
# Data Integration Platform Cloud Hands-on Lab

Mack Bell
Senior Product Manager

Sourav Bhattacharya
Consulting Solution Architect A-Team

Julien Testut
Senior Principal Product Manager

ORACLE

# HOL6278 - Operational Data Stores, Enterprise Data Warehouses, and Data Marts in the Cloud

The rapid adoption of enterprise cloud-based solutions brings with it a new set of challenges. However, the age-old goal of maximizing value from data does not change. Join this hands-on lab to learn how Oracle Data Integration Platform Cloud powered by Oracle Data Integrator, Oracle Enterprise Data Quality solutions, and Oracle GoldenGate ensures your data solution is built from the ground up on a foundation of best-of-breed data integration, big data integration, data governance, data management, and data automation technologies. Explore the opportunities of developing next-generation operational data stores, enterprise data warehouses, and data marts in cloud using Oracle Data Integration Platform Cloud.

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

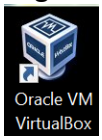
Before you begin this tutorial, you should

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

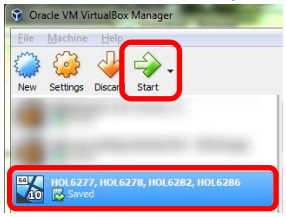
# Task 0: Preparation Steps

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.

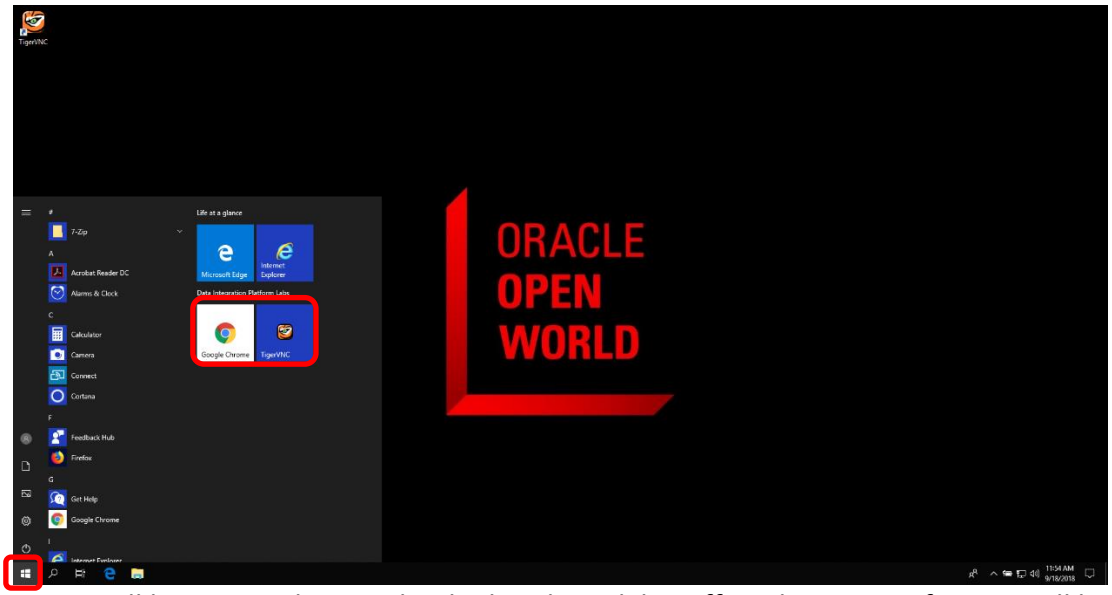
1. Log into your machine and open up Oracle VM VirtualBox from the Desktop




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

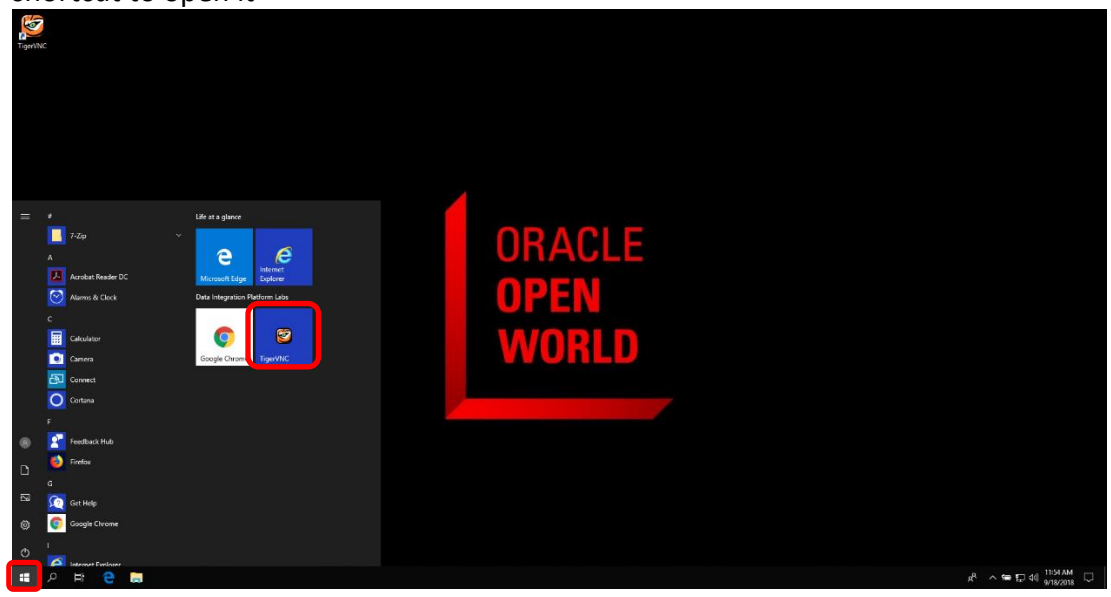



- When the VirtualBox Image is started, you will be logged into a Windows 10 machine. The Start Menu contains shortcuts to access Chrome and TigerVNC that will be used throughout this lab

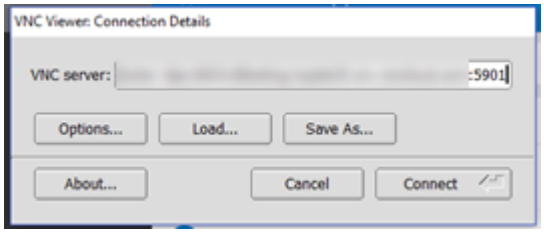



3. A URL will be assigned to you by the hands-on lab staff, make a note of it as it will be your environment for the entire lab:
    a. <hostname>:8001/dicloud (URL for Chrome)
    b. <hostname>:5901 (URL for TigerVNC)

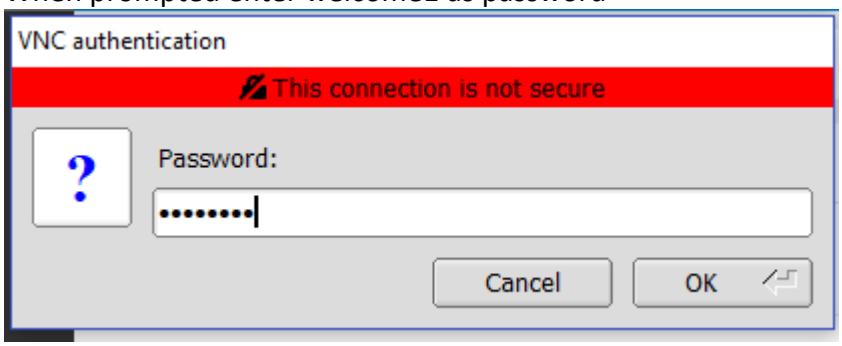
4. We will start with connecting to the DIPC host and starting a DIPC Agent.
Open the Windows 10 Start menu running in the VirtualBox VM and click on TigerVNC shortcut to open it



