

Cloudera ODBC Driver for Impala



Important Notice

© 2010-2013 Cloudera, Inc. All rights reserved.

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

Date: April 14, 2014

Table of Contents

INTRODUCTION	1
WINDOWS DRIVER.....	1
SYSTEM REQUIREMENTS.....	1
INSTALLING THE DRIVER.....	2
CONFIGURING ODBC CONNECTIONS	2
CONFIGURING AUTHENTICATION	5
<i>Using No Authentication.....</i>	<i>5</i>
<i>Using Kerberos</i>	<i>5</i>
<i>Using User Name</i>	<i>6</i>
<i>Using User Name and Password</i>	<i>6</i>
<i>Using User Name and Password (SSL)</i>	<i>6</i>
LINUX DRIVER	7
SYSTEM REQUIREMENTS.....	7
INSTALLATION.....	7
<i>Setting the LD_LIBRARY_PATH Environment Variable</i>	<i>8</i>
MAC OS X DRIVER.....	9
SYSTEM REQUIREMENTS.....	9
INSTALLATION.....	9
<i>Setting the DYLD_LIBRARY_PATH Environment Variable.....</i>	<i>9</i>
CONFIGURING ODBC CONNECTIONS FOR LINUX AND MAC OS X	10
FILES.....	10
SAMPLE FILES	10
CONFIGURING THE ENVIRONMENT.....	10
CONFIGURING THE ODBC.INI FILE.....	11
CONFIGURING THE ODBCINST.INI FILE.....	12
CONFIGURING THE CLOUDERA.IMPALAODBC.INI FILE	13
CONFIGURING AUTHENTICATION	14
<i>Using No Authentication.....</i>	<i>14</i>
<i>Using Kerberos</i>	<i>14</i>
<i>Using User Name</i>	<i>14</i>
<i>Using User Name and Password</i>	<i>15</i>

<i>Using User Name and Password (SSL)</i>	15
DATA TYPES	15
CATALOG AND SCHEMA SUPPORT	16
SQL TRANSLATION	16
ACTIVE DIRECTORY	16
CONTACT US	16
APPENDIX A: CONFIGURING KERBEROS AUTHENTICATION FOR WINDOWS	18
ACTIVE DIRECTORY	18
MIT KERBEROS.....	18
<i>Download and install MIT Kerberos for Windows 4.0.1</i>	18
<i>Set up the Kerberos configuration file in the default location</i>	18
<i>Set up the Kerberos configuration file in another location</i>	18
<i>Set up the Kerberos credential cache file</i>	19
<i>Obtain a ticket for a Kerberos principal using password</i>	19
<i>Obtain a ticket for a Kerberos principal using a keytab file</i>	20
<i>Obtain a ticket for a Kerberos principal using the default keytab file</i>	20
APPENDIX B: DRIVER CONFIGURATION OPTIONS	22
APPENDIX C: ODBC API CONFORMANCE LEVEL	27

Introduction

Welcome to the Cloudera ODBC Driver for Impala. 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.

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. Impala SQL is a subset of SQL-92. If an application is Impala-aware, then the driver is configurable to pass the query through. 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, refer to the section "SQL Translation" on page 16.

Cloudera ODBC Driver for Impala is available for Microsoft Windows, Linux, and Mac OS X. It complies with the ODBC 3.52 data standard and adds important functionality such as Unicode and 32- and 64-bit support for high-performance computing environments on all platforms. Any version of the ODBC driver will connect to an Impala server irrespective of the server's host OS.

The 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 Cloudera ODBC Driver for Impala on client computers accessing data in a Hadoop cluster with the Impala server 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
 - Windows® Server 2008 R2
- 25 MB of available disk space

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

Important:

To install the driver, you need Administrator privileges on the computer.

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 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 the **Change** button, 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**.
7. If you are installing a driver with an evaluation license and you have purchased a perpetual license, then copy the License.lic file you received via e-mail into the \lib subfolder in the installation folder you selected in step 4.

