

Extras/Connecting DV Desktop4.0 a...

Tutorial: Connecting Oracle Data Visualization Desktop 4.0 and BDCS-CE

FINISHED

This tutorial was built for BDCS-CE version 17.4.1 and Data Visualization Desktop 4.0 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 use DVD's "Oracle Big Data Cloud (beta)" Connection type to connect to your BDCS-CE instance.

NOTE: This tutorial shows you how **to** connect using the "Oracle Big Data Cloud (beta)" connection type. Alternatively, you can still also connect via the "Spark" **or** "Hive" connection type. **To** connect using one of these older methods, look **for** the other notebooks **in** the Extras folder containing the instructions **for** Connecting DV Desktop3.0.

Took 0 sec. Last updated by anonymous at November 16 2017, 9:03:47 AM.

About DVD4.0's "Oracle Big Data Cloud (beta)" Connection Type

FINISHED

DVD4.0 introduces a new connection type called "Oracle Big Data Cloud (beta)". This connection type is built for BDCS-CE.

Specifically, the connection type is built to use BDCS-CE's default configuration of running the Spark Thrift Server in http transport mode behind the standard BDCS-CE firewall/proxy running on port 1080. This provides security and load-balancing features.

Note: The DVD connection type **for** Oracle Big Data Cloud does **not** currently allow you **to** specify a http proxy. So, please make sure there is no firewall between where you run DVD **and** port 1080 on the BDCS-CE instance.

To use this connection type, we require that you setup the DVD client with a "trust store" that contains the BDCS-CE server's self-signed certificate.

For full details, see <https://docs.oracle.com/en/cloud/paas/big-data-compute-cloud/csspc/accessing-thrift.html> (<https://docs.oracle.com/en/cloud/paas/big-data-compute-cloud/csspc/accessing-thrift.html>)

This note will walk you through the exact procedure...

Took 0 sec. Last updated by anonymous at November 16 2017, 9:31:17 AM.

Create the Trust Store file

FINISHED

The following shell script will create the trust store file that is needed on the DVD client. Simply run the following paragraph to continue.

Took 0 sec. Last updated by anonymous at November 16 2017, 9:05:10 AM. (outdated)

Script to create Trust Store file for your BDSC-CE instance

FINISHED

```
%sh
#script to create a truststore file with your BDSC-CE self-signed certificate

rm -r /tmp/dvd
mkdir /tmp/dvd
cd /tmp/dvd

echo "Gathering the self-signed certificate"
# it is also stored at /u01/bdcsce/etc/nginx/nginx.crt
echo | \
  openssl s_client -connect 127.0.0.1:1080 2>/dev/null | \
  openssl x509 >nginx.crt

echo "Creating bdcsce.jks trust store file"
/usr/java/default/bin/keytool -import -trustcacerts \
  -keystore /tmp/dvd/bdcsce.jks \
  -storepass welcome -noprompt \
  -alias bdcsce \
  -file nginx.crt;

echo "Validating bdcsce.jks trust store contents"
/usr/java/default/bin/keytool \
  -keystore /tmp/dvd/bdcsce.jks \
  -storepass welcome \
  -list -v

#copy the truststore to somewhere we can download it
echo "Copying bdcsce.jks trust store to browser accessible location"
sudo cp /tmp/dvd/bdcsce.jks /data/tmp/spocui/spoccs/webapp/css/images/bdcsce/bdcsce.jks

echo "done"
```

Took 1 sec. Last updated by anonymous at November 15 2017, 3:53:06 PM. (outdated)

Download the Trust Store file to the laptop where DVD 4.0 is installed.

FINISHED

Now that we have created the necessary Trust Store file, you will need to download it to where DVD is installed. To download it, click on this link: [Trust Store File \(/css/images/bdcsce/bdcsce.jks\)](/css/images/bdcsce/bdcsce.jks)

When prompted, save the file as bdcsce.jks

Took 0 sec. Last updated by anonymous at November 16 2017, 9:05:54 AM.

FINISHED

Copy the downloaded Trust Store File (bdcscs.jks) to the proper DVD directory

Now, (on the computer where DVD is installed) copy the downloaded bdcscs.jks file to this location: C:/Users/USERNAME/AppData/Local/DVDDesktop/components/OBIS/bdcscs/bdcscs.jks

Where USERNAME is your Windows username

NOTE: You will need to create the bdcscs subdirectory the first time you do this.

