL(C2) ... VAL(Cn))
```

VAL(C1) is the literal value for the first column in the table, and VAL(Cn) is the literal value for the nth column in the table.

Note:

The INSERT statement is only supported for temporary tables.

Get Tables With Query

The Get Tables With Query configuration option allows you to choose whether to use the SHOW TABLES query or the GetTables API call to retrieve table names from a database.

Hive Server 2 has a limit on the number of tables that can be in a database when handling the GetTables API call. When the number of tables in a database is above the limit, the API call will return a stack overflow error or a timeout error. The exact limit and the error that appears depends on the JVM settings.

As a workaround for this issue, enable the Get Tables with Query configuration option or the GetTablesWithQuery key to use the query instead of the API call.

Active Directory

The Cloudera ODBC Driver for Apache Hive supports Active Directory Kerberos on Windows. There are two prerequisites for using Active Directory Kerberos on Windows:

- MIT Kerberos is not installed on the client Windows machine.
- The MIT Kerberos Hadoop realm has been configured to trust the Active Directory realm, according to Cloudera's documentation, so that users in the Active Directory realm can access services in the MIT Kerberos Hadoop realm.

Write-back

The Cloudera ODBC Driver for Apache Hive supports translation for INSERT, UPDATE, and DELETE syntax when connecting to a Hive Server 2 instance that is running Hive 0.14 or later.

Dynamic Service Discovery using ZooKeeper

The Cloudera ODBC Driver for Apache Hive can be configured to discover Hive Server 2 services via the ZooKeeper service.

Features

For information about configuring this feature in the Windows driver, see "Creating a Data Source Name" on page 6 or "Configuring a DSN-less Connection" on page 9. For information about configuring this feature when using the driver on a non-Windows platform, see "Configuring Service Discovery Mode" on page 41.

Contact Us

If you are having difficulties using the driver, our [Community Forum](#) may have your solution. In addition to providing user to user support, our forums are a great place to share your questions, comments, and feature requests with us.

If you are a Subscription customer you may also use the [Cloudera Support Portal](#) to search the Knowledge Base or file a Case.

Important:

To help us assist you, prior to contacting Cloudera Support please prepare a detailed summary of the client and server environment including operating system version, patch level, and configuration

Driver Configuration Options

Driver Configuration Options lists the configuration options available in the Cloudera ODBC Driver for Apache Hive alphabetically by field or button label. Options having only key names, that is, not appearing in the user interface of the driver, are listed alphabetically by key name.

When creating or configuring a connection from a Windows machine, the fields and buttons are available in the Cloudera Hive ODBC Driver Configuration tool and the following dialog boxes:

- Cloudera ODBC Driver for Apache Hive DSN Setup
- Advanced Options
- Server Side Properties
- SSL Options
- HTTP Properties

When using a connection string or configuring a connection from a Linux/Mac OS X/AIX/Debian machine, use the key names provided.

Note:

You can pass in configuration options in your connection string, or set them in your `odbc.ini` and `cloudera.hiveodbc.ini` files if you are using a non-Windows version of the driver. Configuration options set in a `cloudera.hiveodbc.ini` file apply to all connections, whereas configuration options passed in in the connection string or set in an `odbc.ini` file are specific to a connection. Configuration options passed in using the connection string take precedence over configuration options set in `odbc.ini`. Configuration options set in `odbc.ini` take precedence over configuration options set in `cloudera.hiveodbc.ini`.

