cloudera
Ask Bigger Questions

Cloudera ODBC Driver for Impala
Version 2.5.26

Important Notice

Cloudera, the Cloudera logo, Cloudera Impala, Impala, 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: 2.5.26

Date: April 9, 2015

Table of Contents

Introduction

The Cloudera ODBC Driver for Impala is used for direct SQL and Impala SQL access to Apache Hadoop / Impala distributions, enabling Business Intelligence (BI), analytics, and reporting on Hadoop / Impala-based data. The driver efficiently transforms an application's SQL query into the equivalent form in Impala SQL, which is a subset of SQL-92. If an application is Impala-aware, then the driver is configurable to pass the query through to the database for processing. The driver interrogates Impala to obtain schema information to present to a SQL-based application. Queries, including joins, are translated from SQL to Impala SQL. For more information about the differences between Impala SQL and SQL, see "Features" on page 29.

The Cloudera ODBC Driver for Impala 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 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 http://www.simba.com/resources/data-access-standards-library. For complete information about the ODBC specification, see the *ODBC API Reference* at 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 Impala from their desktop environment. Application developers may also find the information helpful. Refer to your application for details on connecting via ODBC.

Windows Driver

System Requirements

You install the Cloudera ODBC Driver for Impala on client computers accessing data in a Hadoop cluster with the Impala service installed and running. Each computer where you install the driver must meet the following minimum system requirements:

- One of the following operating systems (32- and 64-bit editions are supported):
  - Windows® XP with SP3
  - Windows® Vista
  - Windows® 7 Professional and Enterprise
  - Windows® 8 Pro and Enterprise
  - Windows® Server 2008 R2
- 25 MB of available disk space

Important:

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

The driver has been tested using Impala 1.0.1 and Apache Thrift 0.9.0

Installing the Driver

On 64-bit Windows operating systems, you can execute 32- and 64-bit applications transparently. You must use the version of the driver matching the bitness of the client application accessing data in Hadoop / Impala:

- **ClouderaImpalaODBC32.msi** for 32-bit applications
- **ClouderaImpalaODBC64.msi** for 64-bit applications

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

> **Note:**
>
> For an explanation of how to use ODBC on 64-bit editions of Windows, see http://www.simba.com/wp-content/uploads/2010/10/HOW-TO-32-bit-vs-64-bit-ODBC-Data-Source-Administrator.pdf

To install the Cloudera ODBC Driver for Impala:

1. Depending on the bitness of your client application, double-click to run **ClouderaImpalaODBC32.msi** or **ClouderaImpalaODBC64.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**

Verifying the Version Number

If you need to verify the version of the Cloudera ODBC Driver for Impala 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. Click the **Start** button , then click **All Programs**, then click the **Cloudera ODBC Driver for Impala 2.5** program group corresponding to the bitness of the client application accessing data in Hadoop / Impala, and then click **ODBC Administrator**
2. In the ODBC Data Source Administrator, click the **Drivers** tab and then find the Cloudera ODBC Driver for Impala in the list of ODBC drivers that are installed on your system. The version number is displayed in the Version column.

Creating a Data Source Name

After installing the Cloudera ODBC Driver for Impala, you need to create a Data Source Name (DSN).

To create a Data Source Name:

1. Click the **Start** button , then click **All Programs**, then click the **Cloudera ODBC Driver for Impala 2.5** program group corresponding to the bitness of the client application accessing data in Hadoop / Impala, and then click **ODBC Administrator**
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 Impala appears in the alphabetical list of ODBC drivers that are installed on your system.
3. 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.
4. Click **Add**
5. In the Create New Data Source dialog box, select **Cloudera ODBC Driver for Impala** and then click **Finish**
6. Use the options in the Cloudera ODBC Driver for Impala DSN Setup dialog box to configure your DSN:
   a) In the Data Source Name field, type a name for your DSN.
   b) Optionally, in the Description field, type relevant details about the DSN.
   c) In the Host field, type the IP address or host name of the network load balancer (NLB) or one of the Impala nodes if you are deployed without an NLB.
   d) In the Port field, type the number of the TCP port on which the Impala server is listening.

      Note:

      The default port number for the Impala service is 21050.

   e) 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 Impala command prompt.

   f) In the Authentication area, configure authentication as needed. For more information, see "Configuring Authentication" on page 25

      Note:

      The default configuration of Impala requires the Cloudera ODBC Driver for Impala to be configured to use the No Authentication mechanism.

g) Optionally, if the operations against Impala 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.

h) To configure advanced driver options, click **Advanced Options**. For more information, see "Configuring Advanced Options" on page 12.

i) To configure server-side properties, click **Advanced Options** and then click **Server Side Properties**. For more information, see "Configuring Server-Side Properties" on page 13.

j) To configure logging behavior for the driver, click **Logging Options**. For more information, see "Configuring Logging Options" on page 14.

7. 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 Impala DSN Setup dialog box are correct. Contact your Impala server administrator as needed.

8. To save your settings and close the Cloudera ODBC Driver for Impala DSN Setup dialog box, click **OK**
9. To close the ODBC Data Source Administrator, click **OK**

Configuring Authentication

The Impala server supports multiple authentication mechanisms. You must determine the authentication type your server is using and configure your DSN accordingly. The available authentication methods are as follows:

- No Authentication
- Kerberos
- SASL User Name
- SASL User Name and Password
- SASL User Name and Password (SSL)
- No Authentication (SSL)
- NOSASL User Name and Password

Using No Authentication

For this authentication mechanism, you do not need to configure any additional settings.