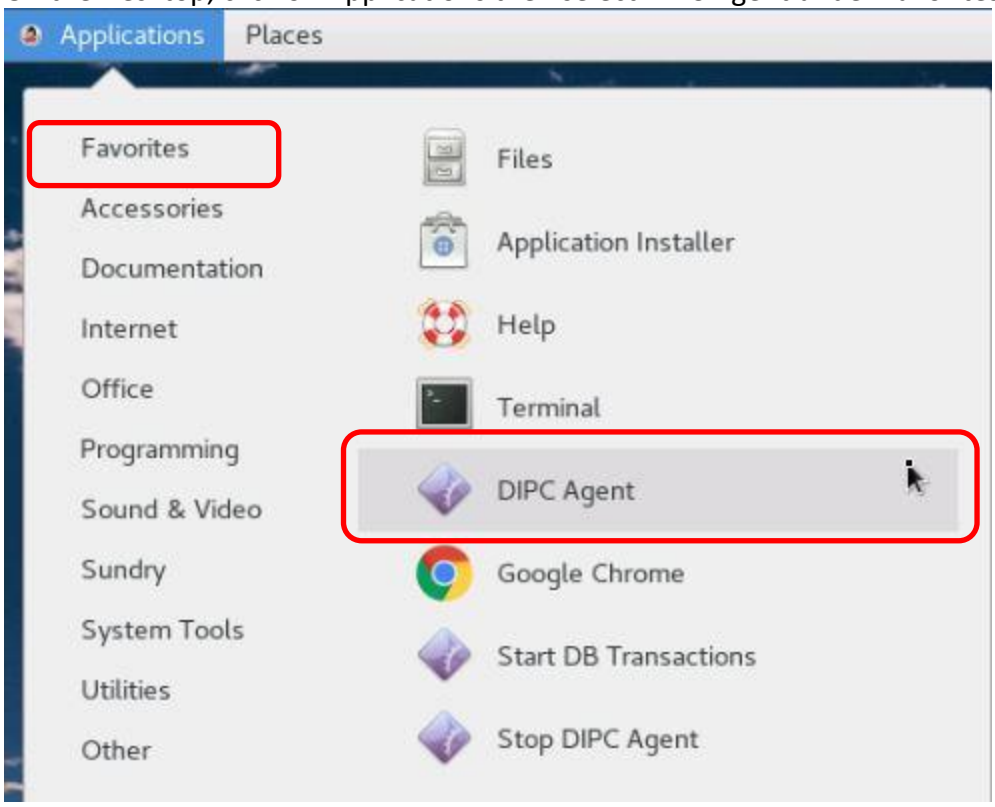

5. In the TigerVNC window enter the URL (<hostname>:5901) given to you by the hands-on lab staff and click Connect



6. When prompted enter welcome1 as password



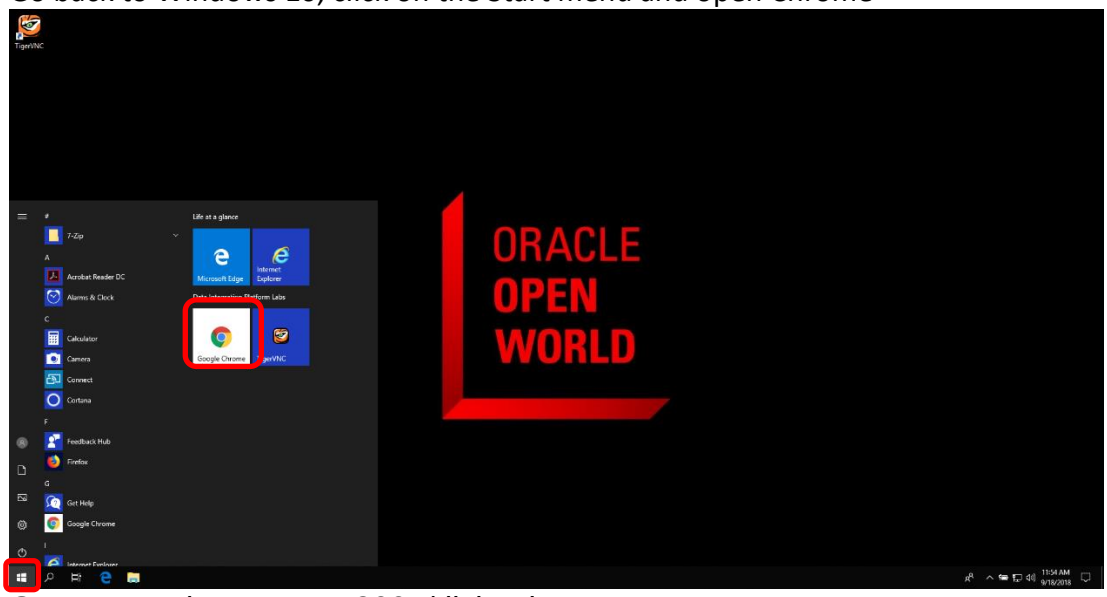

7. On the Desktop, click on Applications then select DIPC Agent under Favorites




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

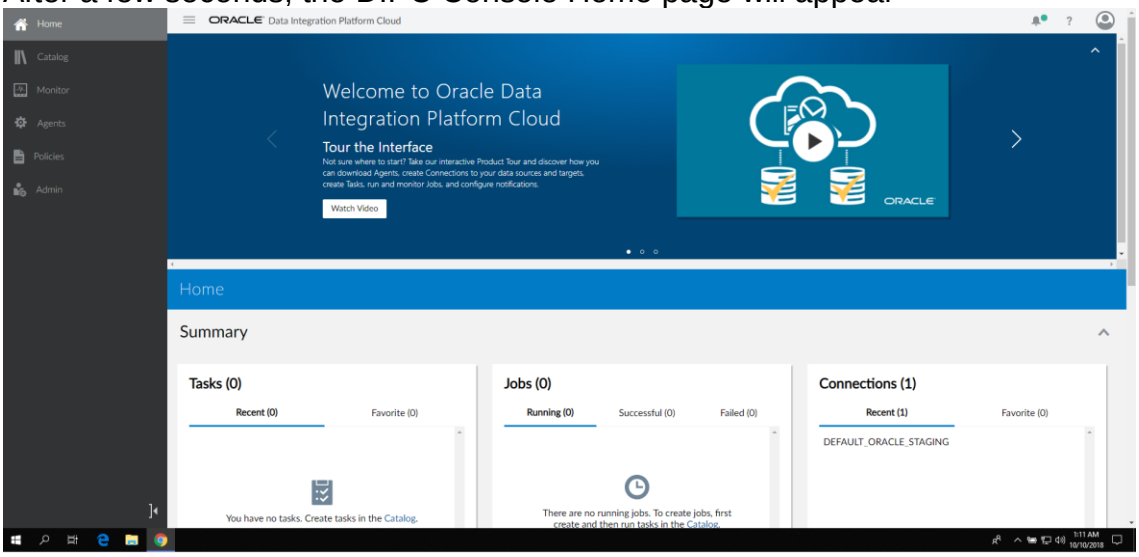
a. Go back to Windows 10, click on the Start menu and open Chrome



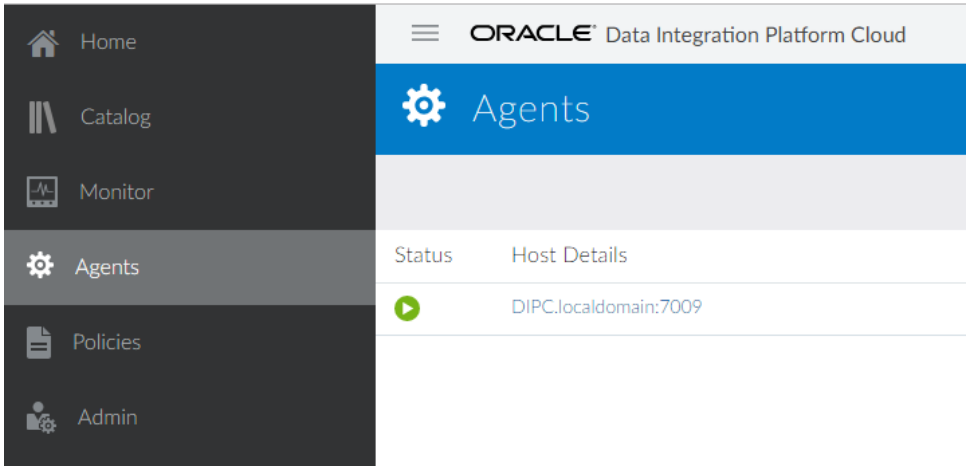

b. Go to your <hostname>:8001/dicloud


c. Login with **weblogic/welcome1**

After a few seconds, the DIPC Console Home page will appear –

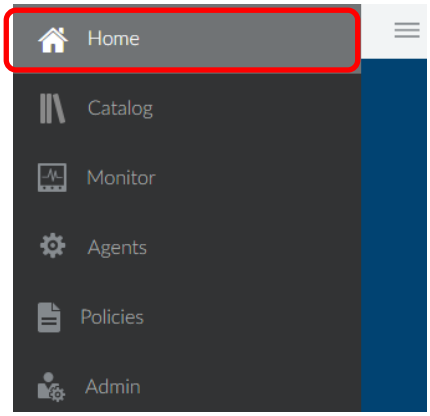



10. Click on Agents to display the list of Agents available in this environment. The Agent (DIPC.localdomain:7009) we just started will be displayed and should be in Running (green) status

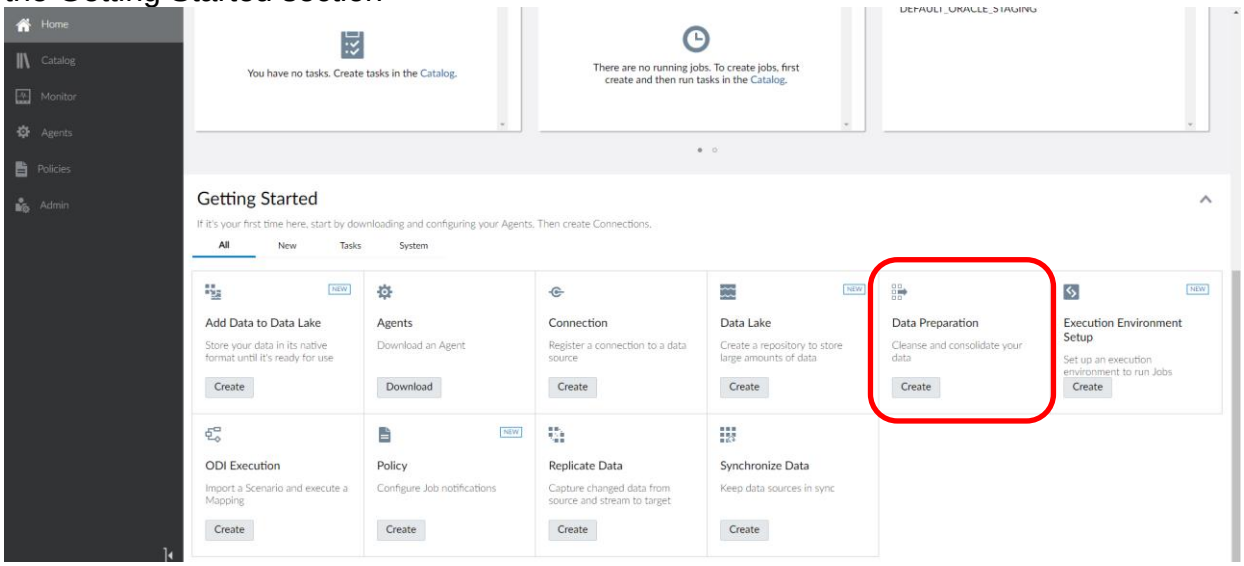



# Task 1: Prepare External Website Data

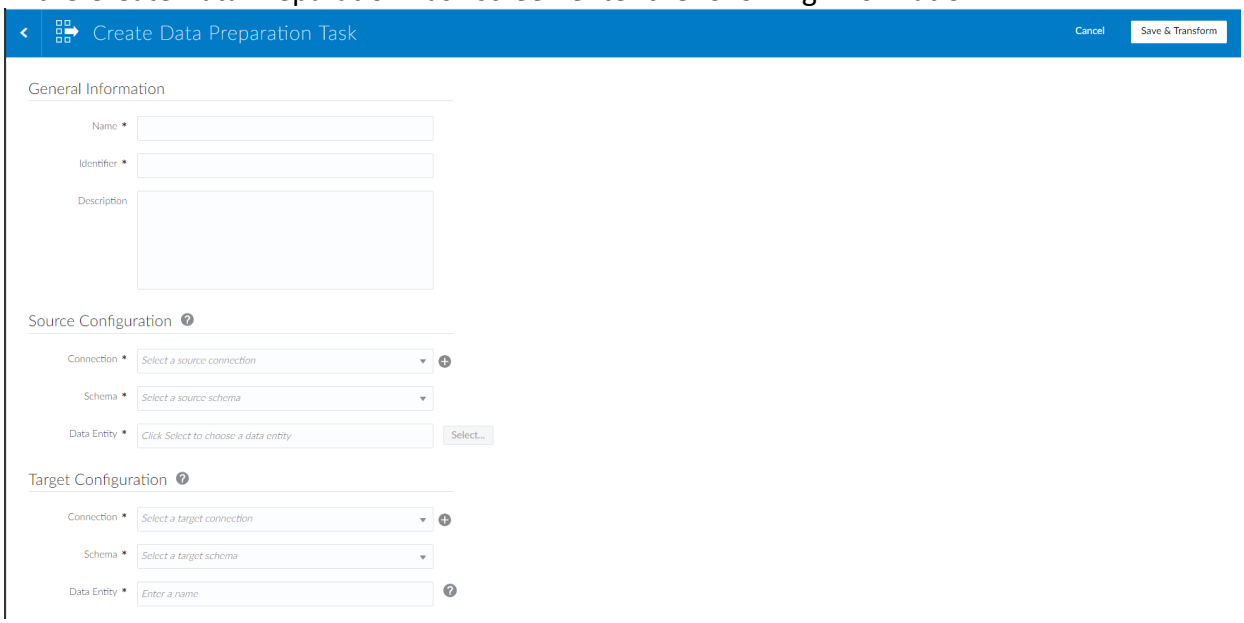
1. Click Home in the left-hand navigation bar to go back to the Home page



