cloudera®

Cloudera Connector for Netezza

Important Notice

Cloudera, Inc.
395 Page Mill Road
Palo Alto, CA 94306
info@cloudera.com
US: 1-888-789-1488
Intl: 1-650-362-0488
www.cloudera.com

Release Information

Version: Connectors 1.x
Date: October 28, 2020

Table of Contents

Cloudera Connector for Netezza Release Notes

This section summarizes the high level changes and most important new features in the Cloudera Connector for Netezza.

Current Release Versions: 1.5.1c6. For an explanation of version label conventions, see Cloudera Connector for Netezza, Version Scheme.

CDH Versions Supported: CDH 5.2.0 and above.

To download Cloudera Connector for Netezza, go to http://www.cloudera.com/downloads.html.

CDH6 Compatible Version of Cloudera Connector for Netezza 1.5.1c6 Available

Cloudera Connector Powered by Netezza 1.5.1c6 is compatible with CDH 6 but not supported in CDH 6.3.x. It does not contain new features or changes.

Note: Cloudera Connector Powered by Netezza 1.5.1c5 and lower are not compatible with CDH 6.

New Features in Cloudera Connector for Netezza

The following new features are included in Cloudera Connector for Netezza.

New Features in Cloudera Connector for Netezza 1.5.1c5

- Adds support for SLES 12.

New Features in Cloudera Connector for Netezza 1.5c5

- Fixed import in direct mode with multi-schema by removing unnecessary object owner check.
- Added fail fast scenario for import with `--as-parquetfile` in direct mode.
- Added fail fast scenario for import with `--as-avrodatafile` in direct mode.

New Features in Cloudera Connector for Netezza 1.4c5

- Added the `escapeColName` method to handle, for example, column names with lowercase characters.

New Features in Cloudera Connector for Netezza 1.3c5

New argument added:

- `--schema`: Defines the schema Sqoop uses for both import and export. The schema is used for metadata queries, and for record extraction and loading.

New Features in Cloudera Connector for Netezza 1.2c4, 1.2c5

New argument added:

- `--nz-uploaddir`: Generated Netezza logs will be stored in the HDFS directory specified for this parameter.

New Features in Cloudera Connector for Netezza 1.1.1

- Direct mode now supports extra argument `--nz-ctrlchars`, that will allow processing data with ASCII characters of value 31 and less.

- Direct mode now supports Sqoop parameters `--null-string` and `--input-null-string` that will user allow to override the default `NULL` substitution character.

New Features in Cloudera Connector for Netezza 1.1.0

- This release updates the connectors package to be compatible with Sqoop 1.4, the version released with CDH4.0.

New Features in Cloudera Connector for Netezza 1.0.5

- This release updated the connectors package to be compatible with Sqoop 1.3, the version released with CDH3u3.
- The connector now supports the `LOGDIR` directive for Netezza. Refer to the Cloudera Connector for Netezza User Guide for details on how to use this feature.
- The problem of jobs hanging when running import or export against views has been resolved. The connector now displays an error message when used for a job that operates against objects other than regular tables in Netezza.
- Support for the `NCHAR` datatype has been added in this release.

New Features in Cloudera Connector for Netezza 1.0.1

- This release updated the connectors package to be compatible with Sqoop 1.2, the version released with CDH3u0.
- Updates the Netezza connector to better handle long strings on exports.
- The connector now supports the `NVARCHAR` type.
- The connector now supports the `MAXERRORS` directive. Refer to the Cloudera Connector for Netezza User Guide for details on how to use this feature.

Known Issues and Workarounds

The following section describes the current known issue for Cloudera Connector for Netezza.

The Boolean data type format is incompatible between import and export when using direct mode.

When using the Netezza connector in direct mode to import data of type Boolean, the format used by the connector is 'T' or 'F' to represent True or False values. However, when doing an export of data from HDFS to Netezza using the connector in direct mode, the expected Boolean format is 'TRUE' or 'FALSE'.

Workaround: Either convert the data to use 'TRUE'/'FALSE' style representation of Boolean data values prior to doing a direct mode export, or use the regular mode of export instead of direct mode for this type of data.

Expected Resolution: Planned to be fixed in a future release of the connector.

