Quest® Data Transporter for Hive 1.1

User Guide

If you have any questions regarding your potential use of this material, contact:

Quest Software World Headquarters
LEGAL Dept
5 Polaris Way
Aliso Viejo, CA 92656
email: legal@quest.com

Refer to our Web site (www.quest.com) for regional and international office information.

Trademarks

Quest Data Transporter for Hive 1.1
User Guide
Monday, 30 May 2011

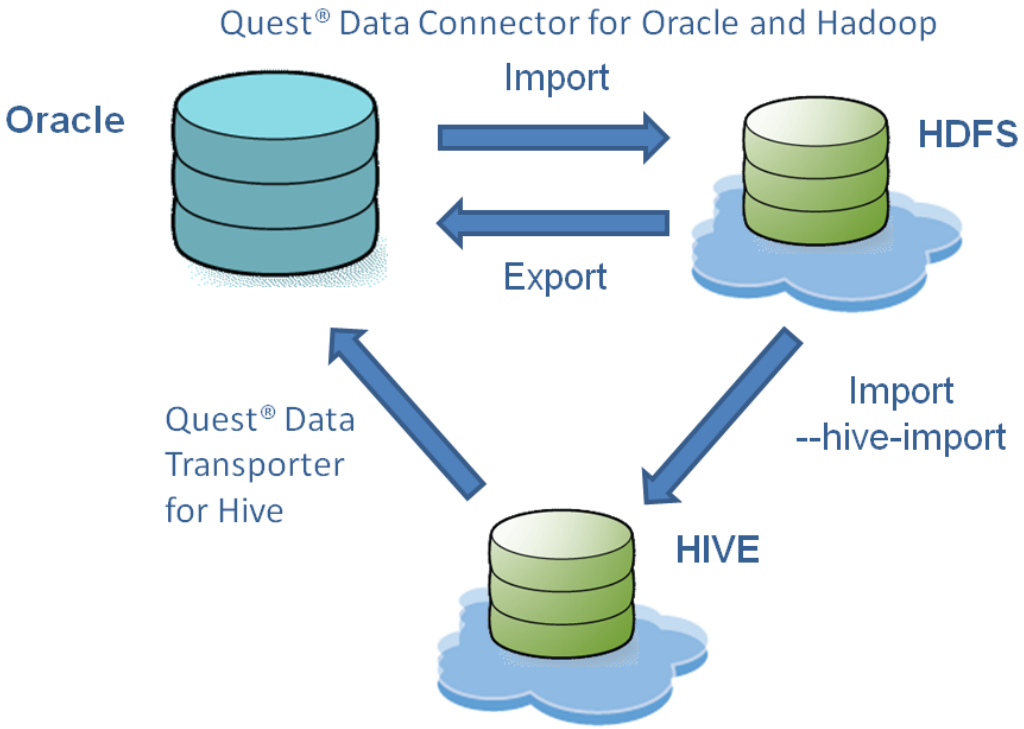
Table of Contents

1

What Is Quest Data Transporter for Hive?

Quest Data Transporter for Hive is a Java command-line utility that allows you to execute a Hive query and insert the results into an Oracle table. Quest Data Transporter for Hive is distributed with Quest Data Connector for Oracle and Hadoop.

To Extract From ...	**And Import Into...**	**Do ...**
The Results of a Hive Query	An Oracle Table	See this user guide for more information.
HDFS	An Oracle Table	See the *Quest Data Connector for Oracle and Hadoop User Guide* for more information.
An Oracle Table	HDFS for analysis by Hive	See the *Quest Data Connector for Oracle and Hadoop User Guide* for more information.
An Oracle Table	HDFS for analysis by Hadoop	See the *Quest Data Connector for Oracle and Hadoop User Guide* for more information.


