Extras/Connecting DV Desktop3.0 a...

xtra tutorial: Working with Oracle Data Visualization Desktop and Hive

FINISHED

This tutorial was built for BDCS-CE version 17.3.5-20 and Data Visualization Desktop 3.0 as part of the New Data Lake User Journey: here (https://github.com/oracle/learning-library/tree/master/workshops/journey2-new-data-lake). Questions and feedback about the tutorial: david.bayard@oracle.com (mailto:david.bayard@oracle.com)

Oracle Data Visualization Desktop (here (https://docs.oracle.com/middleware/bidv1221/desktop/index.html)) is a lightweight, single-file download tool to easily analyze data. Data Visualization Desktop can connect to a variety of data sources. In this tutorial, we will show you how you can configure the HiveServer2 process in BDCS-CE so that DVD can connect.

```
NOTE: As of DVD4, you can now use the native DVD connection type "Oracle Big Data Cloud" to connect to BDCS-CE.  You do not need to use these instructions.
These instructions are left in case you want to configure BDCS-CE's hive to work with the older "Hive" connection type.
```

Took 0 sec. Last updated by anonymous at November 16 2017, 11:19:04 AM.

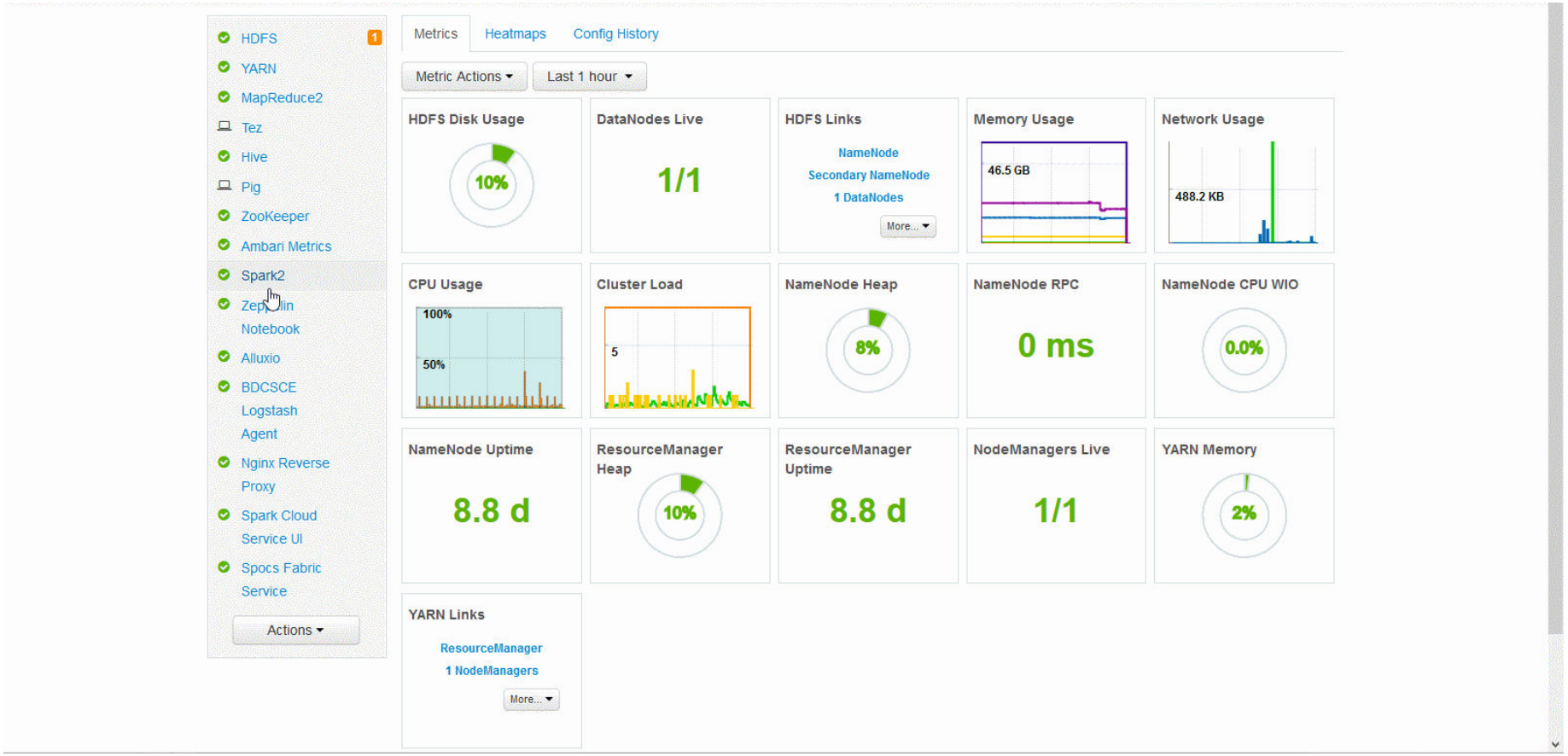
Configuring the HiveServer2 process to use binary transport