Cloudera Connector for Netezza User Guide

The Cloudera Connector for Netezza is designed to use Netezza high-throughput data-transfer mechanisms to import and export data to HDFS. Cloudera Connector for Netezza is a standard Sqoop extension that allows Sqoop to interoperate with Netezza Data Appliance through Netezza JDBC drivers found in the Netezza client v5.0 package. After installation, this connector allows various Sqoop tools such as `sqoop-import` and `sqoop-export` to operate in highly efficient direct modes and exposes some Netezza-specific options.

This document describes how to install and configure this connector in a Sqoop installation, and provides reference information for connector operation. This document is intended for:

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

Version Scheme

This topic describes the versioning scheme used for Sqoop-based connectors such as Cloudera Connector for Netezza. The version string consists of the following parts:

`$MAJOR_VERSION.$MINOR_VERSIONc$MAJOR_CDH_VERSION`

- `MAJOR VERSION and MINOR VERSION`: Identify the major and minor version of the connector project.
- `MAJOR CDH VERSION`: The major Cloudera version for which the connector has been compiled and tested.

For example:

- 1.2c5 - Second revision of a Sqoop 1-based connector that is compatible with CDH 5.

Prerequisites

To use Cloudera Connector for Netezza, you must have a functioning Sqoop CDH installation. Depending on the way Sqoop is installed, you might need administrative privileges to create or modify configuration files.

Netezza connector might need to create external tables to import data. Check with your Netezza administrators or operators to determine if you have the required privileges.

Cloudera Connector for Netezza is compatible with Netezza 7.x.

Cloudera Connector for Netezza is not compatible with Sqoop2.

Both CDH 6 and CDH 5 compatible Cloudera Connectors support Netezza 7.x.

For more information on how to install, configure, and use Sqoop, see the Sqoop documentation in the CDH 5 Installation Guide.

Installing the Cloudera Connector for Netezza

Use one of the following methods to install the Sqoop connectors for Netezza:

- If your CDH 5 cluster is managed by Cloudera Manager 5, see Installation with CDH 5 and Cloudera Manager 5 on page 10.
- If your cluster is not managed by Cloudera Manager, see Installation without Cloudera Manager on page 10.

Installation with CDH 5 and Cloudera Manager 5

Step 1: Adding the Sqoop service

The Sqoop1 Client Gateway sets up and manages configuration for the connector for hosts where you execute Sqoop1 commands. If you do not already have the Sqoop1 Client Gateway deployed in your cluster, deploy it before proceeding. See Managing the Sqoop1 Client for information on adding the gateway using Cloudera Manager Admin Console.

Step 2: Download, Distribute, and Activate the Sqoop Parcels

Note: Copy the Netezza JDBC version 3.0 driver (`nzjdbc3.jar`) into the `/var/lib/sqoop` directory of the Sqoop installation. You can obtain this driver from the Netezza Client distribution for your operating system. Without this driver, the connector will not function correctly.

Parcels for Sqoop connectors are prefixed by SQOOP_, followed by the name of the connector.

Follow the instructions in Managing Parcels to download, distribute, and activate Sqoop parcels.

Installation without Cloudera Manager

1. To install the Cloudera Connector for Netezza, open the distribution archive in a convenient location such as `/usr/lib`. This creates a directory (such as `/usr/lib/sqoop-nz-connector-1.4c5`) that contains the jar file of the compiled version of the connector. Note the path to this jar file (for example, for version 1.4c5 you would use `/usr/lib/sqoop-nz-connector-1.4c5/sqoop-nz-connector-1.4c5.jar`).
2. Copy the Netezza JDBC driver to the `lib` directory of the Sqoop installation. You can obtain this driver from the Netezza Client distribution for your operating system. Without this driver, the connector will not function correctly.
3. Create a text file called `connectors` in a directory named `managers.d` within the Sqoop configuration directory. Create the `managers.d` directory if it does not exist.

Note: Depending on how Sqoop is installed, its configuration directory might be in `/etc/sqoop/conf`, `/usr/lib/sqoop/conf`, or elsewhere if Sqoop was installed using the tarball distribution.

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

```
com.cloudera.sqoop.manager.NetezzaManagerFactory=/usr/lib/sqoop-nz-connector-1.4c5/sqoop-nz-connector-1.4c5.jar
```

The `NetezzaManagerFactory` acts as a single point of delegation for invoking the applicable connector bundled with this distribution.