Configuring ODBC Connections

To create a Data Source Name (DSN):

1. Click the **Start** button .
2. Click **All Programs**.
3. Click the **Cloudera ODBC Driver for Impala 2.5 (64-bit)** or the **Cloudera ODBC Driver for Impala 2.5 (32-bit)** program group. If you installed both versions of the driver, you will see two program groups.

Because DSNs are bit-specific, select the version that matches the bitness of your application. For example, a DSN that is defined for the 32-bit driver will only be accessible from 32-bit applications.

4. Click **64-bit ODBC Administrator** or **32-bit ODBC Administrator**. The ODBC Data Source Administrator window opens.
5. Click the **Drivers** tab and verify that the Cloudera ODBC Driver for Impala appears in the list of ODBC drivers that are installed on your system.
6. Click the **System DSN** tab to create a system DSN or click the **User DSN** tab to create a user DSN.

Note:

A system DSN can be seen by all users that login to a workstation. A user DSN is specific to a user on the workstation. It can only be seen by the user who creates it.

7. Click **Add**. The Create New Data Source window opens.
8. Select **Cloudera ODBC Driver for Impala** and then click **Finish**. The Cloudera ODBC Driver for Impala DSN Setup dialog opens.
9. In the **Data Source Name** field, type a name for your DSN.
10. Optionally, type relevant details related to the DSN in the **Description** field.
11. In the **Host** text box, type the IP address or hostname of the network load balancer or one of the Impala nodes if you are deployed without an NLB.
12. In the **Port** text box, type the listening port for the Impala service (default is 21050).
13. Optionally, configure authentication. For detailed instructions, refer to the section "Configuring Authentication" on page 5.
14. Optionally, click **Advanced Options**. In the Advanced Options window:
 - a) Select the **Use Native Query** checkbox to disable translating ODBC SQL to Impala SQL.

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.

- b) In the **Rows Fetched Per Block** field, type the number of rows to be fetched per block.

Note:

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.

- c) 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.

- d) To allow the common name of a CA issued SSL certificate to not match the hostname of the Impala server, select the **Allow Common Name Hostname Mismatch** checkbox.

Note:

This setting is only applicable to User Name and Password (SSL) authentication mechanism and will be ignored by other authentication mechanisms.

- e) Enter the path of the file containing the trusted certificates (e.g. certificate from the Impala Server) in the **Trusted Certificates** edit box to configure the driver to load the certificates from the specified file to authenticate the Impala server when using SSL.

Note:

This is only applicable to **User Name and Password (SSL)** authentication mechanisms, and will be ignored by other authentication mechanisms.

Note:

SSL certificates in the trusted certificates file has to be in the **PEM** format.

Note:

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

- f) Click **OK**.
15. 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, enter the user name of the user to be delegated in the **Delegation UID** text box.
 16. Click **Test** to test the connection. Review the results of testing the connection as needed, and then click **OK** in the Test Results dialog.
 17. In the Cloudera ODBC Driver for Impala DSN Setup dialog, click **OK**.

For details on the keys involved in configuring authentication, see “Appendix B: Driver Configuration Options” on page 22.

Configuring Authentication

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

- No Authentication
- User Name
- Kerberos

Using No Authentication

No additional details are required when using **No Authentication**.

Using Kerberos

To use **Kerberos** authentication, Kerberos must be configured prior to use. See “Appendix A: Configuring Kerberos Authentication for Windows” on page 18 for details. After Kerberos has been installed and configured, then configure your DSN to use Kerberos.