Quest® Data Connector for Oracle and Hadoop
Oracle
Import
Export
HDFS
Quest® Data
Transporter
for Hive
Import
--hive-import
HIVE

2

Prerequisite Programs

Ensure the following is installed prior to running Quest Data Transporter for Hive:

- Hive 0.5 or 0.7 (version 0.5.0+32 or 0.7.0)
- The thin Oracle 11.2 JDBC driver (ojdbc6.jar)

Strongly recommended:

- Install a fully functional database such as MySQL or Oracle and set it up to store Hive table metadata. See "Store Hive Table Metadata In A Database Like MySQL or Oracle" (page 6) for more information.

Ensure The Oracle Database JDBC Driver Is Setup Correctly

You may want to ensure the Oracle Database 11g Release 2 JDBC driver is setup correctly on your system. This driver is required.

The Oracle Database 11g Release 2 JDBC driver file is **ojdbc6.jar** (3.2Mb).

If this file is not on your system then download it from:
http://www.oracle.com/technetwork/database/features/jdbc/index-091264.html

This location of this file is defined during configuration. See "Configure Quest Data Transporter for Hive" (page 10) for more information.

Store Hive Table Metadata In A Database Like MySQL or Oracle

By default Hive uses a database called Derby to store metadata about the tables contained within Hive. Derby is a single-user database.

Use of Derby with Quest Data Transporter for Hive can result in confusion. For example:

- You may see the following types of errors: *Error in semantic analysis ... table not found ...*
- You cannot run the Hive Command-Line-Interface (CLI) service while the Hive JDBC Service is running. The CLI is used to interactively query Hive, while the JDBC service is used by applications such as Quest Data Transporter for Hive to programmatically query Hive.

- You need to ensure both services are started from the same directory on the file system to ensure they see the same data.

We strongly recommend you configure Hive to use a database such as MySQL or Oracle, instead of Derby **before** you use Quest Data Transporter for Hive.

Instructions are available at:
http://wiki.apache.org/hadoop/Hive/AdminManual/MetastoreAdmin#Local_Metastore.

3

Download and Unpack Quest Data Transporter for Hive

Quest Data Transporter for Hive is distributed with Quest Data Connector for Oracle and Hadoop.

Download Quest Data Connector for Oracle and Hadoop from the Quest Data Connector for Oracle and Hadoop web site: http://www.quest.com/hadoop

Extract The Quest Data Transporter for Hive Archive

The Quest Data Transporter for Hive archive is within the Quest Data Connector for Oracle and Hadoop download. The Quest Data Transporter for Hive archive file name is of the form **orahive-*version*.tgz** where *version* is the numeric identifier of the Quest Data Transporter for Hive release, such as *1.1.0.xxx*.

» Extract the Quest Data Transporter for Hive archive with the command:
`tar xzf orahive-`*`version`***`.tgz`**

The Quest Data Transporter for Hive archive files are extracted into a sub-directory called **orahive-*version***.

Quest Data Transporter for Hive Archive Files

File	Description
configure.sh	The script file used to configure Quest Data Transporter for Hive.
hql.txt	A sample Hive query in HQL.
LICENSE.txt	A text file containing the license for Quest Data Transporter for Hive.
log4j.txt	A file used to configure the logging mechanism of Quest Data Transporter for Hive.
orahive.jar	The Java archive (jar) file containing the Quest Data Transporter for Hive application code.
orahive.sh	The shell script that runs Quest Data Transporter for Hive.
orahiveuserguide.pdf	An Adobe PDF version of this Quest Data Transporter for Hive User Guide.
version.txt	A text file containing the version string for this Quest Data Transporter for Hive release.

