**Steps to establish connection to the HIVE environment**

1. **Install HDP 2.2 HIVE ODBC driver (version 2.1.5):**

If you have a driver prior to the version above, you need to uninstall it using "programs and features" in the control panel before installing the one below. Note that the latest version is required to connect securely using Knox.

Download the driver here:

<http://community.fyiblue.com/sites/GPDATIntak/Shared%20Documents/DDO/Onboarding/Hadoop/HortonworksHiveODBC32_2_1_5.zip>

Unzip and install the driver.

1. **Access to the DataLake servers (HIVE)**

**2.1 Check if you can access the Data Lake servers**

Open a command Window (using "cmd" on the start menu).

Type or copy the following to check your connection to the dev server:

telnet dwauslenapp02.app.dev.hcscint.net 8443

Type or copy the following to check your connection to the test server:

telnet twauslaenapp03.app.test.hcscint.net  8443

telnet twauslaenapp04.app.test.hcscint.net  8443

Type or copy the following to check your connection to the prod server:

telnet pwauslaenapp03.app.hcscint.net 8443

telnet pwauslaenapp04.app.hcscint.net 8443

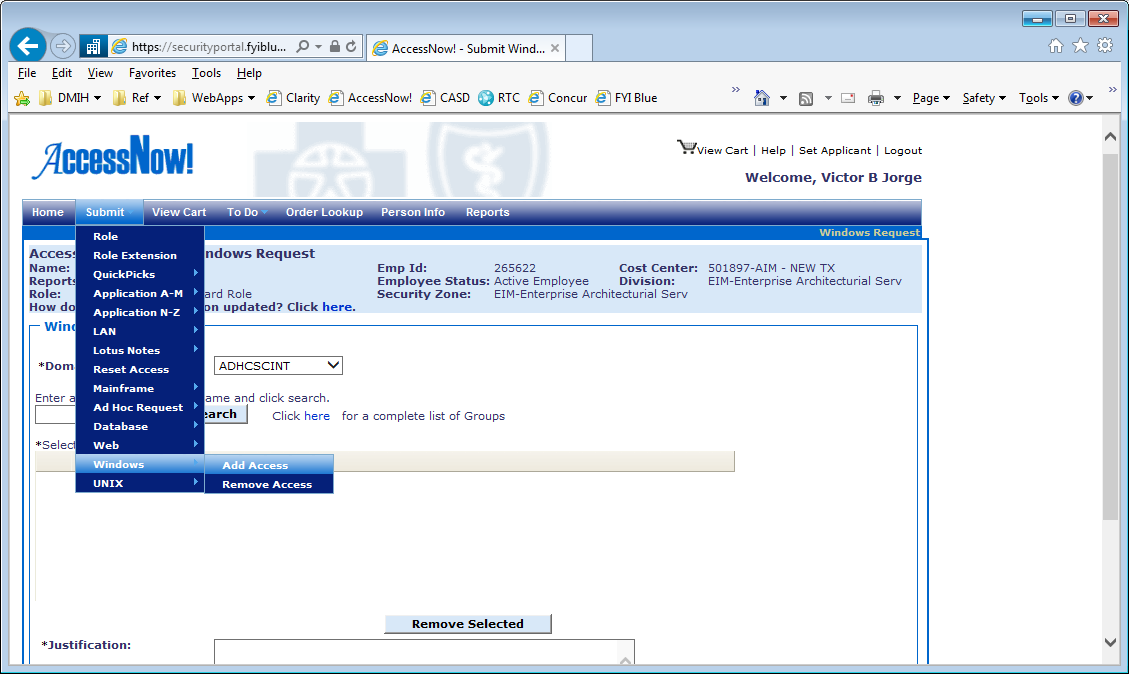
If any of the steps above failed that is, the command hangs or shows a failure, then go to step 2.2 below to request the proper firewall access.

If they return a blank/clear screen then skip step 2.2 (that is, you can access the server already).

**2.2 Firewall Access**

To access Hive thru ODBC, you may need firewall access.