Took 0 sec. Last updated by anonymous at November 16 2017, 9:06:37 AM.

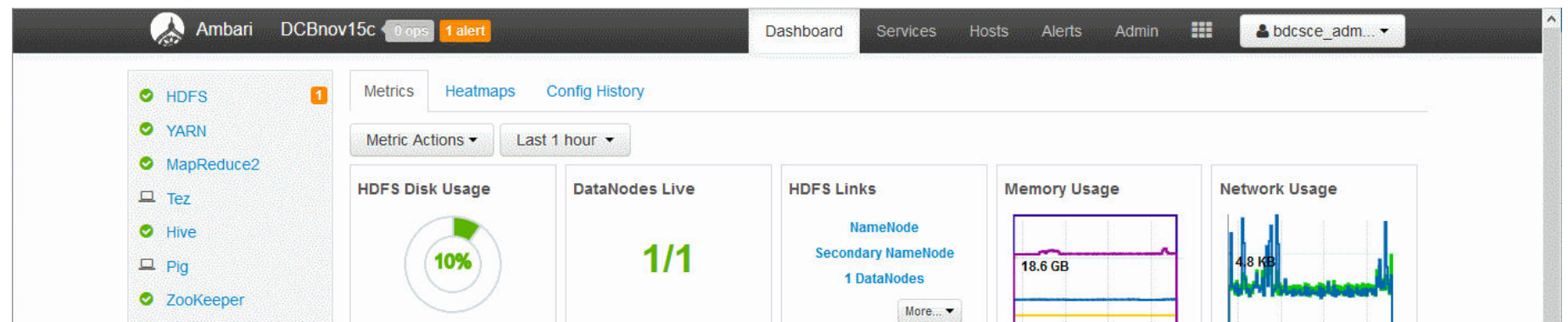
Configuring the memory settings for the Spark Thrift Server process