Note:

The default configuration of Impala requires the Cloudera ODBC Driver for Impala to be configured to use the No Authentication mechanism.

To configure a connection without authentication:

1. To access authentication options, open the ODBC Data Source Administrator where you created the DSN, then select the DSN, and then click **Configure**
2. In the Mechanism list, select **No Authentication**
3. 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 Appendix B "Configuring Kerberos Authentication for Windows" on page 34.

To configure Kerberos authentication:

1. To access authentication options, open the ODBC Data Source Administrator where you created the DSN, then select the DSN, and then click **Configure**
2. In the Mechanism list, select **Kerberos**
3. If your Kerberos setup does not define a default realm or if the realm of your Impala server host is not the default, then type the Kerberos realm of the Impala server host in the **Realm** field.

   OR

   To use the default realm defined in your Kerberos setup, leave the Realm field empty.
4. In the Host FQDN field, type the fully qualified domain name of the Impala server host.
5. In the Service Name field, type the service name of the Impala server.
6. Optionally, in the Transport Buffer Size field, type the number of bytes to reserve in memory for buffering unencrypted data from the network.

   Note:

   In most circumstances, the default value of 1000 bytes is optimal.
7. To save your settings and close the dialog box, click **OK**

Using SASL User Name

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

To configure SASL User Name authentication:

1. To access authentication options, open the ODBC Data Source Administrator where you created the DSN, then select the DSN, and then click **Configure**
2. In the Mechanism list, select **SASL User Name**
3. In the User Name field, type an appropriate user name for accessing the Impala server.
4. Optionally, in the Transport Buffer Size field, type the number of bytes to reserve in memory for buffering unencrypted data from the network.

> **Note:**
>
> In most circumstances, the default value of 1000 bytes is optimal.

5. To save your settings and close the dialog box, click **OK**

Using SASL User Name and Password

This authentication mechanism requires a user name and a password.

> **Note:**
>
> This authentication mechanism should not be used with an Impala configuration that does not have LDAP enabled.

To configure SASL User Name and Password authentication:

1. To access authentication options, open the ODBC Data Source Administrator where you created the DSN, then select the DSN, and then click **Configure**
2. In the Mechanism list, select **SASL User Name and Password**
3. In the User Name field, type an appropriate user name for accessing the Impala server.
4. In the Password field, type the password corresponding to the user name you typed in step 3.
5. Optionally, in the Transport Buffer Size field, type the number of bytes to reserve in memory for buffering unencrypted data from the network.

   > **Note:**
   >
   > In most circumstances, the default value of 1000 bytes is optimal.

6. To save your settings and close the dialog box, click **OK**

Using SASL User Name and Password (SSL)

This authentication mechanism uses SSL and requires a user name and a password. The driver accepts self-signed SSL certificates for this authentication mechanism.

> **Note:**
>
> This authentication mechanism should not be used with an Impala configuration that does not have LDAP enabled.

To configure SASL User Name and Password (SSL) authentication:

1. To access authentication options, open the ODBC Data Source Administrator where you created the DSN, then select the DSN, and then click **Configure**
2. In the Mechanism list, select **SASL User Name and Password (SSL)**
3. In the User Name field, type an appropriate user name for accessing the Impala server.
4. In the Password field, type the password corresponding to the user name you typed in step 3.

5. Optionally, in the Transport Buffer Size field, type the number of bytes to reserve in memory for buffering unencrypted data from the network.

   Note:

   In most circumstances, the default value of 1000 bytes is optimal.

6. Optionally, configure the driver to allow the common name of a CA-issued certificate to not match the host name of the Impala server by clicking **Advanced Options**, and then selecting the **Allow Common Name Host Name Mismatch** check box.

   Note:

   For self-signed certificates, the driver always allows the common name of the certificate to not match the host name.

7. To configure the driver to load SSL certificates from a specific PEM file, click **Advanced Options**, and then type the 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.

8. To save your settings and close the dialog box, click **OK**

Using No Authentication (SSL)

This authentication mechanism uses SSL but does not require a user name or a password. The driver accepts self-signed SSL certificates.

To configure No Authentication (SSL):

1. 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**
2. In the Mechanism list, select **No Authentication (SSL)**
3. Optionally, configure the driver to allow the common name of a CA-issued certificate to not match the host name of the Impala server by clicking **Advanced Options**, and then selecting the **Allow Common Name Host Name Mismatch** check box.

   Note:

   For self-signed certificates, the driver always allows the common name of the certificate to not match the host name.

4. To configure the driver to load SSL certificates from a specific PEM file, click **Advanced Options**, and then type the 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.

5. To save your settings and close the dialog box, click **OK**

Using NOSASL User Name and Password

This authentication mechanism requires a user name and a password, but does not use SASL (Simple Authentication and Security Layer).

To configure NOSASL User Name and Password authentication:

1. 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**
2. In the Mechanism list, select **NOSASL User Name and Password**
3. In the User Name field, type an appropriate user name for accessing the Impala server.
4. In the Password field, type the password corresponding to the user name you typed in step 3.
5. 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. To access advanced options, open the ODBC Data Source Administrator where you created the DSN, then select the DSN, then click **Configure**, and then click **Advanced Options**
2. To disable translation from ODBC SQL to Impala SQL, select the **Use Native Query** check box.