Configuration Options Appearing in the User Interface

The following configuration options are accessible via the Windows user interface for the Cloudera ODBC Driver for Apache Hive, or via the key name when using a connection string or configuring a connection from a Linux/Mac OS X//AIX/Debian machine:

- "Allow Common Name Host Name Mis-match" on page 63
- "Allow Self-Signed Server Certificate" on page 64
- "Apply Properties with Queries" on page 64
- "Binary Column Length" on page 64
- "Client Certificate File" on page 65
- "Client Private Key File" on page 65
- "Client Private Key Password" on page 65
- "HTTP Path" on page 71
- "Invalid Session Auto Recover" on page 71
- "Log Level" on page 72
- "Log Path" on page 73
- "Max File Size" on page 73
- "Max Number Files" on page 73
- "Mechanism" on page 74
- "Password" on page 74
- "Port" on page 74

- "Convert Key Name to Lower Case" on page 66
- "Data File HDFS Dir" on page 66
- "Database" on page 66
- "Decimal Column Scale" on page 67
- "Default String Column Length" on page 67
- "Delegate Kerberos Credentials" on page 67
- "Delegation UID" on page 67
- "Driver Config Take Precedence" on page 68
- "Enable SSL" on page 68
- "Enable Temporary Table" on page 68
- "Fast SQLPrepare" on page 69
- "Get Tables With Query" on page 69
- "HDFS User" on page 70
- "Hive Server Type" on page 70
- "Host(s)" on page 70
- "Host FQDN" on page 71
- "Realm" on page 75
- "Rows Fetched Per Block" on page 75
- "Save Password (Encrypted)" on page 75
- "Service Discovery Mode" on page 76
- "Service Name" on page 76
- "Show System Table" on page 76
- "Socket Timeout" on page 76
- "Temp Table TTL" on page 77
- "Thrift Transport" on page 77
- "Trusted Certificates" on page 77
- "Two Way SSL" on page 78
- "Unicode SQL Character Types" on page 79
- "Use Async Exec" on page 79
- "Use Native Query" on page 79
- "Use Only SSPI Plugin" on page 80
- "User Name" on page 80
- "Web HDFS Host" on page 81
- "Web HDFS Port" on page 81
- "ZooKeeper Namespace" on page 81

Allow Common Name Host Name Mismatch

Key Name	Default Value	Required
CAIssuedCertNamesMismatch	Clear (0)	No

Description

This option specifies whether a CA-issued SSL certificate name must match the host name of the Hive server.

- Enabled (1): The driver allows a CA-issued SSL certificate name to not match the host name of the Hive server.
- Disabled (0): The CA-issued SSL certificate name must match the host name of the Hive server.

Note:

This setting is applicable only when SSL is enabled.

Driver Configuration Options

Allow Self-Signed Server Certificate

Key Name	Default Value	Required
AllowSelfSignedServerCert	Clear (0)	No

Description

This option specifies whether the driver allows self-signed certificates from the server.

- Enabled (1): The driver authenticates the Hive server even if the server is using a self-signed certificate.
- Disabled (0): The driver does not allow self-signed certificates from the server.

Note:

This setting is applicable only when SSL is enabled.

Apply Properties with Queries

Key Name	Default Value	Required
ApplySSPWithQueries	Selected (1)	No

Description

This option specifies how the driver applies server-side properties.

- Enabled (1): The driver applies each server-side property by executing a `set SSPKey=SSPValue` query when opening a session to the Hive server.
- Disabled (0): The driver uses a more efficient method for applying server-side properties that does not involve additional network round-tripping. However, some Hive Server 2 builds are not compatible with the more efficient method.

Note:

When connecting to a Hive Server 1 instance, this option is always enabled.

Binary Column Length

Key Name	Default Value	Required
BinaryColumnLength	32767	No

Description

The maximum data length for BINARY columns.

By default, the columns metadata for Hive does not specify a maximum data length for BINARY columns.

Client Certificate File

Key Name	Default Value	Required
ClientCert	None	No

Description

The full path to the .pem file containing the client's SSL certificate.

Note:

This setting is applicable only when two-way SSL is enabled.

Client Private Key File

Key Name	Default Value	Required
ClientPrivateKey	None	Yes, if two-way SSL verification is enabled.

Description

The full path to the .pem file containing the client's SSL private key.

If the private key file is protected with a password, then provide the password using the driver configuration option "Client Private Key Password" on page 65.

Note:

This setting is applicable only when two-way SSL is enabled.

Client Private Key Password

Key Name	Default Value	Required
ClientPrivateKeyPassword	None	Yes, if two-way SSL verification is enabled and the client's private key file is protected with a password.

Description

The password of the private key file that is specified in the Client Private Key File field (ClientPrivateKey).

Driver Configuration Options

Convert Key Name to Lower Case

Key Name	Default Value	Required
LCaseSspKeyName	Selected (1)	No

Description

This option specifies whether the driver converts server-side property key names to all lower-case characters.

- Enabled (1): The driver converts server-side property key names to all lower-case characters.
- Disabled (0): The driver does not modify the server-side property key names.

Data File HDFS Dir

Key Name	Default Value	Required
HDFSTempTableDir	/tmp/simba	No

Description

The HDFS directory that the driver uses to store the necessary files for supporting the Temporary Table feature.

