Connecting power bi

to hive

**Date**: Dec-14-2018

**Course:** Big Data Technologies

**Students**: Victor Sebastian Valencia, ID: 986775

Carlos Alberto Hernandez, ID: 986636

**Facilitator**: Mrudula Mukadam

**University:** Maharishi University of Management

# Goal

To explain in detail all steps needed to connect a visualization tool like Microsoft Power BI, using ODBC to Hive to query information in an HBase table over Hadoop Cloudera.

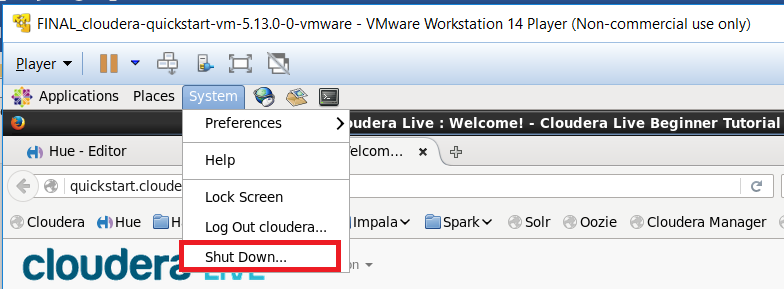
# Challenges

1. Having the Cloudera inside a Virtual Machine (like VMWare or Virtual Box) meanwhile MS Power BI in the guest machine: Network connectivity.
2. Creating Hive table to wrap HBase table, straightforward
3. Finding a driver to support the ODBC connectivity from Windows
4. Configuration and testing

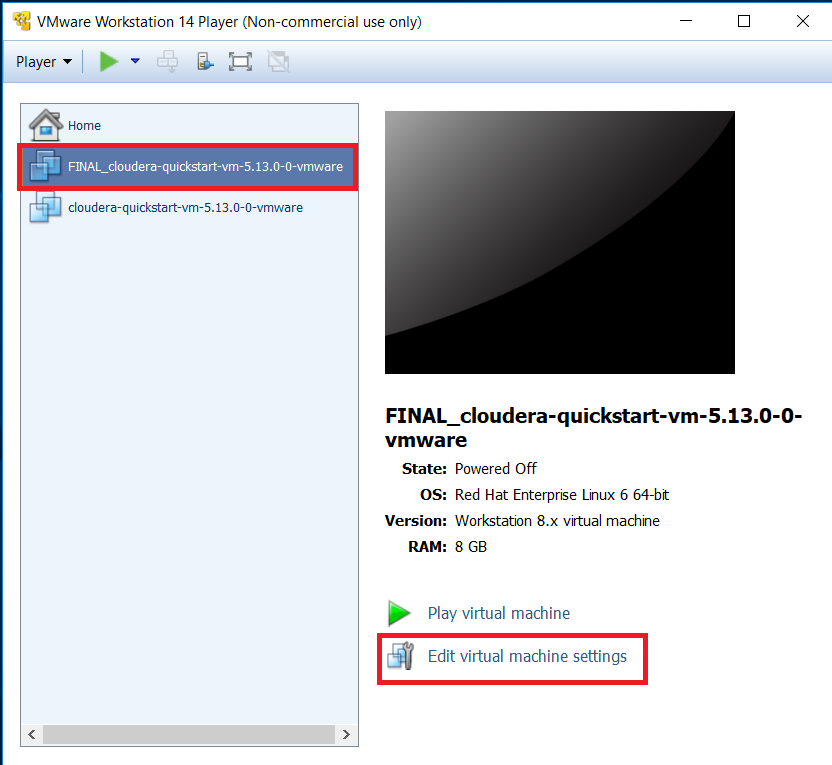
# Process

**Step 1: Network Connectivity:** Since usually, for academic and experimental purposes you use a virtual machine in any technology like Virtual Box or VMWare, and in this case, the Hadoop is inside the virtual machine, while the visualization tool is outside in the guest physical machine; the first challenge is to interconnect them. We are showing here how to do it in VMWare, but we have done too in Virtual Box and can say is quite similar:

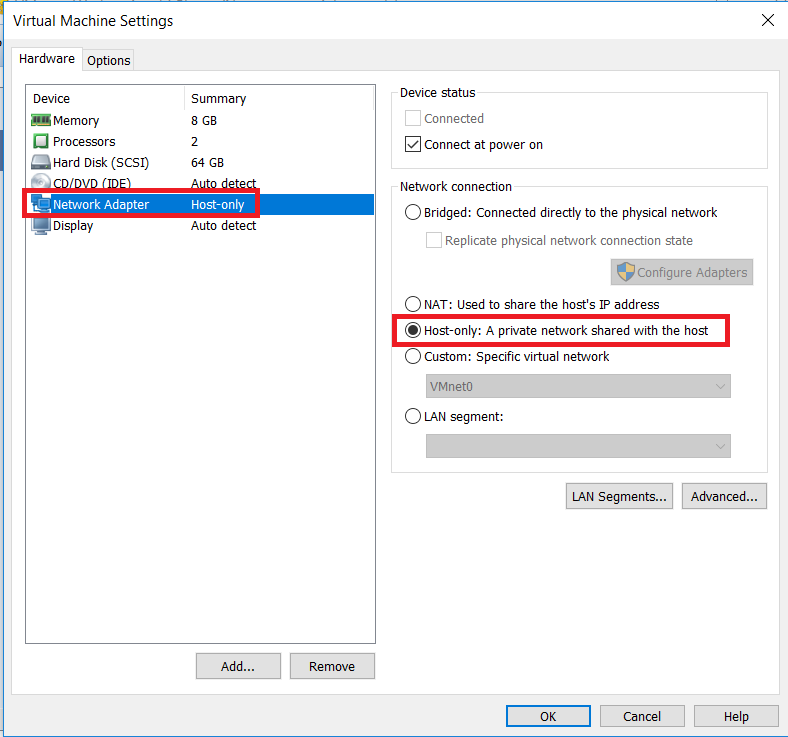
1. Be sure your Virtual Machine is turned off:



1. Open the VMWare software, choose your virtual machine and pick “Edit Virtual Machine Settings”

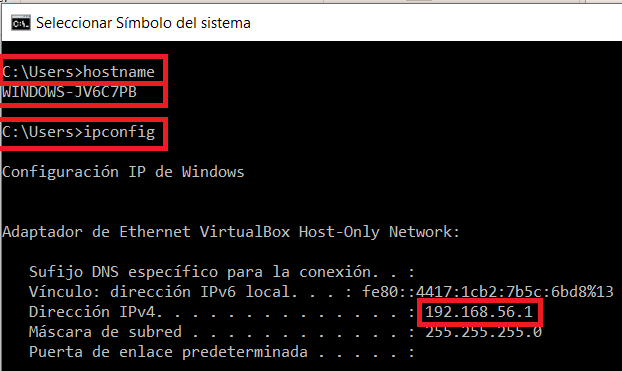


1. Choose “Network Adapter” and “Host-only: A private network shared with the host” radio button option. Note: In virtual box this works with “Bridged connection too”, it is up to to your preferences, but always test it, like we show in next point

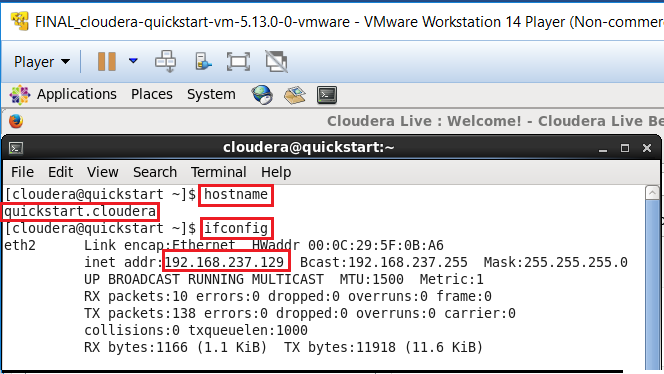


1. Start your virtual machine and waits until all services are up.
2. Do a test of connectivity. First: find out the IP address:

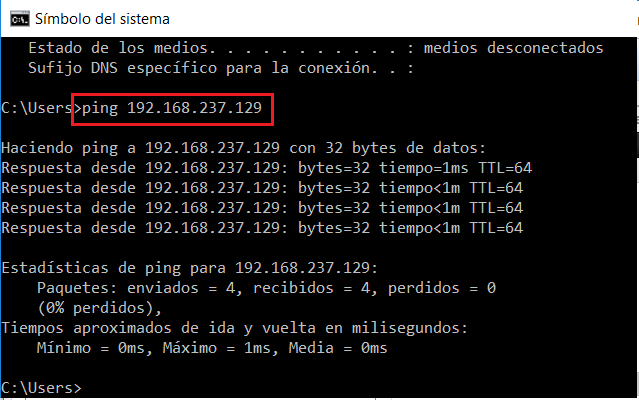
IP of Guest machine (if it is windows):