   Note:

   By default, the driver applies transformations to the queries emitted by an application to convert the queries into an equivalent form in Impala SQL. If the application is Impala-aware and already emits Impala SQL, then turning off the translation avoids the additional overhead of query transformation.
3. To enable the driver to successfully run queries that contain transaction statements, select the **Enable Simulated Transactions** check box.

   Note:

   The transaction statements will not be executed, because ODBC does not support them. Enabling this option allows the driver to run the query without returning error messages.
4. 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 **Use SQL Unicode Types** check box.
5. To handle Kerberos authentication using the SSPI plugin instead of MIT Kerberos by default, select one or both of the check boxes under the Use Only SSPI Plugin option:
   - To configure the current DSN to use the SSPI plugin by default, select **Enable for this DSN**

- To configure all DSN-less connections to use the SSPI plugin by default, select **Enable for DSN-less connections**
- To configure all connections that use the Cloudera ODBC Driver for Impala to use the SSPI plugin by default, select both check boxes.

6. In the Rows fetched per block field, type the number of rows to be fetched per block.
7. In the Socket timeout field, type the number of seconds after which Impala closes the connection with the client application if the connection is idle.

   Note:

   Setting the Socket timeout value to 0 disables the timeout feature.

8. In the String Column Length field, type the maximum data length for STRING columns.
9. To allow the common name of a CA-issued SSL certificate to not match the host name of the Impala server, select the **Allow Common Name Host Name Mismatch** check box.

   Note:

   This option is applicable only to the SASL User Name and Password (SSL) and No Authentication (SSL) authentication mechanisms.

10. To configure the driver to load SSL certificates from a specific PEM file, type the 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.

    Note:

    This option is applicable only to the SASL User Name and Password (SSL) and No Authentication (SSL) authentication mechanisms.

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

Configuring Server-Side Properties

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

To configure server-side properties:

1. To configure server-side properties, 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 **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**
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 force the driver to convert server-side property key names to all lower case characters, select the **Convert Key Name to Lower Case** check box.
6. 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 Impala, the ODBC Data Source Administrator provides tracing functionality.

Important:

Only enable logging long enough to capture an issue. Logging 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. Table 1 lists the logging levels provided by the Cloudera ODBC Driver for Impala, in order from least verbose to most verbose.

Table 1. Cloudera ODBC Driver for Impala Logging Levels

Logging Level	Description
OFF	Disables all logging.
FATAL	Logs very severe error events that will lead the driver to abort.
ERROR	Logs error events that might still 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 more detailed information than the DEBUG level.

To enable the logging functionality available in the Cloudera ODBC Driver for Impala:

1. In the Cloudera Impala ODBC Driver DSN Setup dialog box, click **Logging Options**
2. In the Log Level list, select the desired level of information to include in log files.
3. In the Log Path field, type 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. Click **OK**

The Cloudera ODBC Driver for Impala produces a log file named ImpalaODBC_driver.log at the location you specify using the Log Path field.

To disable Cloudera ODBC Driver for Impala logging:

1. In the Cloudera Impala ODBC Driver DSN Setup dialog box, click **Logging Options**
2. In the Log Level 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 the article *How to Generate an ODBC Trace with ODBC Data Source Administrator* at http://support.microsoft.com/kb/274551

Linux Driver

System Requirements

You install the Cloudera ODBC Driver for Impala on client computers accessing data in a Hadoop cluster with the Impala service installed and running. Each computer where you install the driver must meet the following minimum system requirements:

- One of the following distributions (32- and 64-bit editions are supported):
  - Red Hat® Enterprise Linux® (RHEL) 5.0 or 6.0
  - CentOS 5.0 or 6.0
  - SUSE Linux Enterprise Server (SLES) 11
- 50 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 Cloudera ODBC Driver for Impala has been tested using Impala 1.0.1 and Apache Thrift 0.9.0

Installing the Driver

There are two versions of the driver for Linux:

- **ClouderaImpalaODBC-32bit-*Version-Release.LinuxDistro*.i686.rpm** for the 32-bit driver
- **ClouderaImpalaODBC-*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 / Impala-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:

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

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

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

To install the Cloudera ODBC Driver for Impala:

1. 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 Impala 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 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.

Verifying the Version Number

If you need to verify the version of the Cloudera ODBC Driver for Impala that is installed on your Linux 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:

  ```
  yum list | grep ClouderaImpalaODBC
  ```

  OR

  Run the following command:

  ```
  rpm -qa | grep ClouderaImpalaODBC
  ```

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

Setting the LD_LIBRARY_PATH Environment Variable

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 Impala, see "Configuring ODBC Connections for Non-Windows Platforms" on page 20.

Mac OS X Driver

System Requirements

You install the Cloudera ODBC Driver for Impala on client computers accessing data in a Hadoop cluster with the Impala service installed and running. Each computer where you install the driver must meet the following minimum system requirements:

- Mac OS X version 10.6.8 or later
- 100 MB of available disk space
- iODBC 3.52.7 or later

The Cloudera ODBC Driver for Impala has been tested using Impala 1.0.1 and Apache Thrift 0.9.0. The driver supports both 32- and 64-bit client applications.