Note:

Due to a problem in Hive (see <https://issues.apache.org/jira/browse/HIVE-4554>), HDFS paths with space characters do not work with versions of Hive prior to 0.12.0.

This option is not applicable when connecting to Hive 0.14 or later.

Database

Key Name	Default Value	Required
Schema	default	No

Description

The name of the database schema to use when a schema is not explicitly specified in a query. You can still issue queries on other schemas by explicitly specifying the schema in the query.

Note:

To inspect your databases and determine the appropriate schema to use, at the Hive command prompt, type `show databases`.

Decimal Column Scale

Key Name	Default Value	Required
DecimalColumnScale	10	No

Description

The maximum number of digits to the right of the decimal point for numeric data types.

Default String Column Length

Key Name	Default Value	Required
DefaultStringColumnLength	255	No

Description

The maximum number of characters that can be contained in STRING columns.

By default, the columns metadata for Hive does not specify a maximum length for STRING columns.

Delegate Kerberos Credentials

Key Name	Default Value	Required
DelegateKrbCreds	0	No

Description

This option specifies whether your Kerberos credentials are forwarded to the server and used for authentication.

Note:

This option is only applicable when Authentication Mechanism is set to Kerberos (AuthMech=1).

Delegation UID

Key Name	Default Value	Required
DelegationUID	None	No

Driver Configuration Options

Description

If a value is specified for this setting, the driver delegates all operations against Hive to the specified user, rather than to the authenticated user for the connection.

Note:

This option is applicable only when connecting to a Hive Server 2 instance that supports this feature.

Driver Config Take Precedence

Key Name	Default Value	Required
DriverConfigTakePrecedence	Clear (0)	No

Description

This option specifies whether driver-wide configuration settings take precedence over connection and DSN settings.

- Enabled (1): Driver-wide configurations take precedence over connection and DSN settings.
- Disabled (0): Connection and DSN settings take precedence instead.

Enable SSL

Key Name	Default Value	Required
SSL	Clear (0)	No

Description

This option specifies whether the client verifies the Hive server using SSL.

- Enabled (1): The client verifies the Hive server using SSL.
- Disabled (0): SSL is disabled.

SSL is configured independently of authentication. When authentication and SSL are both enabled, the driver performs the specified authentication method over an SSL connection.

Note:

This option is applicable only when connecting to a Hive server that supports SSL.

Enable Temporary Table

Key Name	Default Value	Required
EnableTempTable	Clear (0)	No

Description

This option specifies whether the driver supports the creation and use of temporary tables.

- Enabled (1): The driver supports the creation and use of temporary tables.
- Disabled (0): The driver does not support temporary tables.

Important:

When connecting to Hive 0.14 or later, the Temporary Tables feature is always enabled and you do not need to configure it in the driver.

Fast SQLPrepare

Key Name	Default Value	Required
FastSQLPrepare	Clear (0)	No

Description

This option specifies whether the driver defers query execution to SQLExecute.

- Enabled (1): The driver defers query execution to SQLExecute.
- Disabled (0): The driver does not defer query execution to SQLExecute.

Note:

When using Native Query mode, the driver executes the HiveQL query to retrieve the result set metadata for SQLPrepare. As a result, SQLPrepare might be slow. If the result set metadata is not required after calling SQLPrepare, then enable Fast SQLPrepare.

Get Tables With Query

Key Name	Default Value	Required
GetTablesWithQuery	Clear (0)	No

Description

This option specifies whether the driver uses the SHOW TABLES query or the GetTables Thrift API call to retrieve table names from the database.

- Enabled (1): The driver uses the SHOW TABLES query to retrieve table names.
- Disabled (0): The driver uses the GetTables Thrift API call to retrieve table names.

Note:

This option is applicable only when connecting to a Hive Server 2 instance.

Driver Configuration Options

HDFS User

Key Name	Default Value	Required
HDFSUser	hdfs	No

Description

The name of the HDFS user that the driver uses to create the necessary files for supporting the Temporary Tables feature.

This option is not applicable when connecting to Hive 0.14 or later.

Hive Server Type

Key Name	Default Value	Required
HiveServerType	Hive Server 2 (2)	No

Description

This option specifies the type of Hive server.