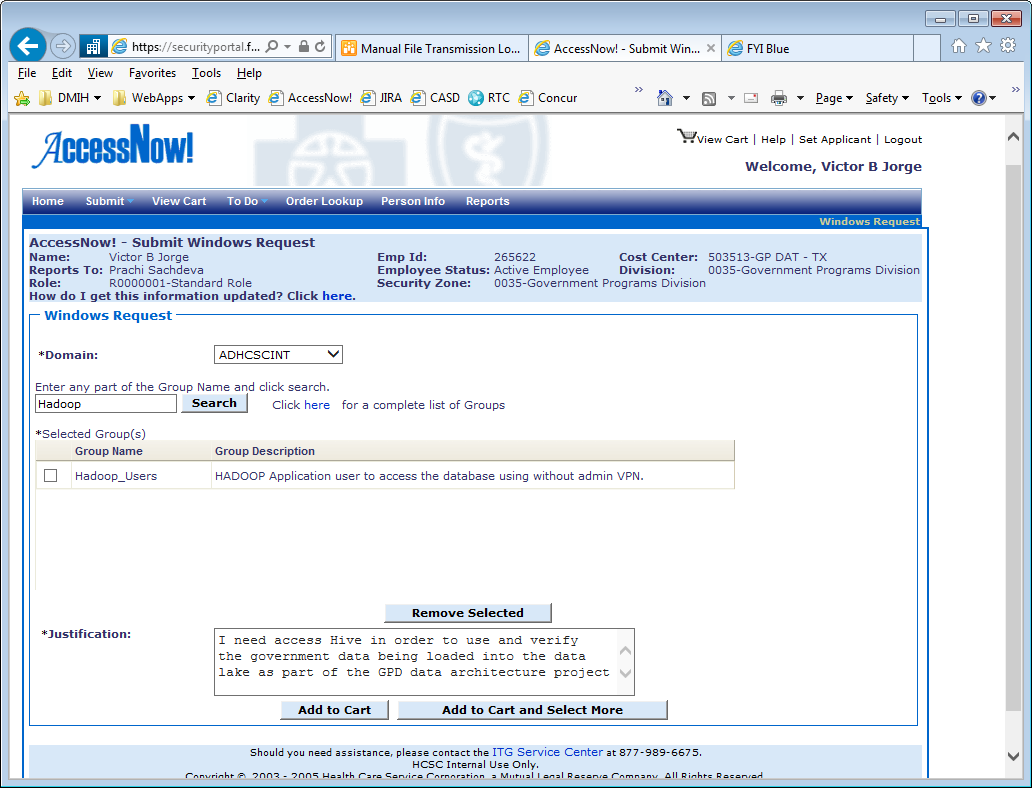
Go to AccessNow, and request the following for each applicable environment (Prod, Test, Dev):



- Select Domain as ADHCSCINT (Prod), ADHCSCTST (Test), ADHCSCDEV (Dev)

- Select the “Hadoop\_Users” group

- Enter the justification, add to cart, and submit the request



These are the roles that need to be requested:



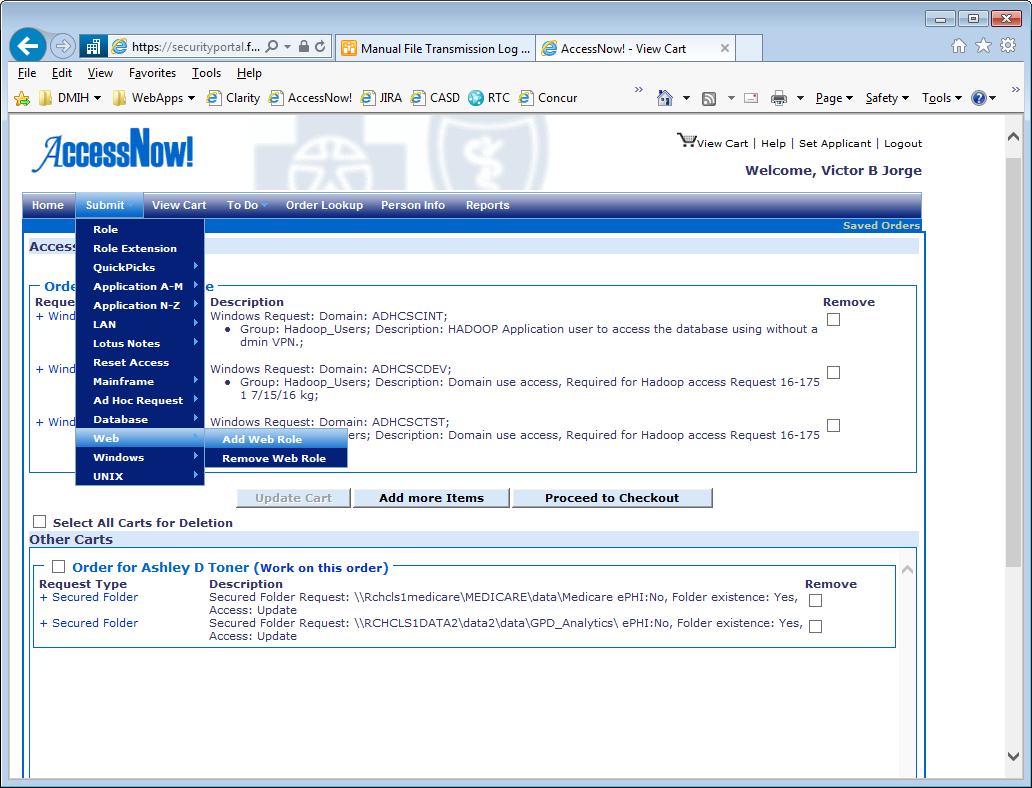
Business Justification:

As part of the GPD data analytics team, I need access to <…> in order to use and verify the government data being loaded into the data lake, to generate reports, research, and provide guidance to other team members.

1. **Entitlements Requests**

After the implementation of Knox, we need to request specific security roles to gain access to HIVE databases. Please follow the procedure below to request web roles indicated below for GPD in all applicable environments (Dev, Test, Prod – note prod is not ready as of Aug/2016):

Select “Web” > “Add Web Role”:



Select “Hadoop” as the Application in the next screen.

First, you need to request authentication roles for each environment. You need to request roles for Knox (used by Hive connection), HUE (if you want to use HUE), and Datameer (to access Datameer).

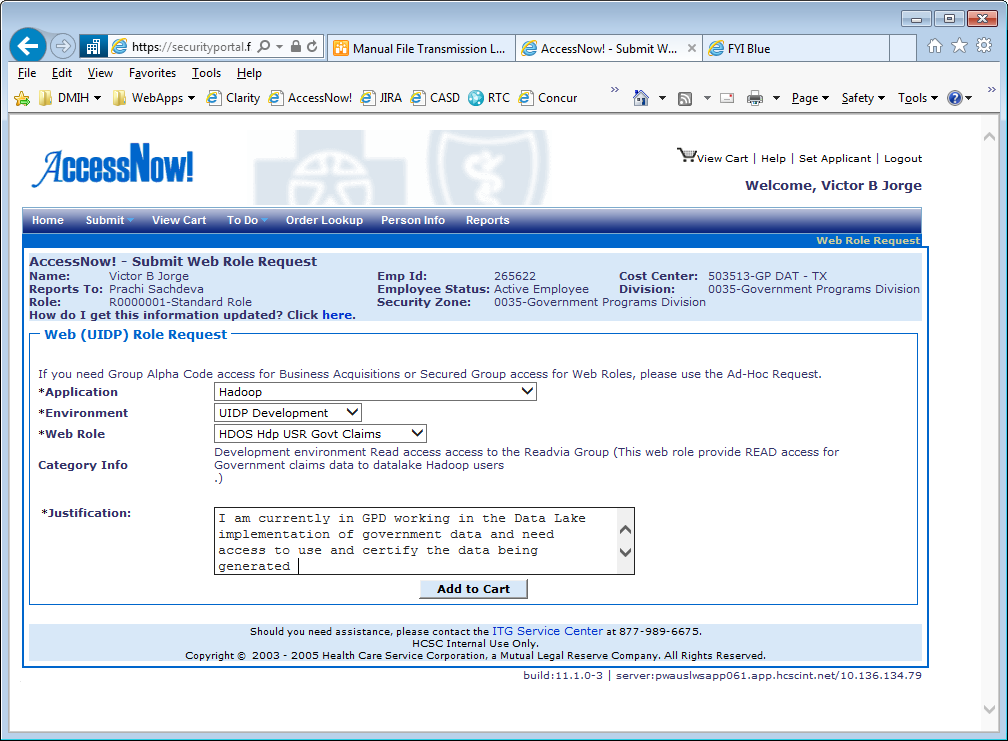
Here are the roles for Dev and Test:



Production:



You also need to request roles for each specific subject area based on the table below. These roles need to be requested for each environment as well. Adding all up, you should request a total of 6 roles per environment (3 authentication roles + 3 subject area roles below).



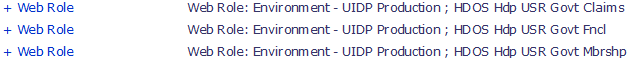
Roles to be requested (pick one side depending on your need):

|  |  |
| --- | --- |
| **IT Stewardship (ETL and Data Consumption)** | **Analyst (Data Consumption only)** |
| HDOS Hdp ETL Govt Fncl  HDOS Hdp ETL Govt Claims  HDOS Hdp ETL Govt Mbrshp | HDOS Hdp USR Govt Fncl  HDOS Hdp USR Govt Claims  HDOS Hdp USR Govt Mbrshp |

Here are the roles for Dev and Test:

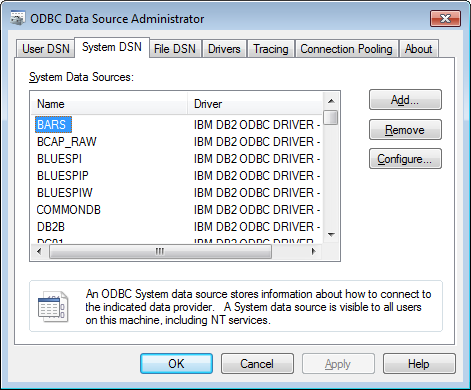


Production:

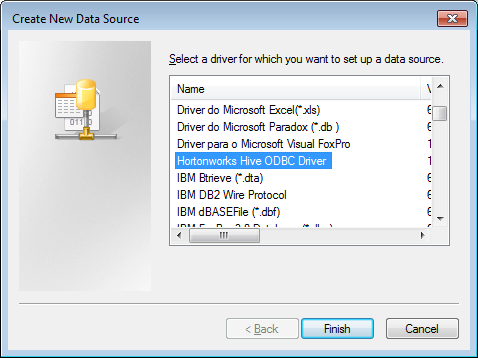


1. **ODBC data source set up**

Open ODBC administrator 32 bit Administrator (if you don’t know where it is search for ‘ODBC 32’ on the start menu)