Installing the Driver

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

- **/opt/cloudera/impalaodbc** contains release notes and the *Cloudera ODBC Driver for Impala Installation and Configuration Guide* in PDF format.
- **/opt/cloudera/impalaodbc/ErrorMessages** contains error messages required by the driver.
- **/opt/cloudera/impalaodbc/Setup** contains sample configuration files named odbc.ini and odbcinst.ini
- **/opt/cloudera/impalaodbc/lib/universal** contains the driver binaries and the clouder-a.impalaodbc.ini configuration file.

To install the Cloudera ODBC Driver for Impala:

1. Double-click **ClouderaImpalaODBC.dmg** to mount the disk image.
2. Double-click **ClouderaImpalaODBC.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**

For information about creating ODBC connections using the Cloudera ODBC Driver for Impala, see "Configuring ODBC Connections for Non-Windows Platforms" on page 20.

Verifying the Version Number

If you need to verify the version of the Cloudera ODBC Driver for Impala 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 cloudera.impalaodbc
```

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

AIX Driver

System Requirements

You install the Cloudera ODBC Driver for Impala on client computers accessing data in a Hadoop cluster with the Impala service installed and running. Each computer where you install the driver 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 Cloudera ODBC Driver for Impala has been tested using Impala 1.0.1 and Apache Thrift 0.9.0

Installing the Driver

There are two versions of the driver for AIX:

- **ClouderaImpalaODBC-32bit-*Version*-*Release*.ppc.rpm** for the 32-bit driver
- **ClouderaImpalaODBC-*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 / Impala-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.

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

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

To install the Cloudera ODBC Driver for Impala:

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
   ```

Verifying the Version Number

If you need to verify the version of the Cloudera ODBC Driver for Impala 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 ClouderaImpalaODBC
  ```

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

Setting the LD_LIBRARY_PATH Environment Variable

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 Impala, see "Configuring ODBC Connections for Non-Windows Platforms" on page 20.

Configuring ODBC Connections for Non-Windows Platforms

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

- "Configuring Logging Options" on page 28
- "Configuring Authentication" on page 25

Files

ODBC driver managers use configuration files to define and configure ODBC data sources and drivers. By default, the following configuration files residing in the user's home directory are used:

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

Also, by default the Cloudera ODBC Driver for Impala is configured using the cloudera.impalaodbc.ini file, which is located in one of the following directories depending on the version of the driver that you are using:

- **/opt/cloudera/impalaodbc/lib/32** for the 32-bit driver on Linux/AIX
- **/opt/cloudera/impalaodbc/lib/64** for the 64-bit driver on Linux/AIX
- **/opt/cloudera/impalaodbc/lib/universal** for the driver on Mac OS X

The cloudera.impalaodbc.ini file is required.

Note:

The cloudera.impalaodbc.ini file in the /lib subfolder provides default settings for most configuration options available in the Cloudera ODBC Driver for Impala.

You can set driver configuration options in your odbc.ini and cloudera.impalaodbc.ini files. Configuration options set in a cloudera.impalaodbc.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.impalaodbc.ini. For information about the configuration options available for controlling the behavior of DSNs that are using the Cloudera ODBC Driver for Impala, see Appendix C "Driver Configuration Options" on page 38.

Sample 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 Impala.

The names of the sample configuration files do not begin with a period (.) so that they will appear in directory listings by default. A filename beginning with a period (.) is hidden. For odbc.ini and odbcinst.ini, if the default location is used, then the filenames 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 CLOUDERAIMPALAINI—to specify different locations for the odbc.ini, odbcinst.ini, and cloudera.impalaodbc.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 CLOUDERAIMPALAINI to point to your cloudera.impalaodbc.ini file.

For example, if your odbc.ini and cloudera.impalaodbc.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 CLOUDERAIMPALAINI=/etc/cloudera.impalaodbc.ini
```

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

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

   Important:

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

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

Configuring the odbc.ini File

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

- **[ODBC]** is optional and used to control global ODBC configuration, such as ODBC tracing.
- **[ODBC Data Sources]** is required, listing DSNs and associating DSNs 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:

```
[ODBC Data Sources]
Sample Cloudera Impala DSN 32=Cloudera Impala ODBC Driver 32-bit
[Sample Cloudera Impala DSN 32]
Driver=/opt/cloudera/impalaodbc/lib/32/libclouderaimpalaodbc32.so
HOST=MyImpalaServer
PORT=21050
```

MyImpalaServer is the IP address or host name of the Impala server.

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

```
[ODBC Data Sources]
Sample Cloudera Impala DSN=Cloudera Impala ODBC Driver
[Sample Cloudera Impala DSN]
Driver=/opt/cloudera/impalaodbc/lib/universal/libclouderaimpalaodbc
.dylib
HOST=MyImpalaServer
PORT=21050
```

MyImpalaServer is the IP address or host name of the Impala server.

To create a Data Source Name:

1. Open the .odbc.ini configuration file in a text editor.
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. In the .odbc.ini file, add a new section with a name that matches the DSN you specified in step 2, and then add configuration options to the section. Specify configuration options as key-value pairs.

   Note:

   The default configuration of Impala requires the Cloudera ODBC Driver for Impala to be configured to use the No Authentication mechanism.
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 Impala, see Appendix C "Driver Configuration Options" on page 38.

Configuring the odbcinst.ini File

ODBC drivers are defined in the odbcinst.ini configuration file. The configuration file is optional because drivers can be specified directly in the odbc.ini configuration file, as described in "Configuring the odbc.ini File" on page 22.