- Hive Server 1 (1): The driver connects to a Hive Server 1 instance.
- Hive Server 2 (2): The driver connects to a Hive Server 2 instance.

Note:

If Service Discovery Mode is enabled, then connections to Hive Server 1 are not supported.

Host(s)

Key Name	Default Value	Required
HOST	None	Yes

Description

If Service Discovery Mode is disabled, specify the IP address or host name of the Hive server.

If Service Discovery Mode is enabled, specify a comma-separated list of ZooKeeper servers in the following format, where *[ZK_Host]* is the IP address or host name of the ZooKeeper server and *[ZK_Port]* is the number of the TCP port that the ZooKeeper server uses to listen for client connections:

[ZK_Host1]:[ZK_Port1], [ZK_Host2]:[ZK_Port2]

Host FQDN

Key Name	Default Value	Required
KrbHostFQDN	None	Yes, if the authentication mechanism is Kerberos.

Description

The fully qualified domain name of the Hive Server 2 host.

You can set the value of Host FQDN to `_HOST` to use the Hive server host name as the fully qualified domain name for Kerberos authentication. If Service Discovery Mode is disabled, then the driver uses the value specified in the Host connection attribute. If Service Discovery Mode is enabled, then the driver uses the Hive Server 2 host name returned by ZooKeeper.

HTTP Path

Key Name	Default Value	Required
HTTPPath	/hive2 if using Windows Azure HDInsight Service (6). / if using non-Windows Azure HDInsight Service with Thrift Transport set to HTTP (2).	No

Description

The partial URL corresponding to the Hive server.

The driver forms the HTTP address to connect to by appending the HTTP Path value to the host and port specified in the DSN or connection string. For example, to connect to the HTTP address `http://localhost:10002/gateway/sandbox/hive/version`, you would set HTTP Path to `/gateway/sandbox/hive/version`.

Note:

By default, Hive servers use `cliservice` as the partial URL.

Invalid Session Auto Recover

Key Name	Default Value	Required
InvalidSessionAutoRecover	Selected (1)	No

Driver Configuration Options

Description

This option specifies whether the driver automatically opens a new session when the existing session is no longer valid.

- Enabled (1): The driver automatically opens a new session when the existing session is no longer valid.
- Disabled (0): The driver does not automatically open new sessions.

Note:

This option is applicable only when connecting to Hive Server 2.

Log Level

Key Name	Default Value	Required
LogLevel	OFF (0)	No

Description

Use this property to enable or disable logging in the driver and to specify the amount of detail included in log files.

Important:

- Only enable logging long enough to capture an issue. Logging decreases performance and can consume a large quantity of disk space.
- This option is not supported in connection strings. To configure logging for the Windows driver, you must use the Logging Options dialog box. To configure logging for a non-Windows driver, you must use the `cloudera.hiveodbc.ini` file.

Set the property to one of the following values:

- OFF (0): Disable all logging.
- FATAL (1): Logs severe error events that lead the driver to abort.
- ERROR (2): Logs error events that might allow the driver to continue running.
- WARNING (3): Logs potentially harmful situations.
- INFO (4): Logs general information that describes the progress of the driver.
- DEBUG (5): Logs detailed information that is useful for debugging the driver.
- TRACE (6): Logs all driver activity.

When logging is enabled, the driver produces a log file named `HiveODBC_driver.log` in the location specified in the Log Path (`LogPath`) property.

Log Path

Key Name	Default Value	Required
LogPath	None	Yes, if logging is enabled.

Description

The full path to the folder where the driver saves log files when logging is enabled.

Important:

This option is not supported in connection strings. To configure logging for the Windows driver, you must use the Logging Options dialog box. To configure logging for a non-Windows driver, you must use the `cloudera.hiveodbc.ini` file.

Max File Size

Key Name	Default Value	Required
LogFileSize	20	No

Description

The maximum size of each log file in megabytes (MB). After the maximum file size is reached, the driver creates a new file and continues logging.

Important:

This option is not supported in connection strings. To configure logging for the Windows driver, you must use the Logging Options dialog box. To configure logging for a non-Windows driver, you must use the `cloudera.hiveodbc.ini` file.

Max Number Files

Key Name	Default Value	Required
LogFileCount	50	No

Description

The maximum number of log files to keep. After the maximum number of log files is reached, each time an additional file is created, the driver deletes the oldest log file.