READY

In order to connect to Hive from DVD, we need to configure BDCS-CE's Hive Thrift Server to use the binary transport protocol. By default, BDCS-CE's hive thrift server is configured to use the http transport protocol. These changes are done using the Ambari web console.

Here are the steps:

1.Follow the note "xtra Connecting to Ambari" to login to Ambari.
2.Once connected to Ambari, click on "Hive" on the left-hand list of services
3.Then click on the "Configs" tab
4.In the search box, type "server2"
5.Click on the Advanced sub-tab
6.In the General section, change the "HiveServer2 Port" to 10002 (it defaults to 10000, but that port will already be in use for something else)
7.Also in the General section, change the "hive.server2.transport.mode" to binary.
8.In the Custom hive-site section, delete the "hive.server2.thrift.bind.host" property (by clicking on the red minus symbol)
9.Click save
10.In the notes field, enter "transport mode"
11.Click save again
12.If you see a "Configurations" pop-up, click "Proceed Anyway"
13.Click OK to acknowledge that changes were made successfully
14.Then click Restart, then Restart All Affected
15.Then click Confirm Restart All



Connecting to the HiveServer2 port

READY

Now, you need to decide how you want to connect to the Hive Thrift Server port, which you set to port 10002. You can either choose to use a SSH tunnel (which is very secure) or choose to open port 10002 to the outside world (which can be less secure).

- If you want to use a SSH tunnel, refer to the note "xtra Connecting to Ambari" which has an example of setting up a SSH tunnel (but you would use port 10002 instead of Ambari's 8080).
- If you want to open up port 10002, you will need to create a new access rule for port 10002. Refer to the note "xtra Connecting via SSH" or "OEHCS Tutorial 1" for examples of working with network access rules.

Note about the jdbc interpreter

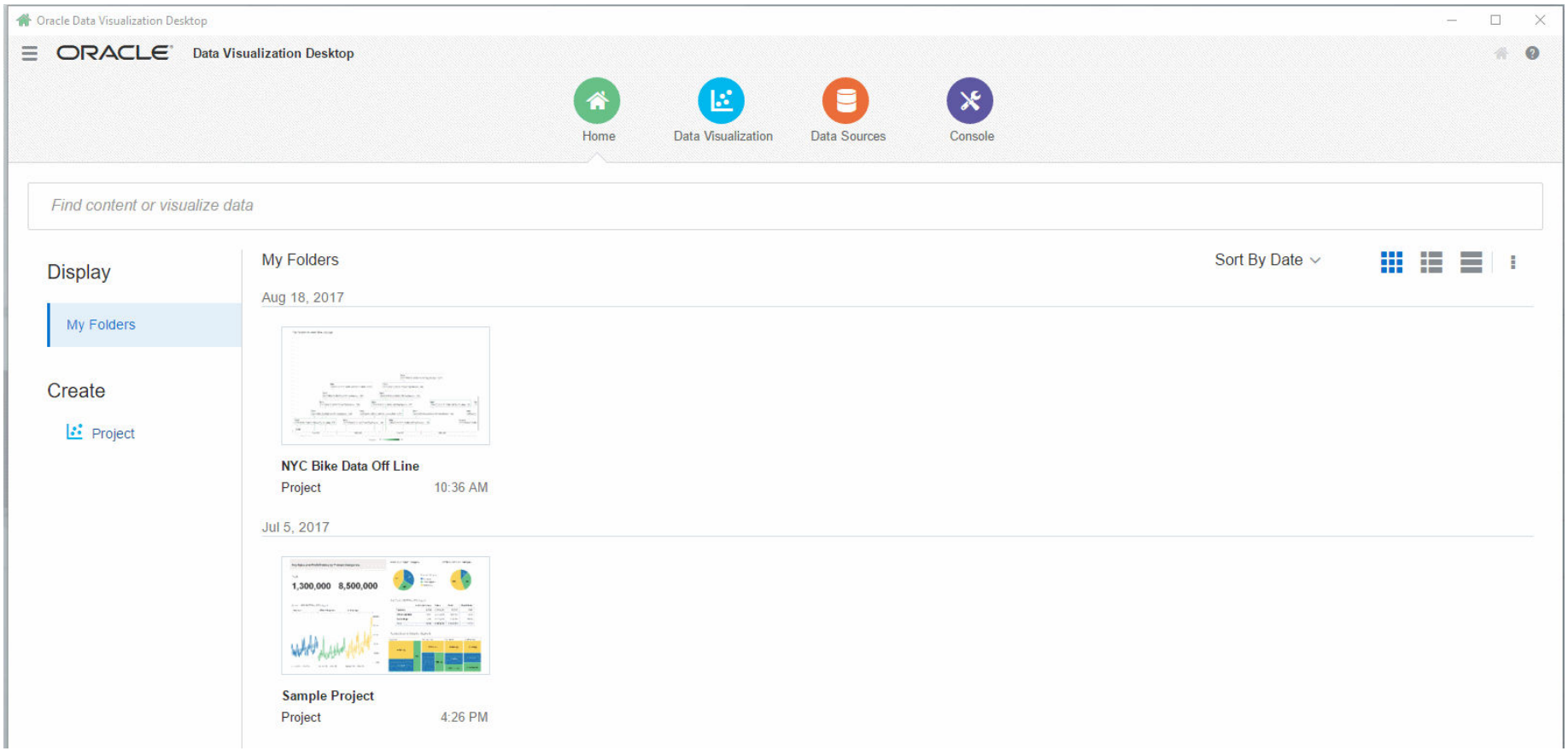
READY

When you make the above changes to the HiveServer2 port and transport mode, you will break the configuration of the jdbc interpreter in Zeppelin. If you want to fix it, navigate to Settings..Notebook..JDBC Interpreter and adjust the hive.url parameter to jdbc:hive2://localhost:10002/