2. Scroll to the bottom of the page and click on Create under Data Preparation in the Getting Started section

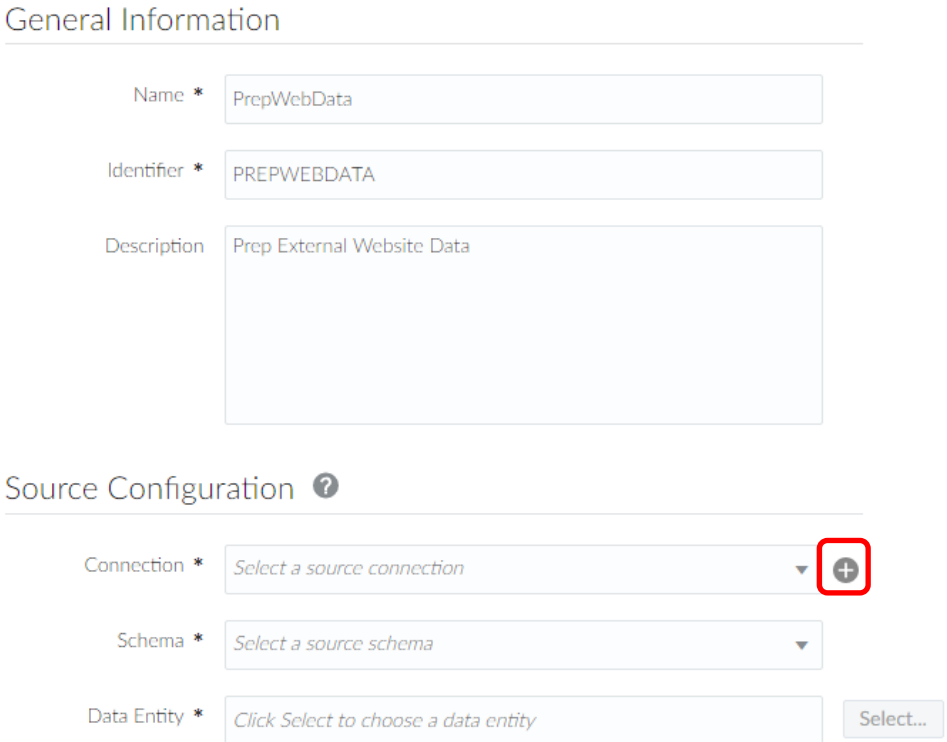


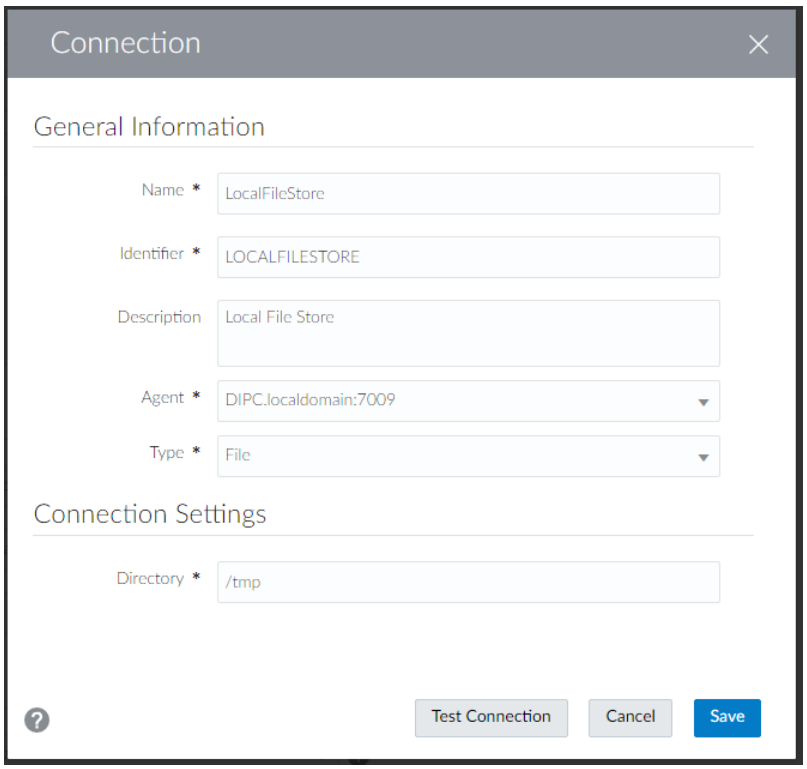

3. In the Create Data Preparation Task screen enter the following information




   a. Name: PrepWebData
   b. Description – Prep External Website Data

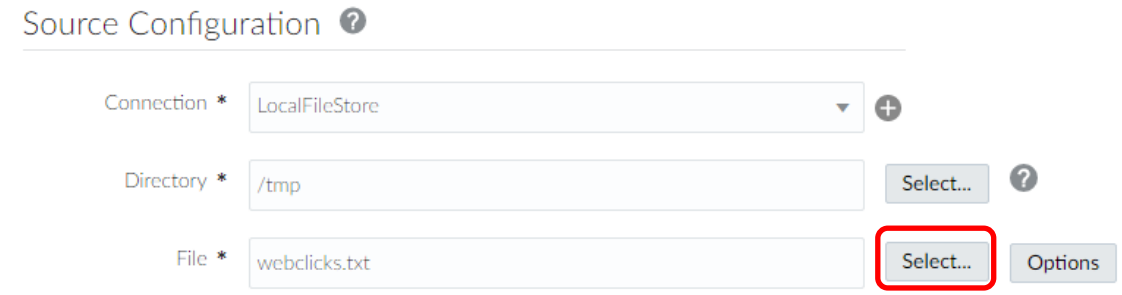
4. For Source Configuration click the add new connection button 

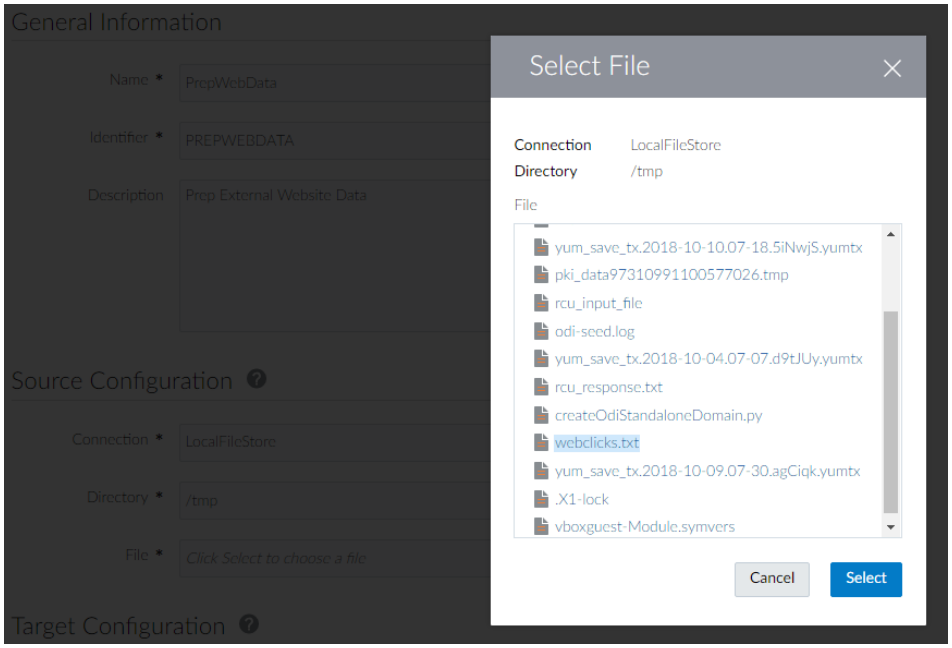


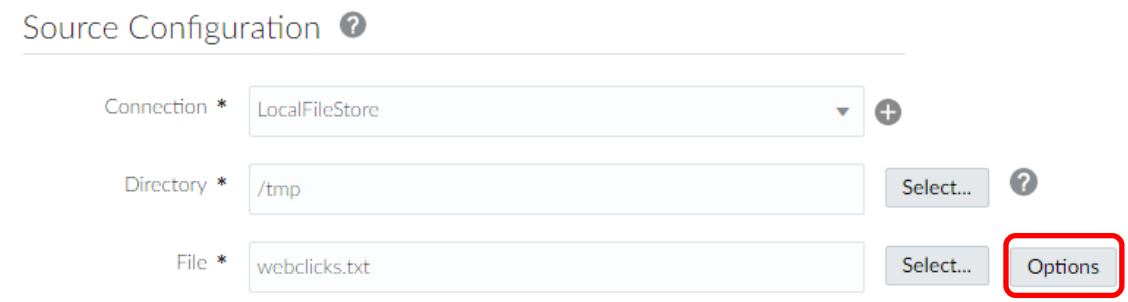

a. Name: LocalFileStore
b. Identifier: use default
c. Description: Local File Store
d. Agent: DIPC.localdomain:7009
e. Type: File
f. Directory: /tmp

5. Review Connections settings.
6. Click Test Connection – Review for Errors.
   a. Click Save if successful
7. Now click on Select… next to File, scroll down in the list and pick webclicks.txt then click Select