Using Cloudera Connector for Netezza

After you have installed the Cloudera Connector for Netezza and copied the required JDBC driver for Netezza to the `lib` directory of the Sqoop installation, use the connector by invoking the Sqoop tools with the appropriate connection string.

The connection string must be of the form `jdbc:netezza://<nz-host>/<nz-instance>`, where:

- `<nz-host>` is the hostname of the machine where the Netezza server runs.
- `<nz-instance>` is the Netezza database instance name.

To 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 ','
```

The following 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 '\\'
```

The direct mode Netezza connector supports the following Netezza-specific arguments for imports and exports:

- `--nz-maxerrors <n>`: Specifies the number of error records needed to abort an import or export operation. By default, this is set to `1`, which implies that the operation fails on the first bad record encountered. Setting it to a higher value, for example `3`, configures Sqoop to continue despite two bad records but abort on the third. If you set the value to `0`, the operation never aborts due to bad records.
- `--nz-logdir <path>`: Specifies the location of a directory on the local filesystem where Sqoop places the Netezza transport specific `nzbad` and `nzlog` files. Use these files to debug and tune the overall operation for the most effective usage. This path is local to the hosts on which the Sqoop map jobs run and does not apply to the system from which Sqoop is launched. If the directory corresponding to this path does not exist, Sqoop attempts to create it before initiating the transport.
- `--nz-ctrlchars`: Instructs Sqoop to use the CTRLCHARS parameter when exchanging data with Netezza. This parameter allows processing of data that has ASCII characters of value 31 and less.
- `--nz-uploaddir`: Generated Netezza logs are stored in the HDFS directory specified for this parameter.
- `--schema:` Defines the schema Sqoop uses for both import and export. The schema is used for metadata queries, and for record extraction and loading.

Note: The Netezza-specific arguments for imports and exports must be specified after all other arguments and must be separated from other arguments by `--`. For more information on how to specify connector specific arguments, see the Sqoop user guide.

The following example uses the `--nz-maxerrors` connector-specific argument:

```
$ 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 '\\' -- --nz-maxerrors 0
```

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, are ignored. The backslash (\ ) is the only supported escape character.

Uninstalling the Cloudera Connector for Netezza

You can use one of the following ways to uninstall the Sqoop connectors for Netezza:

- If your CDH 5 cluster is managed by Cloudera Manager 5, see Uninstallation with CDH 5 and Cloudera Manager 5 on page 13.
- If your cluster is not managed by Cloudera Manager, see Uninstallation without Cloudera Manager on page 13.

Uninstallation with CDH 5 and Cloudera Manager 5

Perform the following steps to uninstall the Cloudera Connector for Netezza using Cloudera Manager 5:

1. **Remove the Sqoop Connector parcels:**
   a. In the Cloudera Manager Admin Console, click **Hosts** in the top navigation bar and then go to the **Parcels** tab. Parcels for the Sqoop connectors are listed on this page, prefixed by "SQOOP_", followed by the name of the connector.
   b. The Sqoop connectors are listed as **Activated**. To deactivate a parcel, click **Actions** on an activated parcel and select **Deactivate**.
   c. To remove the parcel, click the down arrow to the right of the **Activate** button and select **Remove from Hosts**.
2. **Redeploy the client configuration:**
   a. In the Cloudera Manager Admin Console, go to the Sqoop Client service.
   b. From the **Actions** menu at the top right of the service page, select **Deploy Client Configuration**.
   c. Click **Deploy Client Configuration** to confirm redeployment of the client configuration.

Uninstallation without Cloudera Manager

To remove the Cloudera Connector for Netezza, delete the `connectors` file from `managers.d` directory located under the Sqoop configuration directory, and remove the files from the connectors distribution. You should also remove the JDBC driver for Netezza that you copied to the `lib` directory of Sqoop installation.

Limitations

This version of the Cloudera Connector for Netezza has several functional limitations.

- Direct mode does not support import into HBase.
- Direct mode does not support export using updates.
- This connector does not support data type TIME.
- Direct mode does not support import-all-tables.
- Direct mode does not support imports from views or materialized views.
- Direct mode does not support import into or export from Avro.
- The number of records transferred, as reported on the console by Sqoop for an import or export operation, reflects the total number of records *processed*, not the number of records that were actually transferred. The number of records transferred may be less than the number of records processed due to rejection of bad records. This can happen if the number of bad records encountered is less than the value of the `--nz-maxerrors` option, and therefore is not sufficient to cause the transfer operation to abort.