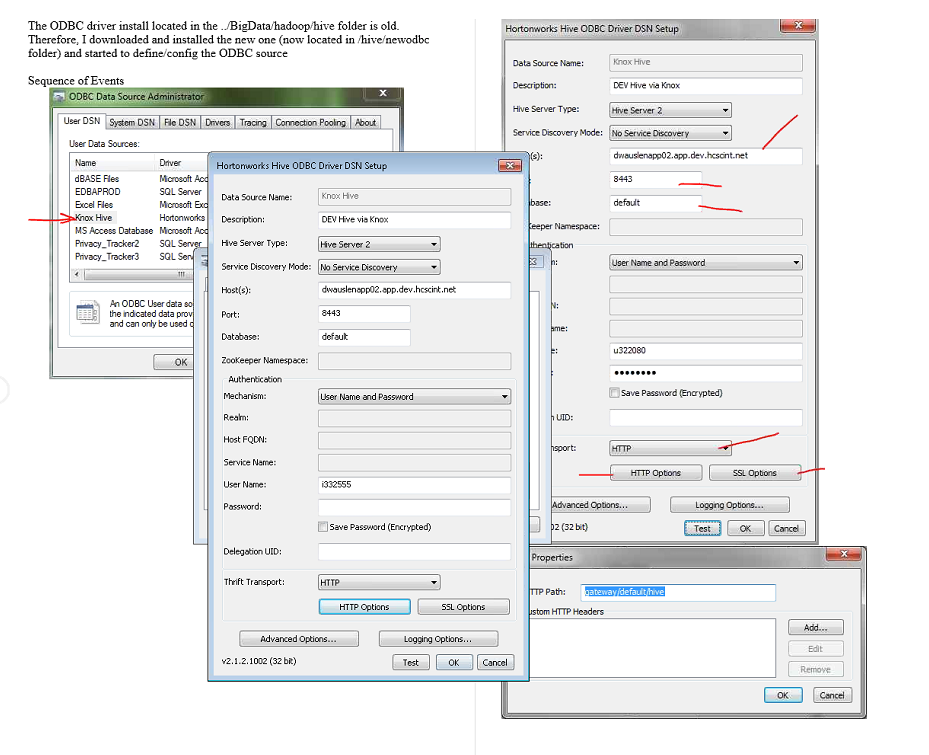


Add new System Data Source, and select the Hortonworks Hive ODBC Driver



* 1. Secured Connection (Knox)

Knox is part of the Data Lake security infrastructure implemented on Aug/2016 for Dev/Test and Oct/16 for Prod. You need to request the entitlements in AccessNow (step 3) before you can access Hive. After that is completed, follow these steps to configure your data sources:



Servers

Production:

Host: pwauslaenapp04.app.hcscint.net

Port: 8443

Test:

Host: twauslaenapp04.app.test.hcscint.net

Port: 8443

Dev:

Host: dwauslenapp02.app.dev.hcscint.net

Port: 8443

Select “User Name and Password” and enter both.

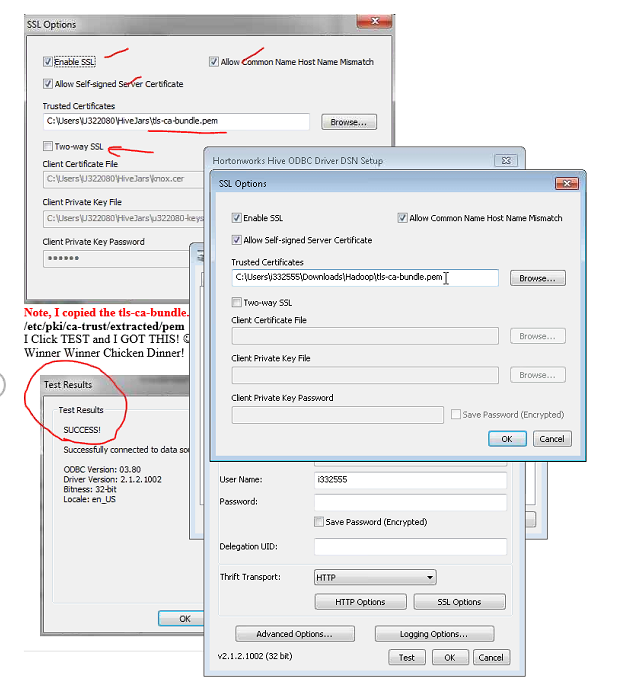
Also select “HTTP” as the transport, and click on "HTTP Options". On HTTP Path, enter: "gateway/default/hive", and click ok.

SSL Setup

Download the certificate file (.pem) from this link and place it somewhere in your computer:

<http://community.fyiblue.com/sites/GPDATIntak/Shared%20Documents/DDO/Onboarding/Hadoop/tls-ca-bundle.pem>

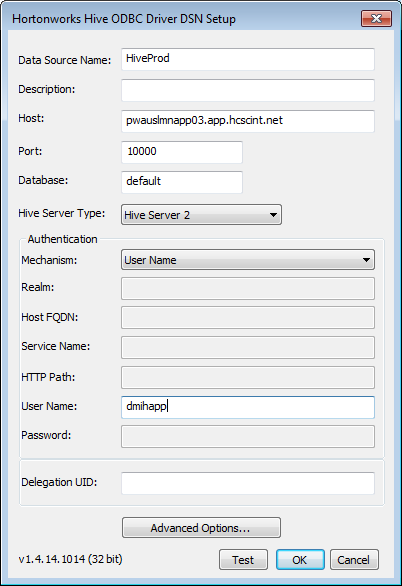
Click on the SSL options. Enable SSL and other options based on the picture below ("Allow Common Name Host..." and "Allow Self-Signed Server …"). Enter the path to your certificate (tls-ca-bundle.pem) file.