4

Configure Quest Data Transporter for Hive

1. Login to the shell as a user who has permission to install files.
2. Change directory to where the Quest Data Transporter for Hive files are extracted: **orahive-*version***
3. Execute the shell script file: `configure.sh`

   Note: At the end of this process the following file is created: ***defaultOraHive.sh***. To specify a different filename execute the shell script:

   `configure.sh -o filename.sh`
4. You will be prompted for the following details which will be saved in the created configuration file (*defaultOraHive.sh*):

Variable	Description
HADOOP_ CORE_JAR	The location of the Hadoop core jar file.
HIVE_ HOME	The location of Hive.
HIVE_URL	The URL of the Hive service.
HIVE_USER	The user name to the Hive service.
HIVE_ PASSWORD	The password to the Hive service. **Notes:** • If omitted, Quest Data Transporter for Hive will prompt you for the password. • The password saved here is not encrypted.
OJDBC6	The location of the Oracle JDBC driver.
ORACLE_ URL	The JDBC connection string to connect to the Oracle database instance.
ORACLE_ USER	The user name to the Oracle database instance.

Variable	Description
ORACLE_ PASSWORD	The password to the Oracle database instance. **Notes:** • If omitted, Quest Data Transporter for Hive will prompt you for the password. • The password saved here is not encrypted.
ORACLE_ TABLE	The Oracle table created as a result of the Hive query. **Notes:** • This table is created by Quest Data Transporter for Hive. Ensure it does not exist prior to running Quest Data Transporter for Hive. • You can override the table name when you run Quest Data Transporter for Hive. See "Run Quest Data Transporter for Hive" (page 13) for more information.
INSERT_ BATCH_ SIZE = 500 COMMIT_ BATCH_ COUNT = 20	The rows are inserted in batches of 500 rows. A commit is performed every 20 batches. **Note:** The smaller the batch size the longer the time Quest Data Transporter for Hive takes to execute the job. If the batch size too large you may see an insufficient memory error. Each batch of data must be able to fit into the JVM's memory.

5

Create the HQL Query

Write the Hive HQL query in a text file. On running Quest Data Transporter for Hive, the result of this query is inserted into the Oracle table.

The HQL query must be one complete statement. Multiple statements are not allowed.

The statement cannot end with a semicolon even though the Hive command line interface requires a semicolon at the end of each command / query.

Notes:

- The HQL query can be included on the command line instead of written to a text file. See "Run Quest Data Transporter for Hive" (page 13) for more information.
- The name of the file is assumed to be *hql.txt*. If a different filename is used then it must be specified on the command line. See "Run Quest Data Transporter for Hive" (page 13) for more information.
- A sample HQL query file can be found in the Quest Data Transporter for Hive archive. See "Download and Unpack Quest Data Transporter for Hive" (page 8) for more information.

6

Run Quest Data Transporter for Hive

1. Login to a shell.
2. Change directory to where the Quest Data Transporter for Hive files are extracted: **orahive-*version***

3. Enter the command to run Quest Data Transporter for Hive: `orahive.sh`

Command Line Options	Description
`-v`	Activate Verbose mode. Display progress and context information on *stdout*.
`-h`	Display basic help on *stdout*.
`-c filename`	Define the Quest Data Transporter for Hive configuration file. See "Configure Quest Data Transporter for Hive" (page 10) for more information. **Note:** If omitted the filename is taken to be *defaultOraHive.sh* in the current directory.
`-t tablename`	Define the Oracle table name. **Notes:** • This option is provided for convenience. Multiple tables can be populated without constantly changing the Quest Data Transporter for Hive configuration file. • If omitted the table name in the configuration file is used.
`-q filename`	Define the file containing the Hive HQL query. See "Create the HQL Query" (page 12) for more information. **Notes:** • If omitted the filename is taken to be *hql.txt* in the current directory. • If used in combination with '*query text*' then the query filename is ignored.
`'text'`	Type the Hive query in single quotes. Use this as an alternative to defining the query in a text file. **Note:** It is important to wrap the query text in single quotes. Do not use double quotes as double quotes may lead to unexpected results.

4. You may be prompted while Quest Data Transporter for Hive is running:

Quest Data Transporter for Hive prompts / actions	Description

Connect to the Hive Service	You will be prompted to enter the Hive password if none was specified in the Quest Data Transporter for Hive configuration file.
	`Error Log File: /home/username/workspace/orahive/stderr.txt`
	`*****`
	`*** Using Quest Data Transporter for Hive 1.1.xxxx ***`
	`*** Copyright 2011 Quest Software, Inc. ***`
	`*** ALL RIGHTS RESERVED ***`
	`*****`
	`Enter the password for the Hive Service:`

Connect to the Oracle Instance	You will be prompted to enter the Oracle password if none was specified in the Quest Data Transporter for Hive configuration file. `Enter the password for the Oracle Database:`
Execute the Hive query	This may take a few minutes to complete.
Create the Oracle Table	If the Oracle table already exists you will be prompted to retry. If you see this prompt then: (1) Drop the Oracle table and (2) Respond "Y". If you respond "N" the job will terminate. `Would you like to retry creating the Oracle table "xxx"? (y/n)` **Notes:** • This prompt may occur several minutes into processing so the most practical approach is to ensure you drop the Oracle table before you run Quest Data Transporter for Hive. • The resultset from Hive is used to determine the structure of the Oracle table. • Oracle does not have a boolean data type. If required a NUMBER column will be used to store boolean data. False is represented by 0. True is represented by 1.
Insert the results from the Hive query into the Oracle table.	Quest Data Transporter for Hive iterates the results returned from Hive and inserts them in the Oracle table. By default these rows are inserted in batches of *500* rows and a commit is performed

Quest Data Transporter for Hive prompts / actions	Description
	every *20* batches. You can customize these default settings. See "Configure Quest Data Transporter for Hive" (page 10) for more information.

7

Troubleshooting Quest Data Transporter for Hive

Check The Error Logs

Check the contents of the stderr file.

During execution, the first line Quest Data Transporter for Hive sends to the console is the name of the stderr text file.

Cannot Connect to The Hive Service

If you cannot connect to the Hive service, it may be due to a blocked port. The Hive service uses port 10000 by default. Port tunneling may be used to overcome this problem.

An example port tunnelling command is:

```
ssh -L 10000:hive-server:10000 user@hive-server
```

Cannot Connect to The Oracle Database Instance

Error: "`Unable to connect to Oracle via the JDBC URL`"

on two Quest Data Transporter for Hive commands submitted concurrently or sequentially without a break between. This kind of error is most likely to occur during batch processing.

The solution is to allow a space of a few seconds between each job.

Insert The Hive Data into Oracle

```
Insufficient memory error
```

The batch size may be too large. Each batch worth of data must be able to fit into the JVM's memory. See "Configure Quest Data Transporter for Hive" (page 10) for more information.

Note: When you reduce the batch size there will be more round-trips from the client machine to the Hive service. This can have a substantial impact on performance. The smaller the batch size the longer the time Quest Data Transporter for Hive takes to execute the job.

Appendix: Contact Quest

Contact Quest Support

If this document is unable to resolve problems you may encounter, please post your questions to the Quest Data Connector for Oracle and Hadoop Forums:

http://toadforcloud.com/forumindex.jspa?categoryID=735

Contact Quest Software

Email	info@quest.com
Mail	Quest Software, Inc. World Headquarters 5 Polaris Way Aliso Viejo, CA 92656 USA
Web site	www.quest.com

See our web site for regional and international office information.

About Quest Software

Now more than ever, organizations need to work smart and improve efficiency. Quest Software creates and supports smart systems management products—helping our customers solve everyday IT challenges easier and faster. Learn more at www.quest.com.

Appendix: Third Party Contributions

Quest Data Transporter for Hive contains some third party components (listed below). Copies of their licenses may be found at http://www.quest.com/legal/third-party-licenses.aspx.

Component	License or Acknowledgement
JUnit 4.8.1	Common Public License - v 1.0
Log4J 1.2.15	Copyright © 2010 The Apache Software Foundation. Developed by the Apache Software Foundation (http://www.apache.org.) Licensed under the Apache Software License, Version 2.0.