Define a connection in DV Desktop for the Hive connection

READY

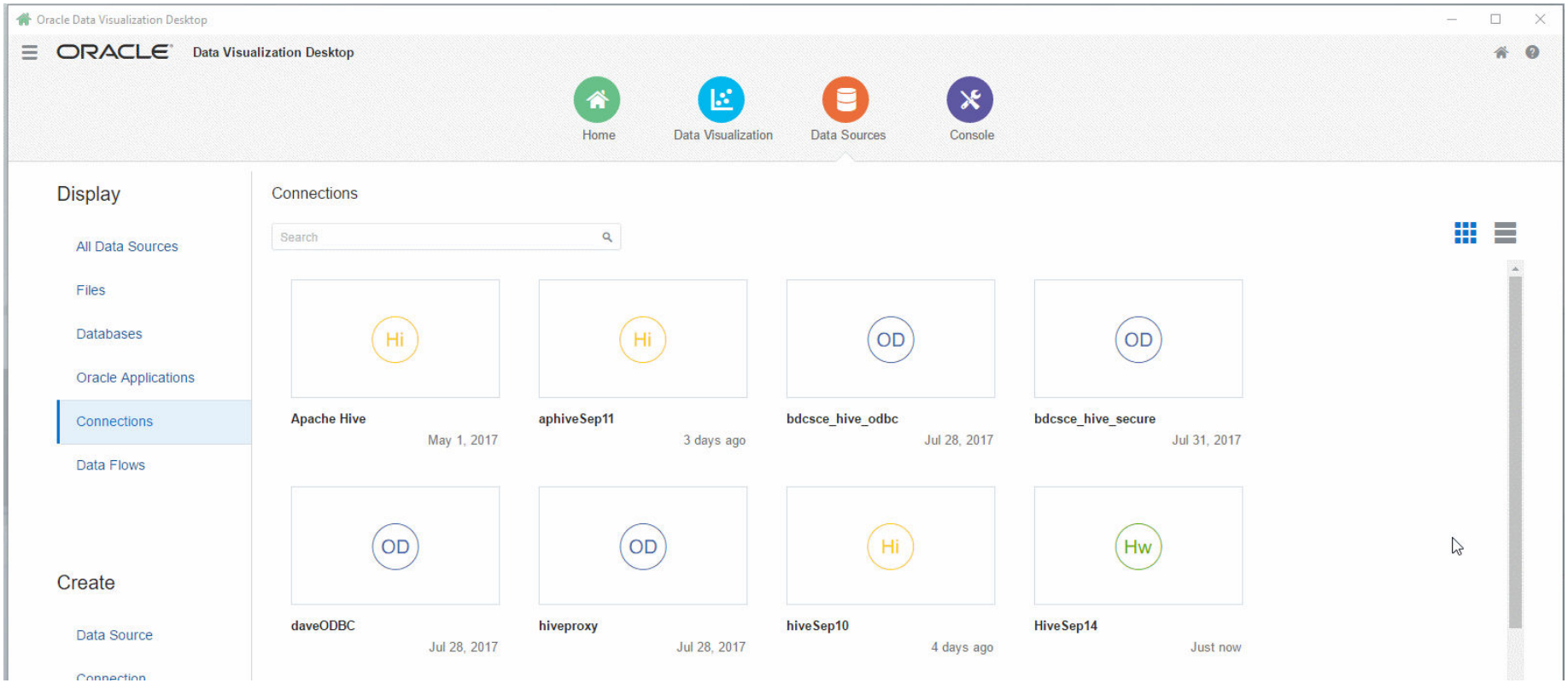
- Open up DV Desktop
- Click on Data Sources
- Click on Connection (Under Create)
- Click on Hortonworks Hive
- Enter the Connection Name
- Enter the Hostname. If you chose to use SSH tunneling, enter "127.0.0.1" for the hostname. If you chose to open port 10002 to the outside, enter the hostname or ip address of your BDCS-CE server.
- Enter to port, which is 10002
- Set the Username to hive
- Set the Password to x

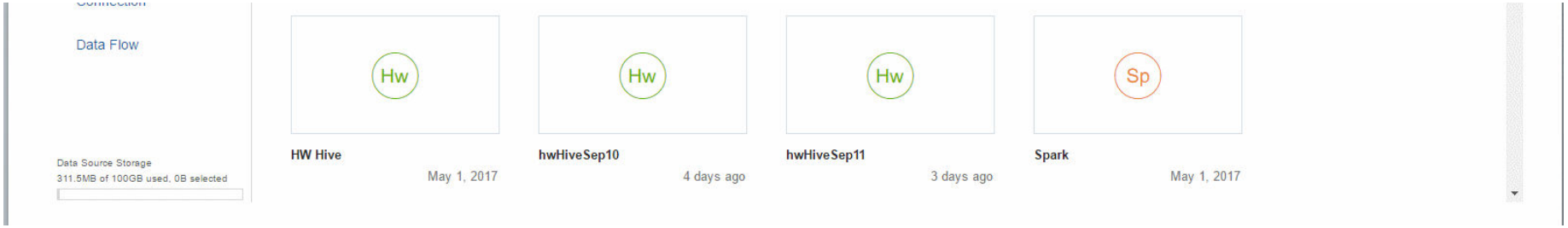


Create a DV Desktop Data Source for your connection

READY

- Double click on your Connection
- Double click on the default database
- Double click on the bike_trips_small table
- Click on Add All
- Now click on the rightmost icon in the dataflow pipeline (it will be the icon after the filter icon). Then, click on the Refresh property. Change this to be "Live - Always use the database". **NOTE: This step is not yet captured in the screenshot below**
- Click on Add
- Click Create Project





Tip - Viewing the Hive queries issued by DVD

READY

Run the following shell paragraph to see the recent SQL commands sent to the Hive Server.

Shell command to see the recent SQL sent to Hive Server

READY

```
%sh
#grep "Parsing command" /data/var/log/hive/hiveserver2.log | tail -300
egrep $'Parsing|\x0d|limit|group' /data/var/log/hive/hiveserver2.log | tail -300
```

```
%md
```

READY