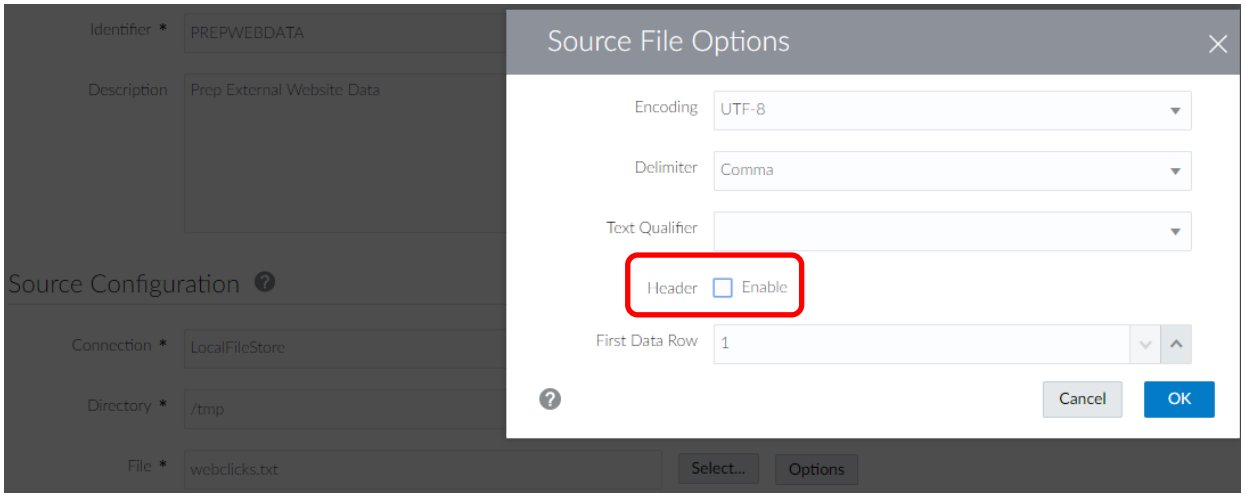


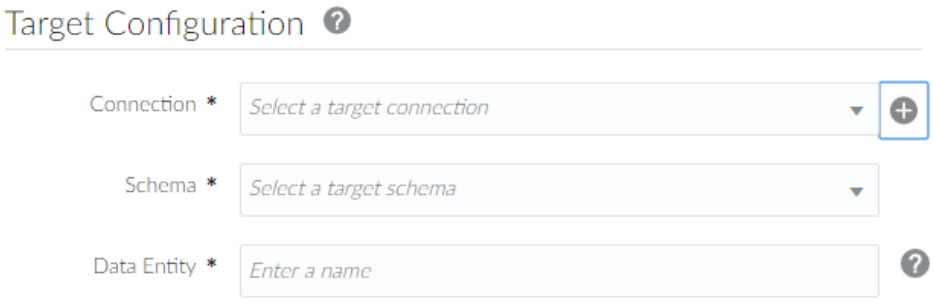

8. Next click on Options next to the File row




Uncheck Header > Enable. Click OK

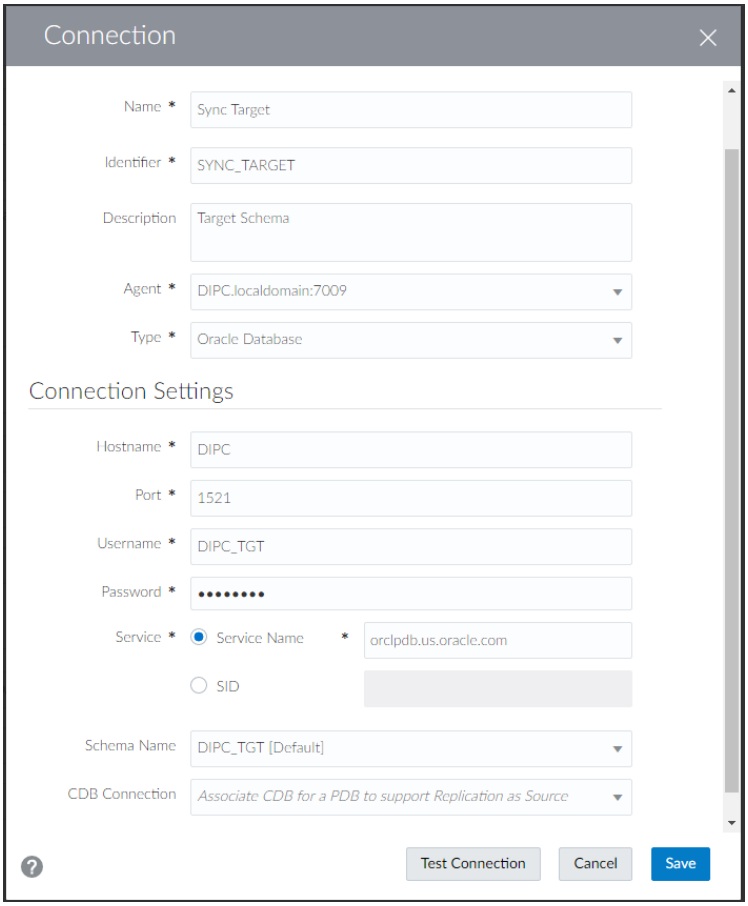



9. In Target Configuration, click on Create button to create a new Oracle Connection



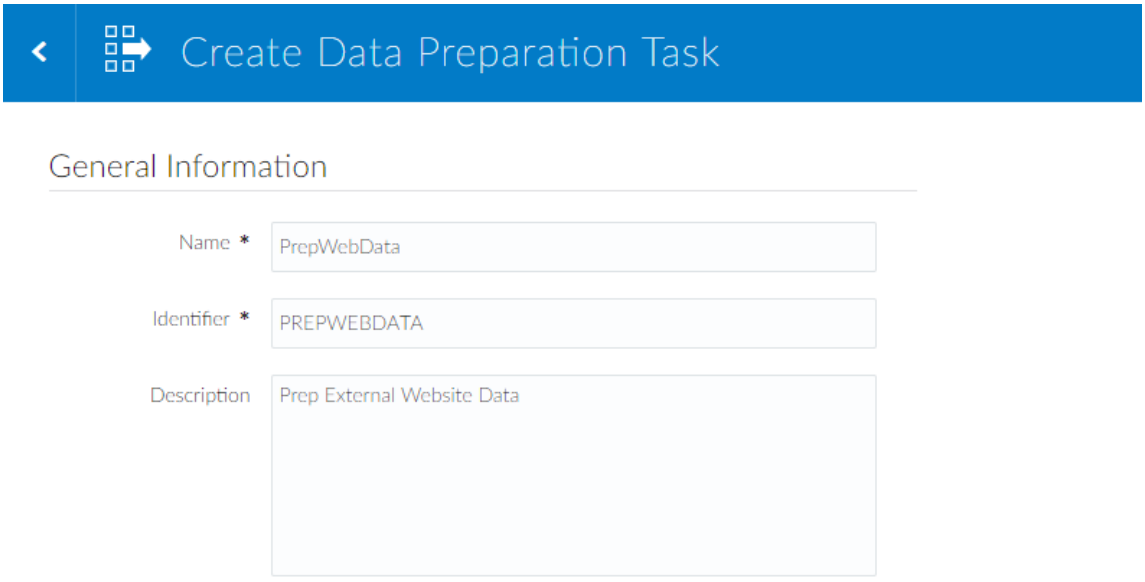

   a. Enter the following information
      - Name: Sync Target
      - Identifier: use default
      - Description: Target Schema
      - Agent: DIPC.localdomain:7009
      - Type: Oracle Database
      - Hostname: DIPC
      - Port: 1521
      - Username: DIPC_TGT
      - Password: welcome1
      - Service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

- CDB Connection: leave blank



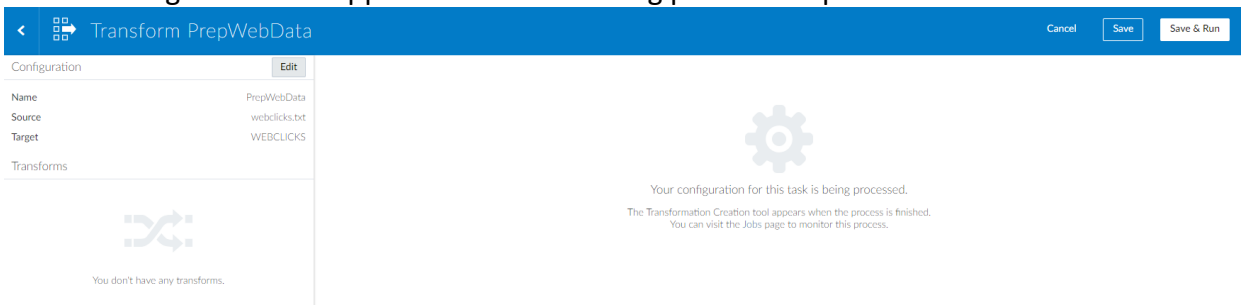

b. Click Test Connection, Review for errors
c. Click Save

10. Enter WEBCLICKS next to Data Entity under Target Configuration
    - Connection: Sync Target
    - Schema: DIPC_TGT
    - Data Entity: WEBCLICKS

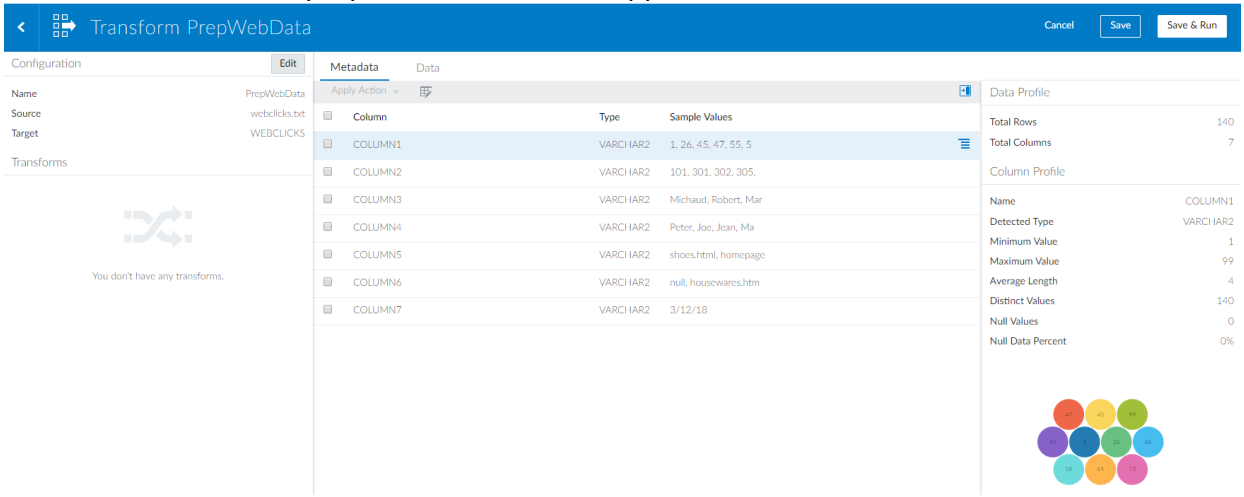



11. Click Save & Transform

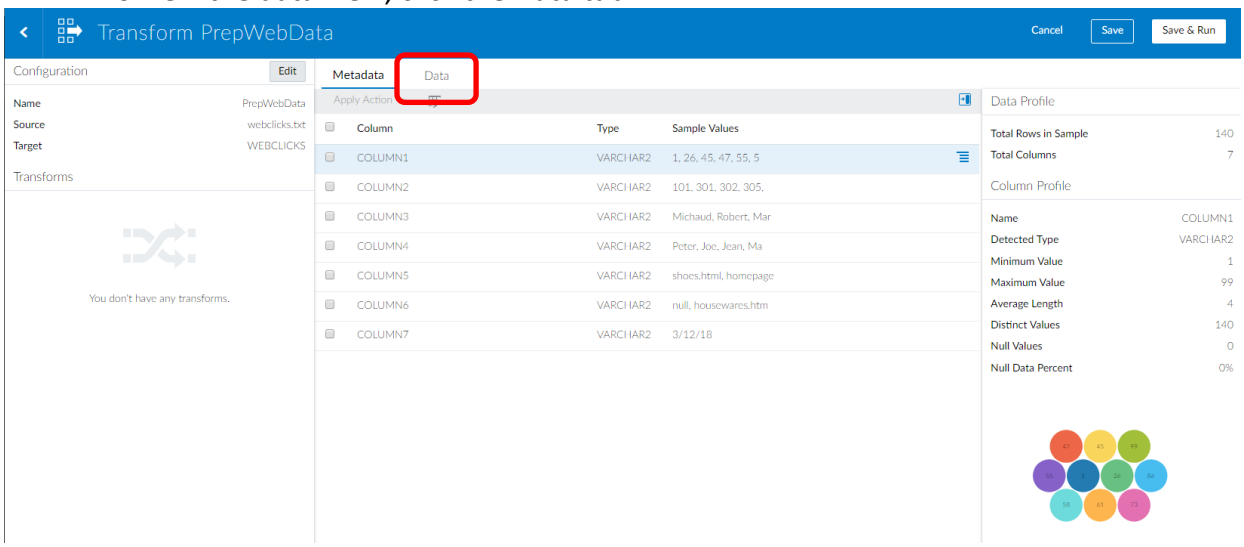

12. The following screen will appear as the file is being parsed and profiled



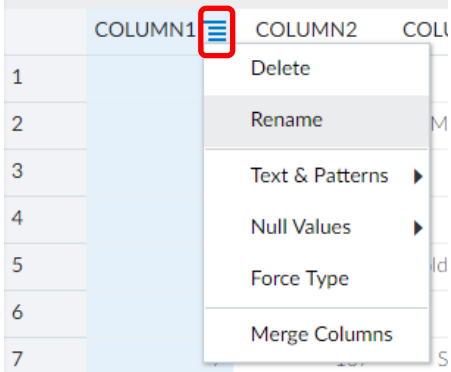
13. Once finished the data preparation screen will appear




14. Begin prepping/transforming the file
    a. The profiling process has captured advanced profiling information as the flat file was ingested. Click each column to review the profiling results in the right hand data profile drawer.
    b. There are also two views in which data can be prepared and transformed. To view the data view, click the Data tab

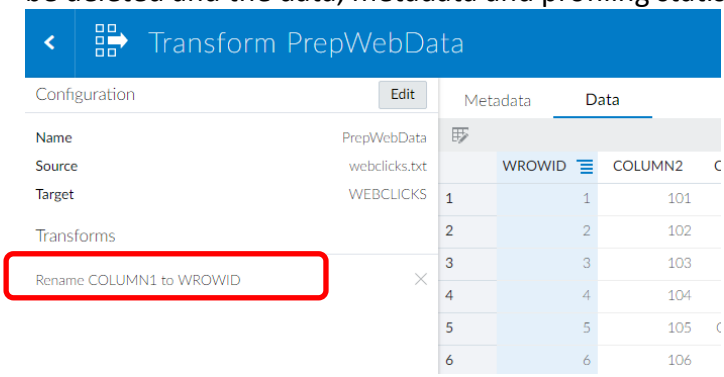



15. To transform data click the menu bar on the column

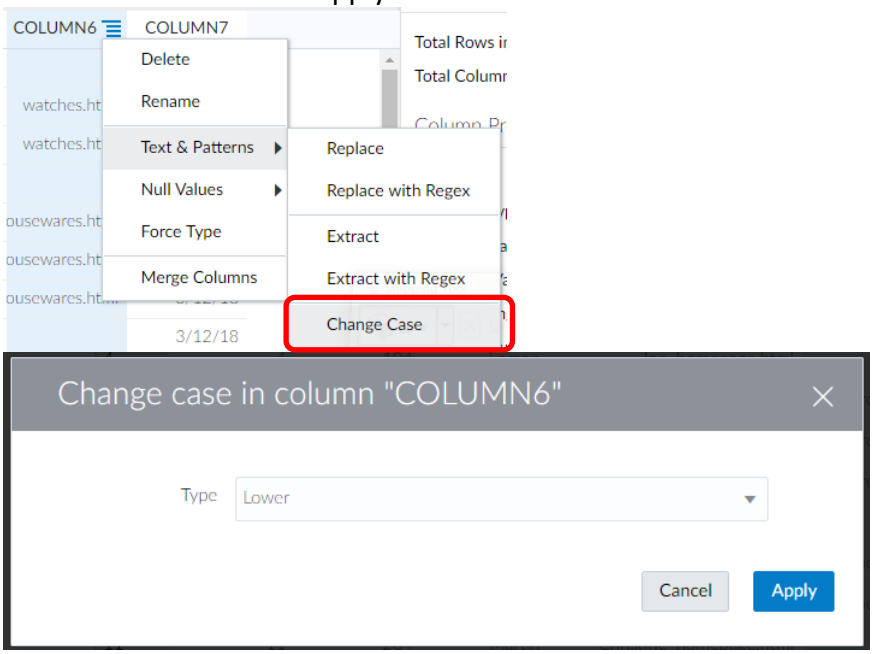



16. For Column 1 rename the column into WROWID

17. Click Apply. Clicking apply updates, the data, metadata and profiling statistics. Also note the transform is saved and displayed in the left-hand drawer. This transform can be deleted and the data, metadata and profiling statistics will be updated accordingly.