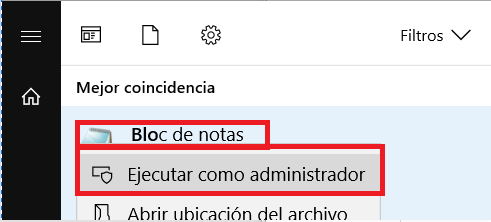
IP and name of Virtual Machine (Linux):



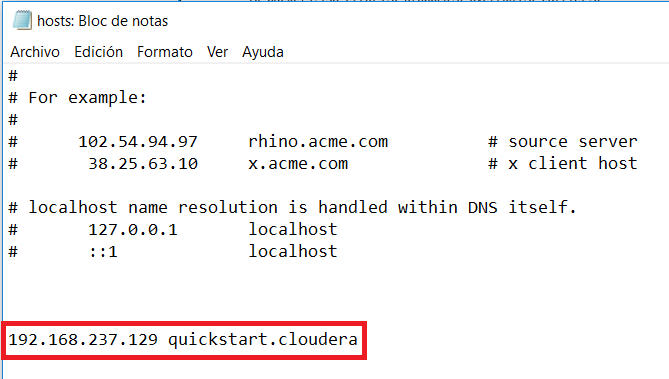
Now it is time to do the test: From the guest Windows machine to the Virtual Machine:



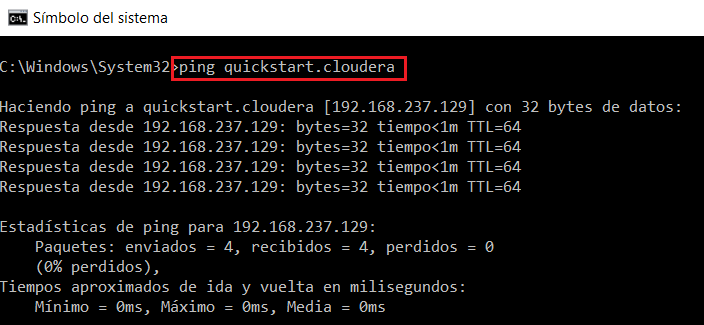
1. In order everything works well is mandatory the ping replies using the name of the Virtual Machine (“quickstart.cloudera” in this case). One quick way to achieve this is to edit in Windows the hosts file, usually located in C:\Windows\System32\drivers\etc. You have to open this file with an account with administrator privileges:



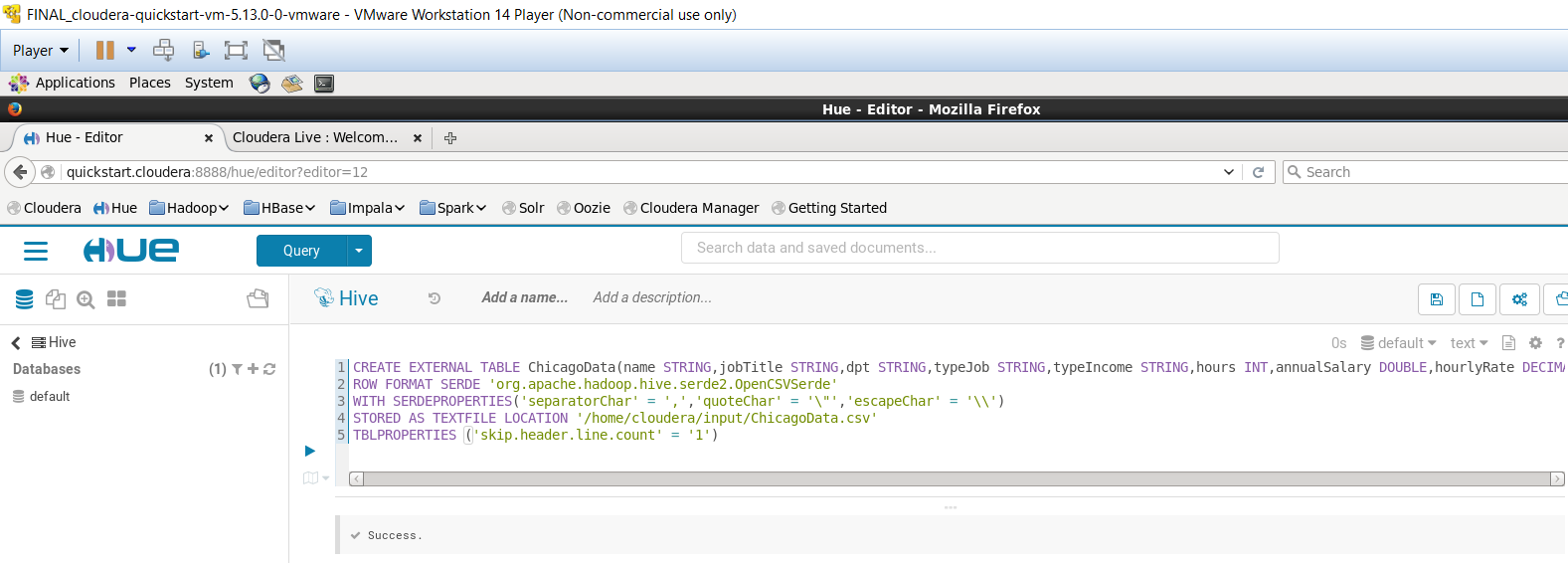
And add your VM IP address and the name as follows:



Save, close and try a ping using the machine’s name:

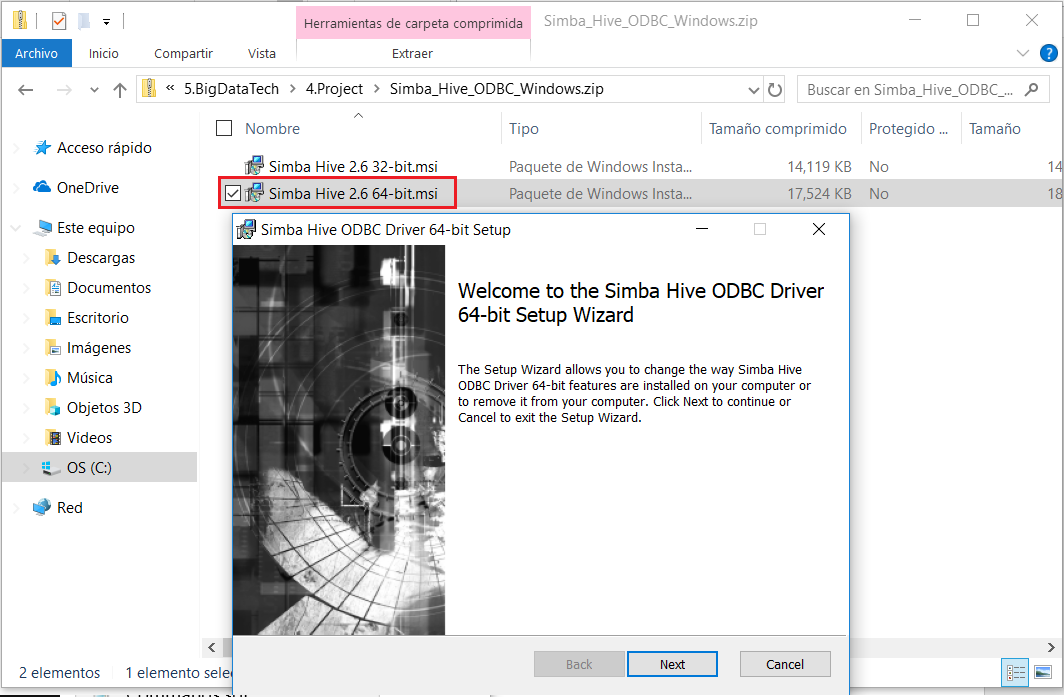


**Step 2: Create external table in Hive:** In your virtual machine open Hive console and execute the create table to map your Hbase table or your HDFS file according to your needs:



**Step 3: Driver:** In order to communicate your visualizer (in this case MS Power BI, but could be any supporting ODBC) is needed to acquire a driver for the ODBC. In this example we are using a trial from Simba (<https://www.simba.com/drivers/hive-odbc-jdbc/>):

1. Register and download the driver according to your platform, for example for Windows:

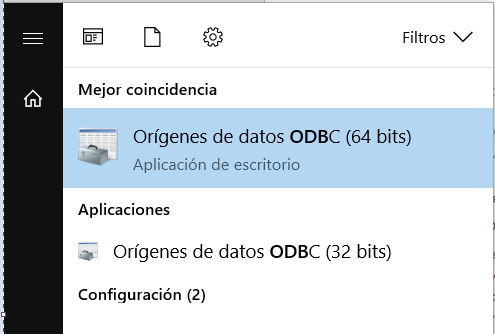


Take care about select the correct platform (32 or 64 bits).

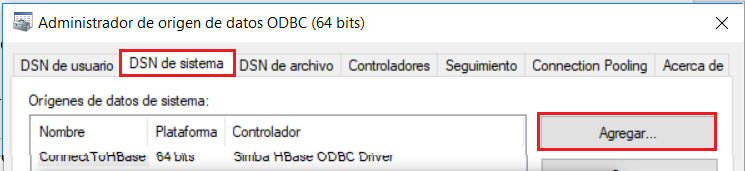
End the installation following the wizard instructions.

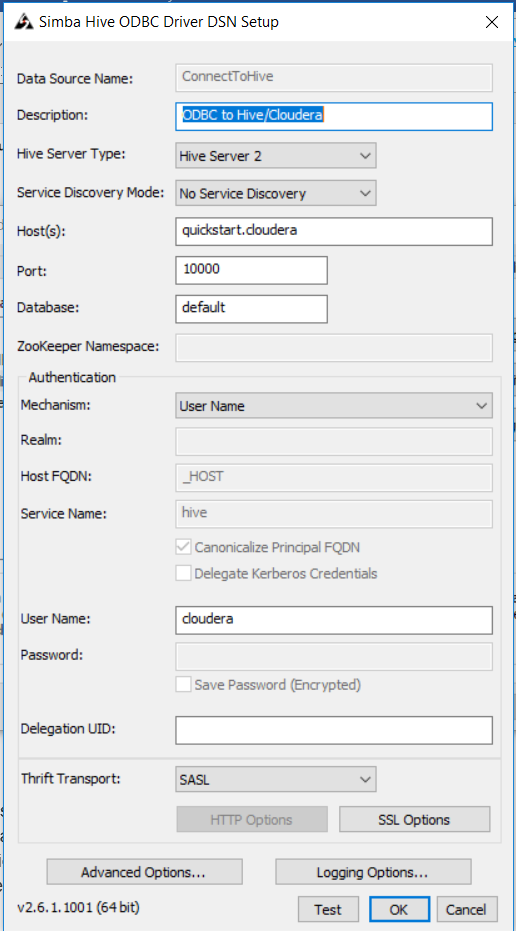
**Note:** Simba send to you a license file to the email registered. You have to copy that file to your installation directory, for example: C:\WINDOWS\system32\SimbaHBaseODBCDriver.lic

1. Configure the new ODBC using the new driver. Take care opening the correct ODBC (remember exists ODBC 32 bits and ODBC 64 Bits too):

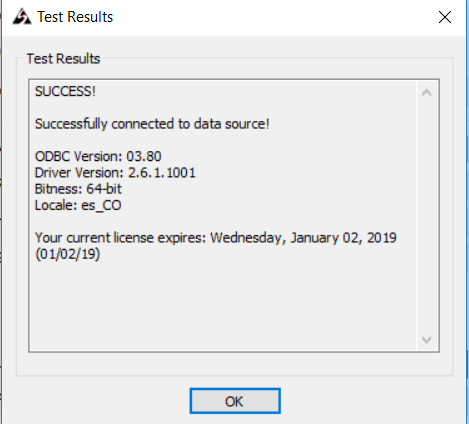


1. Add the new ODBC: In “System DSN” choose “Add…” and configure it as follows:





1. Click on Test button:

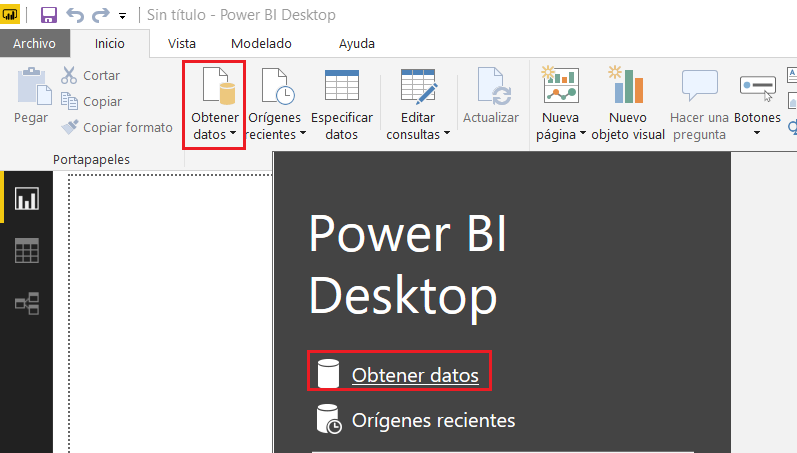


1. Click OK, OK, and close ODBC Administrator.

**Note:** Simba has a driver to connect directly to HBase too, if it is your necessity.

**Step 4: Test in your Visualization Tool:** In your visualization tool you can try to import data from your Hive table, in this case we will show how to do it in Microsoft Power BI Desktop:

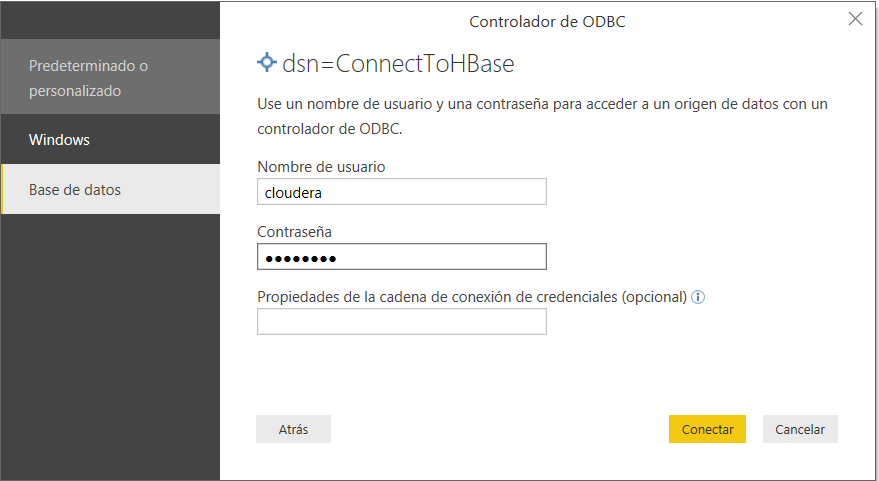
1. Open Power BI Desktop and select “Obtain external data”



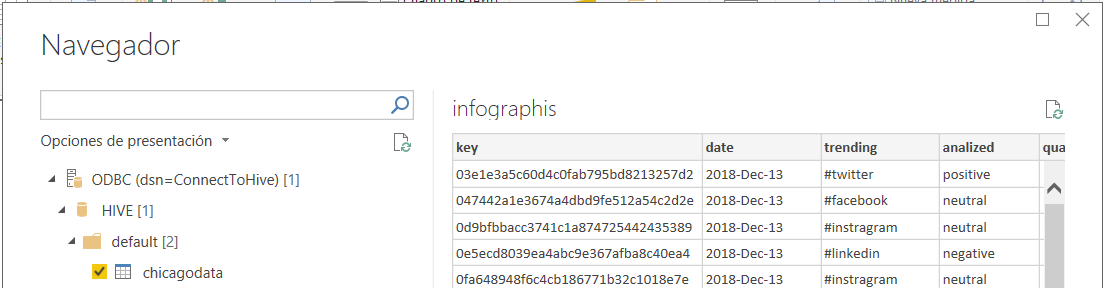
1. Choose “Other” and “ODBC”. Click in “Connect” button.
2. Choose your ODBC and click on Accept:



1. Provide user and password and click on “Connect”:



1. Now you can connect to your Hive tables and start building your dashboards:



# References

More about connecting Power BI to Hive: <https://www.simba.com/webinar/connect-power-bi-hive/>

To download MS Power BI: <https://powerbi.microsoft.com/en-us/downloads/>

Other drivers <https://www.simba.com/drivers/>