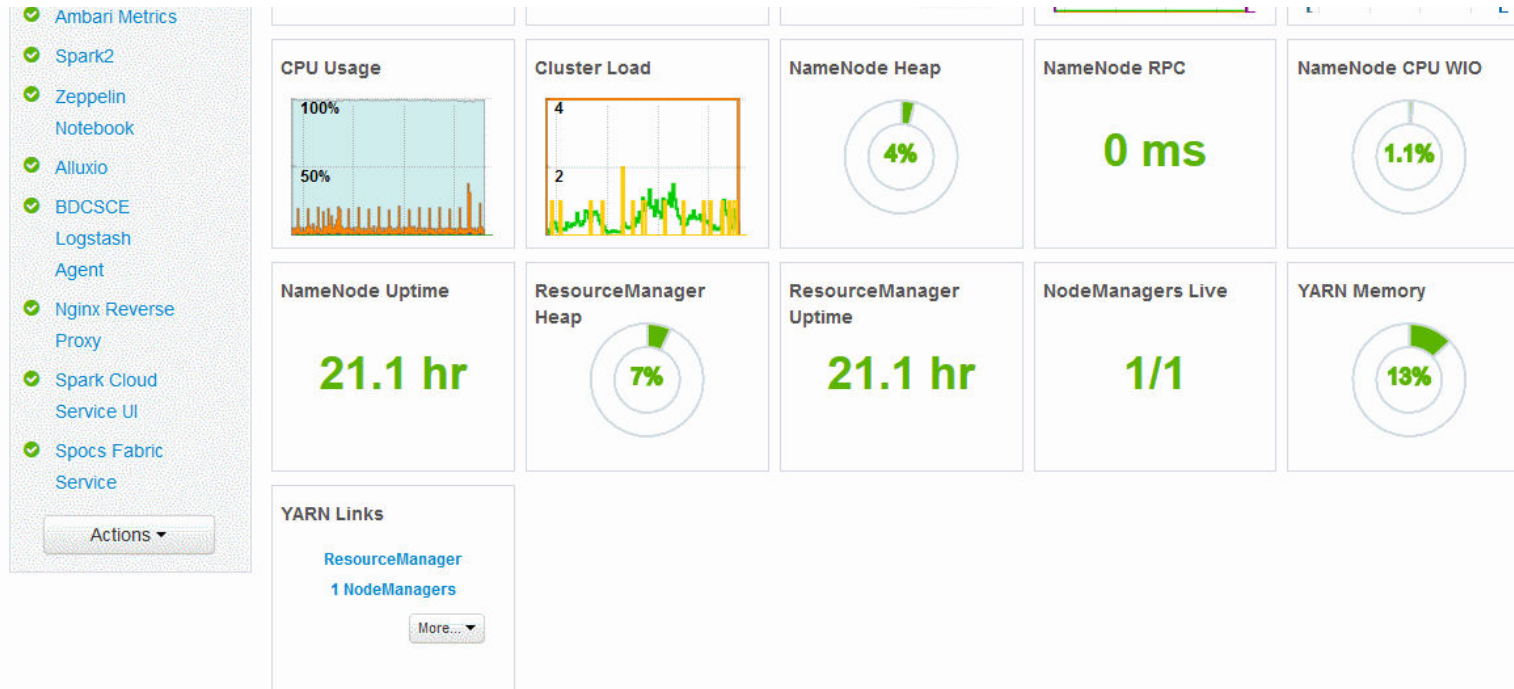
FINISHED

The default memory settings for the Spark Thrift Server are too small for tables like the Citi bike data. In order to query these size of tables effectively, we need to increase a few memory settings on the Spark Thrift Server. 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 "Spark2" on the left-hand list of services
3. Then click on the "Configs" tab
4. Navigate down to the **Custom spark2-thrift-sparkconf** section
5. In the Custom spark2-thrift-sparkconf section, click on the "Add Property..." link and then add the property "spark.sql.shuffle.partitions=4" and click Add.
6. Expand the **Advanced spark2-env** section and change spark_daemon_memory to 4096 MB.
7. Also in the Advanced spark2-env section in the "content" field, edit and uncomment the line about SPARK_EXECUTOR_MEMORY. When finished, it should read: SPARK_EXECUTOR_MEMORY="2G"
8. Click Save at the top of the screen.
9. In the notes field, enter "memory"
10. Click save again
11. If you see a "Configurations" pop-up, click "Proceed Anyway"
12. Click OK to acknowledge that changes were made successfully
13. Then click Restart, then Restart All Affected
14. Then click Confirm Restart All

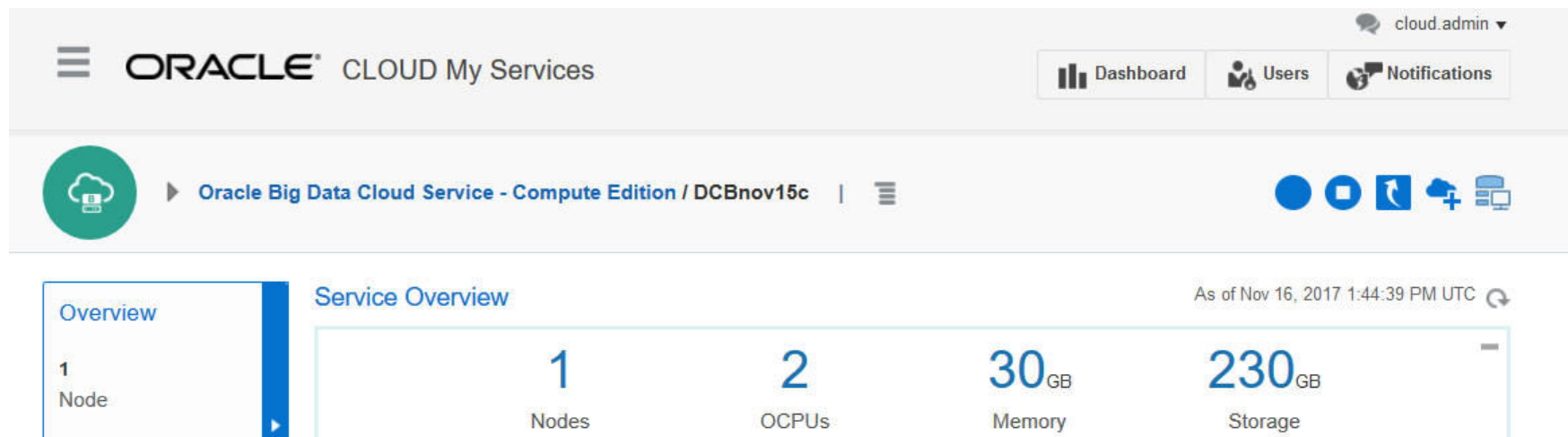




Look up the Public IP address for your BDCS-CE instance

FINISHED

If you have not written down the Public IP address of your BDCS-CE instance, you will need to do so now. You can find this information on Oracle Cloud My Services detail page for your BDCS-CE instance.