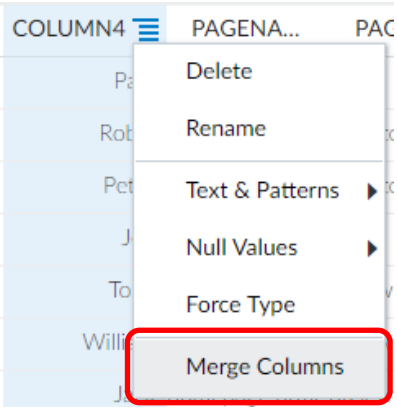



18. Add the following column transforms
    a. Column 2: Rename into CUSTID
    b. Column 5: Rename into PAGENAME
    c. Column 6: transform Shoes.html to shoes.html using Text & Patterns > Change Case to Lower and click Apply

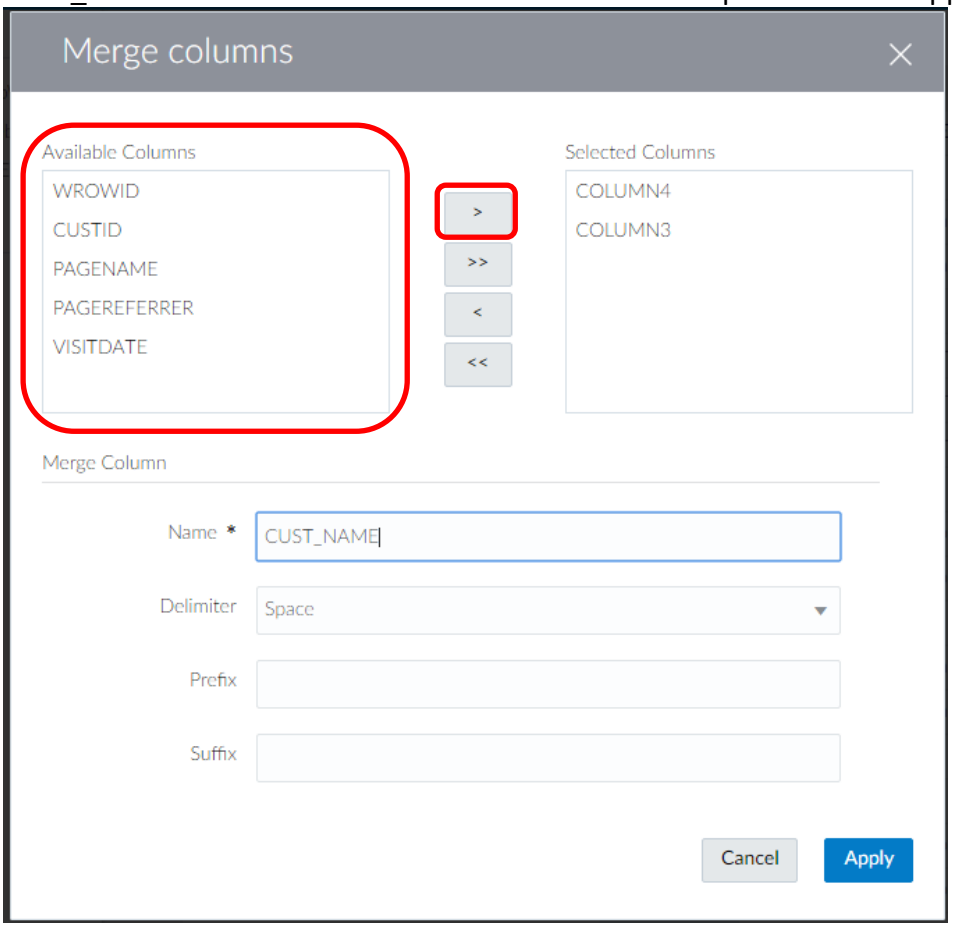



Then rename COLUMN6 into PAGEREFERRER

    d. Column 7: Rename into VISITDATE
    e. Merge COLUMN3 and COLUMN4 to create a new CUST_NAME column
    Click on COLUMN4 menu and select Merge Columns

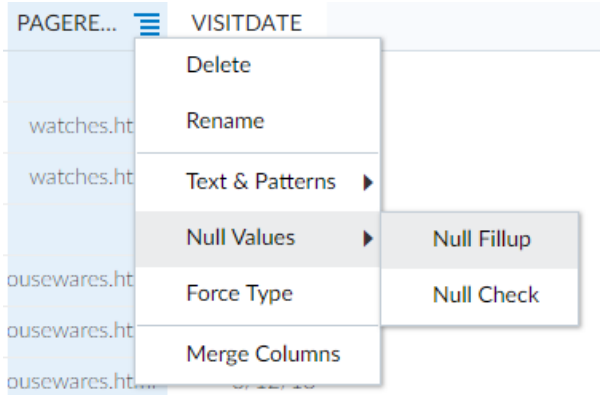



Select COLUMN3 under Available Columns on the left and click on > then enter CUST_NAME in the Name field. Leave Delimiter set to Space and click Apply

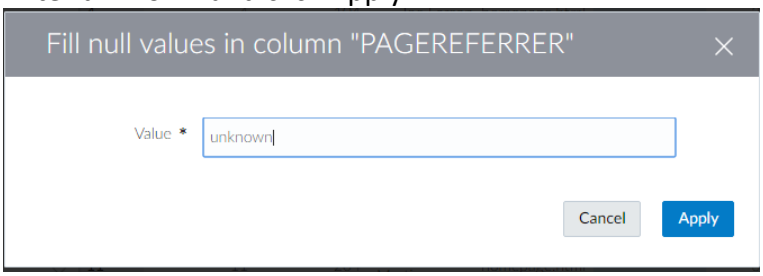



f. Delete COLUMN3 and COLUMN4 (click menu item and select Delete then Yes, Delete)

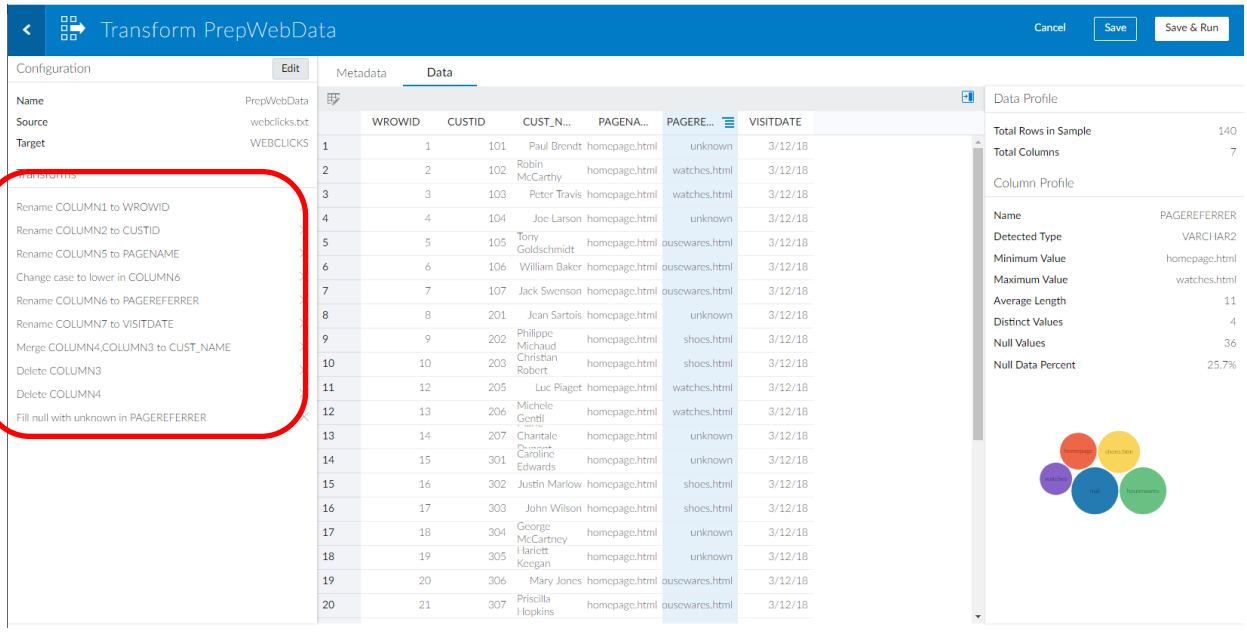
g. PAGEREFERRER: Click on menu then select Null Values > Null Fillup




Enter unknown and click Apply

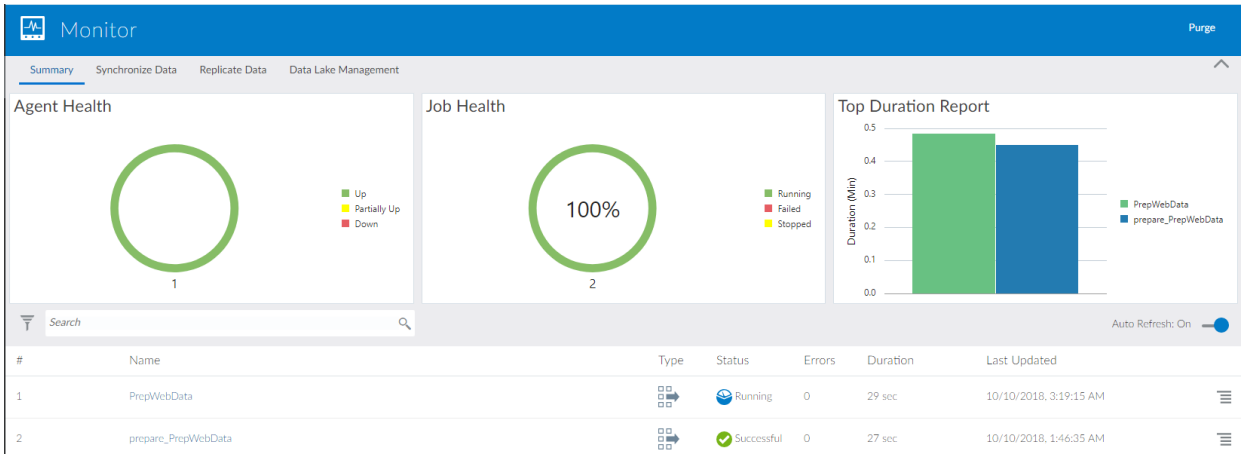


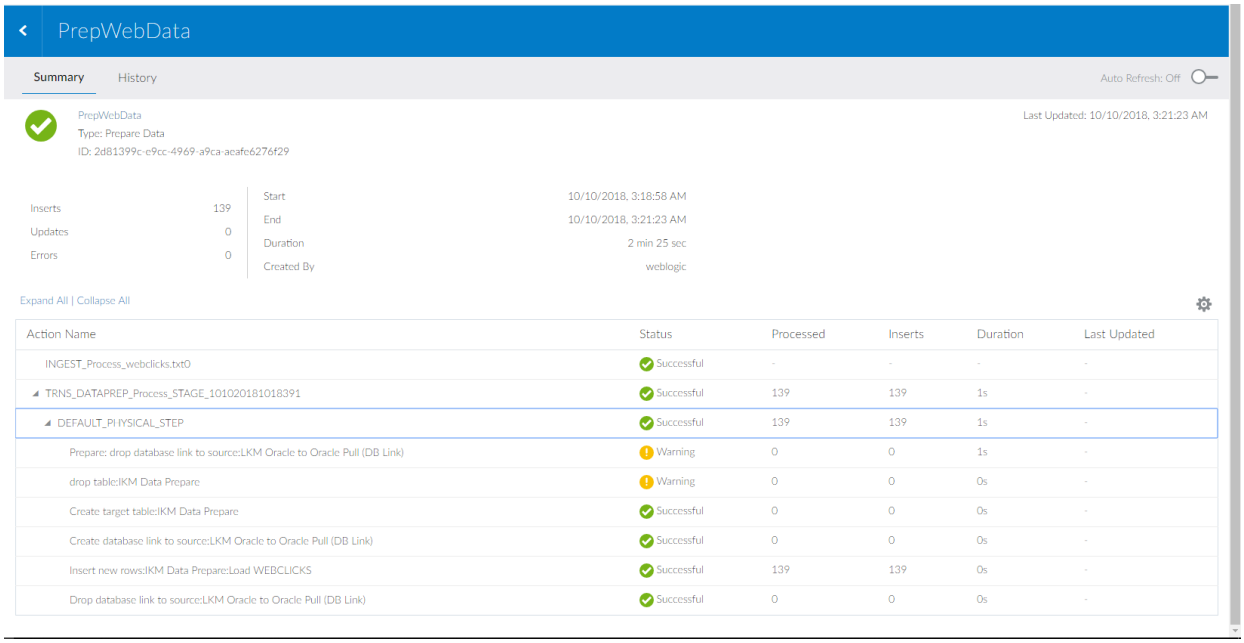

19. Review Transforms and Data



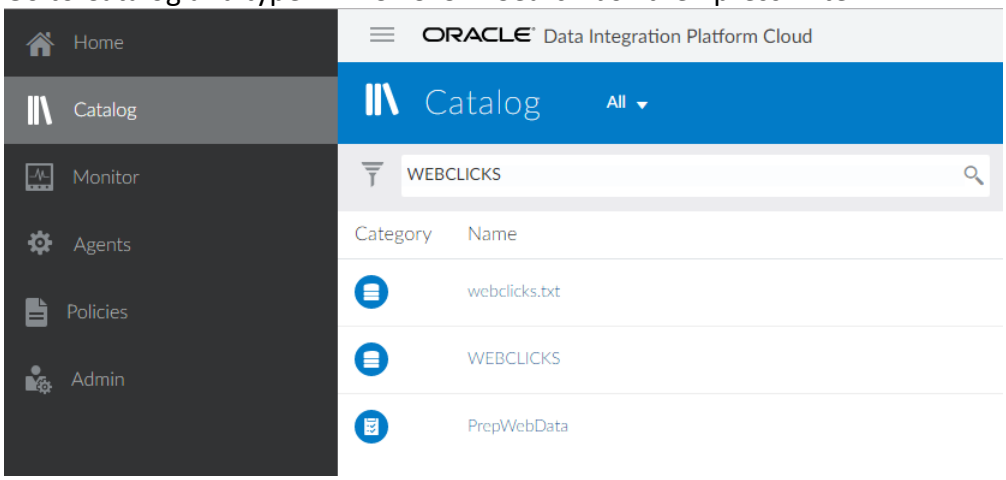

20. Click Save & Run


21. Review Job Execution




22. [Optional] Go to Catalog and use Search functionality to find WEBCLICKS and review the prepared data.
    Go to Catalog and type WEBCLICKS in Search box then press Enter

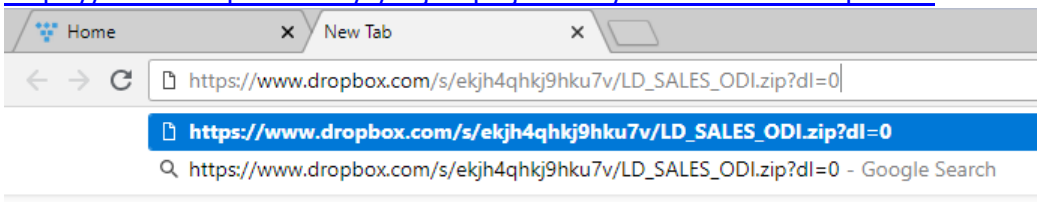


23. [Optional] Click on WEBCLICKS to open it up then on Data tab to view the data prepared by DIPC

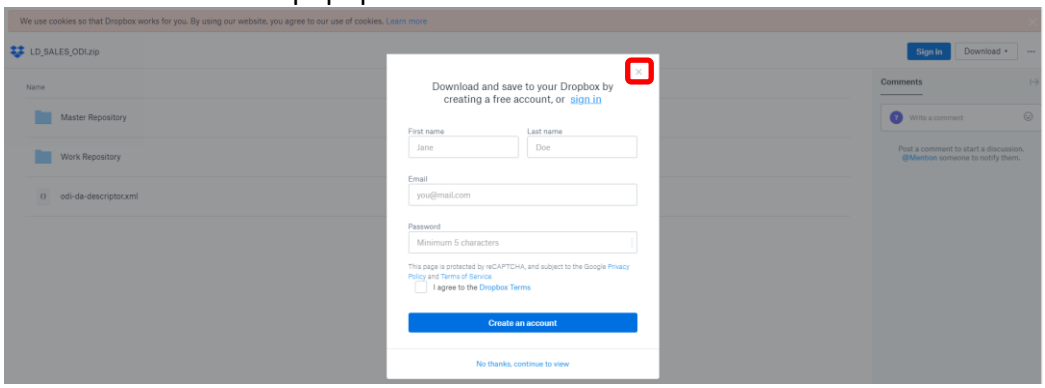

| WROWID | CUSTID | CUST_NAME | PAGENAME | PAGEREFERRER | VISITDATE |
|---|---|---|---|---|---|
| 2 | 102 | Robin McCarthy | homepage.html | watches.html | 3/12/18 |
| 3 | 103 | Peter Travis | homepage.html | watches.html | 3/12/18 |
| 4 | 104 | Joe Larson | homepage.html | unknown | 3/12/18 |
| 5 | 105 | Tony Goldschmidt | homepage.html | housewares.html | 3/12/18 |
| 6 | 106 | William Baker | homepage.html | housewares.html | 3/12/18 |
| 7 | 107 | Jack Swenson | homepage.html | housewares.html | 3/12/18 |
| 8 | 201 | Jean Sartois | homepage.html | unknown | 3/12/18 |
| 9 | 202 | Philippe Michaud | homepage.html | shoes.html | 3/12/18 |

# Task 2: Create ODI Execution Task

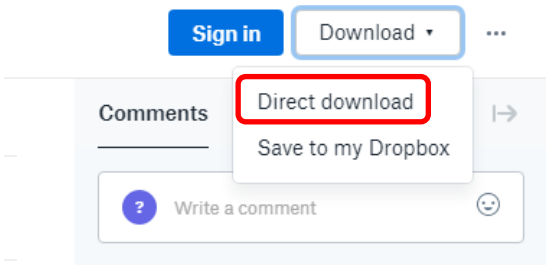
1. First, we will download a deployment archive previously created in Oracle Data Integrator (ODI).
   Open up a new tab in Chrome and go to
   https://www.dropbox.com/s/ekjh4qhkj9hku7v/LD_SALES_ODI.zip?dl=0



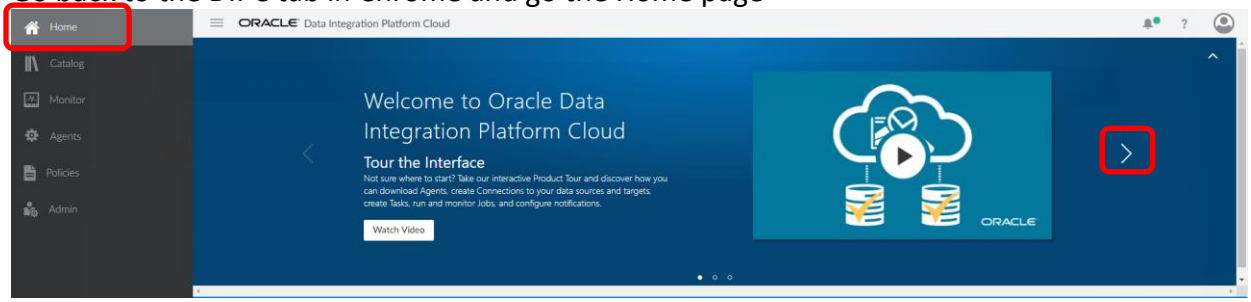

Click on x to close the popup



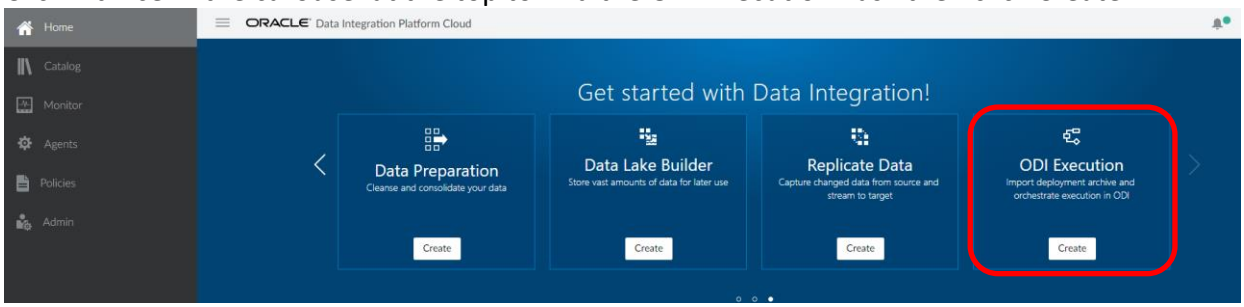

Then click Download and select Direct Download



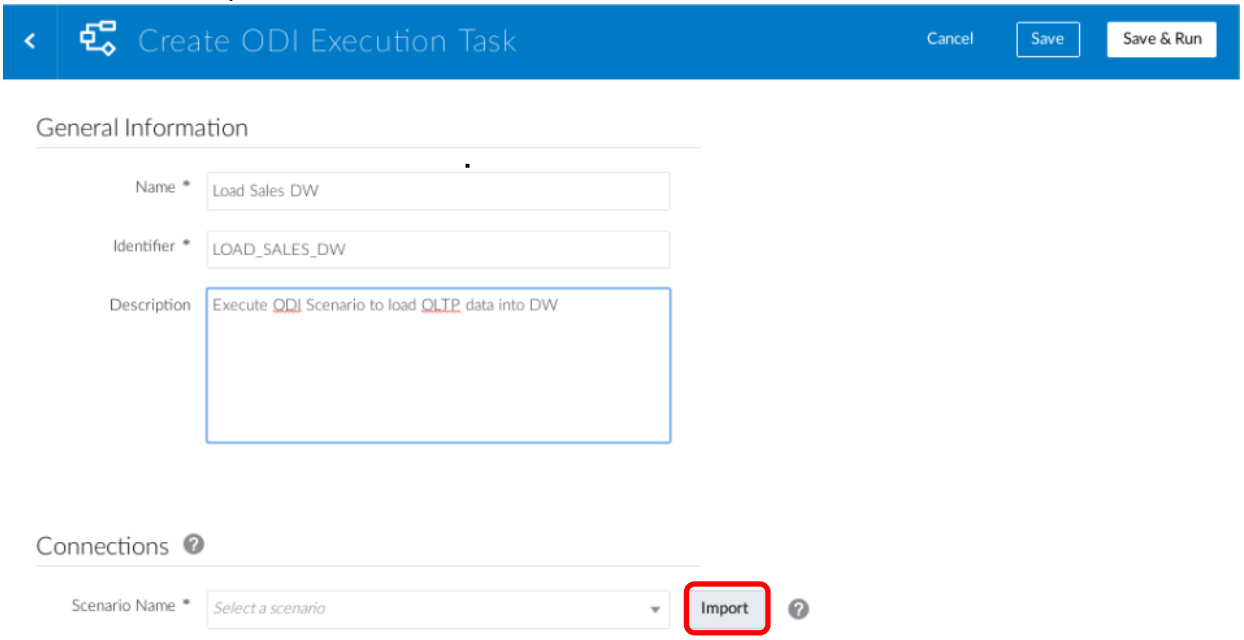
2. Go back to the DIPC tab in Chrome and go the Home page




3. Click > twice in the carousel at the top to find the ODI Execution Task then click Create

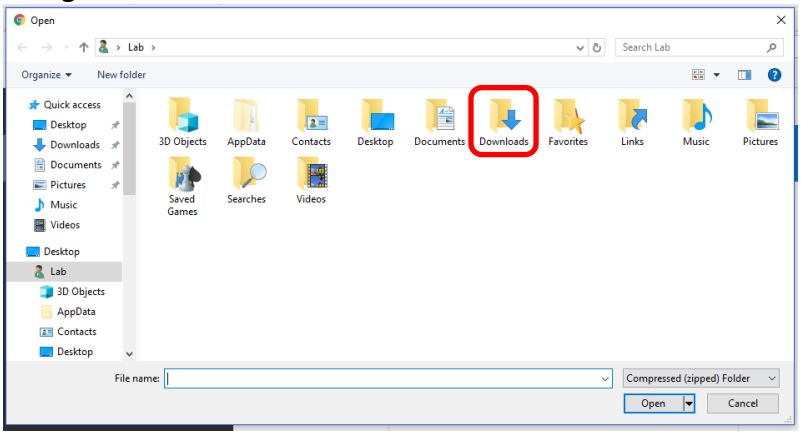



4. Enter
   - Name: Load Sales DW
   - Description: Execute ODI Scenario to load OLTP data into DW

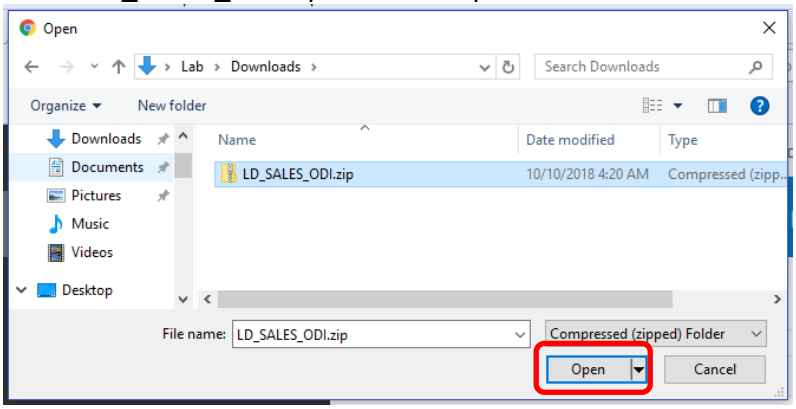



5. Under Connections click on Import to import a deployment archive created in ODI Studio that contains the Scenario we want to execute

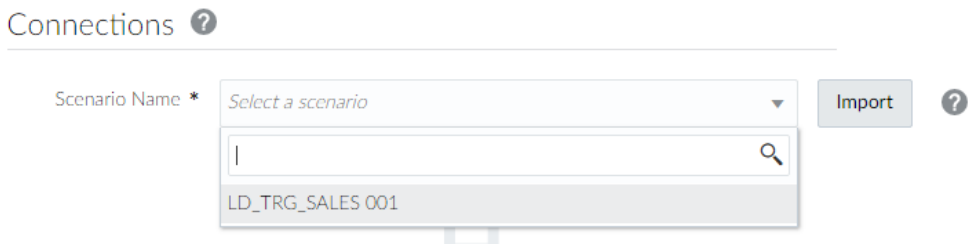
a. Navigate to Downloads




b. Select LD_SALES_ODI.zip and click Open

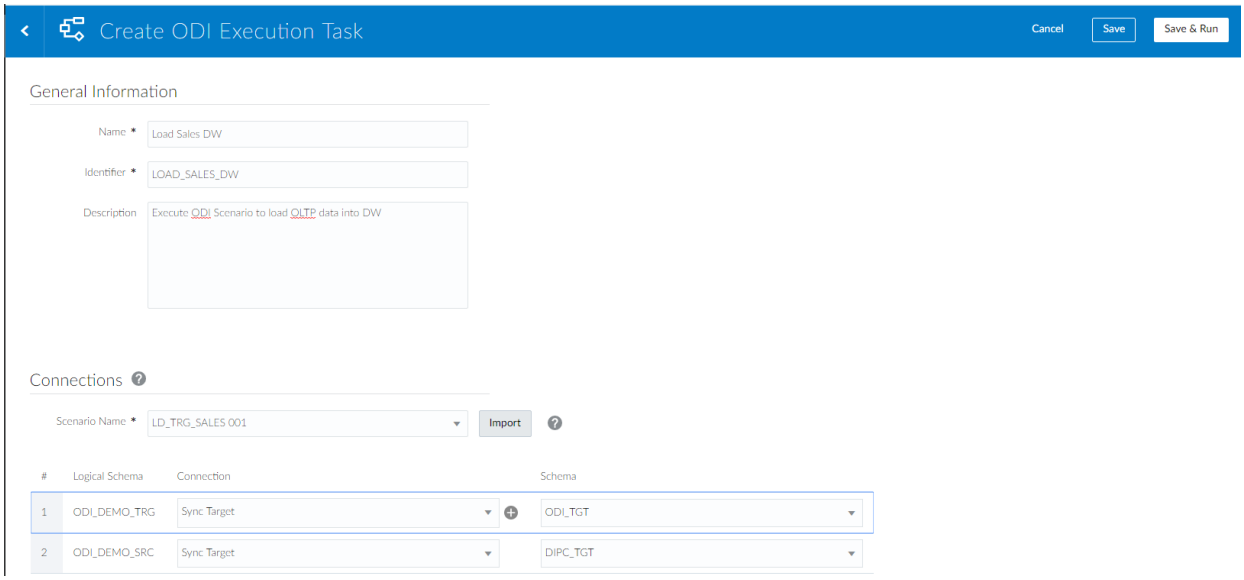



c. Click on the Scenario Name drop-down and select LD_TRG_SALES 001




This scenario joins SRC_ORDERS and SRC_ORDER_LINES, aggregates the data, filters for ORDERS with Status of 'CLO' as well as performs an incremental update (insert new rows or update existing rows when needed).
So only rows that have a status of 'CLO' (closed), will be loaded to the target Sales DW.

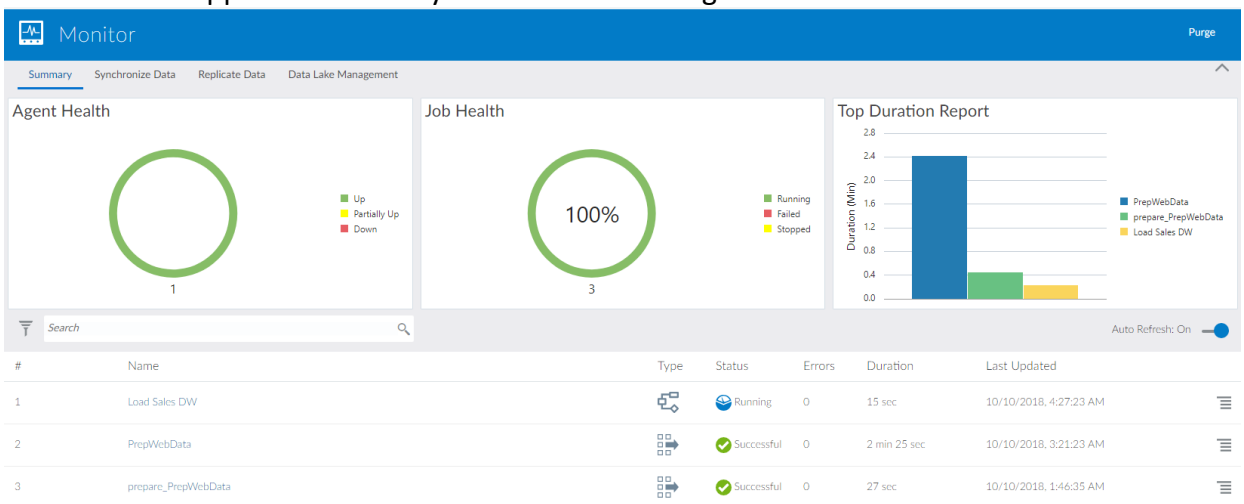
6. In the Connection table pick the following Connections and Schemas:
   a. ODI_DEMO_TRG:
      i. Connection: Sync Target
      ii. Schema: ODI_TGT
   b. ODI_DEMO_SRC
      i. Connection: Sync Target
      ii. Schema: DIPC_TGT



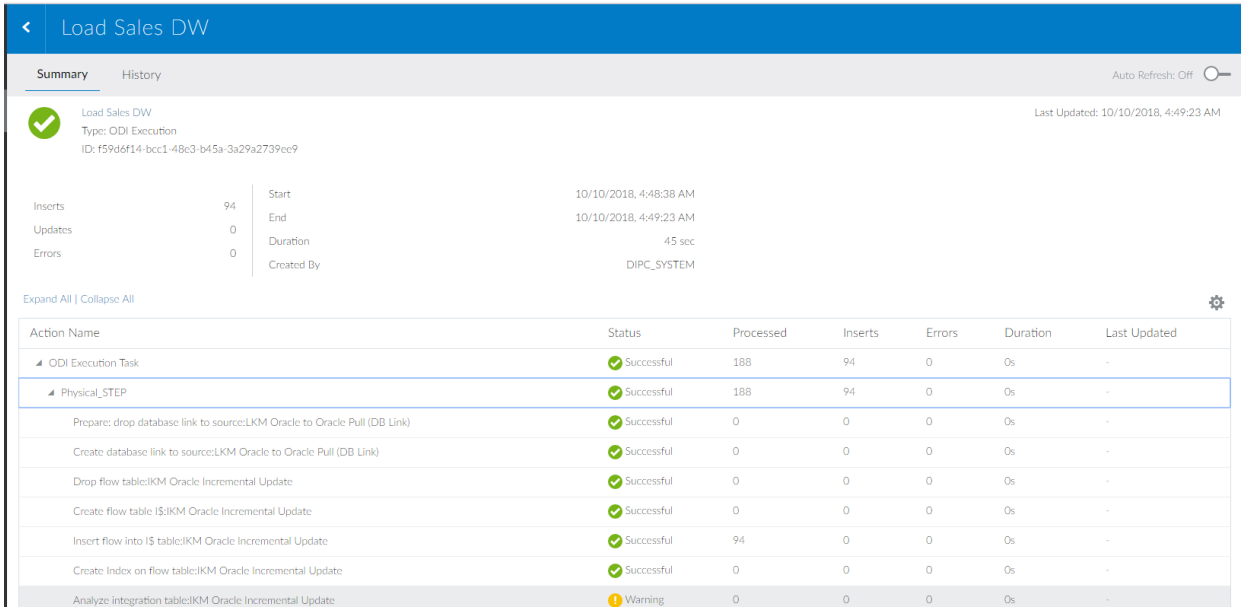

7. Click on Save & Run to execute the Task


8. You will be redirected to the Jobs page and you will see a notification that a new Job execution started

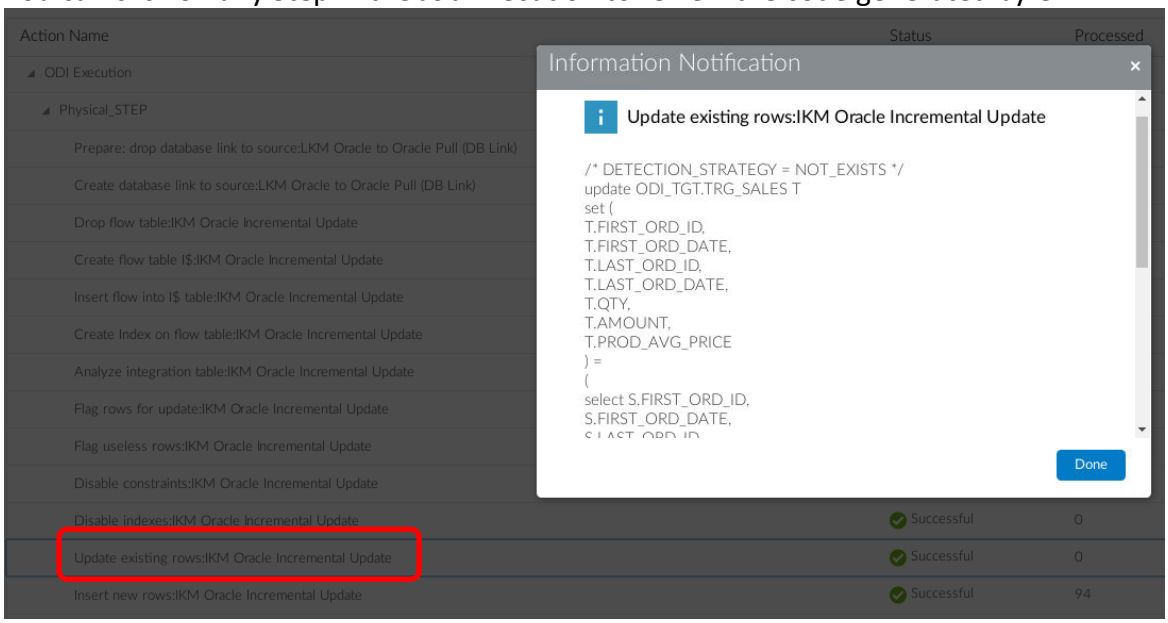
9. When the Job appears in the list you can click on it to get more details