Important:

This option is not supported in connection strings. To configure logging for the Windows driver, you must use the Logging Options dialog box. To configure logging for a non-Windows driver, you must use the `cloudera.hiveodbc.ini` file.

Mechanism

Key Name	Default Value	Required
AuthMech	No Authentication (0) if you are connecting to Hive Server 1. User Name (2) if you are connecting to Hive Server 2.	No

Description

The authentication mechanism to use.

Select one of the following settings, or set the key to the corresponding number:

- No Authentication (0)
- Kerberos (1)
- User Name (2)
- User Name And Password (3)

Password

Key Name	Default Value	Required
PWD	None	Yes, if the authentication mechanism is User Name And Password (3).

Description

The password corresponding to the user name that you provided in the User Name field (the UID key).

Port

Key Name	Default Value	Required
PORT	10000	Yes, if Service Discovery Mode is disabled.

Description

The number of the TCP port that the Hive server uses to listen for client connections.

Realm

Key Name	Default Value	Required
KrbRealm	Depends on your Kerberos configuration.	No

Description

The realm of the Hive Server 2 host.

If your Kerberos configuration already defines the realm of the Hive Server 2 host as the default realm, then you do not need to configure this option.

Rows Fetched Per Block

Key Name	Default Value	Required
RowsFetchedPerBlock	10000	No

Description

The maximum number of rows that a query returns at a time.

Valid values for this setting include any positive 32-bit integer. However, testing has shown that performance gains are marginal beyond the default value of 10000 rows.

Save Password (Encrypted)

Key Name	Default Value	Required
N/A	Selected (1)	No

Description

This option specifies whether the password is saved in the registry.

- Enabled (1): The password is saved in the registry.
- Disabled (0): The password is not saved in the registry.

This option is available only in the Windows driver. It appears in the Cloudera ODBC Driver for Apache Hive DSN Setup dialog box and the SSL Options dialog box.

Important:

The password is obscured (not saved in plain text). However, it is still possible for the encrypted password to be copied and used.

Service Discovery Mode

Key Name	Default Value	Required
ServiceDiscoveryMode	No Service Discovery (0)	No

Description

This option specifies whether the driver uses the ZooKeeper service.

- Enabled (1): The driver discovers Hive Server 2 services via the ZooKeeper service.
- Disabled (0): The driver connects to Hive without using the ZooKeeper service.

Service Name

Key Name	Default Value	Required
KrbServiceName	None	Yes, if the authentication mechanism is Kerberos.

Description

The Kerberos service principal name of the Hive server.

Show System Table

Key Name	Default Value	Required
ShowSystemTable	Clear (0)	No

Description

This option specifies whether the driver returns the `hive_system` table for catalog function calls such as `SQLTables` and `SQLColumns`.

- Enabled (1): The driver returns the `hive_system` table for catalog function calls such as `SQLTables` and `SQLColumns`.
- Disabled (0): The driver does not return the `hive_system` table for catalog function calls.

Socket Timeout

Key Name	Default Value	Required
SocketTimeout	60	No

Description

The number of seconds that an operation can remain idle before it is closed.

Note:

This option is applicable only when asynchronous query execution is being used against Hive Server 2 instances.

Temp Table TTL

Key Name	Default Value	Required
TempTableTTL	10	No

Description

The number of minutes a temporary table is guaranteed to exist in Hive after it is created.

This option is not applicable when connecting to Hive 0.14 or later.

Thrift Transport

Key Name	Default Value	Required
ThriftTransport	Binary (0) if you are connecting to Hive Server 1. SASL (1) if you are connecting to Hive Server 2.	No

Description

The transport protocol to use in the Thrift layer.

Select one of the following settings, or set the key to the number corresponding to the desired setting:

- Binary (0)
- SASL (1)
- HTTP (2)

Trusted Certificates

Key Name	Default Value	Required
TrustedCerts	The cacerts.pem file in	No

Driver Configuration Options

Key Name	Default Value	Required
	the \lib subfolder within the driver's installation directory. The exact file path varies depending on the version of the driver that is installed. For example, the path for the Windows driver is different from the path for the Mac OS X driver.	

Description

The full path of the .pem file containing trusted CA certificates for verifying the server when using SSL.