To configure your DSN to use Kerberos authentication:

1. In the **Cloudera ODBC Driver for Impala DSN Setup** dialog, click the drop-down arrow next to the **Mechanism** field, and then select **Kerberos**.
2. If there is no default realm configured for your Kerberos setup, then type the value for the Kerberos realm of the Impala server host in the **Realm** field. Otherwise, leave the field blank. You only need to provide the realm if your Kerberos setup does not define a default realm or if the realm of your Impala server is not the default.
3. In the **Host FQDN** field, type the value for the fully qualified domain name of the Impala host.
4. In the **Service Name** field, type the value for the service name of the Impala server. For example, if the principle for the Impala server is "impala/[fully.qualified.domain.name@YOUR-REALM.COM](#)", then the value in the service name field should be **impala**. If you are unsure of the correct service name to use for your particular Hadoop deployment, see your Hadoop administrator.
5. 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.

Using User Name

Authenticating by user name does not use a password. The user name labels the session, facilitating database tracking.

To configure your DSN for user name authentication:

1. In the **Cloudera ODBC Driver for Impala DSN Setup** dialog box, select **User Name** in the **Mechanism** field, and then type an appropriate credential in the **User Name** field.
2. 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.

Using User Name and Password

To configure your DSN for **User Name and Password** authentication:

1. In the Cloudera ODBC Driver for Impala DSN Setup dialog, click the drop-down arrow next to the **Mechanism** field, and then select **User Name and Password**.
2. In the **User Name** field, type an appropriate credential.
3. In the **Password** field, type the password corresponding to the user name you typed in step 2.

Using User Name and Password (SSL)

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

1. Click the drop-down arrow next to the **Mechanism** field, and then select **User Name and Password (SSL)**.
2. In the **User Name** field, type an appropriate credential.
3. In the **Password** field, type the password corresponding to the user name you typed in step 2.

Note:

The driver always accepts the use of self-signed SSL certificate.

Note:

Optionally you can configure the **Allow Common Name Host Name Mismatch** setting to control whether the driver allows the common name of a CA issued certificate to not match the host name of the Impala server. For self-signed certificates the driver always allows the common name of the certificate to not match the host name.

Note:

Optionally you can configure the **Trusted Certificates** setting in the **Advanced Options** to allow the driver to load SSL certificates from the specified file. The driver will default to using the trusted CA certificates PEM file install with the driver if this setting is not set.

Linux Driver

System Requirements

- Red Hat® Enterprise Linux® (RHEL) 5.0/6.0, CentOS 5.0/6.0 or SUSE Linux Enterprise Server (SLES) 11. Both 32- and 64-bit editions are supported.
- 50 MB of available disk space.
- An installed ODBC driver manager. Cloudera ODBC Driver for Impala has been tested against:
 - iODBC 3.52.7
 - The following versions of unixODBC:
 - 2.3.0
 - 2.3.1

Cloudera ODBC Driver for Impala requires a Hadoop cluster with the Impala server installed and running. Cloudera ODBC Driver for Impala has been tested using Impala 1.0.1 and Apache Thrift 0.9.0.

Installation

There are two versions of the driver for Linux:

- **ClouderaImpalaODBC-32-bit-BuildNumber-ReleaseNumber.i686.rpm** for 32-bit
- **ClouderaImpalaODBC-BuildNumber-ReleaseNumber.x86_64.rpm** for 64-bit

The version 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.

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

- `/opt/cloudera/impalaodbc/ErrorMessage` – Error messages files directory
- `/opt/cloudera/impalaodbc/Setup` – Sample configuration files directory
- `/opt/cloudera/impalaodbc/lib/32` – 32-bit shared libraries directory
- `/opt/cloudera/impalaodbc/lib/64` – 64-bit shared libraries directory

To install Cloudera ODBC Driver for Impala:

1. In Red Hat Enterprise Linux 5.0/6.0 or CentOS 5.0/6.0, 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 11, 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
```

2. If you are installing a driver with an evaluation license and you have purchased a perpetual license, then copy the License.lic file you received via e-mail into the `/opt/cloudera/impalaodbc/lib/32` or `/opt/cloudera/impalaodbc/lib/64` folder, depending on the version of the driver you installed.

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.

Setting the LD_LIBRARY_PATH Environment Variable

The LD_LIBRARY_PATH environment variable must include the paths to:

- Installed ODBC driver manager libraries
- Installed Cloudera ODBC Driver for Impala shared libraries

Important:

While you can have both 32- and 64-bit versions of the driver installed at the same time on the same computer, do *not* include the paths to both 32- and 64-bit shared libraries in LD_LIBRARY_PATH at the same time. Only include the path to the shared libraries corresponding to the driver matching the bitness of the client application used.

For example, if you are using a 64-bit client application and ODBC driver manager libraries are installed in /usr/local/lib, then set LD_LIBRARY_PATH as follows:

```
export LD_LIBRARY_PATH=/usr/local/lib:/opt/cloudera/impalaodbc/lib/64
```

Refer to your Linux shell documentation for details on how to set environment variables permanently.

For details on creating ODBC connections using Cloudera ODBC Driver for Impala, see “Configuring ODBC Connections for Linux and Mac OS X” on page 10.

Mac OS X Driver

System Requirements

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

Cloudera ODBC Driver for Impala requires a Hadoop cluster with the Impala server installed and running.

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.

Installation

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

- /opt/cloudera/impalaodbc/ErrorMessage – Error messages files directory
- /opt/cloudera/impalaodbc/Setup – Sample configuration files directory
- /opt/cloudera/impalaodbc/lib/universal – Binaries directory

To install Cloudera ODBC Driver for Impala:

1. Double-click to mount the **ClouderaImpalaODBC.dmg** disk image.
2. Double-click **ClouderaImpalaODBC.pkg** to run the Installer.
3. Follow the instructions in the Installer to complete the installation process.
4. When the installation completes, click **Close**.
5. If you are installing a driver with an evaluation license and you have purchased a perpetual license, then copy the License.lic file you received via e-mail into the /opt/cloudera/impalaodbc/lib/universal folder.

Setting the DYLD_LIBRARY_PATH Environment Variable

The DYLD_LIBRARY_PATH environment variable must include the paths to:

- Installed ODBC driver manager libraries

Configuring ODBC Connections for Linux and Mac OS X

- Installed Cloudera ODBC Driver for Impala shared 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=/usr/local/lib:/opt/cloudera/impalaodbc/lib/universal
```

Refer to your Mac OS X shell documentation for details on how to set environment variables permanently.

For details on creating ODBC connections using Cloudera ODBC Driver for Impala, see “Configuring ODBC Connections for Linux and Mac OS X” on page 10.

Configuring ODBC Connections for Linux and Mac OS X

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** – The file used to define ODBC data sources (required)
- **.odbcinst.ini** – The file used to define ODBC drivers (optional)
- **.cloudera.impalaodbc.ini** – The file used to configure Cloudera ODBC Driver for Impala (required)

Sample Files

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

- **odbc.ini**
- **odbcinst.ini**
- **cloudera.impalaodbc.ini**

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 (.). For `cloudera.impalaodbc.ini`, the filename must begin with a period (.) and must reside in the user’s home directory.

If the configuration files do not already exist in the user’s home directory, then the sample configuration files can be copied to that directory and renamed. If the configuration files already exist in the user’s home directory, then the sample configuration files should be used as a guide for modifying the existing configuration files.

Configuring the Environment

By default, the configuration files reside in the user’s home directory. However, three environment variables, `ODBCINI`, `ODBCSYSINI`, and `SIMBAINI`, can be used to specify different locations for the

