

HIVE ODCI, Users Guide

Version 1.x

Table of Contents

| | |
|-------------------|---|
| License..... | 3 |
| Overview..... | 4 |
| Installation..... | 5 |
| Parameters..... | 7 |
| Interface..... | 7 |
| Examples..... | 7 |

License

Copyright (c) 2016, Metasystems Technologies Inc (MTI), Nicholas Van Wyen
All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Overview

HIVE-ODCI is an Oracle Database Data Cartridge Interface plug-in for making Apache Hive Database SQL functionality natively available from within an Oracle 11g or higher RDBMS, making use of the Oracle Extensibility Architecture framework as documented at: <https://docs.oracle.com/database/121/ADDCI/toc.htm>.

The HIVE-ODCI functionality is built upon the idea of using a JDBC pass-through of SQL to a remote database, to retrieve and execute both DML and DDL statements. While Apache Hive is the target database, and the reason for creating, any Type 4 JDBC Driver can be swapped out and used, for example MySQL, PostgreSQL or any other complaint driver.

The initial release used the Data Direct HIVE JDBC Driver, from Progress Software. More information, including licensing and pricing can be found at: <https://www.progress.com/jdbc/apache-hadoop-hive>

Finally, the initial release was created as a proof-of-concept (POC) for the United States Government, Department of Homeland Security, US Customs and Border Protection, Targeting and Analysis Systems Program Directorate (TASPD) by Metasystems Technologies Inc. (MTI).

This POC was created as a way to allow large datasets in an Oracle RDMS to be loaded into Hadoop, accessible through Hive while keeping applications and PL/SQL unchanged. Using HIVE-ODCI, VIEW objects can be created of the same name as the original tables. SQL can be passed through via INSTEAD OF triggers for DML operations and data accessibility from standard SELECT. Once moved to Hadoop Hive, TASPD was able to achieve virtually 100% compatibility with their existing code base, their applications and PL/SQL, implementing HIVE-ODCI.

HIVE-ODCI is released under the permissive BSD Licensing and available freely for download and usage. This does **not** include a Hive JDBC Type 4 driver as part of the release, only the Java and PL/SQL Code needed to achieve the functionality. A JDBC driver will need to be obtained either through OpenSource or through a proprietary licensing agreement with a vendor.

GitHub Access

Source: <https://github.com/nvanwyen/hive-odci>

Latest Release: <https://github.com/nvanwyen/hive-odci/releases/latest>

Installation

Since HIVE-ODCI is an Oracle Extension, it is obviously installed within the Oracle RDBMS (11g or higher)

Download the latest version from Github (<https://github.com/nvanwyen/hive-odci/releases/latest>) making sure to get both the Tarball (TGZ) and Checksum (SHA) files.

Verify the source before starting

```
$ sha1sum -c hive-odci-${latest}.sha
```

Extract the content of the TGZ into a directory. For the purposes of this documentation the directory will be referred to as \$INSTALL_HOME.

```
$ tar -xvf hive-odci-${latest}.tgz
```

This will produce the following structure.

```
.
|-- doc
|   |-- hive-odci-v1-users-guide.docx
|   |-- hive-odci-v1-users-guide.pdf
|-- jdbc
|   |-- hive.jar
|   |-- load-jdbc.sh
|-- LICENSE
|-- policy
|   |-- jce_policy-6.zip
|   |-- jce_policy-7.zip
|   |-- jce_policy-8.zip
|   |-- load-policy.sh
|   |-- set_policies.sh
|-- README
|-- source
|   |-- attr.typ.sql
|   |-- bind.typ.sql
|   |-- dbms_hive.pkb.sql
|   |-- dbms_hive.pks.sql
|   |-- hive.ctx.sql
|   |-- hive.fnc.sql
|   |-- hive.gnt.sql
|   |-- hive.idx.sql
|   |-- hive.jva.sql
|   |-- hive.par.sql.in
|   |-- hive.pkb.sql
|   |-- hive.pks.sql
|   |-- hive.prm.sql
|   |-- hive.rol.sql
|   |-- hive.syn.sql
|   |-- hive.tbl.sql
|   |-- hive.tst.sql
|   |-- hive.typ.sql
|   |-- hive.usr.sql
|   |-- hive.utl.sql
|   |-- hive.vws.sql
|   |-- impl.pkb.sql
|   |-- impl.pks.sql
|   |-- install_hive.sql
|   |-- jdbc -> ../jdbc
|   |-- policy -> ../policy
```

```
| |-- remove_hive.sql
| |-- ver
| |-- wrap.pls.sql
|-- VERSION
```

6 directories, 39 files

Change into the source directory, where the SQL files exist.

```
$ cd source
```

Installation is done through the standard Oracle SQL*Plus tool.

```
$ sqlplus @install_hive.sql
```

Upon successful installation, the message ...

```
Installation successful
```

... should have been revived. If error messages are present review them in the <db>_install_hive_odci.<date_time>.log file.

If errors need to be fixed, then it is recommended to run the remove_hive.sql script before attempting to install again.

Load JDBC JAR

Once the database changes have been made, you must install the Hive JAR file. In this example, we are using the Progress Datadirect HIVE JDBC Driver, which is found in the file hive.jar using the \$ORACLE_HOME/bin./loadjava utility install the JAR classes, like ...

```
loadjava -order \
         -verbose \
         -resolve \
         -recursivejars \
         -resolver "(((* hive) (* sys) (* public)))" \
         -user sys \
         -schema hive \
         jdbc/hive.jar
```

Password:

```
**
```

```
arguments: '-user' 'sys' '-resolver' '(((* hive) (* sys) (* public)))' '-order' '-verbose'
'-resolve' '-recursivejars' '-schem
a' 'hive' 'jdbc/hive.jar'
```

You can use the installation script jdbc/load-jdbc.sh as a template or to load your JAR file. Once complete, make sure to review the report and verify no errors are encountered during the installation.

```
Classes Loaded: 538
Resources Loaded: 41
Sources Loaded: 0
Published Interfaces: 0
Classes generated: 0
Classes skipped: 0
Synonyms Created: 0
Errors: 0
```

Removal

The HIVE-ODCI can be removed from the database in the same manner it was installed, but using the `remove_hive.sql` script.

```
$ sqlplus @remove_hive.sql
```

Parameters

| Name | Description |
|---------------------------------|-------------|
| application | |
| version | |
| log_level | |
| encrypted_values | |
| hive_users | |
| hive_admin | |
| hive_jdbc_driver | |
| hive_jdbc_url | |
| hive_host | |
| hive_port | |
| hive_auth | |
| hive_user | |
| hive_pass | |
| hive_principal | |
| java.security.krb5.realm | |
| java.security.krb5.kdc | |
| java.security.auth.login.config | |
| bind_owner | |
| bind_role | |
| bind_roles | |
| query_limit | |

Interface

Examples