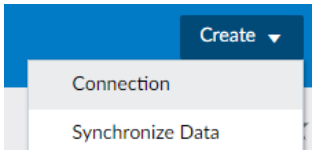
10. The Job Details contains all the details about the ODI scenario execution




You can click on any Step in the Job Execution to review the code generated by ODI

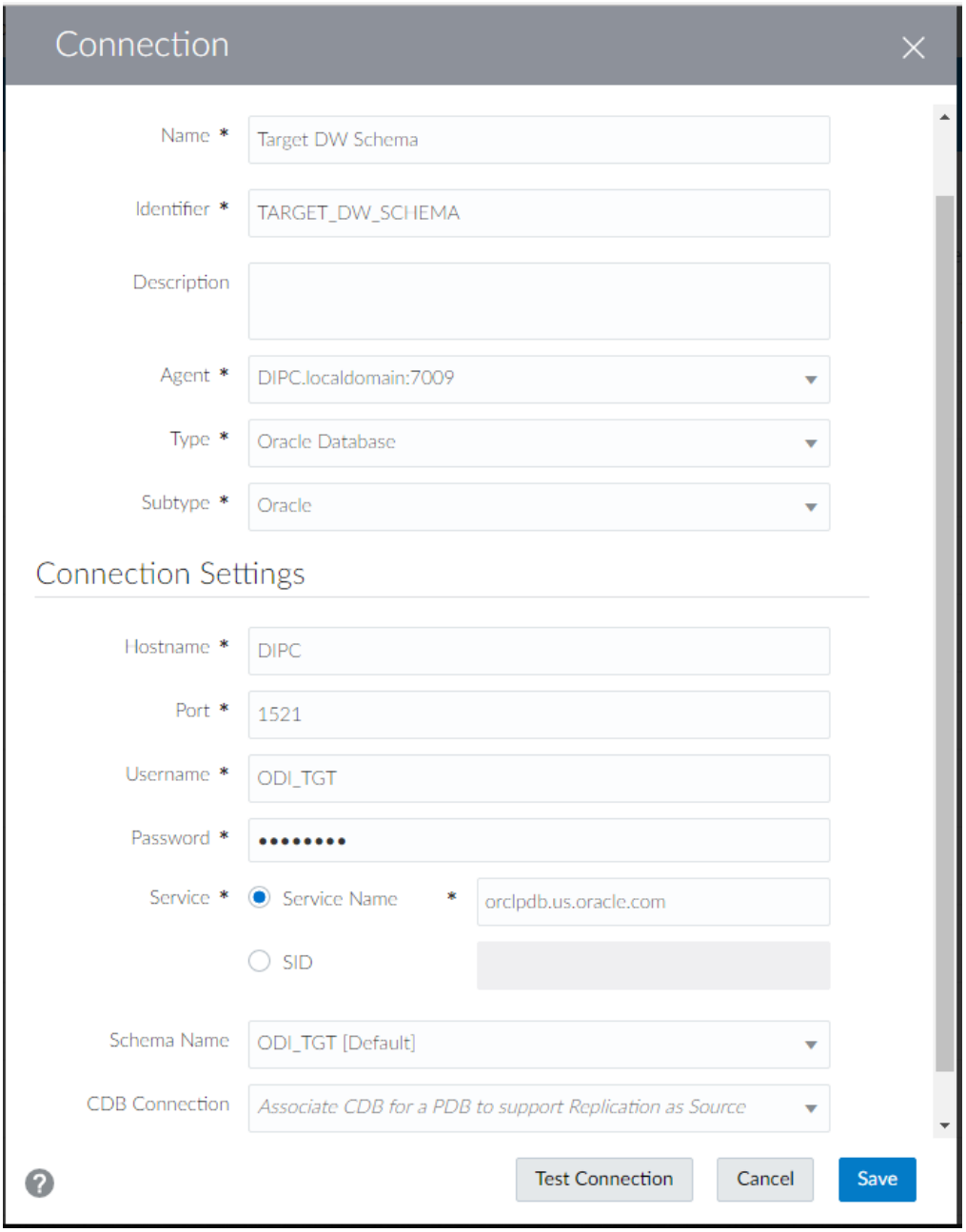



11. [Optional] Create a connection to ODI_TGT schema. Go to Catalog and click Create > Connection



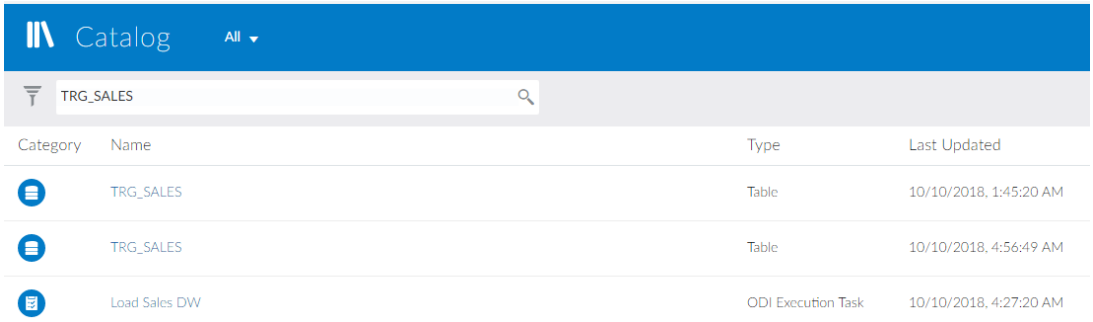

12. [Optional] Enter the following information:
    - Name: Target DW Schema
    - Identifier: use default
    - Description: Target DW Schema
    - Agent: DIPC.localdomain:7009
    - Type: Oracle Database

- Subtype: Oracle
- Hostname: DIPC
- Port: 1521
- Username: ODI_TGT
- Password: welcome1
- Service Name: orclpdb.us.oracle.com
- Schema: ODI_TGT
- CDB Connection: leave blank

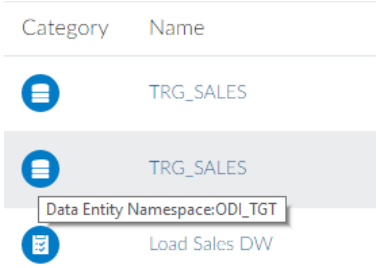



13. [Optional] Click Test Connection and Save

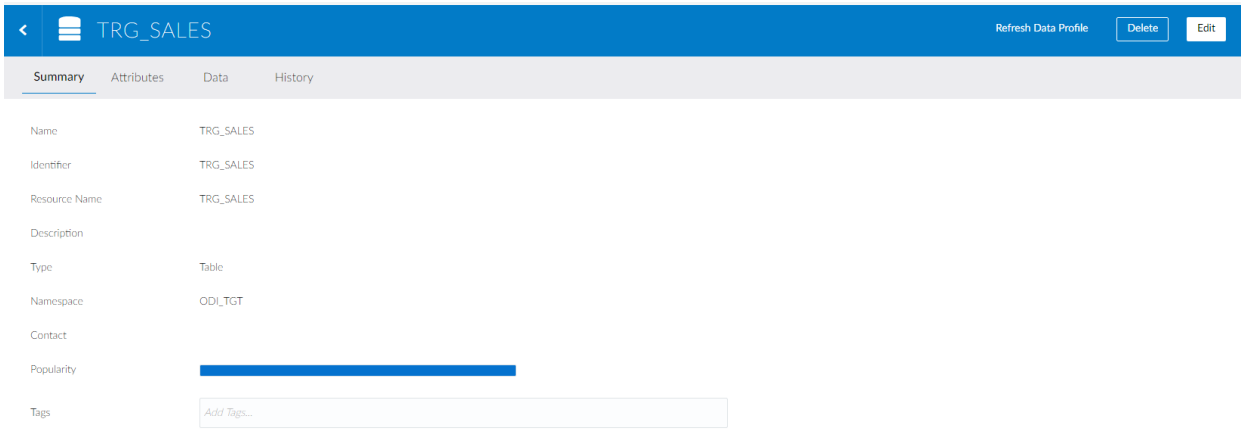
14. [Optional] Wait a minute or so and you should see new Data Entities appearing in the Catalog. Type TRG_SALES in the Search box and press Enter



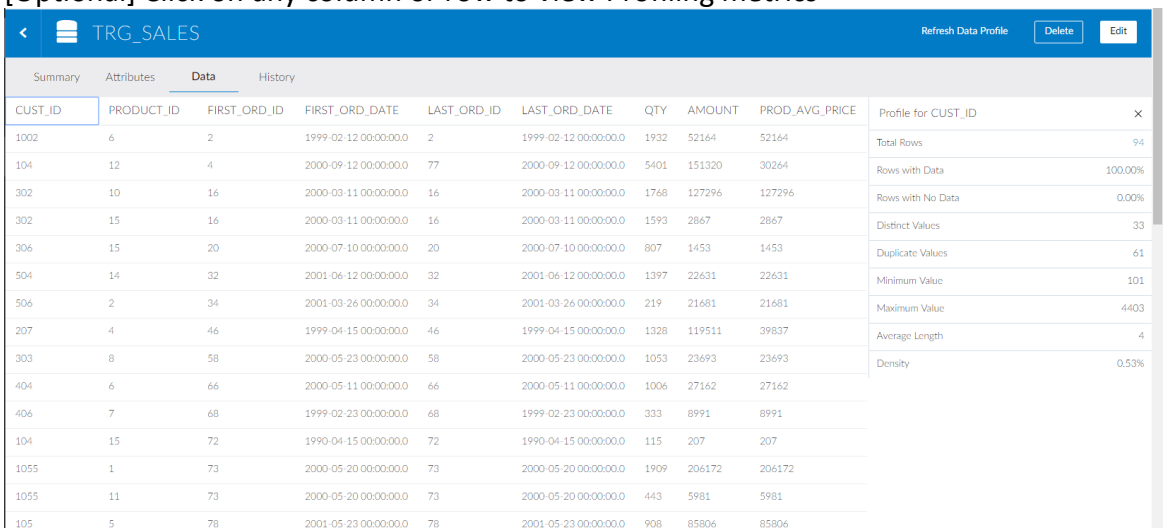

15. [Optional] The Search functionality returns all DIPC objects that are related to TRG_SALES. Click on the 2nd entry that displays Data Entity Namespace: ODI_TGT when hovering your mouse over the Data Entity icon




16. [Optional] Click on Data tab to view the data loaded by the ODI Execution Task




17. [Optional] Click on any column or row to view Profiling metrics



## Summary

You have now successfully completed the Hands on Lab, and have successfully prepared, loaded and transformed data through Oracle's Data Integration Platform Cloud.