You are all set, now you can test the connection.

* 1. Unsecured Hive connection (ignore after 10/10/2016 – only prior to Knox install in prod)

Enter the following information for each environment:



Production:

Host: pwauslmnapp03.app.hcscint.net

Port: 10000

Test:

Host: twauslmnapp01.app.test.hcscint.net

Port: 10000

Dev:

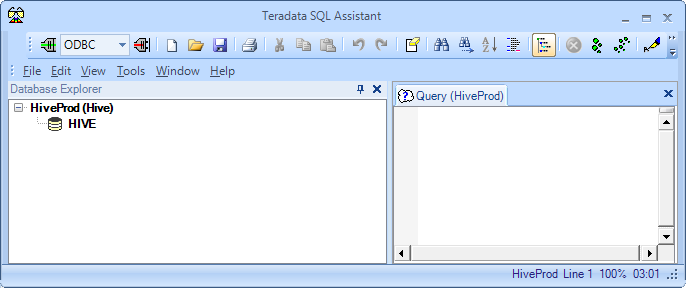
Host: dwauslenapp02.app.dev.hcscint.net

Port: 10000

1. **Open Teradata Sql Assistant, select ODBC, and click on the connect button.**

In the "Machine Data Sources" tab, select either HiveTest or HiveProd as the data source and connect.

If you don't see any of these then there is a configuration problem, and you need to go back to step 3.2.



Test it out by running:

select \* from HIVE\_SYSTEM.HIVE\_SYSTEM;

You should get all Hive configuration data back.

1. **Add databases to SQL Assistant**

Now you can add databases to the left pane in SQL Assistant like any other data sources.

Certain databases you can add are: default, smith, gold, etc.

For a catalog of tables and databases, please go to this URL:

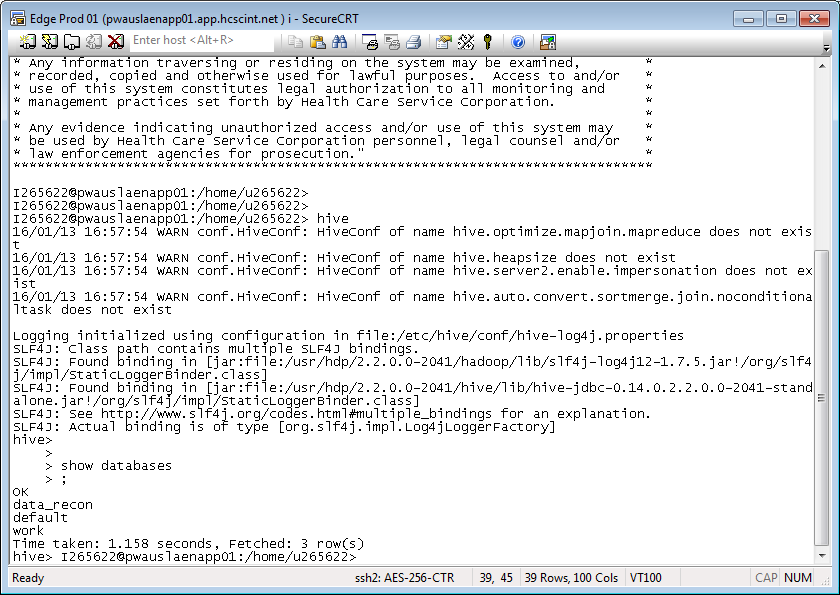
<http://community.fyiblue.com/sites/EntArchSvc/DataArch/Shared%20Pages/Data%20Lake%20Inventory%20Catalog.aspx>

1. **Looking for additional databases or tables (using Hive/Behive - advanced)**

In order to know what databases are available in Hive, you need to first logon to the EDGE server (Production: pwauslaenapp01.app.hcscint.net). There are 2 options to interact with Hive: the Hive CLI or the new beeline client (CLI will be deprecated, so use beehive)

Using Hive CLI

Start Hive CLI by typing ‘hive’. Then simply type ‘show databases;’. The databases will be listed under the “OK” message.