odbc.ini, odbcinst.ini, and cloudera.impalaodbc.ini configuration files. Set ODBCINI to point to your odbc.ini file. Set ODBCYSINI to point to the directory containing the odbcinst.ini file. Set SIMBAINI 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 ODBCYSINI=/usr/local/odbc
export SIMBAINI=/etc/cloudera.impalaodbc.ini
```

The search order for the cloudera.impalaodbc.ini file is as follows:

1. If the SIMBAINI environment variable is defined, then the value of SIMBAINI is used to locate the file. SIMBAINI must contain the full path including the filename.
2. Otherwise, if the SIMBAINI environment variable is not defined, the current working directory of the application will be searched for cloudera.impalaodbc.ini. (Note the lack of a preceding 'dot' in cloudera.impalaodbc.ini.)
3. The next directory that will be searched is ~/ (i.e. \$HOME) for .cloudera.impalaodbc.ini.
4. Finally, the system wide default /etc/cloudera.impalaodbc.ini will be used. (Note the lack of a preceding 'dot' in cloudera.impalaodbc.ini.)

Configuring the odbc.ini File

ODBC Data Sources 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.

Here is an example odbc.ini configuration file for Linux:

```
[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
```

Configuring ODBC Connections for Linux and Mac OS X

Here is an example `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
```

To create a data source:

1. Open the `.odbc.ini` configuration file in a text editor.
2. Add a new entry to the `[ODBC Data Sources]` section. Type the data source name (DSN) and the driver name.
3. To set configuration options, add a new section that has a name that matches the data source name (DSN) you specified in step 2. Specify configuration options as key-value pairs.
4. Save the `.odbc.ini` configuration file.

For details on configuration options available to control the behavior of DSNs using Cloudera ODBC Driver for Impala, see “Appendix B: Driver Configuration Options” on page 22.

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 10.

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.

Here is an example `odbcinst.ini` file for Linux:

```
[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
```

Here is an example `odbcinst.ini` 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. Add a new entry to the `[ODBC Drivers]` section. Type the driver name, and then type the following: **=Installed**

Note:

Assign the driver name as the value of the `Driver` attribute in the data source definition instead of the driver shared library name.

3. In `.odbcinst.ini`, add a new section that has 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 with Cloudera ODBC Driver for Impala in the `Setup` directory. Specify configuration options as key-value pairs.
4. Save the `.odbcinst.ini` configuration file.

Configuring the `cloudera.impalaodbc.ini` File

To configure 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 usually must be **UTF-16** or **UTF-32**, depending on the ODBC driver manager you use. iODBC uses **UTF-32** and unixODBC uses **UTF-16**. Consult your ODBC Driver Manager documentation for the correct setting to use.
3. Edit the `ODBCInstLib` setting. The value is the name of the ODBCInst shared library for the ODBC driver manager you use. The configuration file defaults to the shared library for iODBC. In Linux, the shared library name for iODBC is `libiodbcinst.so`. In Mac OS X, the shared library name for iODBC is `libiodbcinst.dylib`.

Note:

Consult your ODBC driver manager documentation for the correct library to specify. You can specify an absolute or relative filename for the library. If you intend to use the relative filename, then the path to the library must be included in the library path environment variable. In Linux, the library path environment variable is named LD_LIBRARY_PATH. In Mac OS X, the library path environment variable is named DYLD_LIBRARY_PATH.

4. Save the .cloudera.impalaodbc.ini configuration file.

Configuring Authentication

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

- No Authentication
- User Name
- Kerberos

For details on the keys involved in configuring authentication, see “Appendix B: Driver Configuration Options” on page 22.

Using No Authentication

No additional details are required when using **No Authentication**.

Using Kerberos

For information on operating Kerberos, refer to the documentation for your operating system.

To configure a DSN using Cloudera ODBC Driver for Impala to use Kerberos authentication:

1. Set the AuthMech configuration key for the DSN 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 key.
3. Set the KrbFQDN key to the fully qualified domain name of the Impala host.
4. Set the KrbServiceName key to the service name of the Impala server. For example, if the principle for the Impala server "impala/[fully.qualified.domain.name@YOUR-REALM.COM](#)", then the value in the service name field should be **impala**. If you are unsure of the correct service name to use for your particular Hadoop deployment, see your Hadoop administrator

Using User Name

Authenticating by user name does not use a password. The user name labels the session, facilitating database tracking.

