Sqoop Connectors Installation and User Guide (v1.0-beta-u1)

Table of Contents

1. Introduction

Sqoop connectors are standard Sqoop extensions that allow Sqoop to operate with specialized enterprise systems such as Netezza. After installation, these connectors allow various Sqoop tools such as `sqoop-import` and `sqoop-export` to operate directly with these enterprise systems.

This document describes how to install and configure these connectors within a Sqoop installation and provides reference information for the operation of such connectors where necessary. This document is intended for:

- System and application programmers
- System administrators
- Database administrators
- Data analysts
- Data engineers

2. Prerequisites

In order to use Sqoop connectors, you must have a functioning Sqoop installation. Depending on the way Sqoop is installed, you may need administrative privileges in order to create or modify configuration files.

For more information on how to install, configure, and use Sqoop, see the Sqoop documentation at http://docs.cloudera.com/.

3. Installing the Sqoop Connectors

To install the Sqoop connectors, open the distribution archive in a convenient location such as `/usr/lib`. This will create a directory such as `/usr/lib/sqoop-connectors-1.0-beta-u1` that contains the jar file of the compiled version of all connectors. Note the path to this jar file. For example, `/usr/lib/sqoop-connectors-1.0-beta-u1/sqoop-connectors-1.0-beta-u1.jar`.

Next, create a text file called `connectors` in a directory named `managers.d` within the configuration directory of Sqoop. Create the directory `managers.d` if it does not exist.

> **Note**
> Depending on how Sqoop is installed, its configuration directory may be in

The `connectors` file must have the connector class name followed by the complete path to the directory where the connector jar is located. For example:

```
com.cloudera.sqoop.manager.EnterpriseManagerFactory=/usr/lib/sqoop-connectors-X.X/sqoop-connectors-X.X.ja
```

The `EnterpriseManagerFactory` acts as a single point of delegation for invoking the applicable connector bundled with this distribution. However, in order to use a specific connector, there may be additional steps necessary to configure it. Refer to the connector-specific section below in order to complete those steps. You need to complete these steps only for the specific connectors that you intend to use.

4. Uninstalling the Sqoop Connectors

To remove the Sqoop connectors, delete the `connectors` file from `managers.d` directory located under Sqoop configuration directory, and remove the files from the connectors distribution. If you did any additional steps for the installation of specific connectors, refer to the connector-specific section below to continue to any connector-specific uninstall steps as necessary.

5. Available Connectors

5.1. Cloudera Connector for Netezza

> 5.1.1. Installation and Removal
> 5.1.2. Using Cloudera Connector for Netezza

This release of Sqoop Connectors provides the connector for Netezza Database System via Netezza JDBC Version 5.0 drivers. This connector is formally known as the Cloudera connector for Netezza.

5.1. Cloudera Connector for Netezza

5.1.1. Installation and Removal
5.1.2. Using Cloudera Connector for Netezza

The Cloudera connector for Netezza provided with this release of Sqoop Connectors is designed to use Netezza's high-throughput, data transfer system.

5.1.1. Installation and Removal

To ensure that Cloudera connector for Netezza can be used, copy the Netezza JDBC driver in the `lib` directory of Sqoop installation. This driver may be obtained from the Netezza Client distribution for your operating system. Without this driver, the connector will not function correctly.

If you are uninstalling the Sqoop connector, you should also remove the JDBC driver for Netezza that you copied to the `lib` directory of Sqoop installation because Sqoop will no longer need it.

5.1.2. Using Cloudera Connector for Netezza

After Sqoop connectors have been installed and the necessary JDBC driver for Netezza has been copied to the `lib` directory of Sqoop installation, you can use this connector by invoking the Sqoop tools with the appropriate connection string. The Cloudera connector for Netezza will be used by Sqoop only if the connection string being used is of the form `jdbc:netezza://<nz-host>/<nz-instance>` where `<nz-host>` is the host name of the machine where Netezza server runs and `<nz-instance>` is the Netezza database instance name. Also, in order to effectively use the Netezza connector, you must specify the `--direct` option along with a number of mappers greater than one.

For example, the following command invokes the Sqoop import tool with eight mappers and uses the Cloudera connector for Netezza.

```
$ sqoop import --connect jdbc:netezza://localhost/MYDB --username arvind \
--password xxxxx --direct --table MY_TABLE --num-mappers 8 --escaped-by '\\' \
--fields-terminated-by ',' --lines-terminated-by '\n'
```

This command invokes the Sqoop export tool with eight mappers and uses the Cloudera connector for Netezza.

```
$ sqoop export --connect jdbc:netezza://localhost/MYDB --username arvind \
--password xxxxx --direct --export-dir /user/arvind/MY_TABLE --table MY_TABLE_TARGET \
--num-mappers 8 --input-escaped-by '\\' --input-fields-terminated-by ','
```

> **Important**
> The Cloudera Connector for Netezza does not support the specification of `--input-enclosed-by` or `--input-lines-terminated-by` options. Both of these settings when specified during an export operation will be ignored. Also, the only supported escape character is the back-slash (`\`) character.

6. Getting Support

Support for Sqoop Connectors is available via Cloudera Enterprise Support. Refer to http://www.cloudera.com/support for more details.