If this option is not set, then the driver defaults to using the trusted CA certificates .pem file installed by the driver.

Note:

This setting is applicable only when SSL is enabled.

Two Way SSL

Key Name	Default Value	Required
TwoWaySSL	Clear (0)	No

Description

This option specifies whether two-way SSL is enabled.

- Enabled (1): The client and the Hive server verify each other using SSL. See also the driver configuration options "Client Certificate File" on page 65, "Client Private Key File" on page 65, and "Client Private Key Password" on page 65.
- Disabled (0): The server does not verify the client. Depending on whether one-way SSL is enabled, the client might verify the server. For more information, see "Enable SSL" on page 68.

Note:

This option is applicable only when connecting to a Hive server that supports SSL. You must enable SSL before Two Way SSL can be configured. For more information, see "Enable SSL" on page 68.

Unicode SQL Character Types

Key Name	Default Value	Required
UseUnicodeSqlCharacterTypes	Clear (0)	No

Description

This option specifies the SQL types to be returned for string data types.

- Enabled (1): The driver returns SQL_WVARCHAR for STRING and VARCHAR columns, and returns SQL_WCHAR for CHAR columns.
- Disabled (0): The driver returns SQL_VARCHAR for STRING and VARCHAR columns, and returns SQL_CHAR for CHAR columns.

Use Async Exec

Key Name	Default Value	Required
EnableAsyncExec	Clear (0)	No

Description

This option specifies whether to execute queries synchronously or asynchronously.

- Enabled (1): The driver uses an asynchronous version of the API call against Hive for executing a query.
- Disabled (0): The driver executes queries synchronously.

Due to a problem in Hive 0.12.0 (see <https://issues.apache.org/jira/browse/HIVE-5230>), Hive returns generic error messages for errors that occur during query execution. To see the actual error message relevant to the problem, turn off asynchronous query execution and execute the query again.

Note:

This option only takes effect when connecting to a Hive cluster running Hive 0.12.0 or higher.

Use Native Query

Key Name	Default Value	Required
UseNativeQuery	Clear (0)	No

Description

This option specifies whether the driver uses native HiveQL queries, or converts them into an equivalent form in HiveQL.

Driver Configuration Options

- Enabled (1): The driver does not transform the queries emitted by an application, and executes HiveQL queries directly.
- Disabled (0): The driver transforms the queries emitted by an application and converts them into an equivalent form in HiveQL.

Note:

If the application is Hive-aware and already emits HiveQL, then enable this option to avoid the extra overhead of query transformation.

Use Only SSPI Plugin

Key Name	Default Value	Required
UseOnlySSPI	Clear (0)	No

Description

This option specifies how the driver handles Kerberos authentication: either with the SSPI plugin or with MIT Kerberos.

- Enabled (1): The driver handles Kerberos authentication by using the SSPI plugin instead of MIKerberos by default.
- Disabled (0): The driver uses MIT Kerberos to handle Kerberos authentication, and only uses the SSPI plugin if the gssapi library is not available.

Important:

This option is available only in the Windows driver.

User Name

Key Name	Default Value	Required
UID	For User Name (2) authentication only, the default value is anonymous	Yes, if the authentication mechanism is User Name And Password (3). No, if the authentication mechanism is User Name (2).

Description

The user name that you use to access Hive Server 2.

Web HDFS Host

Key Name	Default Value	Required
WebHDFSHost	The Hive server host.	No

Description

The host name or IP address of the machine hosting both the namenode of your Hadoop cluster and the WebHDFS service.

This option is not applicable when connecting to Hive 0.14 or later.

Web HDFS Port

Key Name	Default Value	Required
WebHDFSPort	50070	No

Description

The WebHDFS port for the namenode.

This option is not applicable when connecting to Hive 0.14 or later.

ZooKeeper Namespace

Key Name	Default Value	Required
ZKNamespace	None	Yes, if Service Discovery Mode is enabled.

Description

The namespace on ZooKeeper under which Hive Server 2 znodes are added.