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

- **[ODBC Drivers]** lists the names of all the installed ODBC drivers.
- A section having the same name as the driver name specified in the [ODBC Drivers] section lists driver attributes and values.

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

```
[ODBC Drivers]
Cloudera Impala ODBC Driver 32-bit=Installed
Cloudera Impala ODBC Driver 64-bit=Installed
[Cloudera Impala ODBC Driver 32-bit]
Description=Cloudera Impala ODBC Driver (32-bit)
Driver=/opt/cloudera/impalaodbc/lib/32/libclouderaimpalaodbc32.so
[Cloudera Impala ODBC Driver 64-bit]
Description=Cloudera Impala ODBC Driver (64-bit)
Driver=/opt/cloudera/impalaodbc/lib/64/libclouderaimpalaodbc64.so
```

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

```
[ODBC Drivers]
Cloudera Impala ODBC Driver=Installed
[Cloudera Impala ODBC Driver]
Description=Cloudera Impala ODBC Driver
Driver=/opt/cloudera/impalaodbc/lib/universal/libclouderaimpalaodbc
.dylib
```

To define a driver:

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

   Note:

   Type a symbolic name that you want to use to refer to the driver in connection strings or DSNs.
3. In the .odbcinst.ini file, add a new section with a name that matches the driver name you typed in step 2, and then add configuration options to the section based on the sample odbcinst.ini file provided in the Setup directory. Specify configuration options as key-value pairs.
4. Save the .odbcinst.ini configuration file.

Configuring the cloudera.impalaodbc.ini File

The cloudera.impalaodbc.ini file contains configuration settings for the Cloudera ODBC Driver for Impala. Settings that you define in the cloudera.impalaodbc.ini file apply to all connections that

use the driver.

To configure the Cloudera ODBC Driver for Impala to work with your ODBC driver manager:

1. Open the cloudera.impalaodbc.ini configuration file in a text editor.
2. Edit the DriverManagerEncoding setting. The value is usually **UTF-16** or **UTF-32** if you are using Linux/Mac OS X, depending on the ODBC driver manager you use. iODBC uses **UTF-32**, and unixODBC uses **UTF-16**. To determine the correct setting to use, refer to your ODBC Driver Manager documentation.

   OR

   If you are using AIX and the unixODBC driver manager, then set the value to **UTF-16**. If you are using AIX and the iODBC driver manager, then set the value to **UTF-16** for the 32-bit driver or **UTF-32** for the 64-bit driver.
3. Edit the ODBCInstLib setting. The value is the name of the ODBCInst shared library for the ODBC driver manager you use. To determine the correct library to specify, refer to your ODBC driver manager documentation.

   The configuration file defaults to the shared library for iODBC. In Linux/AIX, the shared library name for iODBC is **libiodbcinst.so**. In Mac OS X, the shared library name for iODBC is **libiodbcinst.dylib**.

   Note:

   You can specify an absolute or relative filename for the library. For Linux/AIX, if you intend to use the relative filename, then the path to the library must be included in the LD_LIBRARY_PATH environment variable.
4. Optionally, configure logging by editing the LogLevel and LogPath settings. For more information, see "Configuring Logging Options" on page 28.
5. Save the cloudera.impalaodbc.ini configuration file.

Configuring Authentication

The Impala server supports multiple authentication mechanisms. You must determine the authentication type your server is using and configure your DSN accordingly. The available authentication methods are as follows:

- No Authentication
- Kerberos
- SASL User Name
- SASL User Name and Password
- SASL User Name and Password (SSL)
- No Authentication (SSL)
- NOSASL User Name and Password

Using No Authentication

For this authentication mechanism, you do not need to configure any additional settings.

Note:

The default configuration of Impala requires the Cloudera ODBC Driver for Impala to be configured to use the No Authentication mechanism.

To configure a connection without authentication:

- Set the AuthMech connection attribute to 0

Using Kerberos

Kerberos must be installed and configured before you can use this authentication mechanism. For more information, refer to the MIT Kerberos documentation.

To configure Kerberos authentication:

1. Set the AuthMech connection attribute to 1
2. If your Kerberos setup does not define a default realm or if the realm of your Impala server is not the default, then set the appropriate realm using the KrbRealm attribute.

   OR

   To use the default realm defined in your Kerberos setup, do not set the KrbRealm attribute.
3. Set the KrbFQDN attribute to the fully qualified domain name of the Impala server host.
4. Set the KrbServiceName attribute to the service name of the Impala server.
5. Optionally, set the TSaslTransportBufSize attribute to the number of bytes to reserve in memory for buffering unencrypted data from the network.

   Note:

   In most circumstances, the default value of 1000 bytes is optimal.

Using SASL User Name

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

To configure SASL User Name authentication:

1. Set the AuthMech connection attribute to 2
2. Set the UID attribute to an appropriate user name for accessing the Impala server.
3. Optionally, set the TSaslTransportBufSize attribute to the number of bytes to reserve in memory for buffering unencrypted data from the network.

   Note:

   In most circumstances, the default value of 1000 bytes is optimal.

Using SASL User Name and Password

This authentication mechanism requires a user name and a password.

Note:

This authentication mechanism should not be used with an Impala configuration that does not have LDAP enabled.

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 Impala server.
3. Set the PWD attribute to the password corresponding to the user name you provided in step 2.
4. Optionally, set the TSaslTransportBufSize attribute to the number of bytes to reserve in memory for buffering unencrypted data from the network.