Administration

0 Patches available

Status: Ready

Administrative User: bdcscs_admin

Cloud Storage Container: https://gse00002281.storage.orac...

Deployment Profile: Full

Show more...

Version: 17.4.1-20

Ambari Server Host: 140.86.34.20

Compute Shape: oc2m

Spark Thrift Server: jdbc:hive2://140.86.34.20:1080/d...

Resources

Host Name: dcbnov15c-bdcscs-1

Public IP: 140.86.34.20

Instance: Runs MASTER-1

OCPUs: 2

Memory: 30 GB

Storage: 230 GB

Took 0 sec. Last updated by anonymous at November 16 2017, 9:09:34 AM. (outdated)

Define a connection in DV Desktop for the Oracle Big Data Cloud

FINISHED

- Open up DV Desktop 4.0
- Click on the menu in the upper left and choose Data
- Click on the Create button and choose Connection
- Click on Oracle Big Data Cloud (Beta)
- Enter the Connection Name
- Enter the Host Name. Use the public IP address of your BDCS-CE master server. This is the same IP address you use to SSH to the cluster.
- Enter the Port. Use 1080
- Enter the Username. Use bdcscs_admin
- Enter the Password. Use the password you defined for the bdcscs_admin user when you created the cluster
- Optionally enter the Database Name. Typically, you would leave this empty.

Oracle Data Visualization Desktop

ORACLE Data Visualization Desktop

Home

Create

Get Started with Oracle Analytics Cloud

Watch Overview

Visualize Data

Prepare Data

Learn More

Explore your data and uncover important insights using interactive and intuitive visualizations

Get your data ready for analysis using visual data flows that transform, enrich and blend different sources

Visit our Academy and video library to learn how you can do more with Oracle Analytics Cloud

What are you interested in?

What's New

Updated Nov 3, 2017

Heart Disease likelihood Training File

0.95 1.42

CART Training - Hea... Project

Updated Nov 3, 2017

Heart Disease Likelihood Prediction File

Linear Regression Ap... Project

Updated Nov 3, 2017

Heart Disease likelihood Predict File

Linear Regression Tr... Project

Sample Project Project

NYC Bike Data Off LI... Project

Bikes and Weather Project

Heart Disease Likelihood Prediction Project

Took 0 sec. Last updated by anonymous at November 16 2017, 10:59:49 AM.

Create a DV Desktop Data Set using your connection

FINISHED

- Click on Create and choose "Data Set"
- Select the connection you just created
- Navigate through the database, tables, and columns to choose the elements you want to add.
 - ++ For instance, click on the Default database then click on the "bike_trips" table. And click on "Add All" to add all the columns.
- Once you have selected your table and columns, 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".
- Name the new data source and Click the Add button. It may take a minute for DVD to inspect the data before continuing.

Oracle Data Visualization Desktop

Data Visualization Desktop

Home

Projects

Data

Machine Learning

ORACLE Data Visualization Desktop

Data

Data Sets
Connections
Data Flows
Sequences

Search Data

Sort By
Date

Type	Name	Description	Modified
	BDCnov16		Just now

6 of 7

11/16/2017 11:00 AM

<div>⌘ Console</div> <div>📊 Active Data Set</div>	<div>Hi</div> Apache Hive	Oct 18, 2017
	<div>Sp</div> sparkOct11	Oct 11, 2017
	<div>Hi</div> Hivesep29	Sep 29, 2017

Took 0 sec. Last updated by anonymous at November 16 2017, 11:00:07 AM.

Tip - Tracking Spark queries

READY

Run the following shell paragraph to peak at queries sent to the Spark thrift server.

Shell command to peak at queries sent to Spark Thrift Server

FINISHED

```
%sh
egrep '$Running|\x0d|limit' /data/var/log/spark2-thrift/spark-hive-org.apache.spark.sql.hive.thriftserver.HiveThriftServer2-1-*--1.out | tail -400
```

Took 0 sec. Last updated by anonymous at November 16 2017, 10:06:04 AM. (outdated)

READY

READY