Configuration Options Having Only Key Names

The following configuration options do not appear in the Windows user interface for the Cloudera ODBC Driver for Apache Hive. They are accessible only when you use a connection string or configure a connection from a Linux/Mac OS X/AIX/Debian machine:

- "Driver" on page 82
- "http.header." on page 82
- "SSP_" on page 83

Driver Configuration Options

Driver

Key Name	Default Value	Required
Driver	Cloudera ODBC Driver for Apache Hive when installed on Windows, or the absolute path of the driver shared object file when installed on a non-Windows machine.	Yes

Description

The name of the installed driver (Cloudera ODBC Driver for Apache Hive) when installed on Windows, or the absolute path of the driver shared object file when installed on a non-Windows machine.

http.header.

Key Name	Default Value	Required
http.header	None	No

Description

Set a custom HTTP header by using the following syntax, where *[HeaderKey]* is the name of the header to set and *[HeaderValue]* is the value to assign to the header:

```
http.header. [HeaderKey]=[HeaderValue]
```

For example:

```
http.header.AUTHENTICATED_USER=john
```

After the driver applies the header, the http.header. prefix is removed from the DSN entry, leaving an entry of *[HeaderKey]=[HeaderValue]*

The example above would create the following custom HTTP header:

```
AUTHENTICATED_USER: john
```

Note:

The http.header. prefix is case-sensitive. This option is applicable only when you are using HTTP as the Thrift transport protocol. For more information, see "Thrift Transport" on page 77.

SSP_

Key Name	Default Value	Required
SSP_	None	No

Description

Set a server-side property by using the following syntax, where *[SSPKey]* is the name of the server-side property and *[SSPValue]* is the value for that property:

```
SSP_[SSPKey]=[SSPValue]
```

For example:

```
SSP_mapred.queue.names=myQueue
```

After the driver applies the server-side property, the SSP_ prefix is removed from the DSN entry, leaving an entry of *[SSPKey]=[SSPValue]*.

Note:

- The SSP_ prefix must be upper case.
- When setting a server-side property in a connection string, it is recommended that you enclose the value in braces ({ }) to make sure that special characters can be properly escaped.

ODBC API Conformance Level

The following table lists the ODBC interfaces that the Cloudera ODBC Driver for Apache Hive implements and the ODBC compliance level of each interface.

ODBC compliance levels are Core, Level 1, and Level 2. These compliance levels are defined in the ODBC Specification published with the Interface SDK from Microsoft.

Interfaces include both the Unicode and non-Unicode versions. For more information, see "Unicode Function Arguments" in the *ODBC Programmer's Reference*:
<http://msdn.microsoft.com/en-us/library/ms716246%28VS.85%29.aspx>.

Conformance Level	INTERFACES		Conformance Level	INTERFACES
Core	SQLAllocHandle		Core	SQLGetStmtAttr
Core	SQLBindCol		Core	SQLGetTypeInfo
Core	SQLBindParameter		Core	SQLNativeSql
Core	SQLCancel		Core	SQLNumParams
Core	SQLCloseCursor		Core	SQLNumResultCols
Core	SQLColAttribute		Core	SQLParamData
Core	SQLColumns		Core	SQLPrepare
Core	SQLConnect		Core	SQLPutData
Core	SQLCopyDesc		Core	SQLRowCount
Core	SQLDescribeCol		Core	SQLSetConnectAttr
Core	SQLDisconnect		Core	SQLSetCursorName

Conformance Level	INTERFACES	Conformance Level	INTERFACES
Core	SQLDriverconnect	Core	SQLSetDescField
Core	SQLEndTran	Core	SQLSetDescRec
Core	SQLExecDirect	Core	SQLSetEnvAttr
Core	SQLExecute	Core	SQLSetStmtAttr
Core	SQLFetch	Core	SQLSpecialColumns
Core	SQLFetchScroll	Core	SQLStatistics
Core	SQLFreeHandle	Core	SQLTables
Core	SQLFreeStmt	Core	SQLBrowseConnect
Core	SQLGetConnectAttr	Core	SQLPrimaryKeys
Core	SQLGetCursorName	Core	SQLGetInfo
Core	SQLGetData	Level 1	SQLProcedureColumns
Core	SQLGetDescField	Level 1	SQLProcedures
Core	SQLGetDescRec	Level 2	SQLColumnPrivileges
Core	SQLGetDiagField	Level 2	SQLDescribeParam
Core	SQLGetDiagRec	Level 2	SQLForeignKeys

ODBC API Conformance Level

Conformance Level	INTERFACES		Conformance Level	INTERFACES
Core	SQLGetEnvAttr		Level 2	SQLTablePrivileges
Core	SQLGetFunctions			