   Note:

   In most circumstances, the default value of 1000 bytes is optimal.

Using SASL User Name and Password (SSL)

This authentication mechanism uses SSL and requires a user name and a password. The driver accepts self-signed SSL certificates for this authentication mechanism.

Note:

This authentication mechanism should not be used with an Impala configuration that does not have LDAP enabled.

To configure SASL User Name and Password (SSL) authentication:

1. Set the AuthMech connection attribute to 4
2. Set the UID attribute to an appropriate user name for accessing the Impala server.
3. Set the PWD attribute to the password corresponding to the user name you provided in step 2.
4. Optionally, set the TSaslTransportBufSize attribute to the number of bytes to reserve in memory for buffering unencrypted data from the network.

   Note:

   In most circumstances, the default value of 1000 bytes is optimal.

5. Optionally, configure the driver to allow the common name of a CA-issued certificate to not match the host name of the Impala server by setting the CAIssuedCertNamesMismatch attribute to 1.

   Note:

   For self-signed certificates, the driver always allows the common name of the certificate to not match the host name.

6. To configure the driver to load SSL certificates from a specific PEM file, set the TrustedCerts attribute to the path of the 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.

Using No Authentication (SSL)

This authentication uses SSL but does not require a user name or a password. The driver accepts self-signed SSL certificates.

To configure No Authentication (SSL):

1. Set the AuthMech connection attribute to 5
2. Optionally, configure the driver to allow the common name of a CA-issued certificate to not match the host name of the Impala server by setting the CAIssuedCertNamesMismatch attribute to 1.

   Note:

   For self-signed certificates, the driver always allows the common name of the certificate to not match the host name.

3. To configure the driver to load SSL certificates from a specific PEM file, set the TrustedCerts attribute to the path of the 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.

Using NOSASL User Name and Password

This authentication mechanism requires a user name and a password, but does not use SASL (Simple Authentication and Security Layer).

To configure NOSASL User Name and Password authentication:

1. Set the AuthMech connection attribute to 6
2. Set the UID attribute to an appropriate user name for accessing the Impala server.
3. Set the PWD attribute to the password corresponding to the user name you provided in step 2.

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. Table 2 lists the logging levels provided by the Cloudera ODBC Driver for Impala, in order from least verbose to most verbose.

Table 2. Cloudera ODBC Driver for Impala Logging Levels

LogLevel Value	Description
0	Disables all logging.
1	Logs very severe error events that will lead the driver to abort.
2	Logs error events that might still 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 more detailed information than LogLevel=5

To enable logging:

1. Open the cloudera.impalaodbc.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. Save the cloudera.impalaodbc.ini configuration file.

The Cloudera ODBC Driver for Impala produces a log file named ImpalaODBC_driver.log at the location you specify using the LogPath key.

To disable logging:

1. Open the cloudera.impalaodbc.ini configuration file in a text editor.
2. Set the LogLevel key to 0
3. Save the cloudera.impalaodbc.ini configuration file.

Features

More information is provided on the following features of the Cloudera ODBC Driver for Impala:

- "Data Types" on page 30
- "Catalog and Schema Support" on page 31

Data Types

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

Table 3 lists the supported data type mappings.

Table 3. Supported Data Types

Impala Type	SQL Type
BIGINT	SQL_BIGINT
BINARY	SQL_VARBINARY
BOOLEAN	SQL_BOOLEAN
CHAR **Note:** Only available in CDH 5.2 or later.	SQL_CHAR **Note:** SQL_WCHAR is returned instead if the Use SQL Unicode Types configuration option (the UseUnicodeSqlCharacterTypes key) is enabled.
DATE	SQL_DATE
DECIMAL **Note:** Only available in CDH 5.2 or later.	SQL_DECIMAL
DOUBLE **Note:** REAL is an alias for DOUBLE.	SQL_DOUBLE
INT	SQL_INTEGER

Impala Type	SQL Type
FLOAT	SQL_REAL
SMALLINT	SQL_SMALLINT
TINYINT	SQL_TINYINT
TIMESTAMP	SQL_TIMESTAMP
VARCHAR **Note:** Only available in CDH 5.2 or later.	SQL_VARCHAR **Note:** SQL_WVARCHAR is returned instead if the Use SQL Unicode Types configuration option (the UseUnicodeSqlCharacterTypes key) is enabled.

Note:

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

Catalog and Schema Support

The Cloudera ODBC Driver for Impala supports both catalogs and schemas in order to make it easy for the driver to work with various ODBC applications. Since Impala only organizes tables into schemas/databases, the driver provides a synthetic catalog called “IMPALA” under which all of the schemas/databases are organized. The driver also maps the ODBC schema to the Impala schema/database.

SQL Translation

The Cloudera ODBC Driver for Impala can parse queries locally before sending them to the Impala server. This feature allows the driver to calculate query metadata without executing the query, support query parameters, and support extra SQL features such as ODBC escape sequences and additional scalar functions that are not available in the Impala-shell tool.

Server-Side Properties

The Cloudera ODBC Driver for Impala 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.

For more information about setting server-side properties when using the Windows driver, see "Configuring Server-Side Properties" on page 13. For information about setting server-side

properties when using the driver on a non-Windows platform, see "Driver Configuration Options" on page 38.

Active Directory

The Cloudera ODBC Driver for Impala 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.

Contact Us

If you have difficulty using the driver, please contact our Technical Support staff. We welcome your questions, comments, and feature requests.

Important:

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

For details on contacting Technical Support, see http://www.cloudera.com/content/cloudera/en/products/cloudera-support.html

Appendix A Authentication Options

Impala supports multiple authentication mechanisms. You must determine the authentication type that your server is using. The authentication methods available in the Cloudera ODBC Driver for Impala are as follows:

- No Authentication
- Kerberos
- SASL User Name
- SASL User Name and Password
- SASL User Name and Password (SSL)
- No Authentication (SSL)
- NOSASL User Name and Password

Note:

The default configuration of Impala requires the Cloudera ODBC Driver for Impala to be configured to use the No Authentication mechanism.

The Impala server uses SASL (Simple Authentication and Security Layer) to support some of the authentication methods. Kerberos is supported with the SASL GSSAPI mechanism. SASL User Name, User Name and Password, and SASL User Name and Password (SSL) are supported with the SASL PLAIN mechanism.

Table 4. Impala Authentication Mechanisms

SASL mechanisms	Non-SASL mechanisms
Kerberos	No Authentication
SASL User Name	No Authentication (SSL)
SASL User Name and Password	NOSASL User Name and Password
SASL User Name and Password (SSL)	

Note:

Thrift (the layer for handling remote process communication between the Cloudera ODBC Driver for Impala and the Impala server) has a limitation where it cannot detect a mix of non-SASL and SASL mechanisms being used between the driver and the server. If this happens, the driver will appear to hang during connection establishment.

Appendix B Configuring Kerberos Authentication for Windows

Active Directory

The Cloudera ODBC Driver for Impala 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.

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 at http://web.mit.edu/kerberos/

To download and install MIT Kerberos for Windows 4.0.1:

1. To download the Kerberos installer for 64-bit computers, 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

   OR

   To download the Kerberos installer for 32-bit computers, 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 in step 1.
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 a .INI file in the default location—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 from your Kerberos administrator.

   OR

   Obtain the configuration file from the /etc/krb5.conf folder on the computer that is hosting the Impala server.
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 from your Kerberos administrator.

   OR

   Obtain the configuration file from the /etc/krb5.conf folder on the computer that is hosting the Impala server.
2. Place the krb5.conf file in an accessible directory and make note of the full path name.
3. Click the **Start** button , then right-click **Computer**, 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 absolute path to the krb5.conf file from step 2.
9. Click **OK** to save the new variable.
10. Ensure 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. Click the **Start** button , then right-click **Computer**, 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 in step 1, and then append the file name krb5cache. For example, if you created the folder C:\temp in step 1, 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, ensure that the krb5cache file does not already exist as a file or a directory.
8. Click **OK** to save the new variable.
9. Ensure 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 ensure that Kerberos uses the new settings, restart your computer.

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. Click the **Start** button , then click **All Programs**, and then click the **Kerberos for Windows (64-bit)** or **Kerberos for Windows (32-bit)** program group.
2. Click **MIT Kerberos Ticket Manager**
3. In the MIT Kerberos Ticket Manager, click **Get Ticket**
4. 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 the MIT Kerberos Ticket Manager.

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

1. Click the Start button , then click All Programs, then click Accessories, and then click Command Prompt
2. In the Command Prompt, type a command using the following syntax:

   `kinit -k -t keytab_path principal`

   keytab_path 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. Click the Start button , then click All Programs, then click Accessories, 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.

Appendix C Driver Configuration Options

Appendix C "Driver Configuration Options" on page 38 lists the configuration options available in the Cloudera ODBC Driver for Impala alphabetically by field or button label. Options having only key names—not appearing in the user interface of the driver—are listed alphabetically by key name.

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

- Cloudera ODBC Driver for Impala DSN Setup
- Advanced Options
- Server Side Properties

When using a connection string or configuring a connection from a Linux, Mac OS X, or AIX computer, 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.impalaodbc.ini files. Configuration options set in a cloudera.impalaodbc.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.impalaodbc.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 Impala, or via the key name when using a connection string or configuring a connection from a Linux/Mac OS X/AIX computer:

Allow Common Name Host Name Mismatch

Key Name	Default Value	Required
CAIssuedCertNamesMismatch	Clear (0)	No

Description

When this option is enabled (1), the driver allows a CA-issued SSL certificate name to not match the host name of the Impala server.

When this option is disabled (0), the CA-issued SSL certificate name must match the host name of the Impala server.

Note:

This setting is applicable only to the SASL User Name and Password (SSL) and No Authentication (SSL) authentication mechanisms.

Convert Key Name to Lower Case

Key Name	Default Value	Required
LCaseSspKeyName	Selected (1)	No

Description

When this option is enabled (1), the driver converts server-side property key names to all lower case characters.

When this option is disabled (0), the driver does not modify the server-side property key names.

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, type the **show databases** command at the Impala command prompt.

Delegation UID

Key Name	Default Value	Required
DelegationUID	None	No

Description

Use this option to delegate all operations against Impala to a user that is different than the authenticated user for the connection.

Enable Simulated Transactions

Key Name	Default Value	Required
EnableSimulatedTransactions	Clear (0)	No

Description

When this option is enabled (1), the driver simulates transactions, enabling queries that contain transaction statements to be run successfully. The transactions will not be executed.

When this option is disabled (0), the driver returns an error if it attempts to run a query that contains transaction statements.

Note:

ODBC does not support transaction statements, so they cannot be executed.

Host

Key Name	Default Value	Required
HOST	None	Yes

Description

The IP address or host name of the Impala server.

Host FQDN

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

Description

The fully qualified domain name of the Impala host.

Mechanism

Key Name	Default Value	Required
AuthMech	No Authentication (0)	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)
- SASL User Name (2)
- SASL User Name and Password (3)
- SASL User Name and Password (SSL) (4)
- No Authentication (SSL) (5)
- NOSASL User Name and Password (6)

Password

Key Name	Default Value	Required
PWD	None	Yes, if the authentication mechanism is one of the following: • SASL User Name and Password • SASL User Name and Password (SSL) • NOSASL User Name and Password

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	21050	Yes

Description

The number of the TCP port on which the Impala server is listening.

Realm

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

Description

The realm of the Impala host.

If your Kerberos configuration already defines the realm of the Impala 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.

Any positive 32-bit integer is a valid value, but testing has shown that performance gains are marginal beyond the default value of 10000 rows.

Service Name

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

Description

The Kerberos service principal name of the Impala server.

Socket Timeout

Key Name	Default Value	Required
SocketTimeout	0	No

Description

The number of seconds after which Impala closes the connection with the client application if the connection is idle.

When this option is set to 0, the connection does not time out.

String Column Length

Key Name	Default Value	Required
StringColumnLength	32767	No

Description

The maximum data length for STRING columns.

Transport Buffer Size

Key Name	Default Value	Required
TSaslTransportBufSize	1000	No

Description

The number of bytes to reserve in memory for buffering unencrypted data from the network.

Note:

In most circumstances, the default value of 1000 bytes is optimal.

Trusted Certificates

Key Name	Default Value	Required
TrustedCerts	The cacerts.pem file in the lib folder or 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.	No

Description

The location of the PEM file containing trusted CA certificates for authenticating the Impala server when using SSL.

If this option is not set, then the driver will default to using the trusted CA certificates PEM file installed by the driver.

This option is applicable only to the following authentication mechanisms:

- SASL User Name and Password (SSL)
- No Authentication (SSL)

Use Native Query

Key Name	Default Value	Required
UseNativeQuery	Clear (0)	No

Description

When this option is enabled (1), the driver does not transform the queries emitted by an application, so the native query is used.

When this option is disabled (0), the driver transforms the queries emitted by an application and converts them into an equivalent from in Impala SQL.

Note:

If the application is Impala-aware and already emits Impala SQL, 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

When the **Enable for this DSN** check box is selected (UseOnlySSPI is set to 1 in the DSN entry in the registry), the setting applies to the DSN only, and the driver handles Kerberos authentication in the DSN connection by using the SSPI plugin instead of Kerberos by default.

When the **Enable for DSN-less connections** check box is selected (UseOnlySSPI is set to 1 in the driver configuration section in the registry), the setting applies to DSN-less connections only, and the driver handles Kerberos authentication in DSN-less connections by using the SSPI plugin instead of Kerberos by default.

If you want all connections that use the Cloudera ODBC Driver for Impala to use the SSPI plugin by default, then enable Use Only SSPI Plugin for both DSN and DSN-less connections.

When this option is 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.

Use SQL Unicode Types

Key Name	Default Value	Required
UseUnicodeSqlCharacterTypes	Clear (0)	No

Description

When this option is enabled (1), the driver returns SQL_WVARCHAR for STRING and VARCHAR columns, and returns SQL_WCHAR for CHAR columns.

When this option is disabled (0), the driver returns SQL_VARCHAR for STRING and VARCHAR columns, and returns SQL_CHAR for CHAR columns.

User Name

Key Name	Default Value	Required
UID	For User Name authentication only, the default value is anonymous	No, if the authentication mechanism is SASL User Name Yes, if the authentication mechanism is one of the following: • SASL User Name and Password • SASL User Name and Password (SSL) • NOSASL User Name and Password

Description

The user name that you use to access the Impala server.

Configuration Options Having Only Key Names

The following configuration options do not appear in the Windows user interface for the Cloudera ODBC Driver for Impala and are only accessible when using a connection string or configuring a connection from a Linux/Mac OS X/AIX computer:

- "Driver" on page 46
- "SSP_" on page 47

Driver

Default Value	Required
The default value varies depending on the version of the driver that is installed. For example, the value for the Windows driver is different from the value of the Mac OS X driver.	Yes

Description

The name of the installed driver (Cloudera ODBC Driver for Impala) or the absolute path of the Cloudera ODBC Driver for Impala shared object file.

SSP_

Default Value	Required
None	No

Description

Set a server-side property by using the following syntax, where *SSPKey* is the name of the server-side property to set and *SSPValue* is the value to assign to the server-side 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.

Appendix D ODBC API Conformance Level

Table 5 lists the ODBC interfaces that the Cloudera ODBC Driver for Impala 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 http://msdn.microsoft.com/en-us/library/ms716246%28VS.85%29.aspx

Table 5. ODBC API Conformance Level

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
Core	SQLDriverconnect		Core	SQLSetDescField
Core	SQLEndTran		Core	SQLSetDescRec
Core	SQLExecDirect		Core	SQLSetEnvAttr
Core	SQLExecute		Core	SQLSetStmtAttr
Core	SQLFetch		Core	SQLSpecialColumns

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