Using Beeline

Unsecured (no Knox):

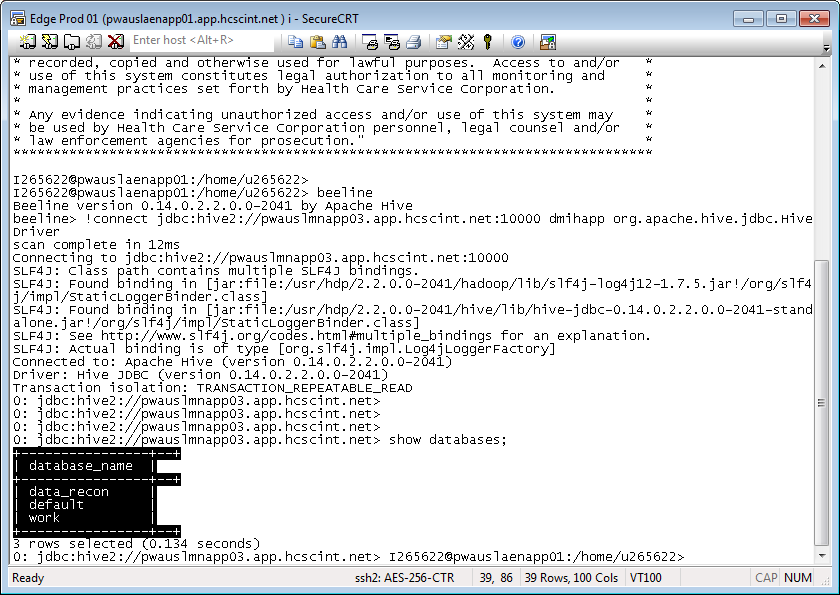
Start Beeline by typing ‘beeline’. Connect to the Hive database by using the command:

!connect jdbc:hive2://pwauslmnapp03.app.hcscint.net:10000 dmihapp org.apache.hive.jdbc.HiveDriver

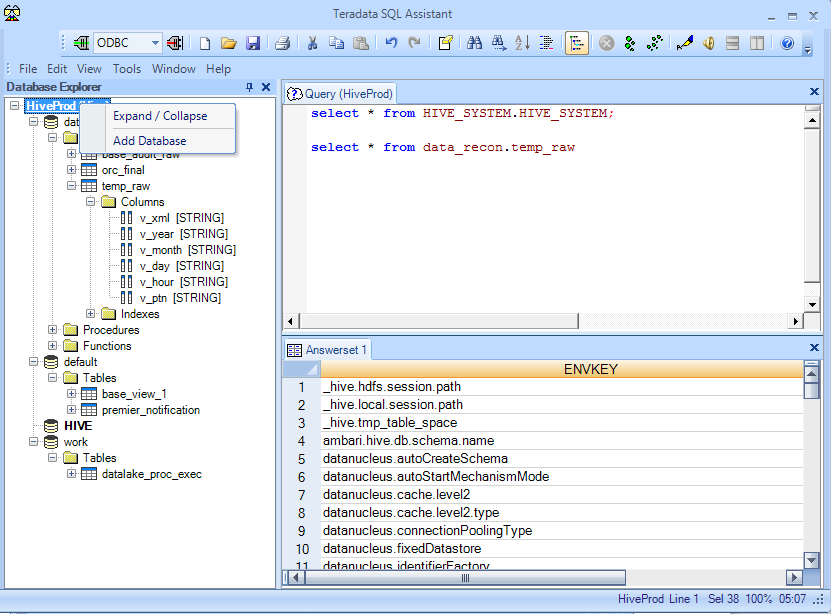
Secured:

!connect jdbc:hive2://twauslaenapp03.app.test.hcscint.net :10001/;principal=hive/\_HOST@ADHCSCTST.NET;transportMode=http;httpPath=cliservice

Now type ‘show databases;’



Go back to SQL Assistant and click on the left panel and select “Add Database” and start adding whatever databases you found in the previous step (data\_recon, default, and work).



Now you should be ready to query the data in the environment.

For additional information on Hive commands and SQL please refer to this cheat sheet:

