Data Integration Platform Cloud Hands-on Lab

Denis Gray
Senior Director of Product Management

Werner He
Senior Director of Product Management

Alex Kotopoulis
Director of Product Management

Julien Testut
Senior Principal Product Manager

ORACLE

HOL6277 - Introduction to Data Integration Platform Cloud

The rapid adoption of enterprise cloud-based solutions brings with it a new set of challenges. Data integration remains one of the greatest challenges of any enterprise cloud-based solution. Join this hands-on lab to gain firsthand experience with the power and simplicity of Oracle Data Integration Platform Cloud powered by Oracle Data Integrator, Oracle Enterprise Data Quality solutions, and Oracle GoldenGate. See how Oracle Data Integration Platform Cloud simplifies the end-to-end creation and execution of the historically arduous data integration tasks of ingesting, loading, preparing, and transforming data as well as performing a real-time synchronization between an on-premises and a cloud database in just a few clicks.

The following lessons will walk us through various steps that are needed to create Data Integration Platform Cloud connections, a Data Preparation task, an ODI Execution Task, and run Jobs to prepare and load data into a target data warehouse.

Overview

Time to Complete

Perform all tasks – 60 Minutes

Prerequisites

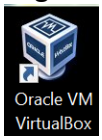
Before you begin this tutorial, you should

- Have a general understanding of RDBMS and data integration concepts
- Have a general understanding of ETL and data synchronization concepts

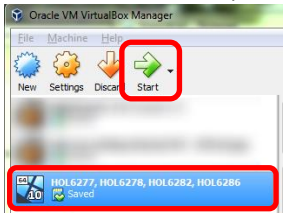
Task 0: Preparation Steps

For this lab, the Data Integration Platform Cloud and the sources and targets are contained within one environment for simplicity. This environment is hosted in the Cloud. All user interactions with Data Integration Platform Cloud will be through a browser (Chrome) and VNC Client (TigerVNC) installed in a VirtualBox image on your machine.

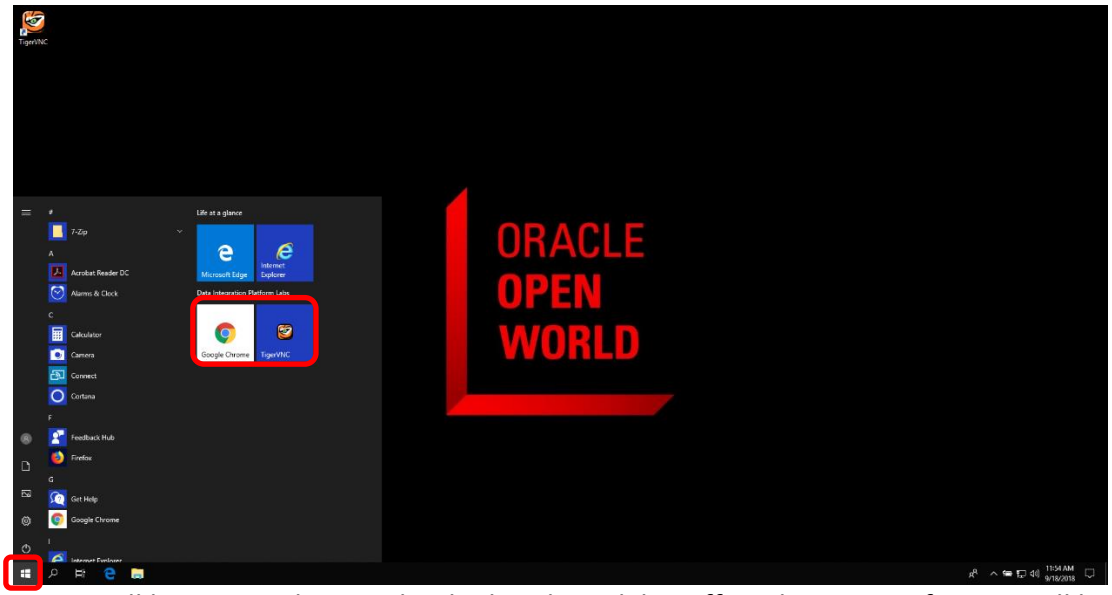
1. Log into your machine and open up Oracle VM VirtualBox from the Desktop



2. Click on HOL6277, HOL6278, HOL6282, HOL6286 and click on Start

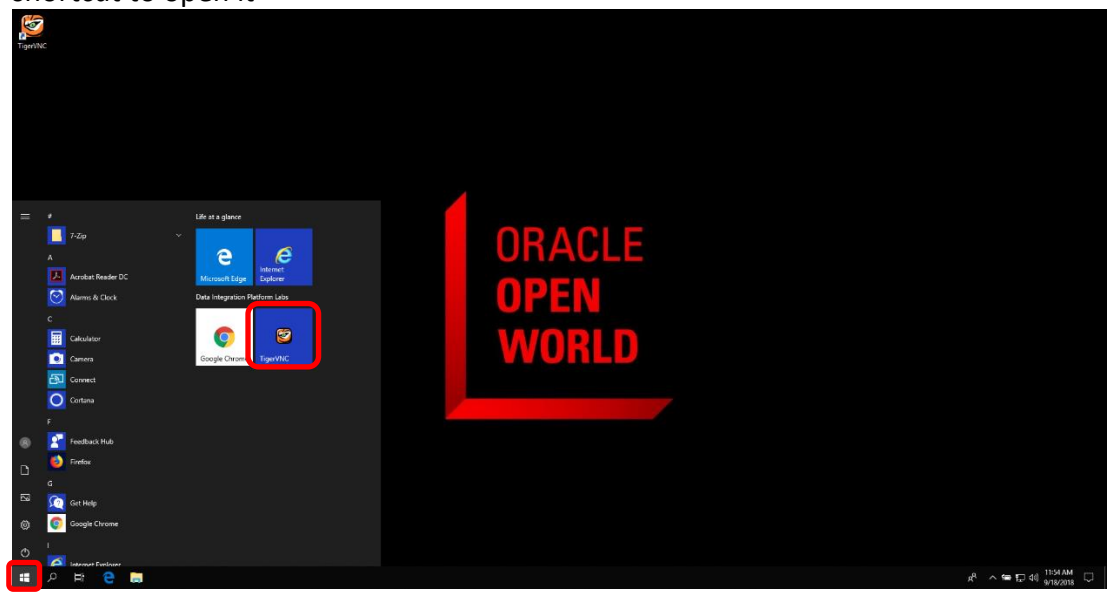


- When the VirtualBox Image is started, you will be logged into a Windows 10 machine. The Start Menu contains shortcuts to access Chrome and TigerVNC that will be used throughout this lab

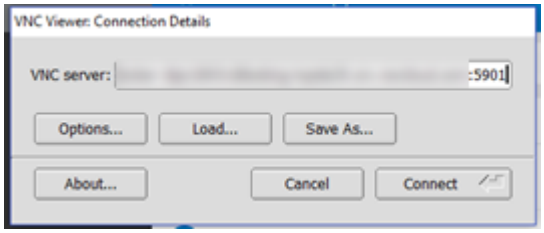


3. A URL will be assigned to you by the hands-on lab staff, make a note of it as it will be your environment for the entire lab:
    a. <hostname>:8001/dicloud (URL for Chrome)
    b. <hostname>:5901 (URL for TigerVNC)

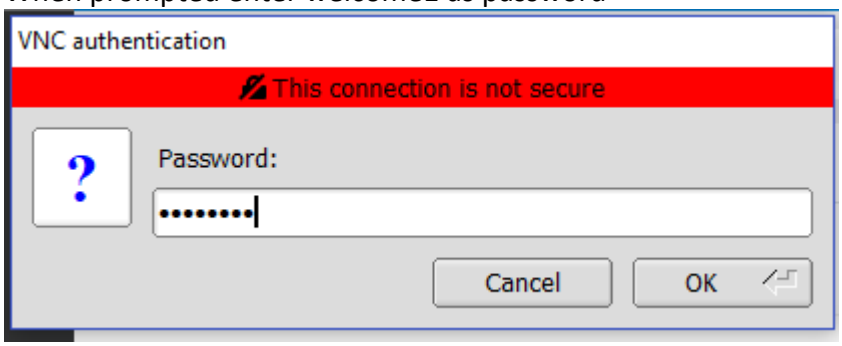
4. We will start with connecting to the DIPC host and starting a DIPC Agent.
Open the Windows 10 Start menu running in the VirtualBox VM and click on TigerVNC shortcut to open it



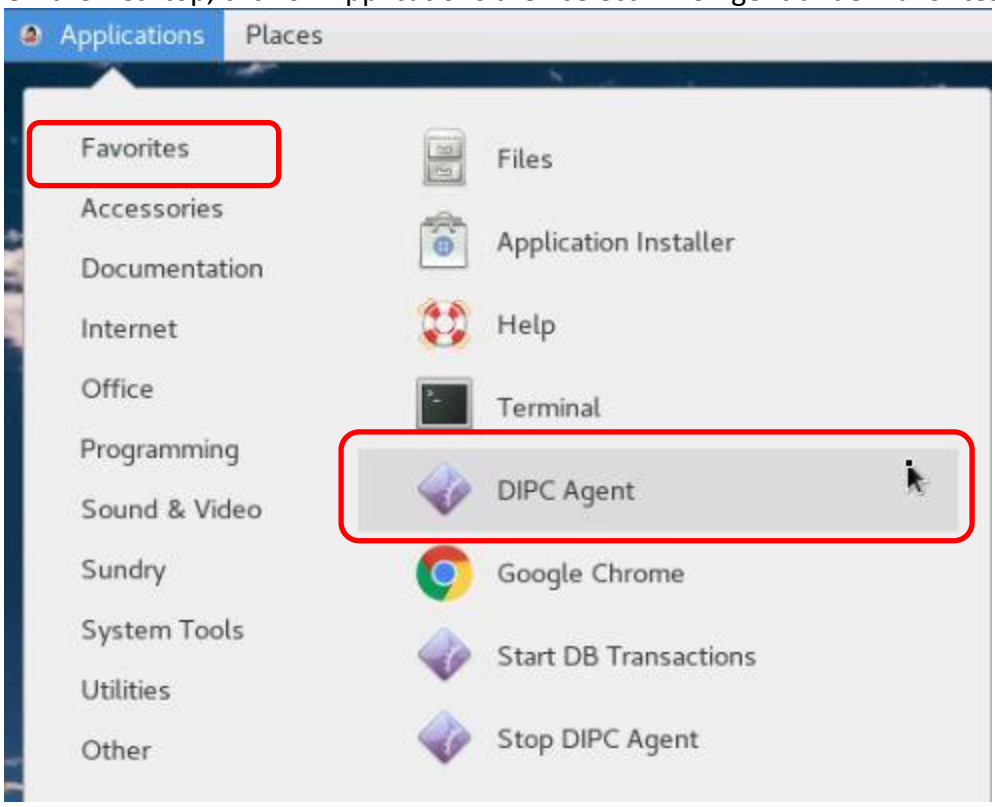
5. In the TigerVNC window enter the URL (<hostname>:5901) given to you by the hands-on lab staff and click Connect



6. When prompted enter welcome1 as password



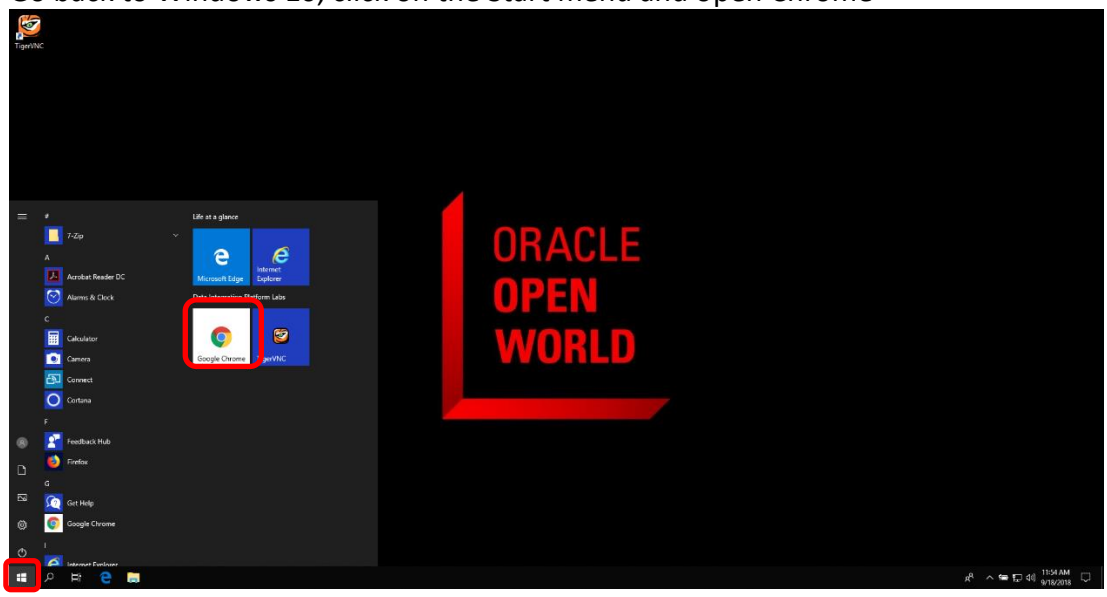
7. On the Desktop, click on Applications then select DIPC Agent under Favorites



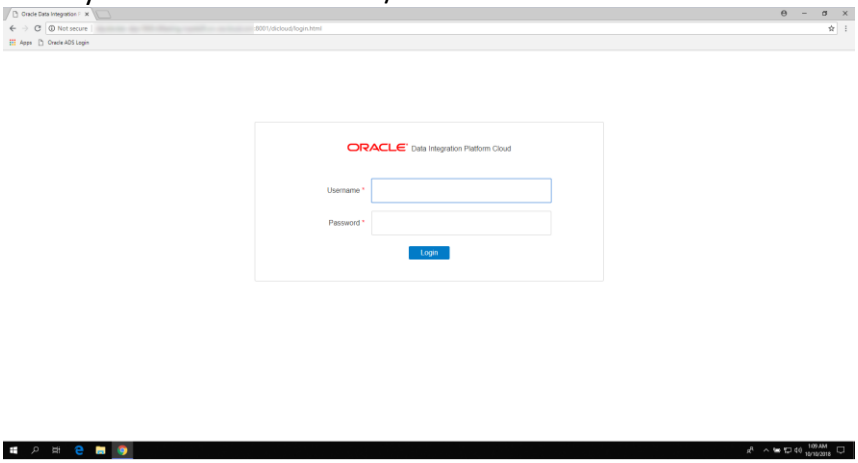
8. A window will appear and will track the DIPC Agent startup. An Agent will be used by DIPC to access the different sources and targets, it will be fully started in a minute or so **DO NOT CLOSE THIS WINDOW, you can minimize it if needed**

9. Log into Data Integration Platform Cloud

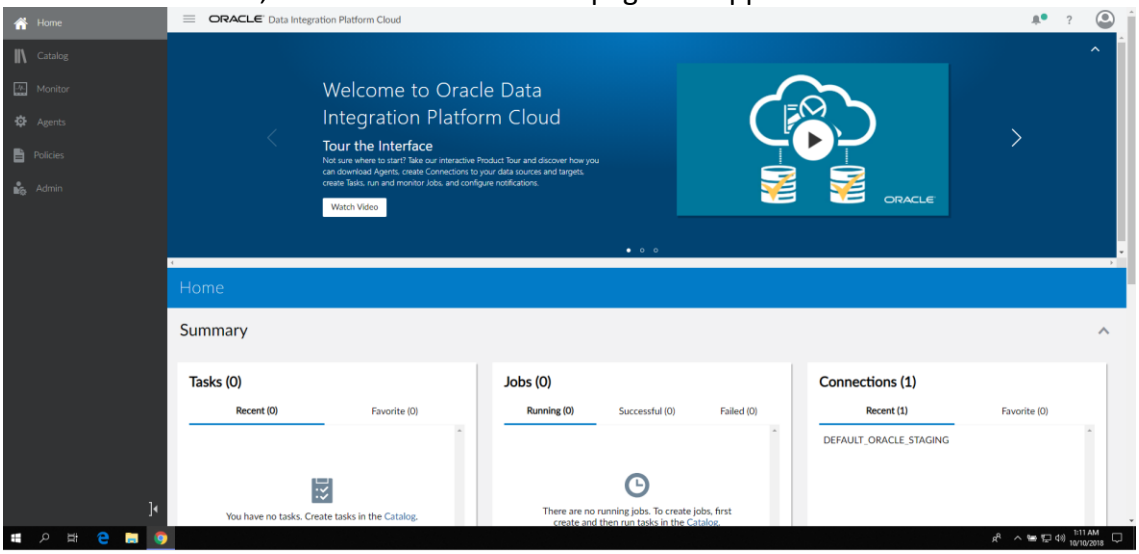
a. Go back to Windows 10, click on the Start menu and open Chrome



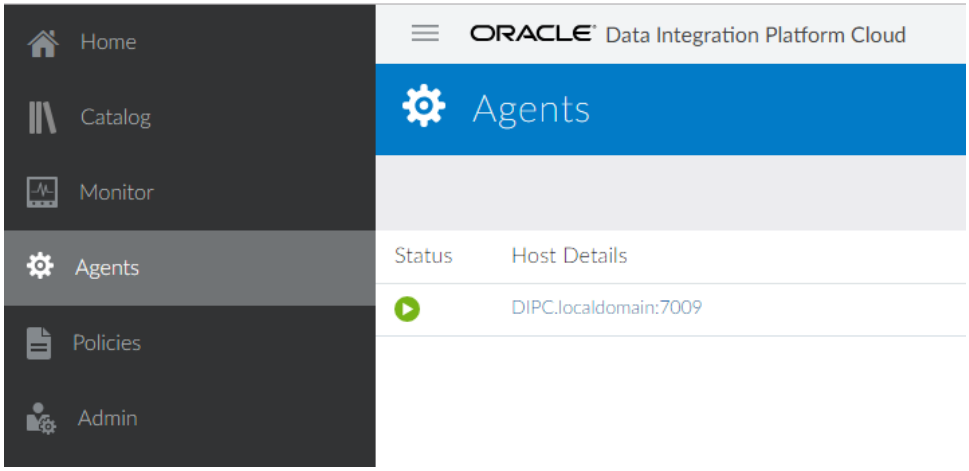
b. Go to your <hostname>:8001/dicloud



c. Login with **weblogic/welcome1**
After a few seconds, the DIPC Console Home page will appear –



10. Click on Agents to display the list of Agents available in this environment. The Agent (DIPC.localdomain:7009) we just started will be displayed and should be in Running (green) status



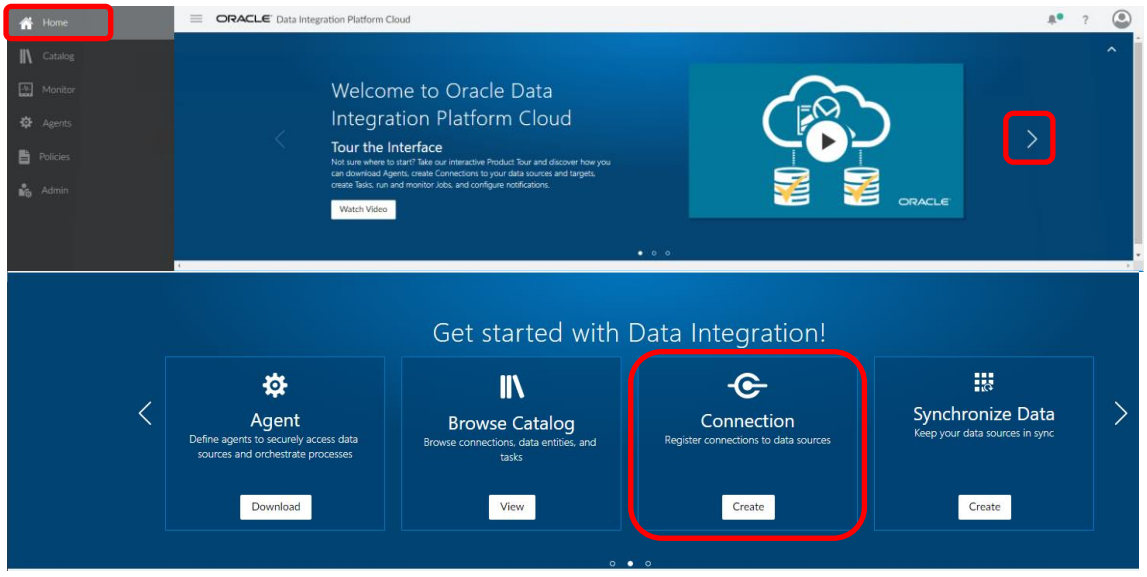
Task 1: Setup DIPC Connections

1. The connectivity information for this hand-on lab is as follows:

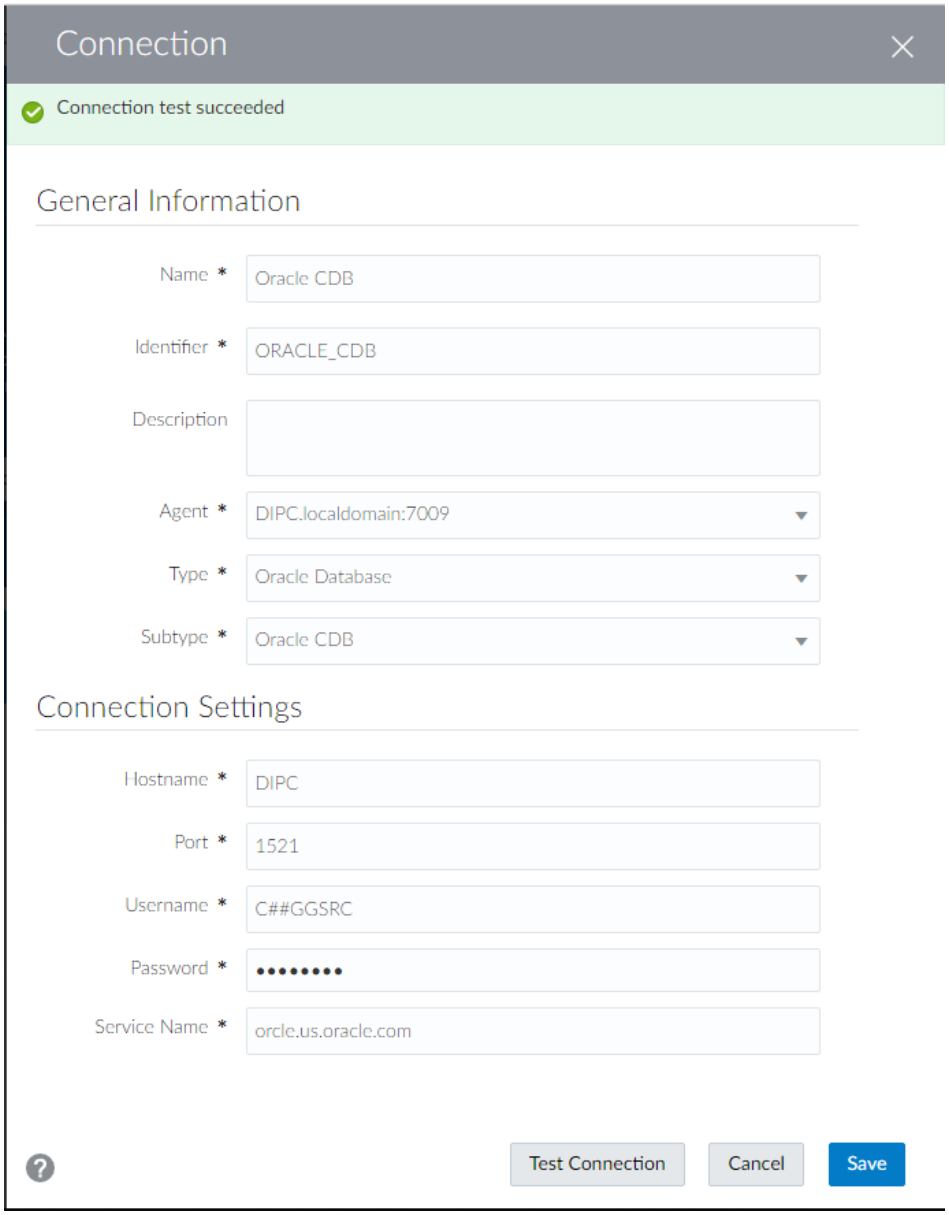
   CDB User: C##GGSRC
   Source Schema/User: DIPC_SRC
   Target Schema/User: DIPC_TGT
   Passwords: welcome1
   Server: DIPC
   Port: 1521
   Services:
   orcle.us.oracle.com (CDB Connection)
   orclpdb.us.oracle.com (PDB Connections)

   Let us first create a new Oracle CDB Connection

   a. Click on Home then click on > in the carousel at the top to locate Connection and click Create



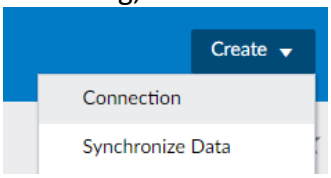
b. Enter the following information
   - Name: Oracle CDB
   - Identifier: use defaut
   - Agent – DIPC.localdomain:7009
   - Type Oracle Database
   - Subtype: Oracle CDB
   - Hostname: DIPC
   - Port: 1521
   - Username: C##GGSRC
   - Password: welcome1
   - Service Name: orcle.us.oracle.com



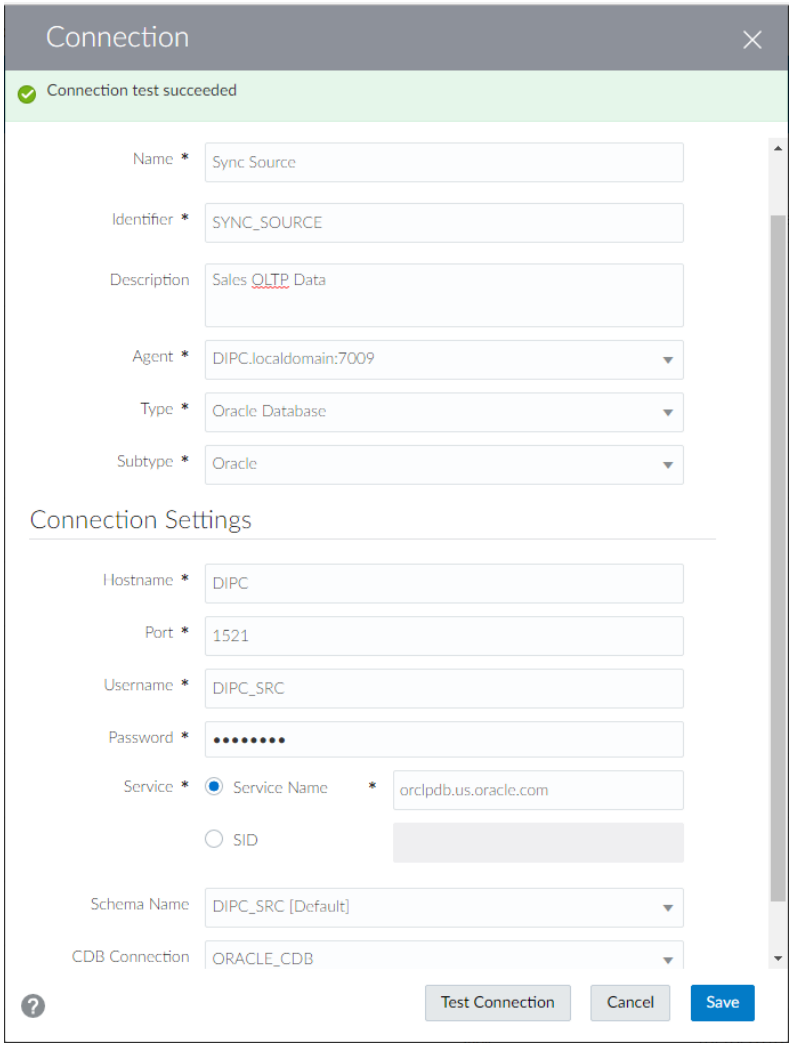
c. Click Test Connection and Save if successful

2. Create Sync Source Connection

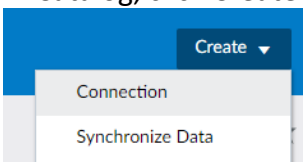
- In Catalog, click Create > Connection



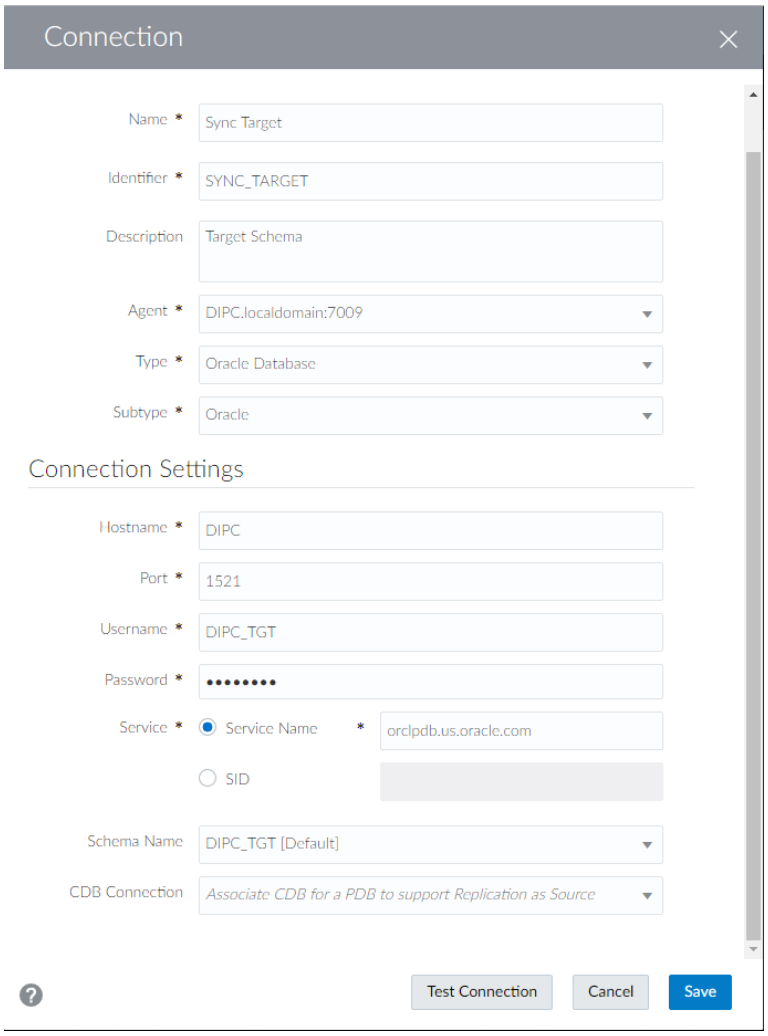
- Enter the following information
  - Name: Sync Source
  - Identifier: use default
  - Description: Sales OLTP Data
  - Agent: DIPC.localdomain:7009
  - Type Oracle Database
  - Subtype: Oracle
  - Hostname: DIPC
  - Port: 1521
  - Username: DIPC_SRC
  - Password: welcome1
  - Service Name: orclpdb.us.oracle.com
  - Schema: DIPC_SRC
  - CDB Connection: ORACLE_CDB



3. Create Sync Target Connection
    - In Catalog, click Create > Connection

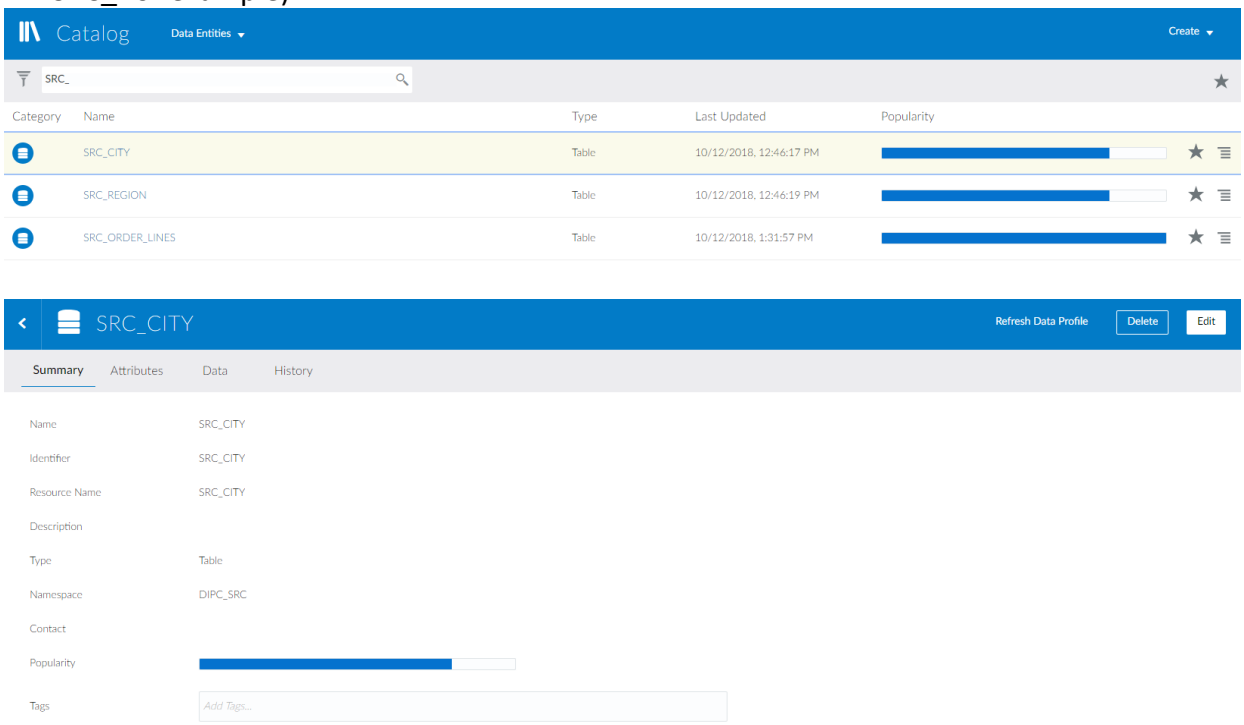


    - Enter the following information
        - Name: Sync Target
        - Identifier: use default
        - Description – Target Schema
        - Agent – DIPC.localdomain:7009
        - Type Oracle Database
        - Subtype: Oracle
        - Hostname: DIPC
        - Port: 1521
        - Username: DIPC_TGT
        - Password: welcome1
        - Service Name: orclpdb.us.oracle.com
        - Schema: DIPC_TGT
        - CDB Connection: leave blank



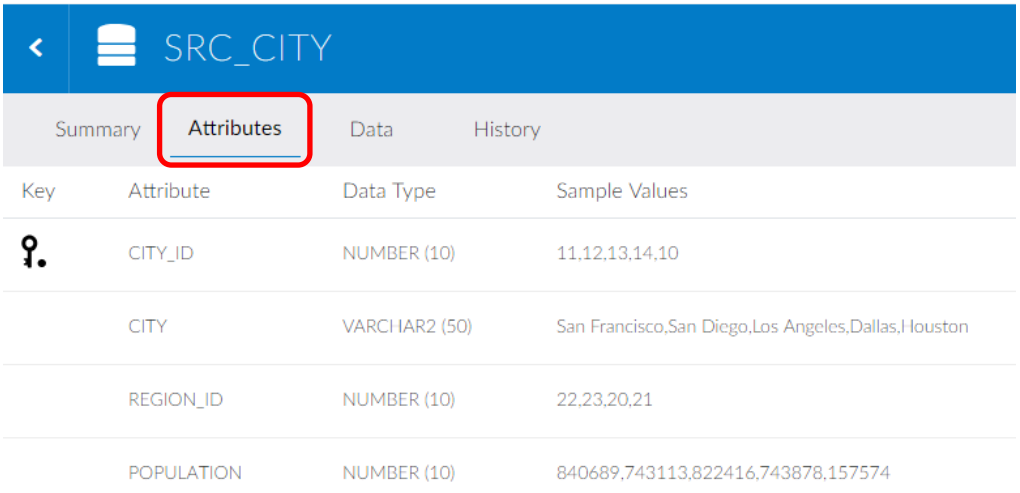
- Click Test Connection and Save if successful

4. Review Catalog after saving –
5. Data Entities are harvested and profiled at Connection creation, their popularity is also calculated by reviewing the DB query logs
   Note: This process may take a little bit of time
6. Click an entity – SRC_CITY
   You can browse the Catalog pages to find it or you can use the Search bar (search for SRC_ for example)

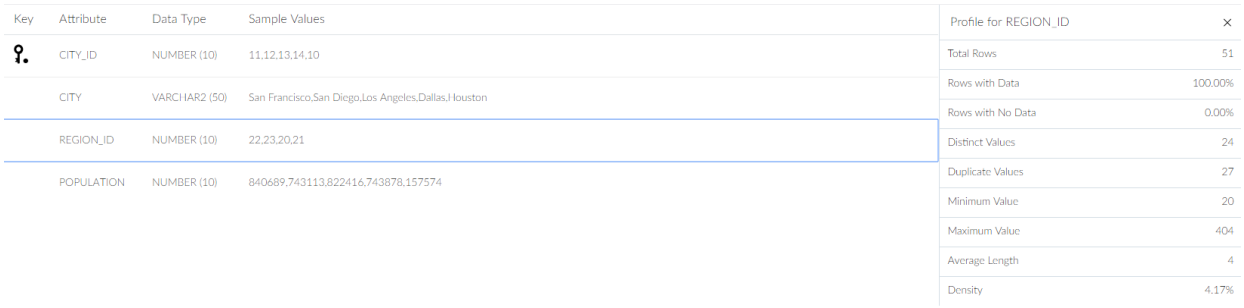


Notice the Popularity score calculated for SRC_CITY, a full bar means that this is one of the Data Entities that has been used the most in queries.Tags can be added as well to group objects together

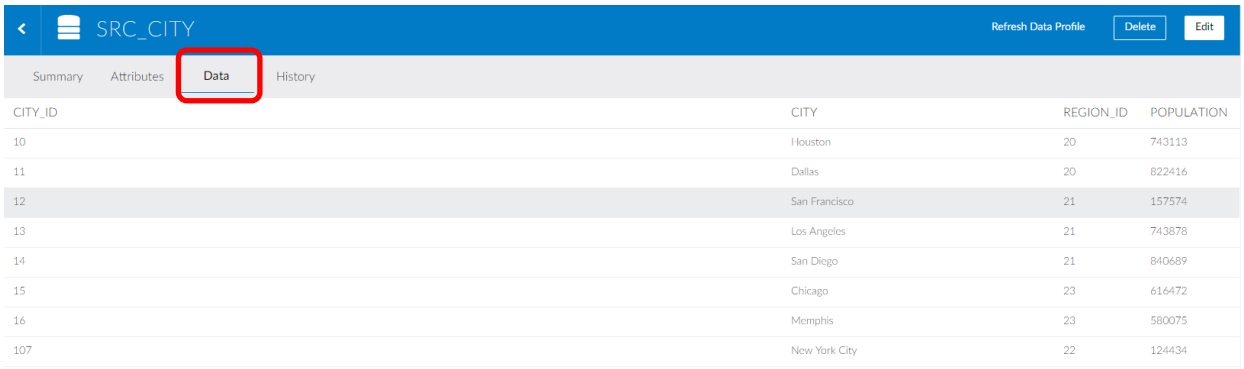
7. Click on the Attributes tab



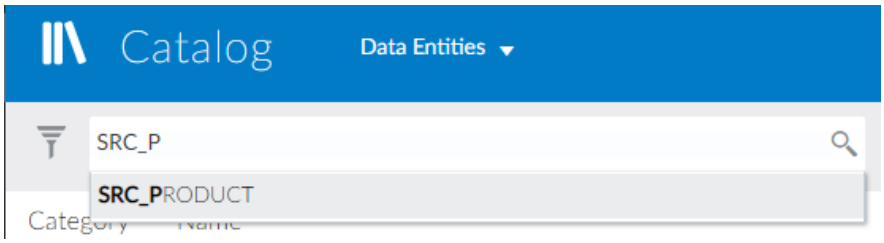
8. Click on an Attribute, REGION_ID for example



9. Notice the Profiling statistics appearing in the right-hand side drawer
10. Click on the Data tab



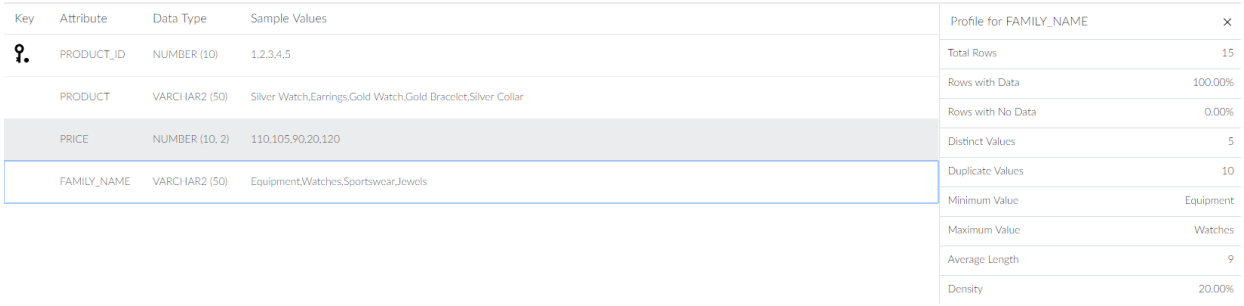
11. Go back to the main Catalog page – Search for Data Entity – SRC_PRODUCT and click on it in the Quick Search results



12. Click Attributes tab

13. Click on a column, for example column – FAMILY_NAME

14. Notice the Profiling statistics



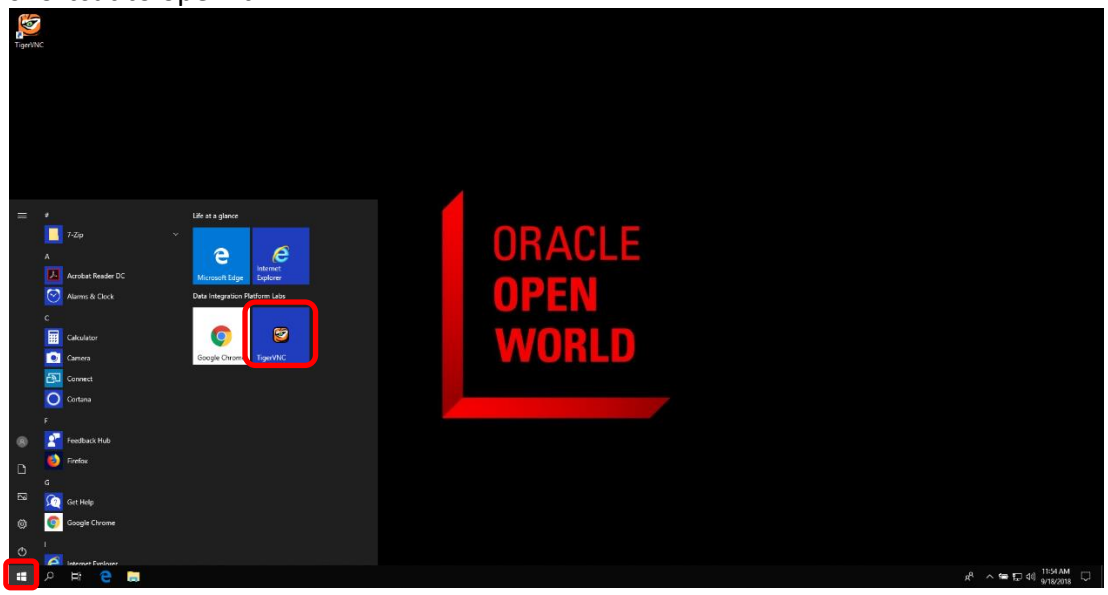
15. Click Data tab
16. Review other entities as needed

Task 2: Create DIPC Synchronize Data Task

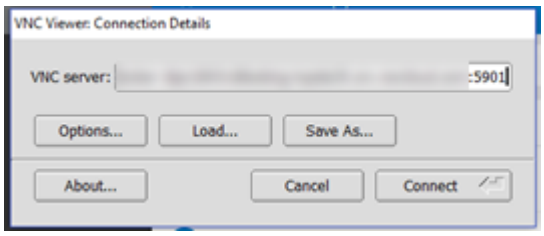
1. This hands-on lab uses a JDBC utility client that was built specifically for this demo. This client is **NOT** part of DIPC, however it does help visualize the Synchronize Data and ODI Execution Job process.

2. Go back to the TigerVNC window that should still be opened in Windows 10.

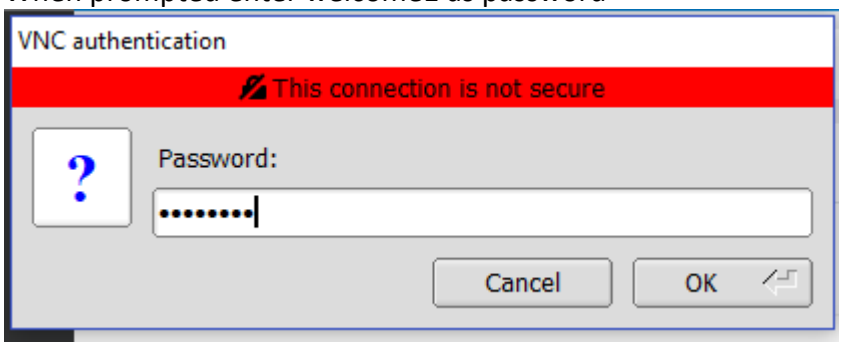
   If you need to start it again: click on the Windows 10 Start menu and click on TigerVNC shortcut to open it



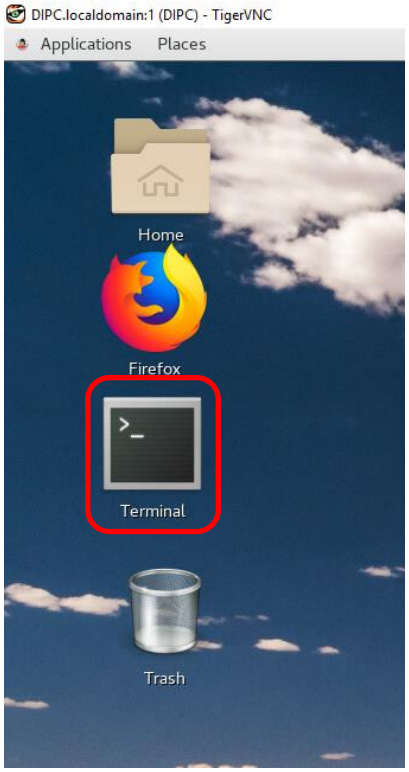
   In the TigerVNC window enter the URL (<hostname>:5901) given to you by the hands-on lab staff and click Connect



When prompted enter welcome1 as password

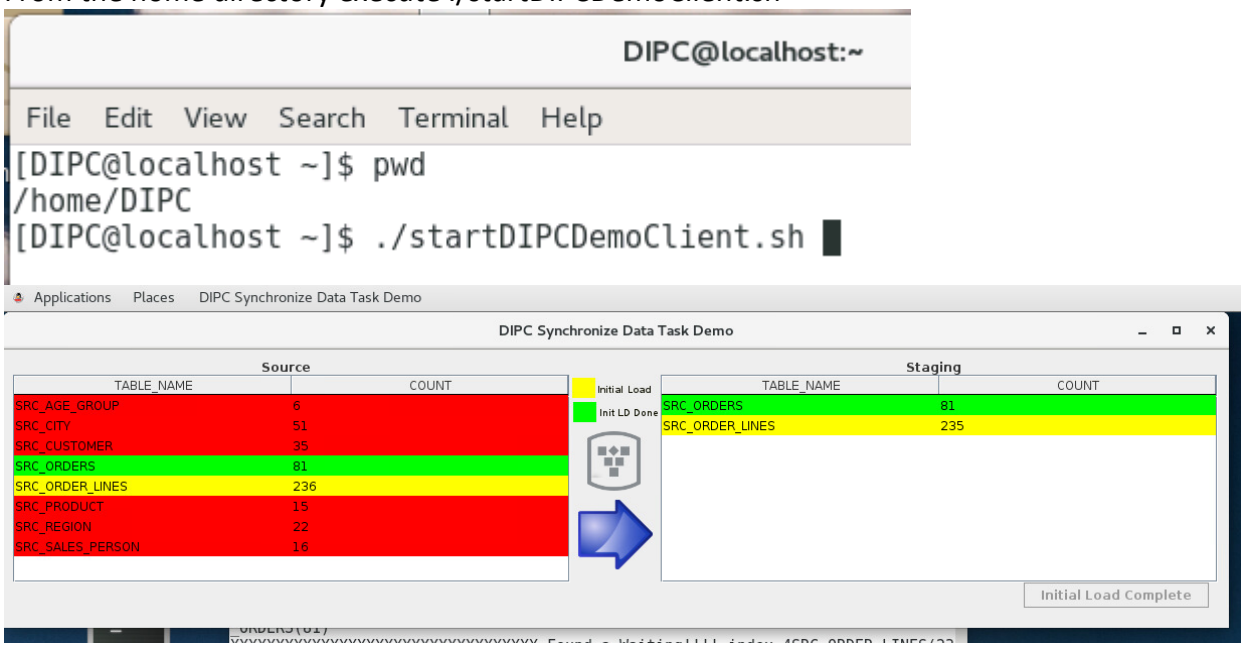


3. Open a terminal window using the Terminal shortcut on the Desktop



If needed press Enter to see the Login window, re-enter the DIPC user password (welcome1) and press Enter or click Unlock

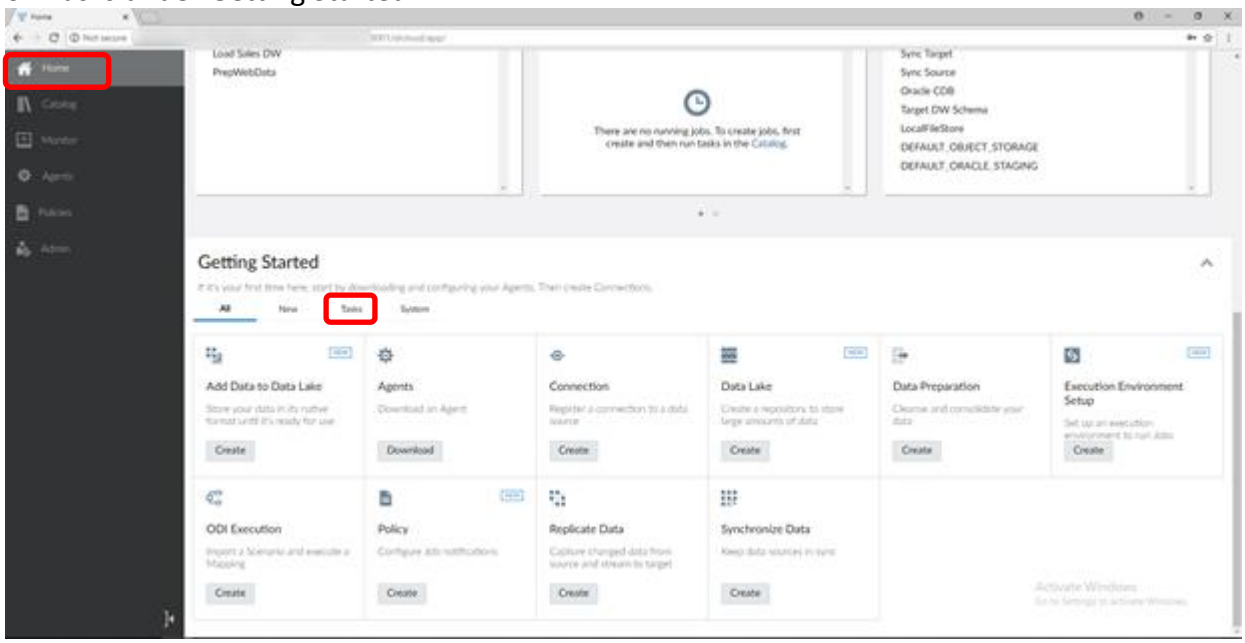
4. From the home directory execute ./startDIPCDemoClient.sh



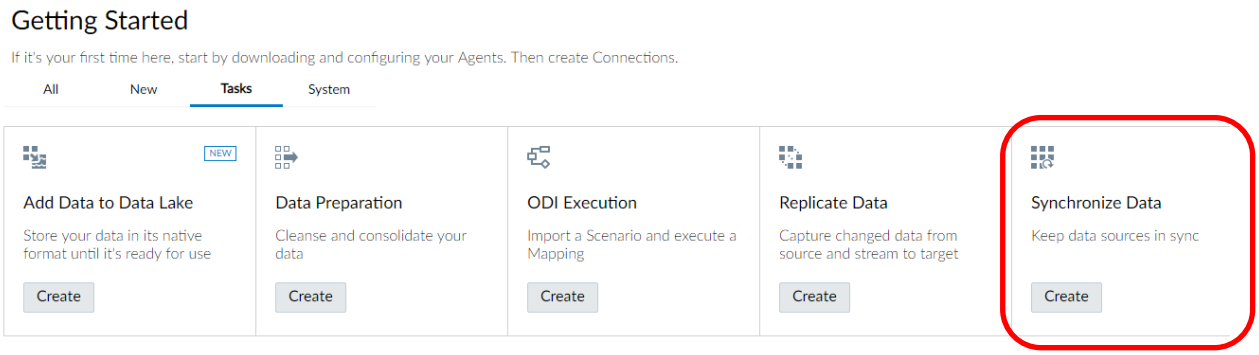
Note: The Staging schema is not empty as it is also being used by another hands-on lab, this is expected and the Synchronize Data Job will first clean up the schema before loading the data

Once the Synchronize Data task is saved and executed this client will be used to visually monitory the Replicated Schema, the tables and their row counts

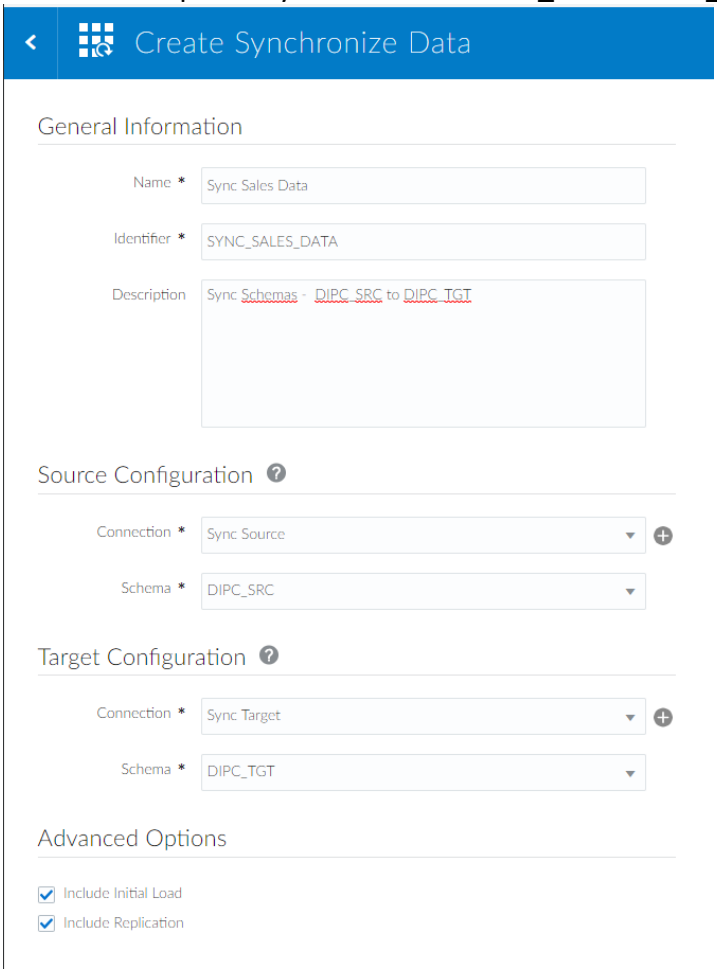
5. Go back to Chrome, click Home in the DIPC Console, scroll all the way down and click on Tasks under Getting Started



6. Click Create under Synchronize Data

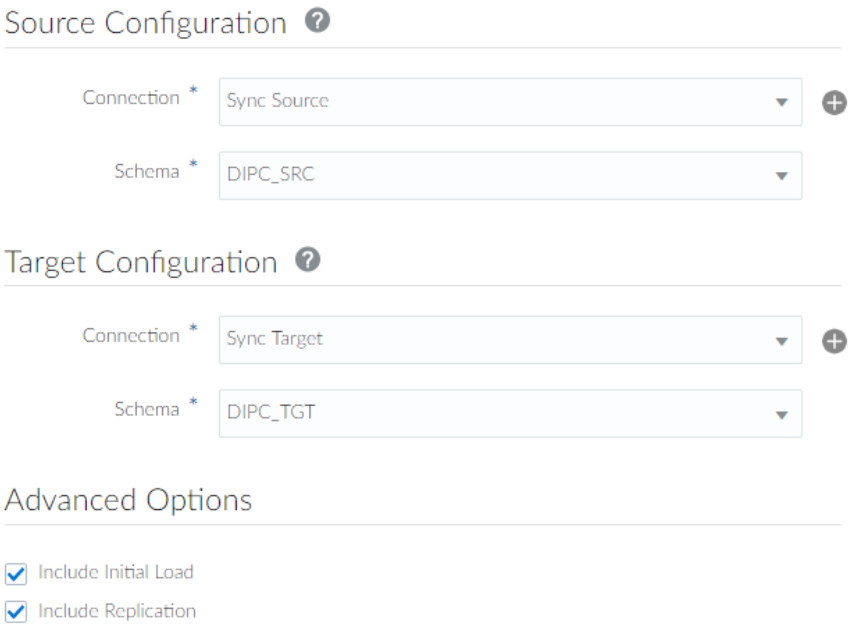


7. Name your task –Sync Sales Data
8. Description: Sync Schemas - DIPC_SRC to DIPC_TGT



9. Select your source connection and schema
    a. Connection: Sync Source
    b. Schema: DIPC_SRC
10. Select your target connection and schema
    a. Connection: Sync Target
    b. Schema: DIPC_TGT
11. Leave 'Include Initial Load' and 'Include Replication' checked under Advanced Options. These options allow you to optionally enable or disable the initial load and/or the on-

going schema replication

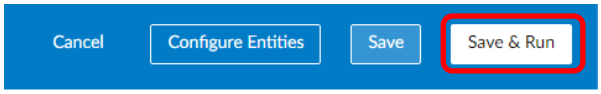


12. Next click on Configure Entities. This page allows you to filter the objects that will be transferred using the Synchronize Data Task

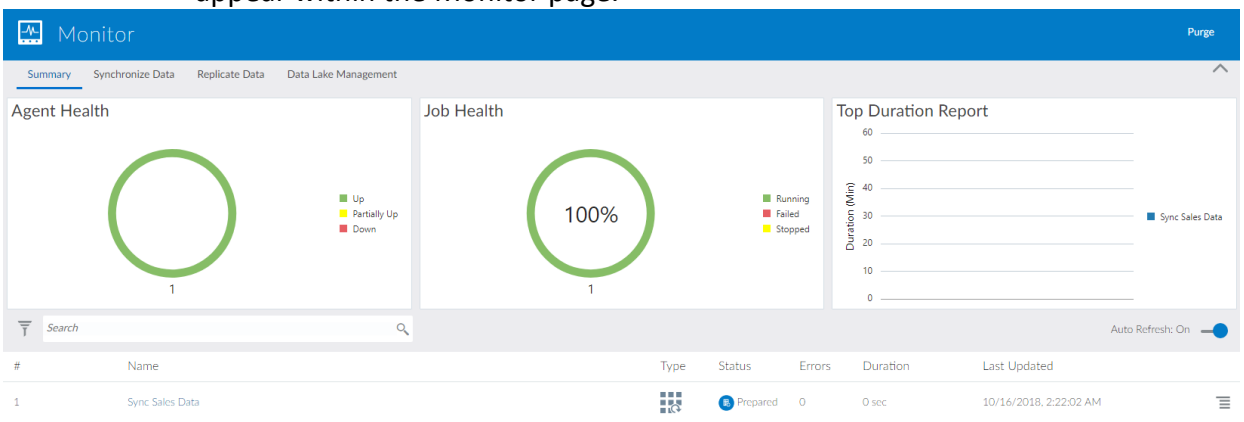
13. The Configure Entities screen helps you create Include or Exclude rules to define precisely which database objects will be transferred to the target schema. By default all Data Entities are transferred with the rule: Include *
Note: The list of Data Entities may take a few seconds to show up

14. Click on < to go back to the main Synchronize Data Task screen

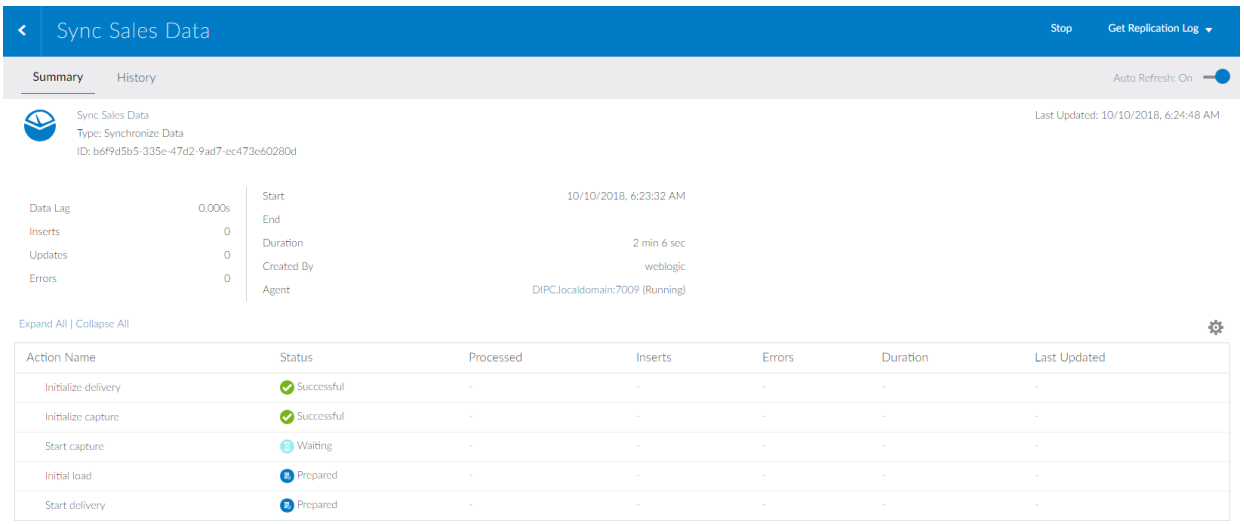
15. Click on Save & Run to start the execution



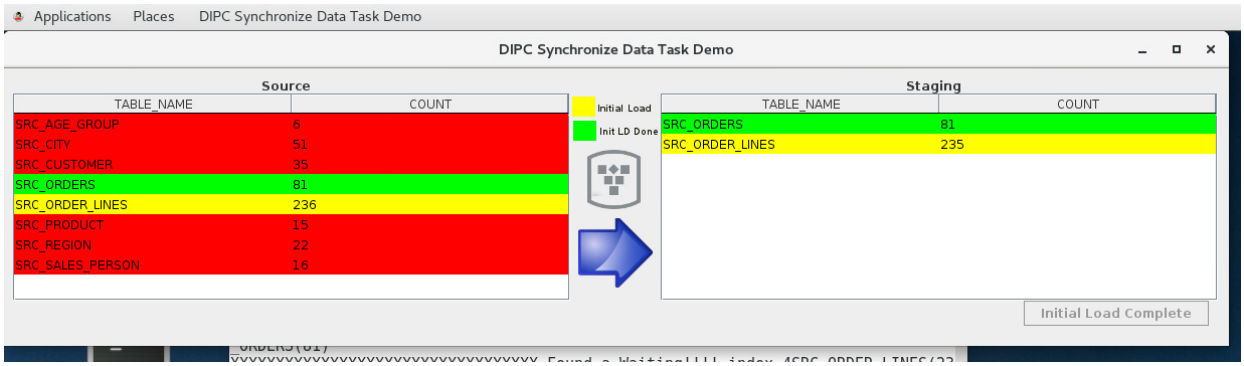
a. A notification will appear mentioning that the job was saved
b. A new DIPC Job will be created to execute the Task.
   A notification will appear in the notification bar and the job will automatically appear within the Monitor page.



c. Click on the Job to see the Job Details

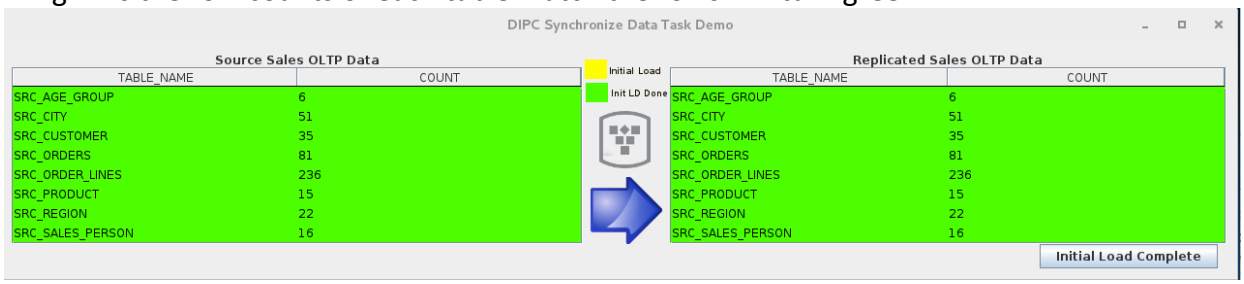


d. Auto-refresh is on, status will be updated frequently
e. As the job executes, the Initial Load process is created in Oracle Data Integrator (ODI) while DIPC configures Oracle GoldenGate (OGG) for the Source Capture and Target Delivery
f. As this job executes, the Replicated Sales OLTP Source Data table will be updated in the Demo Client (in TigerVNC window). As new tables are created they will show up as yellow, when the row counts of the source and replicated schemas match the rows will turn green

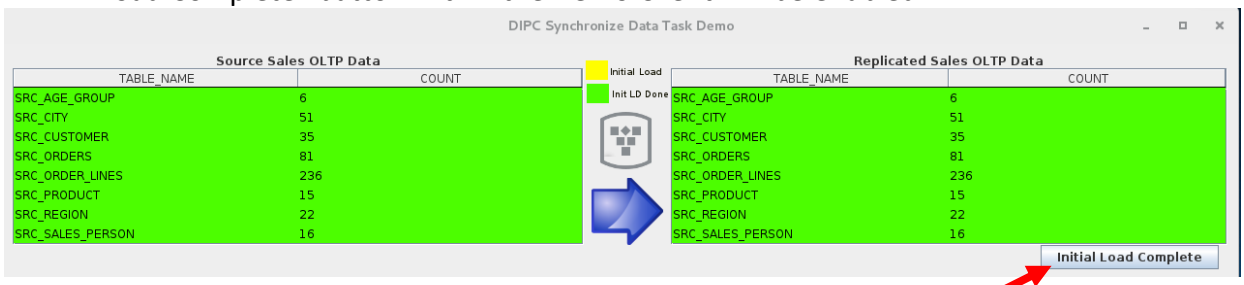


Note: It may take several minutes (10+) for the Replicated Sales OLTP Data side to show anything. This is normal.

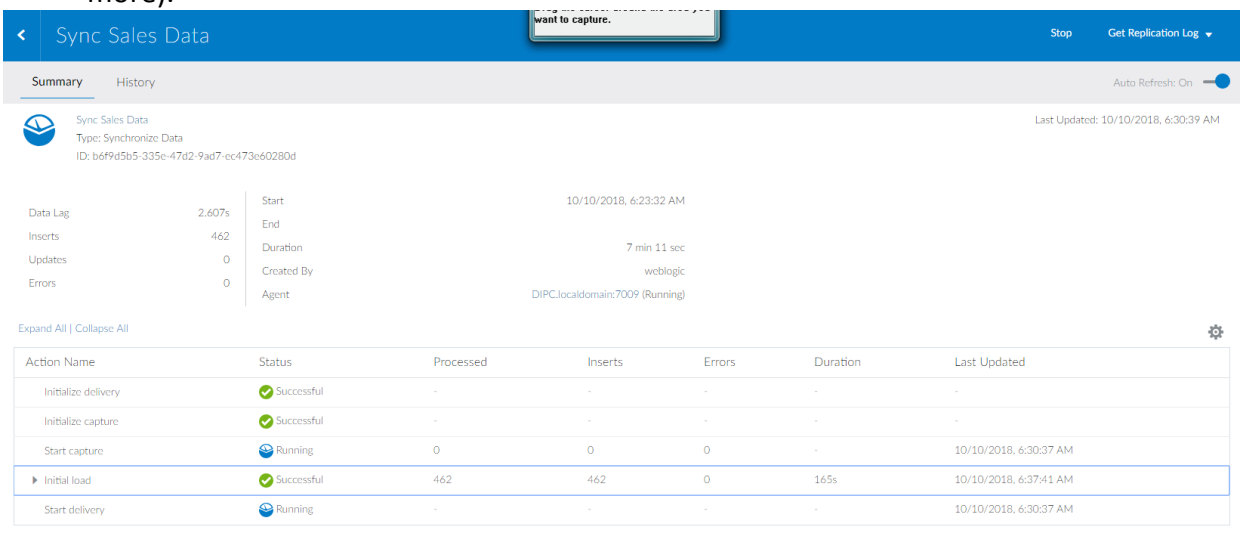
g. As the row counts of each table match the rows will turn green



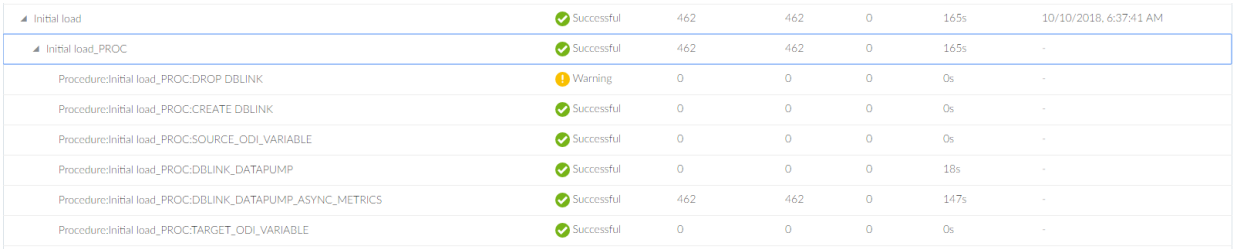
h. Once the row counts match and the Initial Load process is complete the "Initial Load Complete" button within the Demo Client will be enabled.



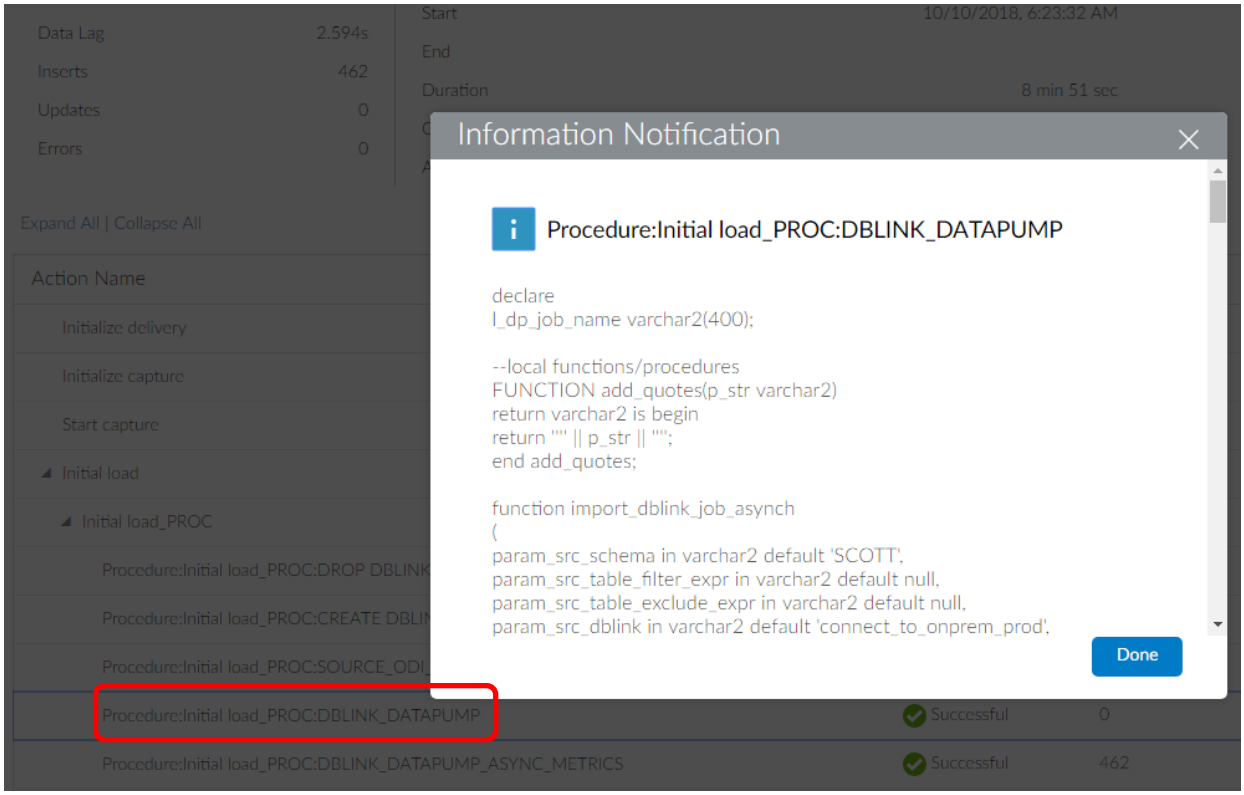
i. Go back to the Job Details in the DIPC App. to review the status there. The Initial load Action will show Successful after a little while (may take 10 minutes or more).



j. Once done, the Initial load Action can be expanded to review the various Steps underneath

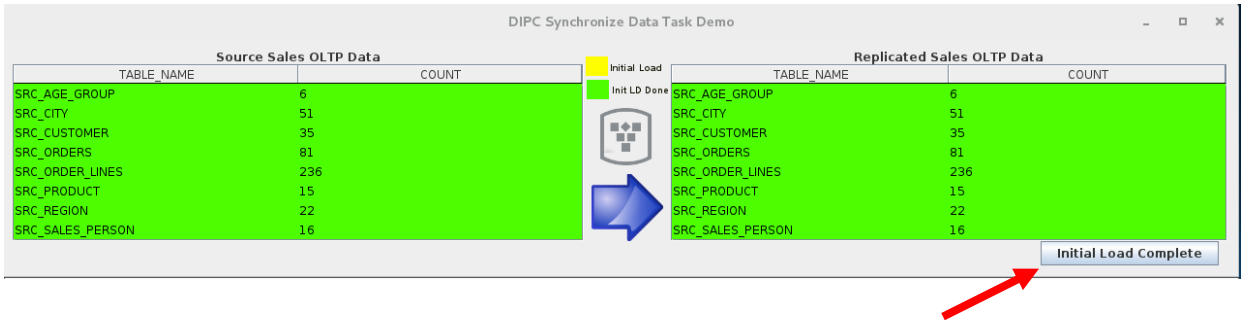


k. Click on Procedure:Initial load_PROC:DBLINK_DATAPUMP to review the Code generated by DIPC for the Initial Load. Click Done when you've completed the code review

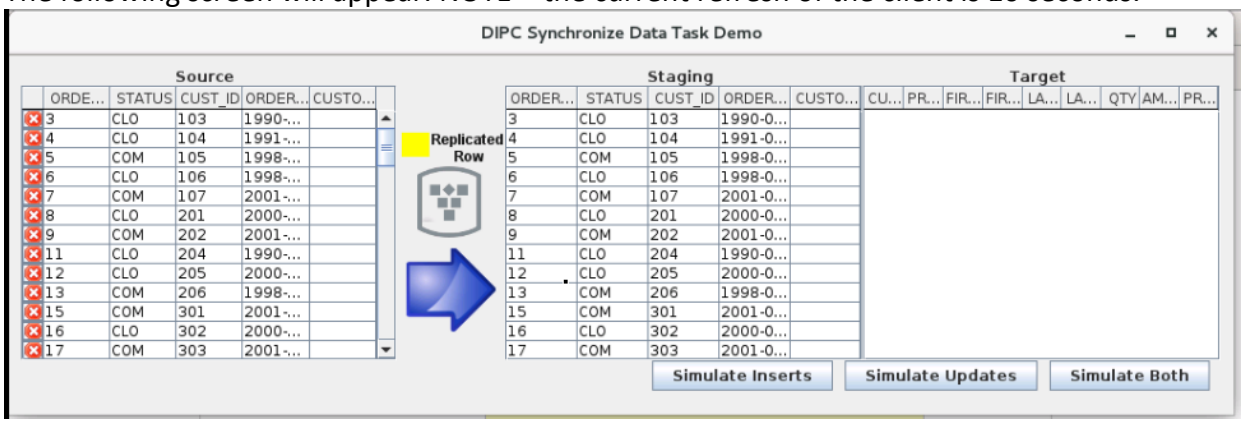


Task 3: Monitor Source Inserts/Updates/Deletes

1. Now that the initial load is complete and the capture and delivery processes have been created and are running, we can simulate insert/updates and deletes on the source and monitor the replicated data through the Demo Client
2. Using the Demo Client opened in TigerVNC click on the "Initial Load Complete" Button.

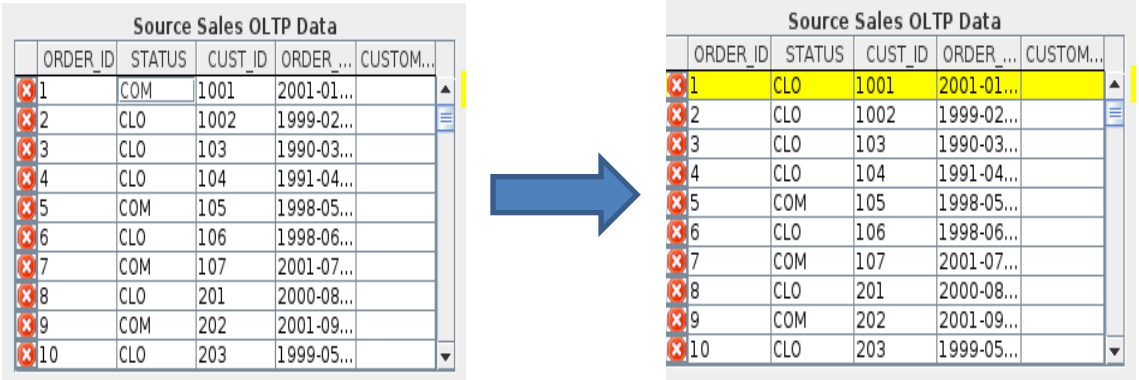


3. The following screen will appear. NOTE – the current refresh of the client is 10 seconds.

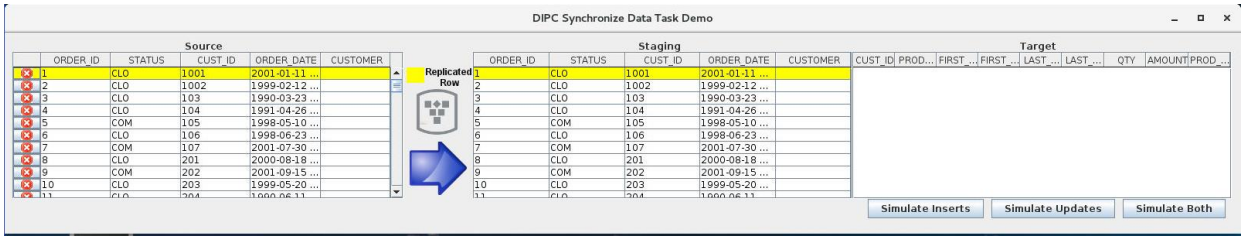


The demo client shows the source Sales OLTP data, the replicated Sales OLTP data, and the target Sales DW. As data is updated, inserted or deleted from the source the data will be automatically synchronized with the replicated schema by the Sync Sales Data Job we created in DIPC.

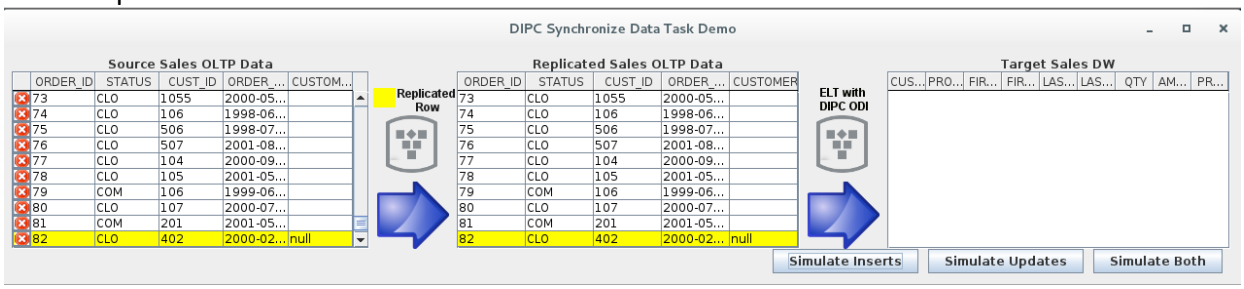
4. Perform a simple update of the source table by editing the data directly within the table grid. Update the first row's status which contains ORDER_ID=1 from COM to CLO, clicking enter will commit the update and turn the column yellow –



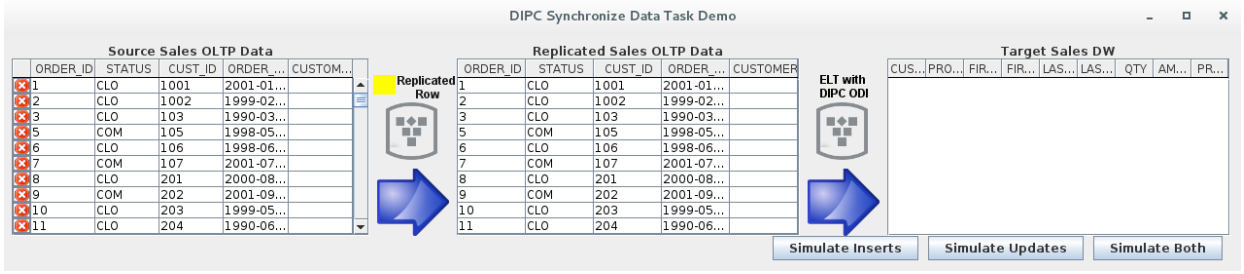
5. This row will be automatically update on the replicated schema as the DIPC Job picks up the change. The Demo Client is set to refresh at 10 seconds, so it will at least take 10 seconds for the replicated table grid to update. Once the Demo Client finds the change both rows will be updated to yellow.



6. The yellow highlights will automatically expire within the client.
7. To perform an insert click on the “Simulate Inserts” Button once.
   This will perform an insert and the demo client will scroll to the row that was inserted.

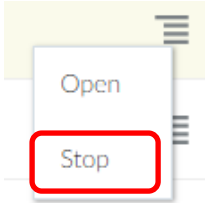


8. Depending on the refresh, the row may directly be replicated to the replicated schema or appear in the next refresh by the Demo Client.
   Both rows should show as yellow
9. To perform a delete click last row and click on the  icon. Notice this record will be delete from the replicated schema as well.

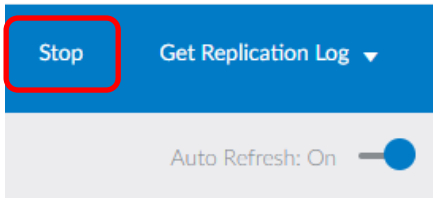


10. In Chrome the Job Details page gets updated in real-time as new data is captured and synchronized into the target

11. **[Optional]** Go to the Monitor page or Jobs Details page and click on the menu next to your Synchronize Data Job and click Stop to stop the overall process.



OR



Summary

You have now successfully completed the Hands on Lab, and have successfully performed an end-to-end data synchronization task through Oracle's Data Integration Platform Cloud.