Troubleshooting

This section provides guidance on problems you may encounter when installing or running Netezza connector.

How do I verify that the Netezza connector has been properly installed?

To verify that the Netezza connector has been loaded, use the `--verbose` option on the Sqoop command line. If the connector is properly installed, you see a message that says `com.cloudera.sqoop.manager.NetezzaManagerFactory` has been loaded.

```
$ sqoop import --connect jdbc:netezza//nzhost/PGTESTDB --username admin \
--password password --table TESTTABLE --num-mappers 1 --escaped-by '\\' \
--verbose ... 11/06/08 12:58:37 DEBUG sqoop.ConnFactory: Loaded manager factory:
com.cloudera.sqoop.manager.NetezzaManagerFactory ...
```

Imports fail with errors such as: ' Caused by: org.netezza.error.NzSQLException: Driver can process one query at one time '

Certain versions of the Netezza JDBC driver (`nzjdbc3.jar`) are susceptible to this problem. You can work around this problem by changing to a different version of the Netezza JDBC driver. To determine the version of the Netazza JDBC driver you are using, open `nzjdbc.jar` and look at the `nzjdbc_version.properties` file. The following is an example of a version that can import successfully.

```
$ jar xvf nzjdbc3.jar nzjdbc_version.properties
$ cat nzjdbc_version.properties
major.version=6
minor.version=0
full.version=Release 6.0
driver [build 16564] driver.name=Netezza JDBC Driver
database.product.name=Netezza NPS
$
```

Operations fail with an "unable to validate object type" error

Operations fail with error: 'java.io.IOException: Unable to validate object type for given table. Please ensure that the given user name and table name is in the correct case. If you are not sure, please use upper case to specify both these values.'

Starting with version 1.0.5 of Cloudera Connector for Netezza, you must specify the username and table name in the correct case. If you are not sure of the exact case of the username or the table name, specify them in UPPERCASE. By default, Netezza converts object names to upper case unless they were specified within quotes during creation. If converting these to UPPERCASE does not resolve this problem, contact your Netezza administrator to find out the exact case of these values.

Because the username and table name must be specified in the correct case, the Sqoop commands that worked with earlier versions might stop working. After these values are specified in the correct case, these commands should work as expected again.

Getting Support

Support for the Cloudera Connector for Netezza is available through Cloudera Enterprise Support. See http://www.cloudera.com/support for more details.

Appendix: Apache License, Version 2.0

SPDX short identifier: Apache-2.0

Apache License
Version 2.0, January 2004
http://www.apache.org/licenses/

TERMS AND CONDITIONS FOR USE, REPRODUCTION, AND DISTRIBUTION

1. Definitions.

"License" shall mean the terms and conditions for use, reproduction, and distribution as defined by Sections 1 through 9 of this document.

"Licensor" shall mean the copyright owner or entity authorized by the copyright owner that is granting the License.

"Legal Entity" shall mean the union of the acting entity and all other entities that control, are controlled by, or are under common control with that entity. For the purposes of this definition, "control" means (i) the power, direct or indirect, to cause the direction or management of such entity, whether by contract or otherwise, or (ii) ownership of fifty percent (50%) or more of the outstanding shares, or (iii) beneficial ownership of such entity.

"You" (or "Your") shall mean an individual or Legal Entity exercising permissions granted by this License.

"Source" form shall mean the preferred form for making modifications, including but not limited to software source code, documentation source, and configuration files.

"Object" form shall mean any form resulting from mechanical transformation or translation of a Source form, including but not limited to compiled object code, generated documentation, and conversions to other media types.

"Work" shall mean the work of authorship, whether in Source or Object form, made available under the License, as indicated by a copyright notice that is included in or attached to the work (an example is provided in the Appendix below).

"Derivative Works" shall mean any work, whether in Source or Object form, that is based on (or derived from) the Work and for which the editorial revisions, annotations, elaborations, or other modifications represent, as a whole, an original work of authorship. For the purposes of this License, Derivative Works shall not include works that remain separable from, or merely link (or bind by name) to the interfaces of, the Work and Derivative Works thereof.

"Contribution" shall mean any work of authorship, including the original version of the Work and any modifications or additions to that Work or Derivative Works thereof, that is intentionally submitted to Licensor for inclusion in the Work by the copyright owner or by an individual or Legal Entity authorized to submit on behalf of the copyright owner. For the purposes of this definition, "submitted" means any form of electronic, verbal, or written communication sent to the Licensor or its representatives, including but not limited to communication on electronic mailing lists, source code control systems, and issue tracking systems that are managed by, or on behalf of, the Licensor for the purpose of discussing and improving the Work, but excluding communication that is conspicuously marked or otherwise designated in writing by the copyright owner as "Not a Contribution."

"Contributor" shall mean Licensor and any individual or Legal Entity on behalf of whom a Contribution has been received by Licensor and subsequently incorporated within the Work.

2. Grant of Copyright License.

Subject to the terms and conditions of this License, each Contributor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable copyright license to reproduce, prepare Derivative Works of, publicly display, publicly perform, sublicense, and distribute the Work and such Derivative Works in Source or Object form.

3. Grant of Patent License.

Subject to the terms and conditions of this License, each Contributor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable (except as stated in this section) patent license to make, have made, use, offer to sell, sell, import, and otherwise transfer the Work, where such license applies only to those patent claims

licensable by such Contributor that are necessarily infringed by their Contribution(s) alone or by combination of their Contribution(s) with the Work to which such Contribution(s) was submitted. If You institute patent litigation against any entity (including a cross-claim or counterclaim in a lawsuit) alleging that the Work or a Contribution incorporated within the Work constitutes direct or contributory patent infringement, then any patent licenses granted to You under this License for that Work shall terminate as of the date such litigation is filed.

4. Redistribution.

You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

1. You must give any other recipients of the Work or Derivative Works a copy of this License; and
2. You must cause any modified files to carry prominent notices stating that You changed the files; and
3. You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
4. If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.

You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

5. Submission of Contributions.

Unless You explicitly state otherwise, any Contribution intentionally submitted for inclusion in the Work by You to the Licensor shall be under the terms and conditions of this License, without any additional terms or conditions. Notwithstanding the above, nothing herein shall supersede or modify the terms of any separate license agreement you may have executed with Licensor regarding such Contributions.

6. Trademarks.

This License does not grant permission to use the trade names, trademarks, service marks, or product names of the Licensor, except as required for reasonable and customary use in describing the origin of the Work and reproducing the content of the NOTICE file.

7. Disclaimer of Warranty.

Unless required by applicable law or agreed to in writing, Licensor provides the Work (and each Contributor provides its Contributions) on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied, including, without limitation, any warranties or conditions of TITLE, NON-INFRINGEMENT, MERCHANTABILITY, or FITNESS FOR A PARTICULAR PURPOSE. You are solely responsible for determining the appropriateness of using or redistributing the Work and assume any risks associated with Your exercise of permissions under this License.

8. Limitation of Liability.

In no event and under no legal theory, whether in tort (including negligence), contract, or otherwise, unless required by applicable law (such as deliberate and grossly negligent acts) or agreed to in writing, shall any Contributor be liable to You for damages, including any direct, indirect, special, incidental, or consequential damages of any character arising as a result of this License or out of the use or inability to use the Work (including but not limited to damages for loss of goodwill, work stoppage, computer failure or malfunction, or any and all other commercial damages or losses), even if such Contributor has been advised of the possibility of such damages.

9. Accepting Warranty or Additional Liability.

While redistributing the Work or Derivative Works thereof, You may choose to offer, and charge a fee for, acceptance of support, warranty, indemnity, or other liability obligations and/or rights consistent with this License. However, in accepting such obligations, You may act only on Your own behalf and on Your sole responsibility, not on behalf of any other Contributor, and only if You agree to indemnify, defend, and hold each Contributor harmless for any liability incurred by, or claims asserted against, such Contributor by reason of your accepting any such warranty or additional liability.

END OF TERMS AND CONDITIONS

APPENDIX: How to apply the Apache License to your work

To apply the Apache License to your work, attach the following boilerplate notice, with the fields enclosed by brackets "[]" replaced with your own identifying information. (Don't include the brackets!) The text should be enclosed in the appropriate comment syntax for the file format. We also recommend that a file or class name and description of purpose be included on the same "printed page" as the copyright notice for easier identification within third-party archives.

```
Copyright [yyyy] [name of copyright owner]

Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License.
```