To configure your DSN for User Name authentication:

1. Set the AuthMech configuration key for the DSN to 2.

2. Set the UID key to the appropriate credential recognized by the Impala server.

Using User Name and Password

To configure User Name and Password authentication:

1. Set the AuthMech configuration key for the DSN to 3.
2. Set the UID key to the appropriate user name recognized by the Impala server.
3. Set the PWD key to the password corresponding to the user name you provided in step 2.

Using User Name and Password (SSL)

To configure User Name and Password (SSL) authentication:

1. Set the AuthMech configuration key for the DSN to 4.
2. Set the UID key to the appropriate user name recognized by the Impala server.
3. Set the PWD key to the password corresponding to the user name you provided in step 2.

Note:

The driver always accepts the use of self-signed SSL certificate.

Note:

Optionally you can configure the **CAIssuedCertNamesMismatch** setting to control whether the driver allows the common name of a CA issued certificate to not match the host name of the Impala server. For self-signed certificates the driver always allows the common name of the certificate to not match the host name. See “Appendix B: Driver Configuration Options” on page 22.

Note:

Optionally you can configure the **TrustedCerts** setting in the **Advanced Options** to allow the driver to load SSL certificates from the specified file. The driver will default to using the trusted CA certificates PEM file install with the driver if this setting is not set. See “Appendix B: Driver Configuration Options” on page 22.

Data Types

The following data types are supported:

- TINYINT
- SMALLINT
- INT

Contact Us

- BIGINT
- FLOAT
- DOUBLE
- BOOLEAN
- STRING
- TIMESTAMP

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

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 schema/database, we have added 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

Cloudera ODBC Driver for Impala is able to parse queries locally prior to 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.

Active Directory

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

1. MIT Kerberos is **not** installed on client Windows machine.
2. 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, you can contact Cloudera Technical Support. 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: Configuring Kerberos Authentication for Windows

Active Directory

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

1. MIT Kerberos is **not** installed on client Windows machine.
2. 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

Download and install MIT Kerberos for Windows 4.0.1

1. For 64-bit computers: <http://web.mit.edu/kerberos/dist/kfw/4.0/kfw-4.0.1-amd64.msi>. The installer includes both 32-bit and 64-bit libraries.
2. For 32-bit computers: <http://web.mit.edu/kerberos/dist/kfw/4.0/kfw-4.0.1-i386.msi>. The installer includes 32-bit libraries only.

Set up the Kerberos configuration file in the default location

1. Obtain a **krb5.conf** configuration file from your Kerberos administrator. The configuration file should also be present at **/etc/krb5.conf** on the machine hosting the Impala Server.
2. The default location is **C:\ProgramData\MIT\Kerberos5** but this is normally a hidden directory. Consult your Windows documentation if you want to view and use this hidden directory.
3. Rename the configuration file from **krb5.conf** to **krb5.ini**.
4. Copy **krb5.ini** to the default location and overwrite the empty sample file.

Consult the MIT Kerberos documentation for more information on configuration.

Set up the Kerberos configuration file in another location

If you do not want to put the Kerberos configuration file in the default location, then you can use another location. The steps required to do this are as follows:

1. Obtain a **krb5.conf** configuration file for your Kerberos setup.
2. Store **krb5.conf** in an accessible directory and make note of the full path name.
3. Click the Windows **Start** menu.
4. Right-click **Computer**.
5. Click **Properties**.
6. Click **Advanced system settings**.
7. Click **Environment Variables**.
8. Click **New** for System variables.

9. In the **Variable Name** field, type **KRB5_CONFIG**.
10. In the **Variable Value** field, type the absolute path to the krb5.conf file you stored in step 2.
11. Click **OK** to save the new variable.
12. Ensure the variable is listed in the **System variables** list.
13. Click **OK** to close Environment Variables Window.
14. Click **OK** to close System Properties Window.

Set up the Kerberos credential cache file

1. Create a new directory where you want to save the Kerberos credential cache file. For example, you may create the C:\temp directory.
3. Click the Windows **Start** menu.
4. Right-click **Computer**.
5. Click **Properties**.
6. Click **Advanced system settings**.
7. Click **Environment Variables**.
8. Click **New** for System variables.
9. In the **Variable Name** field, type **KRB5CCNAME**
10. In the **Variable Value** field, type the path to the directory you created in step 1, and then append the file name **krb5cache**. For example, if you created the C:\temp directory in step 1, then type: C:\temp\krb5cache

Note:

krb5cache is a file—not a directory—managed by the Kerberos software and should not be created by the user. If you receive a permission error when you first use Kerberos, check to ensure that the krb5cache file does not exist as a file or a directory.

11. Click **OK** to save the new variable.
12. Ensure the variable appears in the System variables list.
13. Click **OK** to close Environment Variables Window.
14. Click **OK** to close System Properties Window.
15. Restart your computer to ensure that MIT Kerberos for Windows uses the new settings.

Obtain a ticket for a Kerberos principal using password

Note:


If your Kerberos environment uses keytab files please see the next section.

1. Click the **Start** button .

Appendix A: Configuring Kerberos Authentication for Windows

2. Click **All Programs**.
3. Click the **Kerberos for Windows (64-bit)** or the **Kerberos for Windows (32-bit)** program group.
4. Use **MIT Kerberos Ticket Manager** to obtain a ticket for the principal that will be connecting to the Impala server.

Obtain a ticket for a Kerberos principal using a keytab file

1. Click the **Start** button .
2. Click **All Programs**.
3. Click **Accessories**.
4. Click **Command Prompt**.
5. Type: `kinit -k -t <keytab pathname> <principal>`

<keytab pathname> is the full pathname to the keytab file. For example,
`C:\mykeytabs\hiveserver2.keytab`

<principal> is the Kerberos principal to use for authentication. For example,
[hive/hiveserver2.example.com@EXAMPLE.COM](#)

On some Windows systems, the cache location KRB5CCNAME is either not set or not used. In this case, use the '-c' option of the kinit command to specify the proper ticket cache that needs to be populated.

The command syntax is:

```
kinit -k -t C:\mykeytabs\hive.keytab hive/HOST@HADOOP.NET -c c:\ProgramData\MIT\krbcache  
(krbcache is the kerberos cache file, not a directory.)
```

Note:

The order of the options shown above is important because the '-c' argument must be last.

Obtain a ticket for a Kerberos principal using the default keytab file

A default keytab file can be set for your Kerberos configuration. Consult the MIT Kerberos documentation for instructions on configuring a default keytab file.

1. Click the **Start** button .
2. Click **All Programs**.
3. Click **Accessories**.
4. Click **Command Prompt**.
5. Type: `kinit -k <principal>`

Appendix A: Configuring Kerberos Authentication for Windows

`<principal>` is the Kerberos principal to use for authentication. For example,
`hive/hiveserver2.example.com@EXAMPLE.COM`

On some Windows systems, the cache location `KRB5CCNAME` is either not set or not used. In this case, use the `-c` option of the `kinit` command to specify the proper ticket cache that needs to be populated.

The command syntax is:

```
kinit -k hive/HOST@HADOOP.NET -c c:\ProgramData\MIT\krbcache  
(krbcache is the kerberos cache file, not a directory.)
```

Note:

The order of the options shown above is important because the `-c` argument must be last.

Appendix B: Driver Configuration Options

The configuration options available to control the behavior of Cloudera ODBC Driver for Impala are listed and described in Table 1.

Note:

You can set 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`

Table 1 Driver Configuration Options

Key	Default Value	Description
Driver		The location of the Cloudera ODBC Driver for Impala shared object file
HOST		The IP address or hostname of the Impala server
PORT	10000	The listening port for the service
UseNativeQuery	0	Enabling the UseNativeQuery option using a value of 1 disables the feature of the driver to apply 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 disabling the feature avoids the extra overhead of query transformation.

