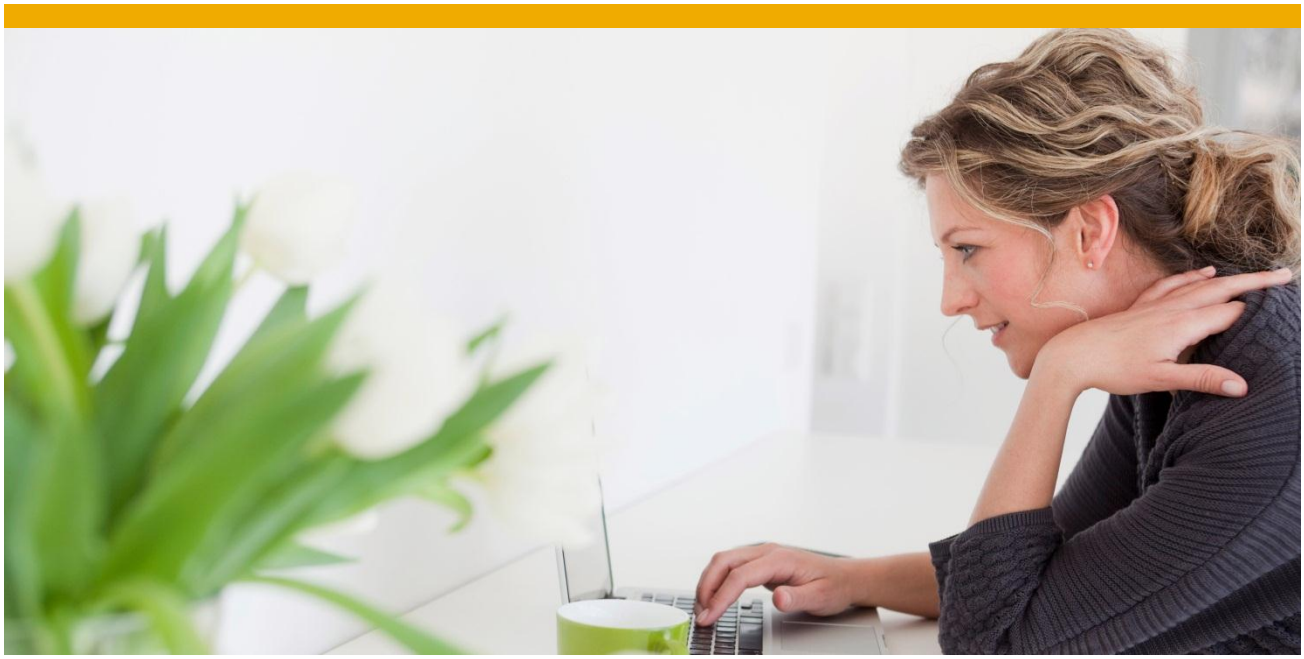


Set Up Hortonworks Hadoop with SQL Anywhere

TABLE OF CONTENTS

1 INTRODUCTION 3

2 INSTALL HADOOP ENVIRONMENT 3

3 SET UP WINDOWS ENVIRONMENT 5
- 3.1 Install Hortonworks ODBC Driver 5
- 3.2 ODBC Driver Setup 5

4 CONNECT USING SYBASE CENTRAL 6
- 4.1 Create Remote Server 8
- 4.2 Create and Access a Proxy Table 11

5 CONNECT USING INTERACTIVE SQL 14
- 5.1 Create Remote Server 14
- 5.2 Access Hadoop from Server 15
- 5.3 Access Hadoop from Proxy Table 16
- 5.4 Create Table 16

1 INTRODUCTION

Welcome to setting up Hadoop and Hive with SQL Anywhere!

Hadoop is an open-source framework designed to handle big data. It allows large amounts of data to be distributed across clusters of computers. Then, when retrieving the data, a MapReduce algorithm is implemented to divide up the work across the clusters (map) and then recombine the data (reduce).

Hive is an infrastructure which can be used on top of Hadoop. It provides the ability to access data without using MapReduce or code files. Hive allows the creation of tables, which are stored as directories in the Hadoop File System (HDFS). These tables can be accessed using HiveQL, a subset of SQL-92, which implicitly makes MapReduce calls.
Note: Since a Hive query must access the server, wait for a MapReduce job and aggregate results, access to a proxy table may be quite slow.

This guide provides directions to set up SQL Anywhere with a Hortonworks Hadoop Hive server and HiveQL. This is not a guide to help with installation, set-up or usage of SQL Anywhere. For SQL Anywhere documentation check: http://dcx.sybase.com/index.html

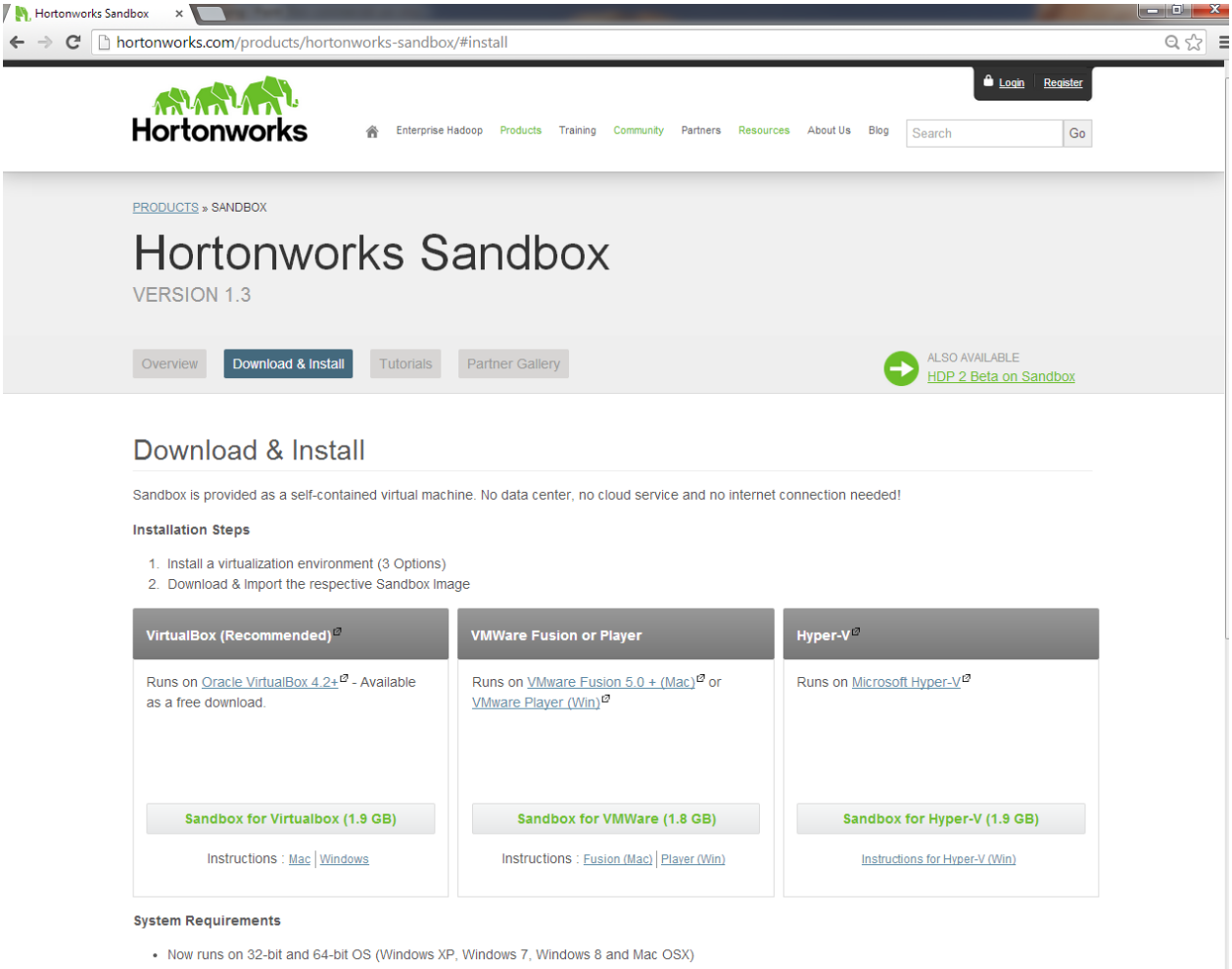
The provided instructions assume an installation of SQL Anywhere on a Windows machine.

2 INSTALL HADOOP ENVIRONMENT

Hortonworks Sandbox is a distribution of the Hadoop environment that installs as a standalone virtual machine. This environment is a portable Hadoop environment, which is smaller and easier to use, but does not allow for a multi-node cluster or some of the functionality of a complete Hadoop installation. For the full platform, consider the Hortonworks Data Platform.

Download the product and follow the installation instructions at:
http://hortonworks.com/products/hortonworks-sandbox/#install

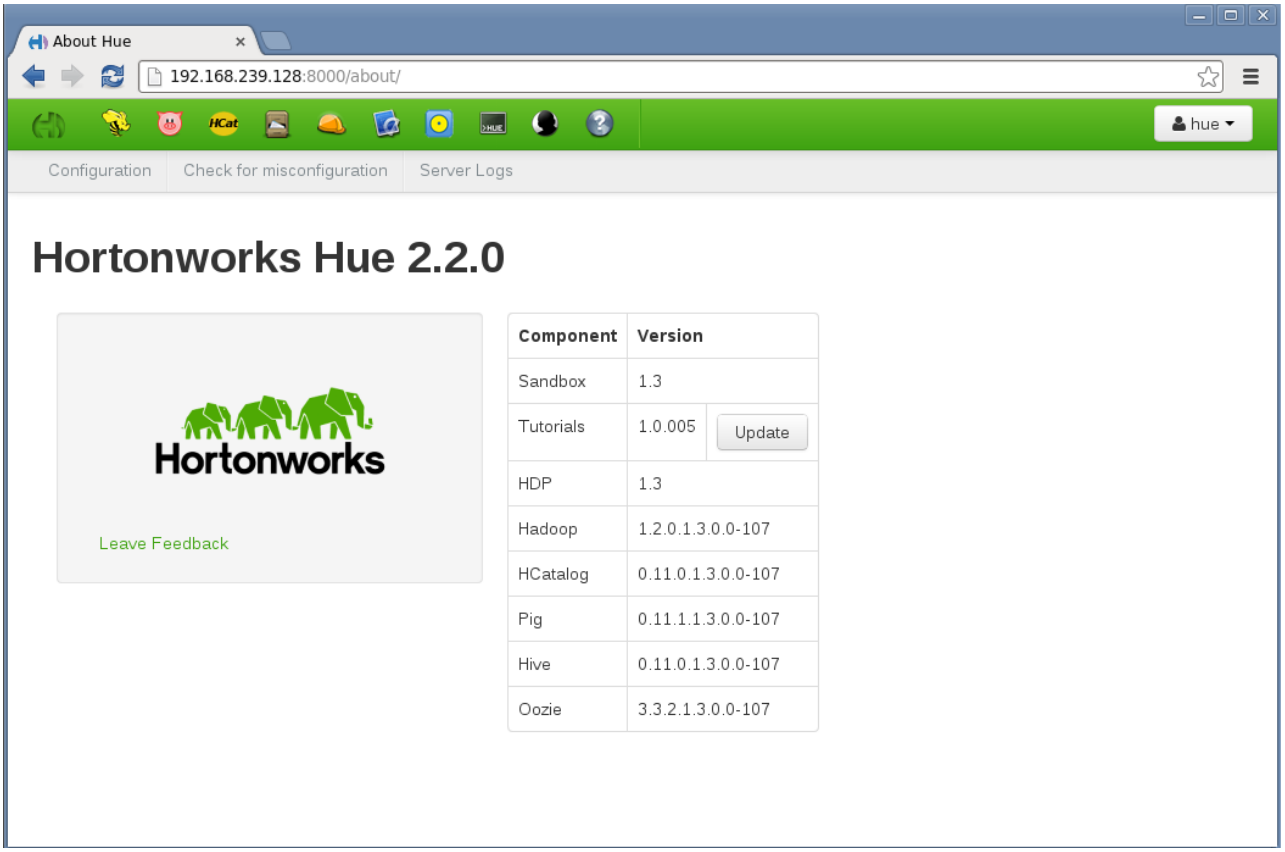
Note: This installation requires an installation of VMware Player, Virtualbox or equivalent software. Download links are available on the Hortonworks website.



The simplest installation process is to use VMware Player, and then open and import the provided file from the player.

Once this is installed and running, the provided IP address can be used to connect to the Hive server or access a Hadoop UI in a web browser.

To ensure the IP address is accessible, enter the given IP address in a web browser. You should be able to continue to the sandbox and access a page similar to the following:



Note: Network security can limit access to the IP address generated for the Hadoop VM. It may only be accessible from other devices in the same subnet. Ensure that the host which IQ is running from can access the IP.

3 SET UP WINDOWS ENVIRONMENT

Note: Screenshots and further instructions for this set up can be found at:

http://hortonworks.com/hadoop-tutorial/how-to-install-and-configure-the-hortonworks-odbc-driver-on-windows-7/

3.1 Install Hortonworks ODBC Driver

In the Windows environment with SQL Anywhere installed, a Hortonworks ODBC driver is required to use to access the Hive server. This can be downloaded at the following link:
http://hortonworks.com/products/hdp/hdp-1-3/#add_ons

Save the appropriate .msi file (32 or 64-bit).

Double-click to run the file.

Finish the installation allowing it to continue at all prompts.

3.2 ODBC Driver Setup

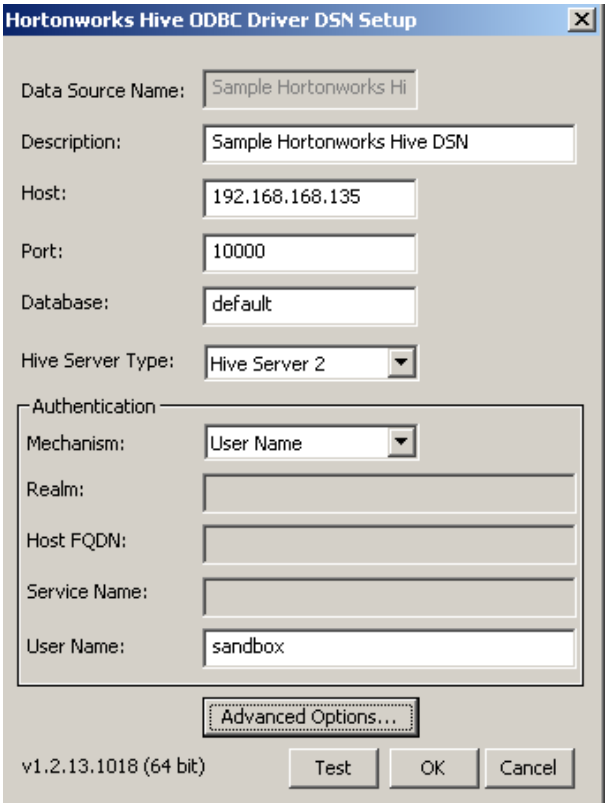
To use the installed ODBC driver, you must set up the ODBC settings.

Open the Control Panel. Then select “Administrative Tools”. Double-click “Data Sources (ODBC)”.

Click the “System DSN” tab.

Select the “Sample Hortonworks Hive DSN” and click “Configure”.

In the window that appears, fill out the information, using the IP address from the Hortonworks installation for Host and “sandbox” for User Name.



Click the “Test” button and ensure it completes successfully.

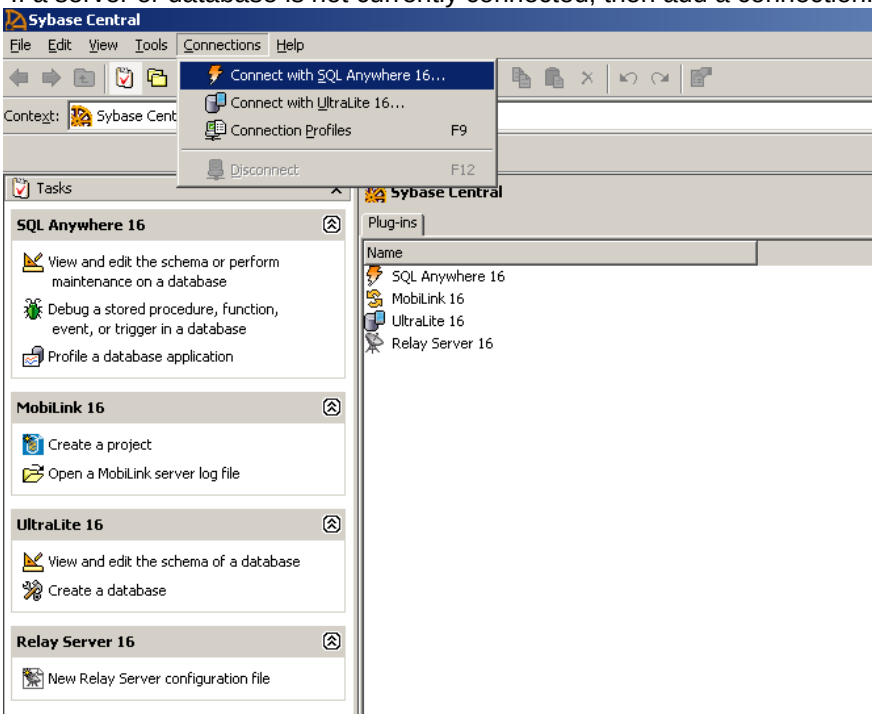
4 CONNECT USING SYBASE CENTRAL

Now that the environment and DSN are set up, we can use them to access the Hive server. If you would rather use SQL statements executed in Interactive SQL, instead of the Sybase Central graphical tool, to access the Hive server, skip this section and go to section 5.

We will access the Hive server from an existing SQL Anywhere server, by creating a remote server and then proxy tables, as necessary.

Open Sybase Central.

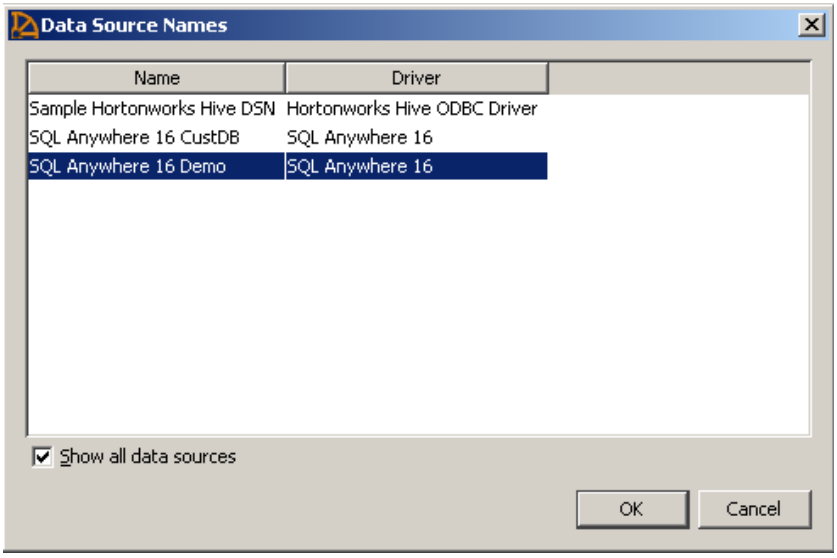
If a server or database is not currently connected, then add a connection.



Click “Connections” in the top menu and choose “Connect with SQL Anywhere 16”.

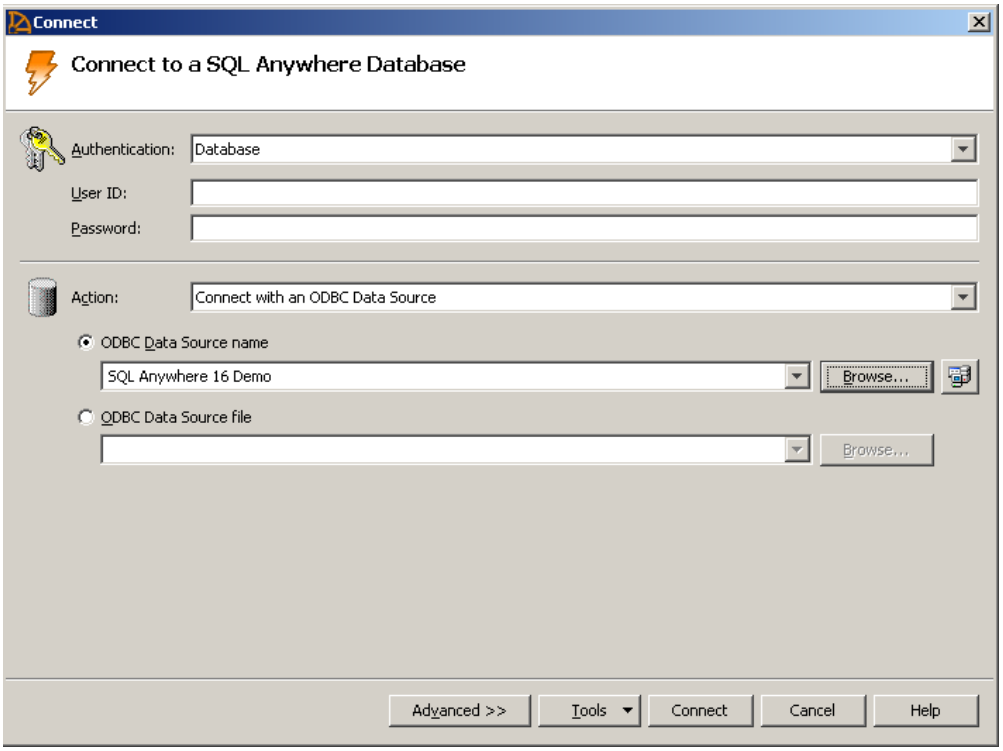
Connect to any created or running database, as applicable. In this example, we will connect to the demo database, that comes with the SQL Anywhere installation.

We will choose “Connect with an ODBC Data Source” . Choosing “ODBC Data Source Name”, click “Browse…” to choose the data source.

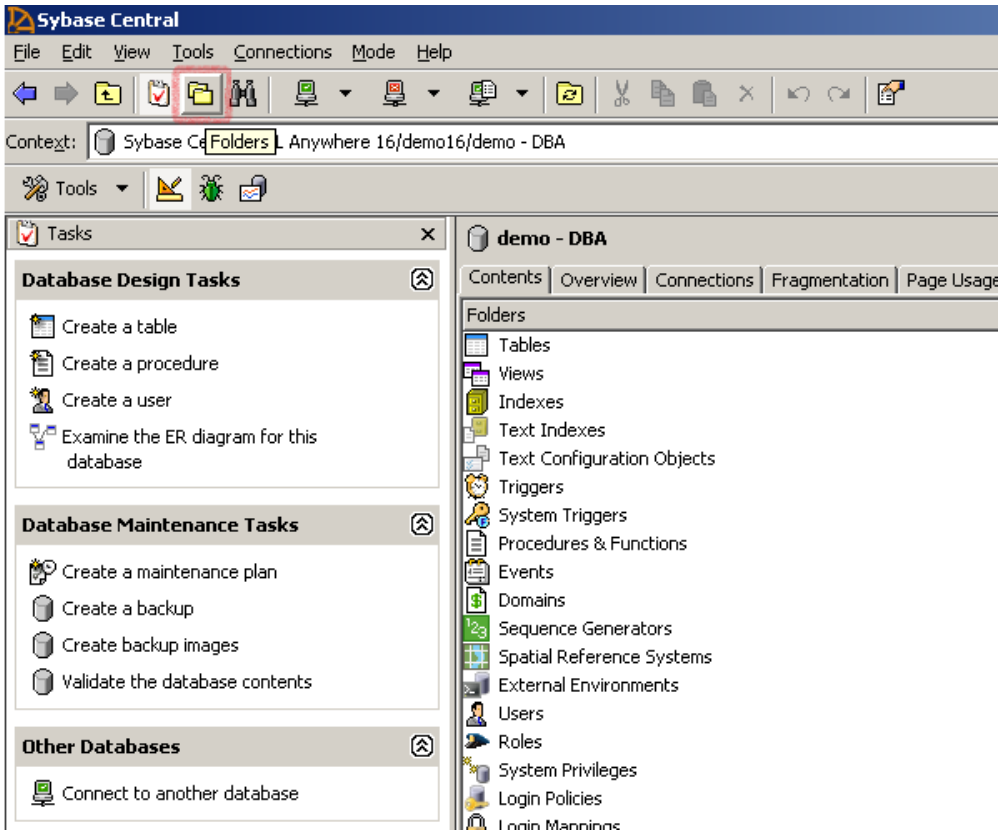


Choose “SQL Anywhere 16 Demo” and click OK.

Note: You can connect to the “Sample Hortonworks Hive DSN” directly, if you want to access the Hive server as a server, without using proxy tables. This, however, does not allow you to access both SQL Anywhere data and Hive data simultaneously.



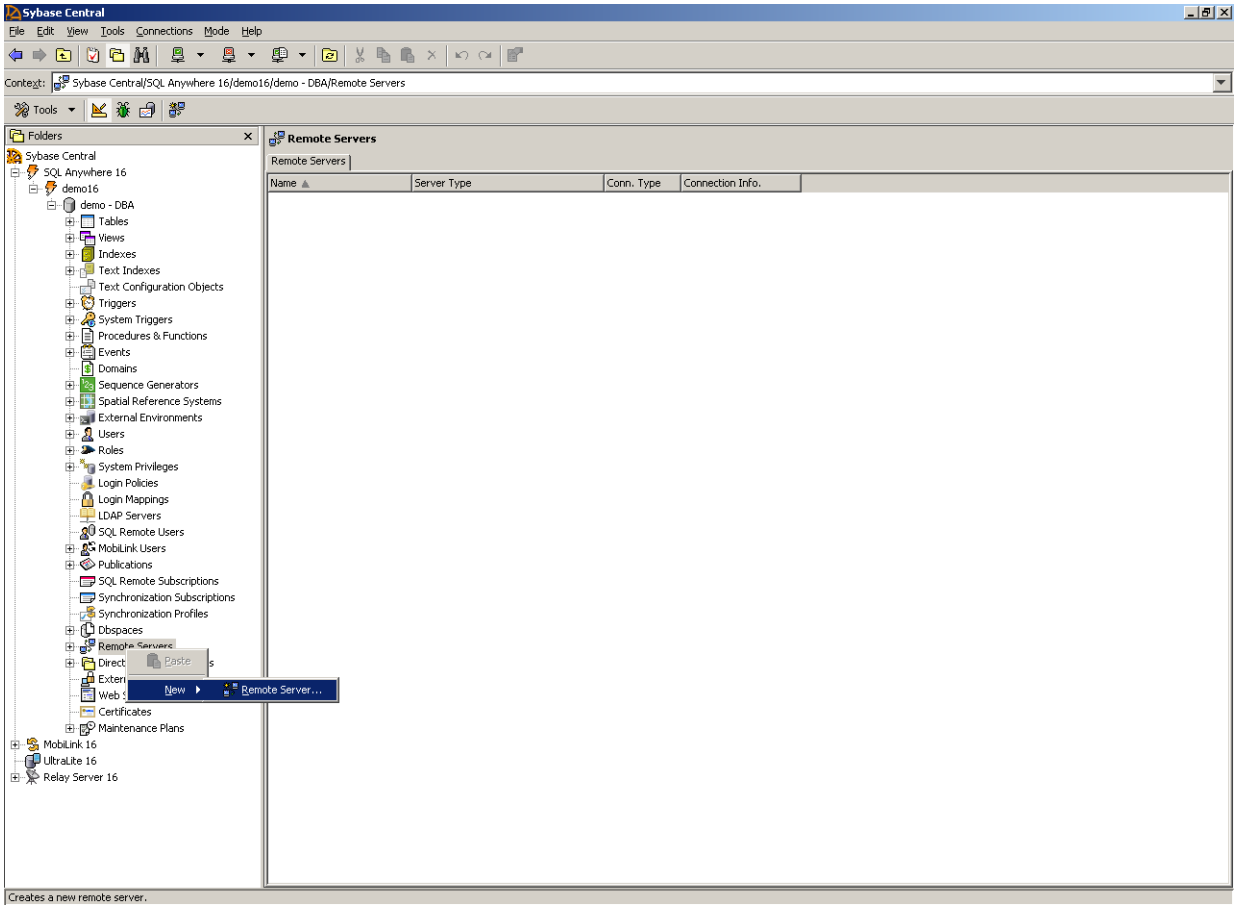
Now, click “Connect”, no User ID or Password is required to connect to the demo database.



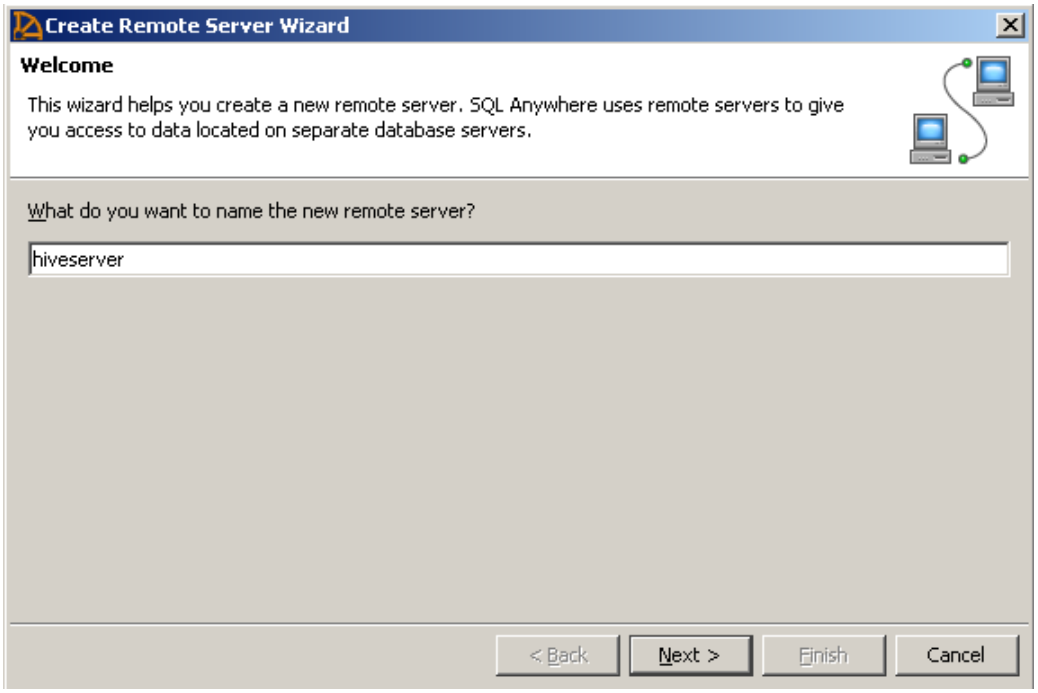
Click the “Folders” icon to change to Folder View.

4.1 Create Remote Server

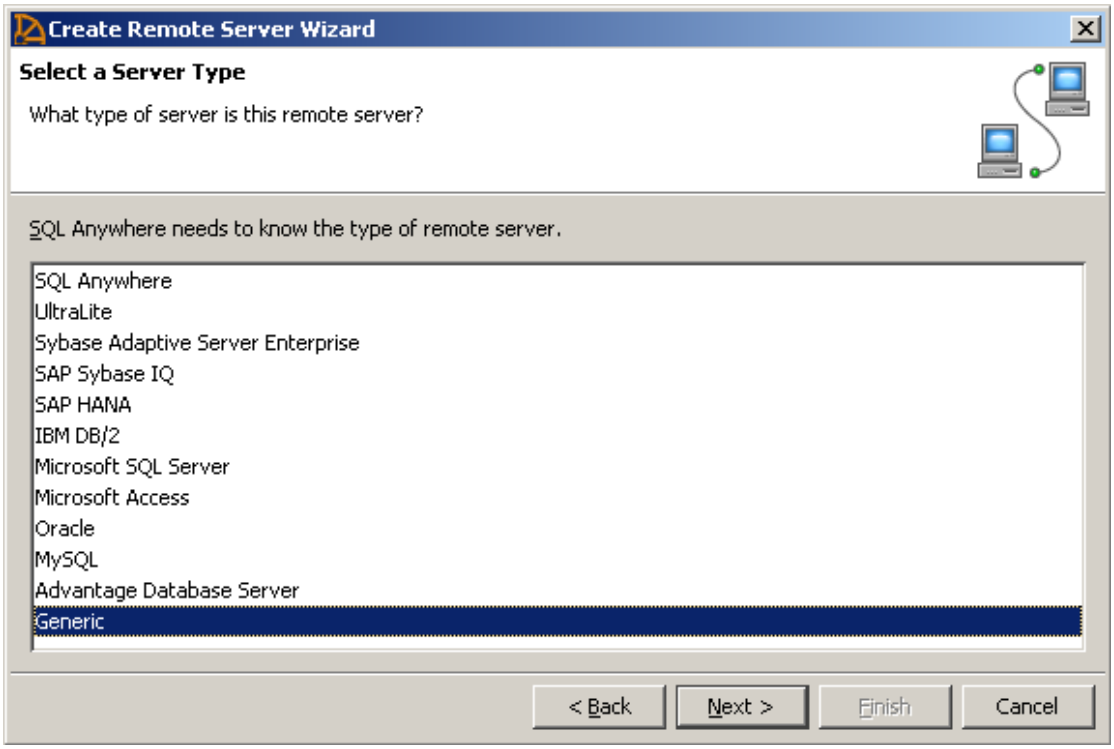
Now, we will create a remote server.



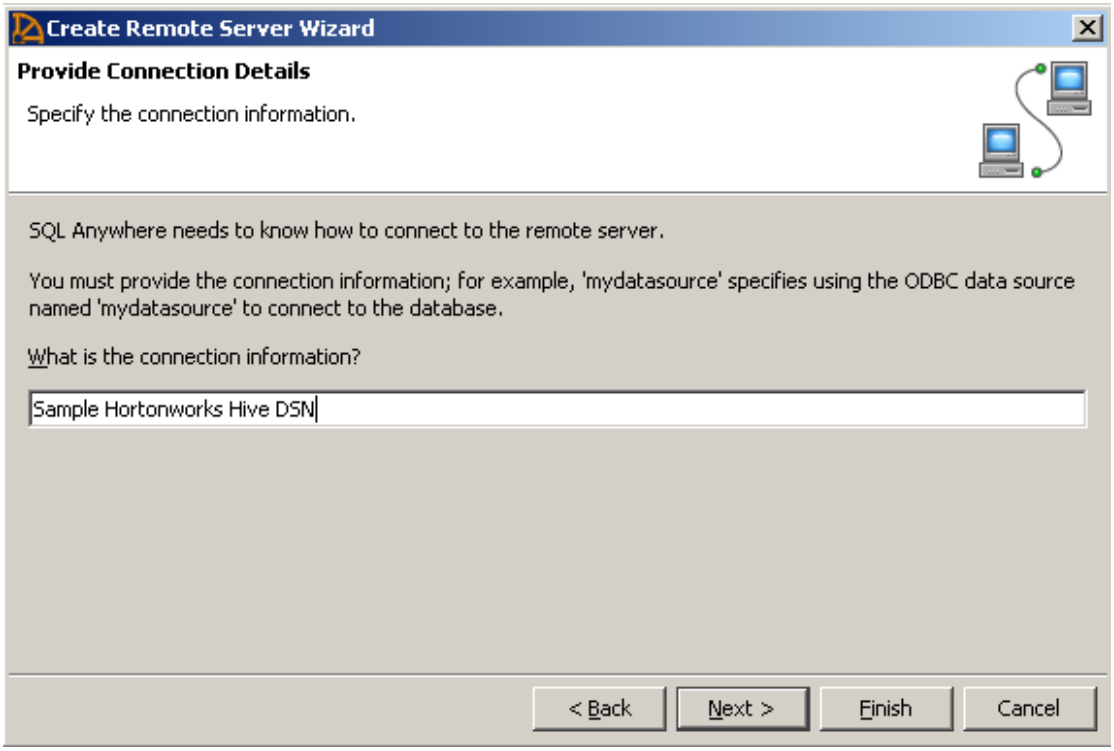
Right-click on “Remote Servers” and choose “New”, “Remote Server…”



Choose a name for the remote server and click “Next”.



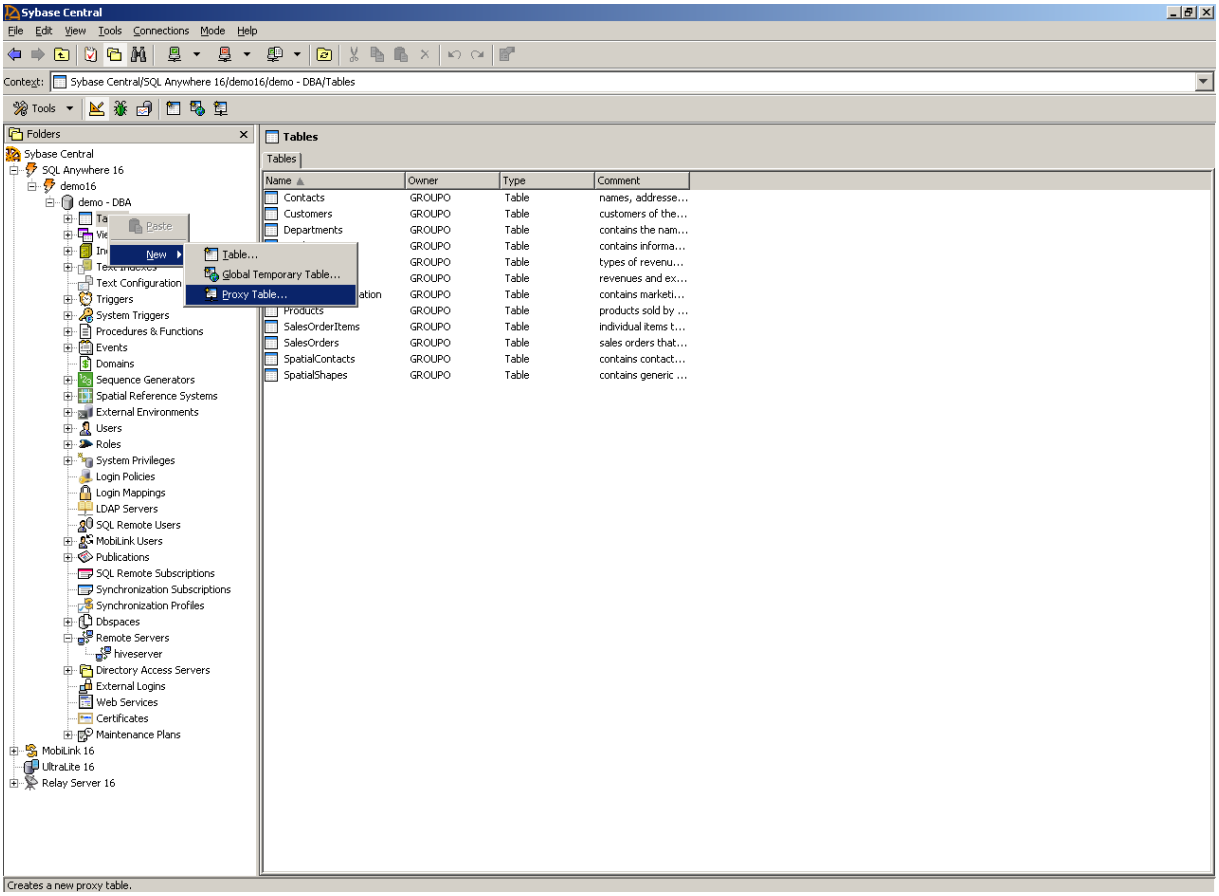
Choose “Generic” for the type of server and click “Next”.



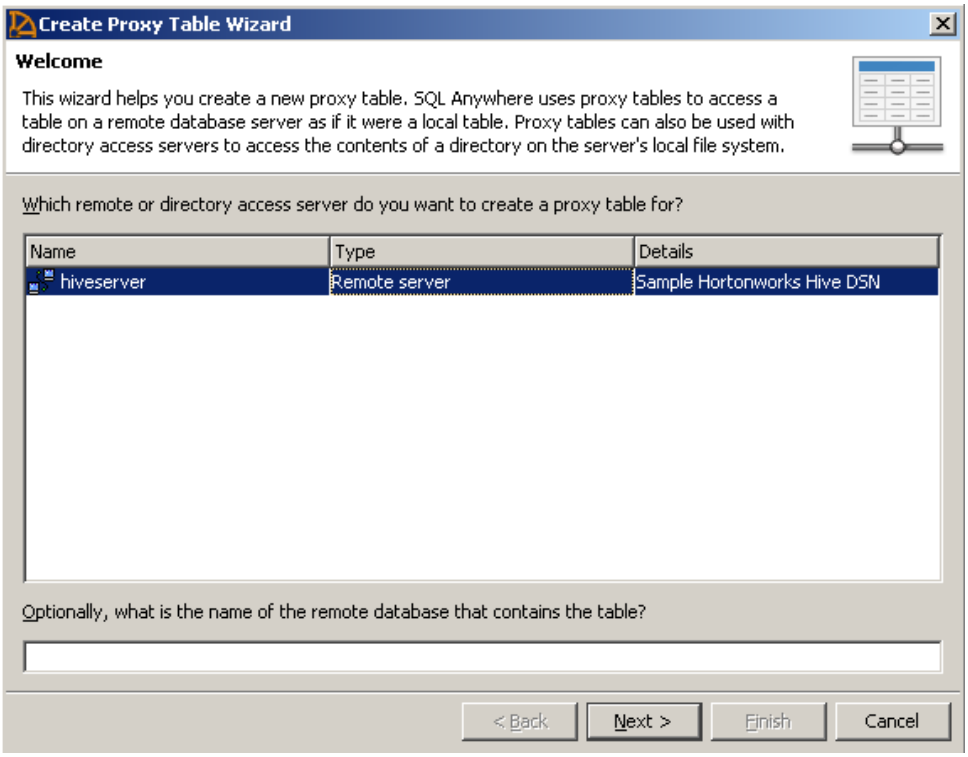
Specify the name of the DSN created in the ODBC Driver Setup for connection information and click “Next”. Continue to click next, accepting the defaults until the remote server is created.

4.2 Create and Access a Proxy Table

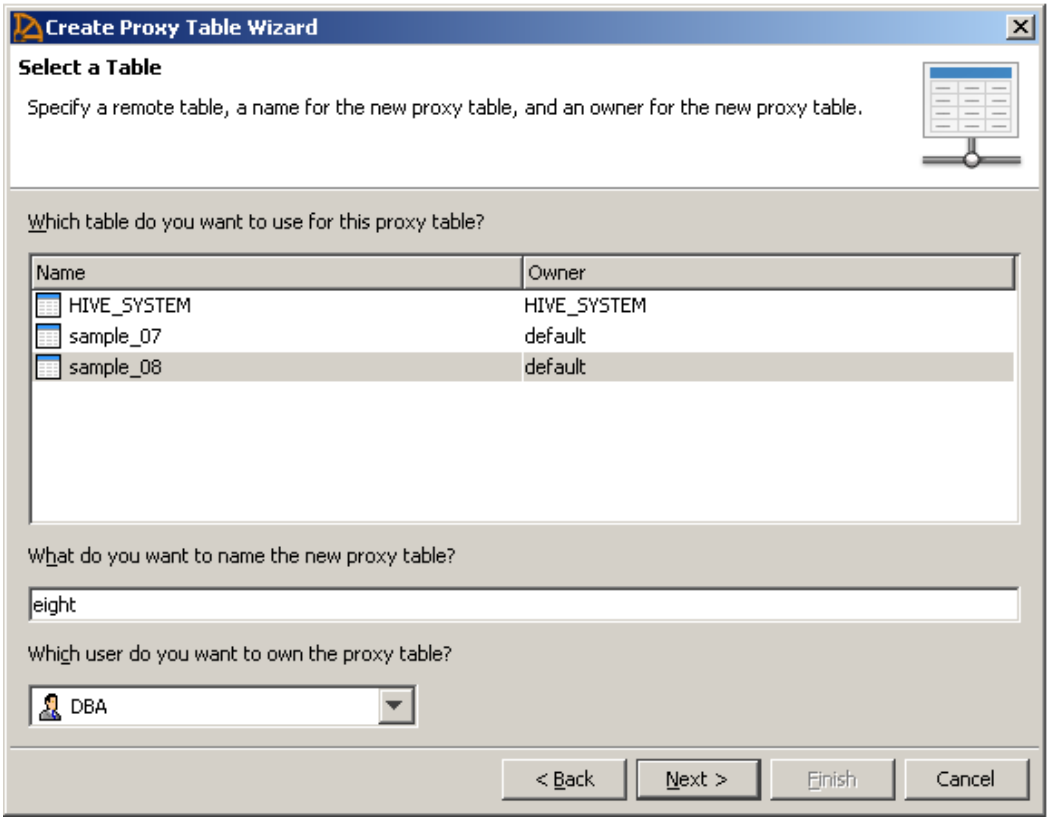
We can now create a proxy table, which is a table in SQL Anywhere pointing to a table on the Hive server.



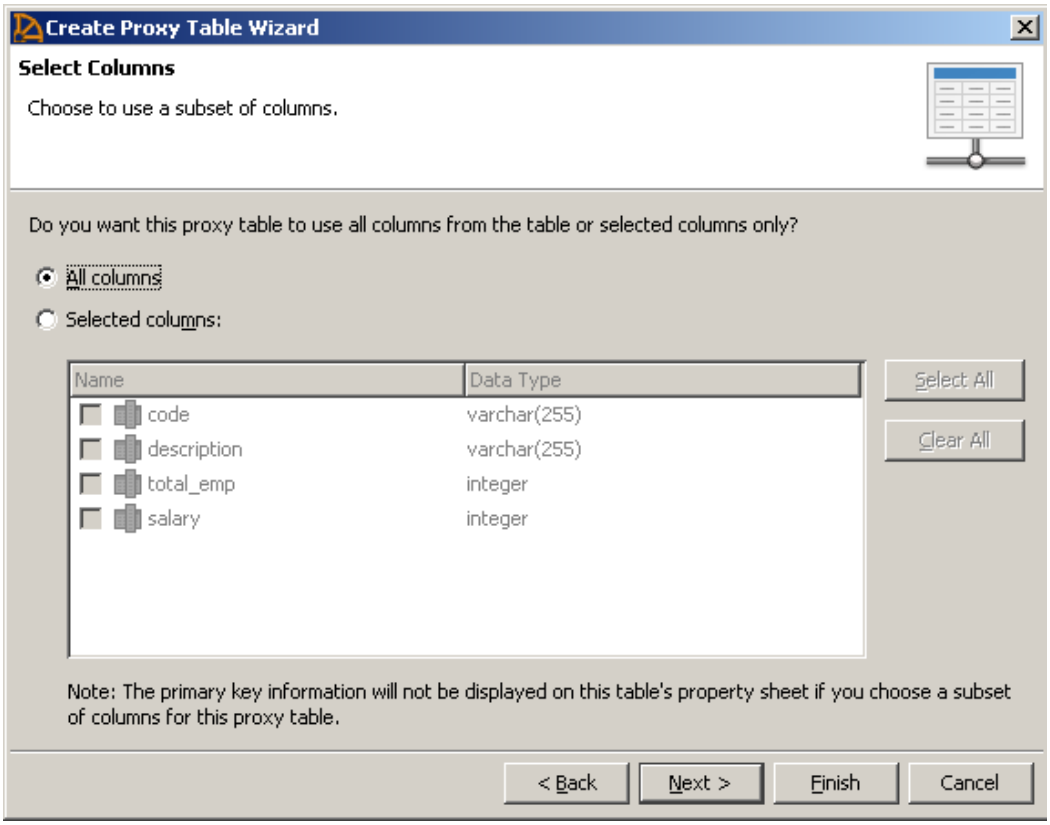
Right-click on “Tables” and click “New”, “Proxy Tables…”



Choose the remote server that was just created and click “Next”.



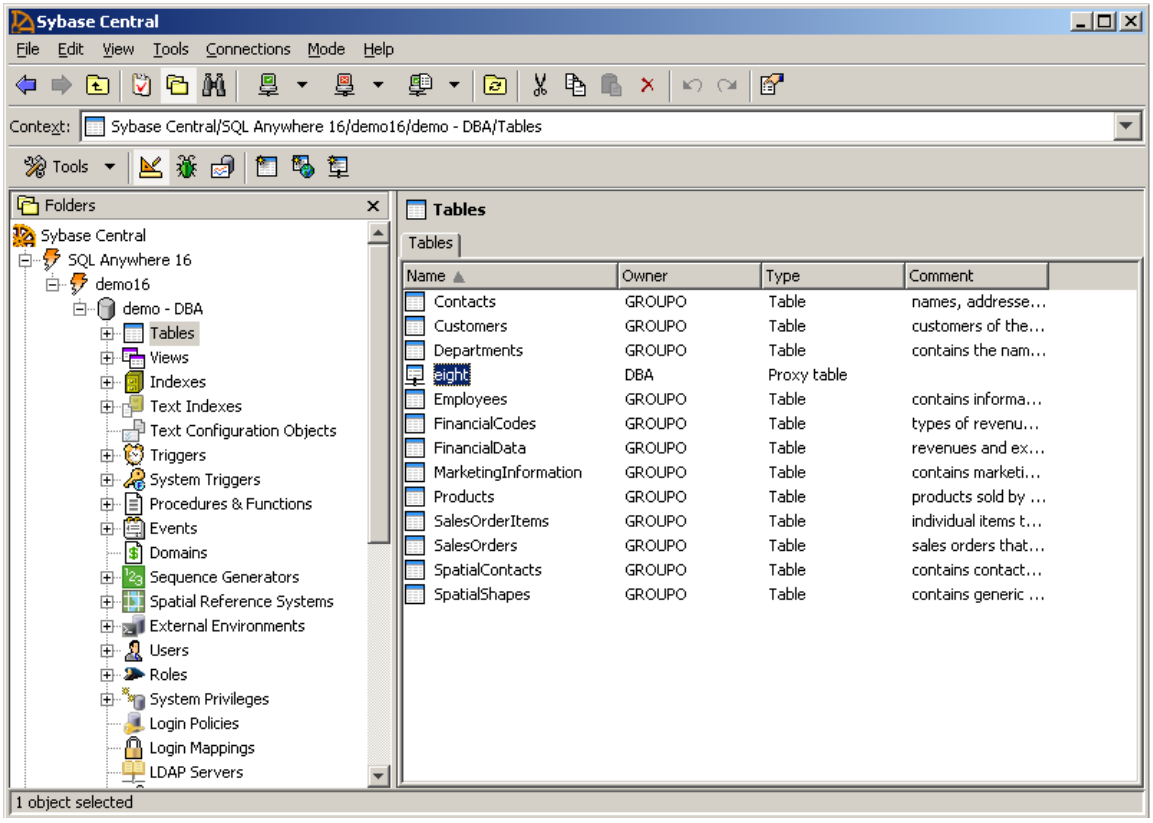
Highlight the table for which to create a proxy of, and choose a new name and user, if desired. Click “Next”.



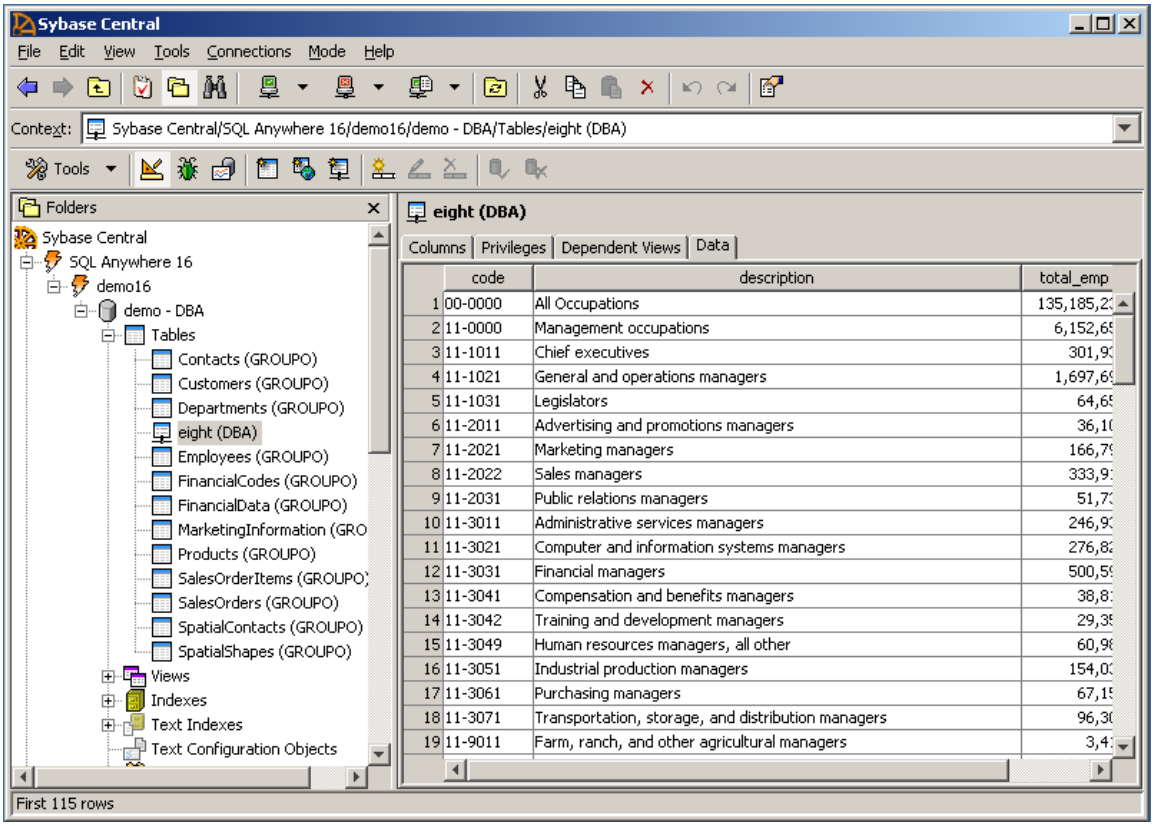
Change column choices, if desired. Click “Next”.

Continue to click next and finish until the proxy table is created.

Now the newly created proxy table will show up in the list of tables.



Double-click on the table and choose the “Data” tab.



All the data from the table in the Hive server will appear.

5 CONNECT USING INTERACTIVE SQL

This section is an alternative to section 4. Instead of connecting using the Sybase Central graphical tool, this section describes how to connect using SQL statements executed from Interactive SQL.

Now that the environment and DSN are set up, we can use them to access the Hive server.

In the Start Menu, open SQL Anywhere 16 -> Administration Tools -> Interactive SQL.

Connect to the demo database by choosing the “SQL Anywhere 16 Demo” ODBC Data Source Name from the “Browse” menu. Connect without any User ID or Password.

Alternatively, connect to any existing database.

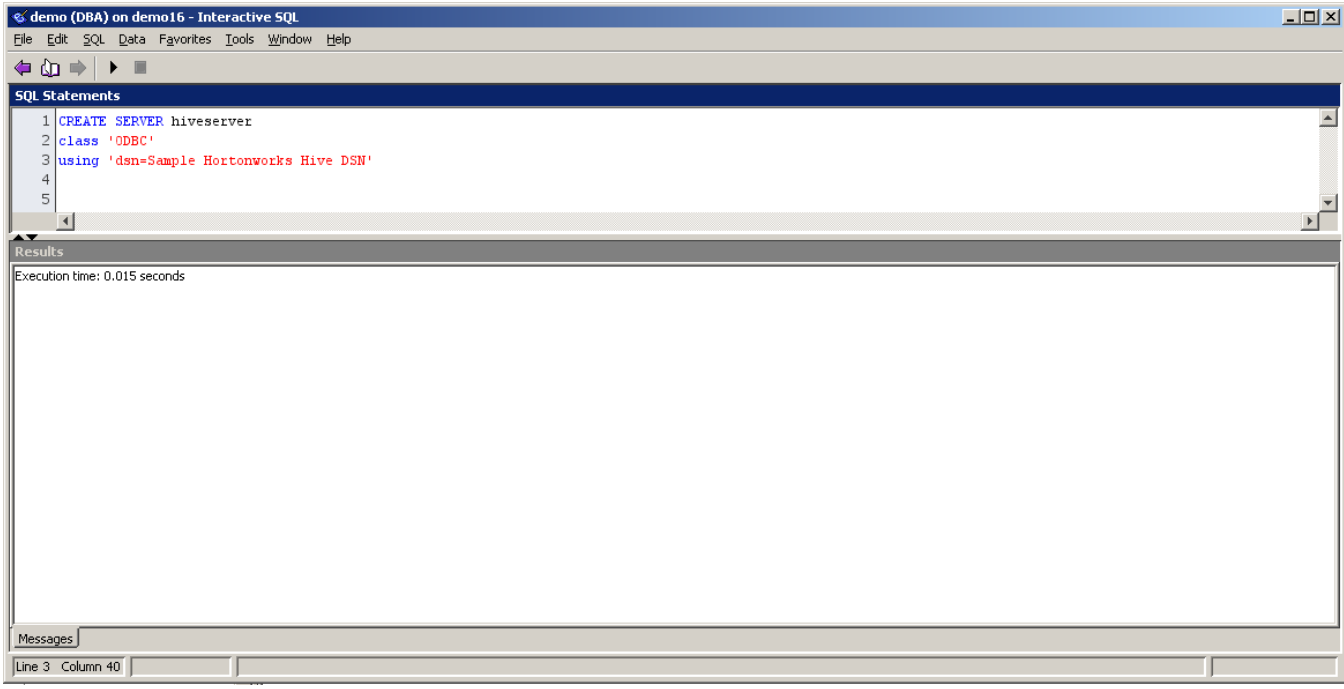
Note: You can connect to the “Sample Hortonworks Hive DSN” directly, if you want to access the Hive server as a server, without using proxy tables. This, however, does not allow you to access both SA data and Hive data simultaneously.

5.1 Create Remote Server

Now, in the SQL Statements window, we will create a remote server.

Run the following command, choosing a server name (hiveserver) and specifying the DSN as created above.

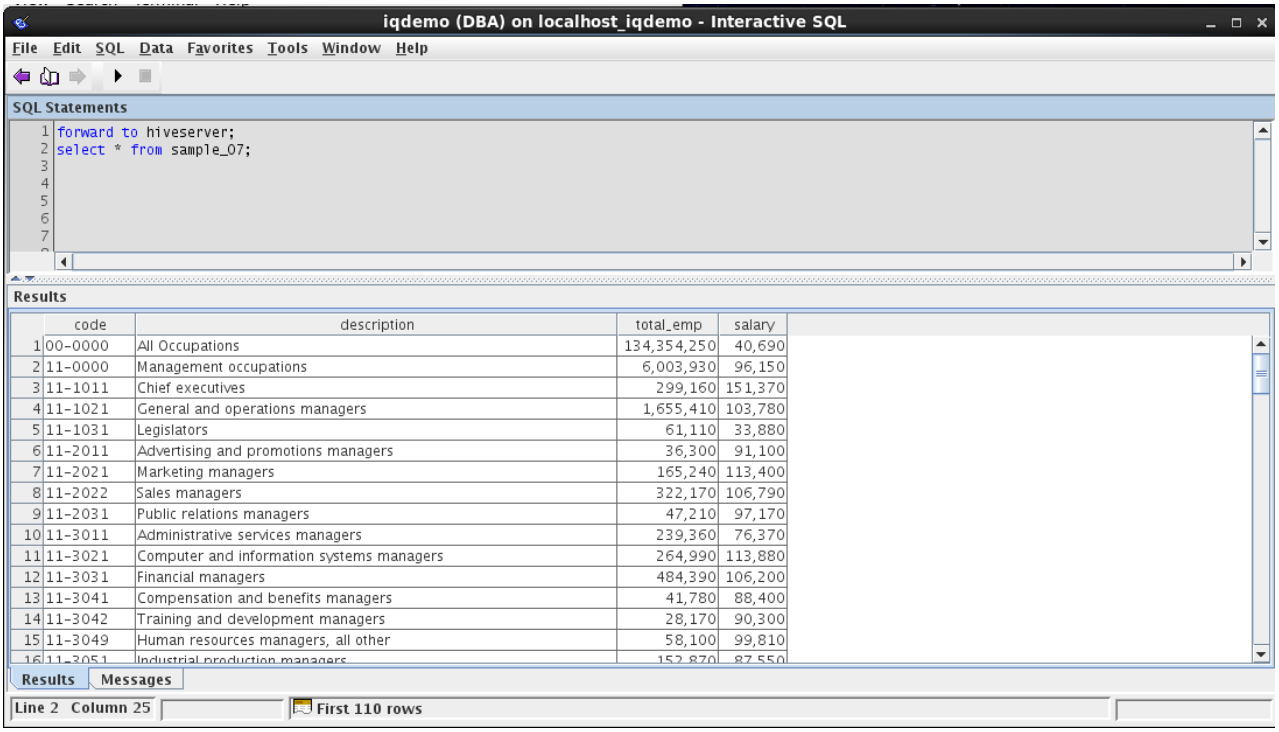
```
CREATE server hiveserver
class 'ODBC'
using 'dsn=Sample Hortonworks Hive DSN'
```



5.2 Access Hadoop from Server

Data can be retrieved from the newly connected Hive server using normal SQL statements, provided that those statements also exist in HiveQL. Select statements can be run with the following commands:

```
Forward to hiveserver;
Select * from sample_07;
```



Note: Existing data, in the Hadoop tables, can also be accessed by accessing the IP address specified in the Hortonworks VM in a web browser.

JOIN also works in the same way.

5.3 Access Hadoop from Proxy Table

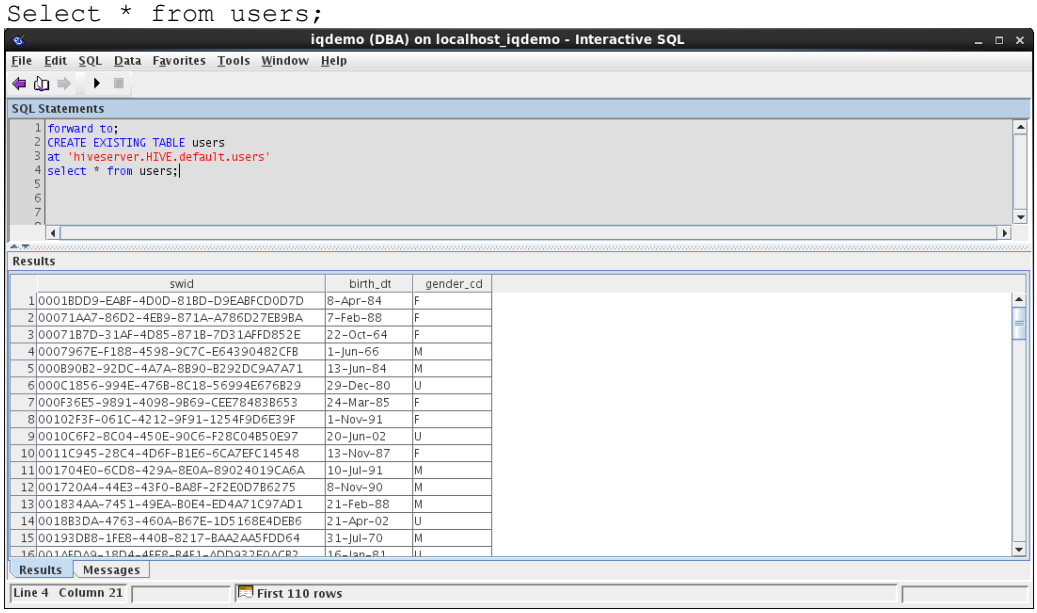
A proxy table can be created to access the data as well. Execute the following command:

```
Forward to;
Create existing table users
At 'hiveserver.HIVE.default.users'
```

Note: The connection string 'hiveserver.HIVE.default.users' is generated by 'servername.databasename.username.tablename'. The users table must already exist in Hive for this to work.

Now the users table is a proxy to the users table on the Hive server. You can access data from that table from IQ:

```
Select * from users;
```



5.4 Create Table

To create a table from IQ, run the following command:

```
Forward to;
Create table hive_t (bee int, nest int)
At 'hiveserver.HIVE.default.hive_t'
```

Then you can access the table both from IQ and from the Hive server directly, with a proxy table already created.

www.sap.com

© 2014 SAP AG. All rights reserved.

SAP, R/3, SAP NetWeaver, Duet, PartnerEdge, ByDesign, SAP BusinessObjects Explorer, StreamWork, SAP HANA, and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP AG in Germany and other countries.

Business Objects and the Business Objects logo, BusinessObjects, Crystal Reports, Crystal Decisions, Web Intelligence, Xcelsius, and other Business Objects products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of Business Objects Software Ltd. Business Objects is an SAP company.

Sybase and Adaptive Server, iAnywhere, Sybase 365, SQL Anywhere, and other Sybase products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of Sybase Inc. Sybase is an SAP company.

Crossgate, m@gic EDDY, B2B 360°, and B2B 360° Services are registered trademarks of Crossgate AG in Germany and other countries. Crossgate is an SAP company.

All other product and service names mentioned are the trademarks of their respective companies. Data contained in this document serves informational purposes only. National product specifications may vary.

These materials are subject to change without notice. These materials are provided by SAP AG and its affiliated companies ("SAP Group") for informational purposes only, without representation or warranty of any kind, and SAP Group shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP Group products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.