Key	Default Value	Description
RowsFetchedPerBlock	10000	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.
SocketTimeout	0	The number of seconds after which Impala closes the connection with the client application if the connection is idle
AuthMech	0	The authentication mechanism to use. Set the value to 0 for no authentication, 1 for Kerberos, 2 for User Name and Password, or 4 for User Name and Password (SSL).
KrbFQDN		The fully qualified domain name of the Impala host used.
KrbServiceName		The Kerberos service principal name of the Impala server. By convention the service name is <code>hive</code> , but the name may be different in your server environment.

Appendix B: Driver Configuration Options

Key	Default Value	Description
KrbRealm		If there is no default realm configured or the realm of the Impala host is different from the default realm for your Kerberos setup, then define the realm of the Impala host using this option.
UID		The user name of the user credential. (Required if AuthMech is User Name and Password or User Name and Password (SSL)) (Optional if AuthMech is User Name . Default Value: anonymous)
PWD		The password of the user credential. (Required if AuthMech is User Name and Password or User Name and Password (SSL))
TSaslTransportBufSize	1000	The number of bytes to reserve in memory for buffering unencrypted data from the network. <div>Note: In most circumstances, the default value of 1000 bytes is optimal.</div>

Key	Default Value	Description
CAIssuedCertNamesMismatch	0	<p>Control whether to allow CA issued SSL certificate's common name to not match the host name of the Impala server.</p> <p>Set to 1 to enable.</p> <p>Set to 0 to disable.</p> <div>Note: This setting is only applicable to User Name and Password (SSL) authentication mechanism and will be ignored by other authentication mechanisms.</div>

Appendix B: Driver Configuration Options

TrustedCerts	<p>For 32 bit driver: /opt/cloudera/impalaodbc/lib/32/cacerts.pem</p> <p>For 64 bit driver: /opt/cloudera/impalaodbc/lib/64/cacerts.pem</p>	<p>Used to specify the location of the file containing trusted CA certificates for authenticating the Impala server when using SSL.</p> <div>Note: This setting is only applicable to User Name and Password (SSL) authentication mechanism, and will be ignored by other authentication mechanisms.</div> <div>Note: If this setting is not set then the driver will default to using the trusted CA certificates file installed by the driver.</div>
DelegationUID		<p>Used to delegate all operation against Hive to a user that is different than the authenticated user for the connection.</p>

Appendix C: ODBC API Conformance Level

Conformance Level ^[1]	INTERFACES ^[2]		Conformance Level ¹	INTERFACES ^[2]
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
Core	SQLFetchScroll		Core	SQLStatistics
Core	SQLFreeHandle		Core	SQLTables
Core	SQLFreeStmt		Core	SQLBrowseConnect
Core	SQLGetConnectAttr		Level 1	SQLPrimaryKeys

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

^[2] Interfaces include both the Unicode and non-unicode versions. See <http://msdn.microsoft.com/en-us/library/ms716246%28VS.85%29.aspx> for more details.

Appendix C: ODBC API Conformance Level

Conformance Level ^[1]	INTERFACES ^[2]		Conformance Level ¹	INTERFACES ^[2]
Core	SQLGetCursorName		Level 1	SQLProcedureColumns
Core	SQLGetData		Level 1	SQLProcedures
Core	SQLGetDescField		Level 1	SQLProcedureColumns
Core	SQLGetDescRec		Level 2	SQLColumnPrivileges
Core	SQLGetDiagField		Level 2	SQLDescribeParam
Core	SQLGetDiagRec		Level 2	SQLForeignKeys
Core	SQLGetEnvAttr		Level 2	SQLTablePrivileges
Core	SQLGetFunctions			
Core	SQLGetInfo			

^[1] ODBC Compliance levels are Core, Level 1 and Level 2. These are defined in the ODBC Specification published by Microsoft.

^[2] Interfaces include both the Unicode and non-unicode versions. See <http://msdn.microsoft.com/en-us/library/ms716246%28VS.85%29.aspx> for more details.