



## Data Integration Platform Cloud Hands-on Lab

<Name>  
<Title>

<Name>  
<Title>

## Hands on Lab - Data Integration Platform Cloud – Data Preparation

The rapid adoption of enterprise cloud-based solutions brings with it a new set of challenges. Data Integration becomes one of the greatest challenges of any enterprise cloud-based solution. Join this hands-on lab to have firsthand experience of the power and simplicity of Oracle Data Integration Platform Cloud. See how DIPC allows you to prepare data for downstream consumption in just a few of clicks.

The following lessons will walk us through various steps that are needed to create Data Integration Platform Cloud connections, a Data Preparation task and job to prepare and load data into a target data warehouse

HANDS ON LAB - DATA INTEGRATION PLATFORM CLOUD – DATA PREPARATION .....	2
OVERVIEW .....	3
<i>Time to Complete</i> .....	3
<i>Prerequisites</i> .....	3
<i>Lab Environment</i> .....	3
TASK 0: PREPARATION STEPS.....	4
TASK 1: PREPARE EXTERNAL WEBSITE DATA.....	7
SUMMARY .....	18

## Overview

### Time to Complete

Perform all tasks – 20 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

### Lab Environment

For this lab, the Data Integration Platform Cloud and the client environment are contained within an environment for simplicity. Most user interactions with Data Integration Platform Cloud will be through a browser installed on your local machine (Chrome preferred, Firefox is also supported).

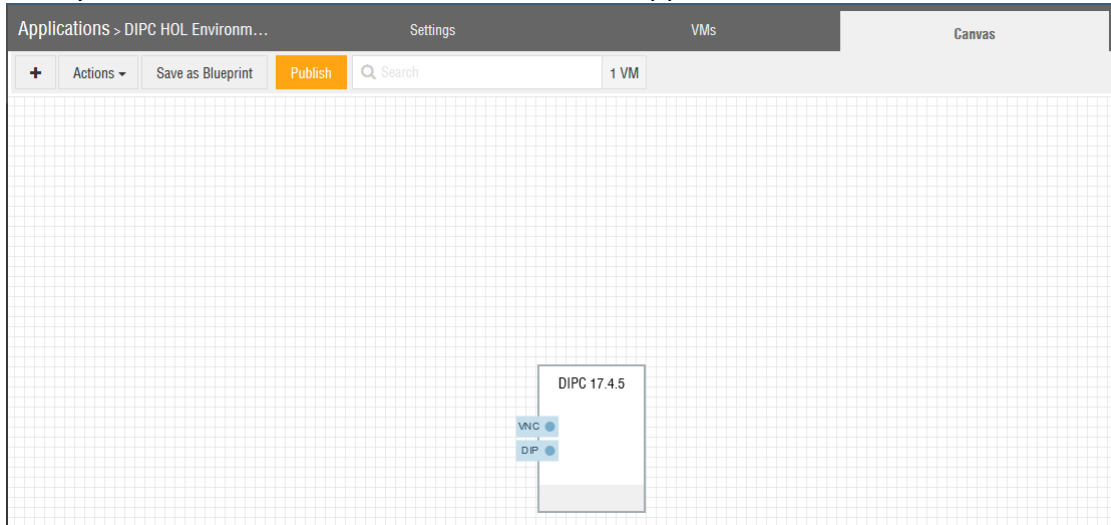
### Ravello Setup

1. Log into Ravello and click on Applications then click on '+Create Application'

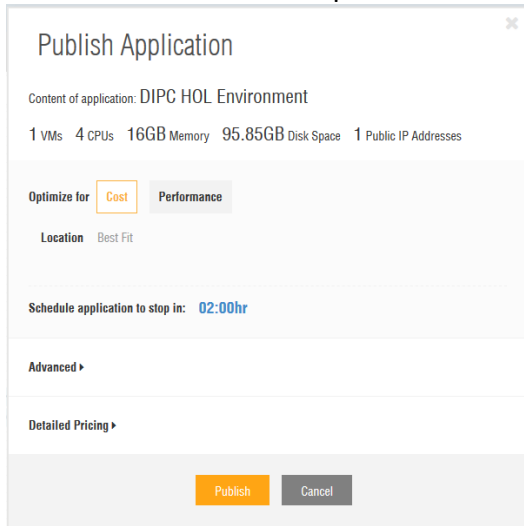


2. Enter a Name and pick the DIPC Blueprint that was shared with you

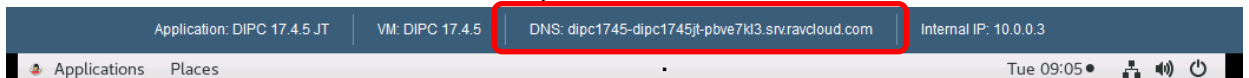
- Next you should land in the Canvas for this new Application, click Publish



- Enter the information required and click Publish



- You will need to get the hostname of the Ravello Application you have started. When the VM shows as Started on the Canvas, go to Console to get access to the VM and write down the hostname listed at the top of the screen:



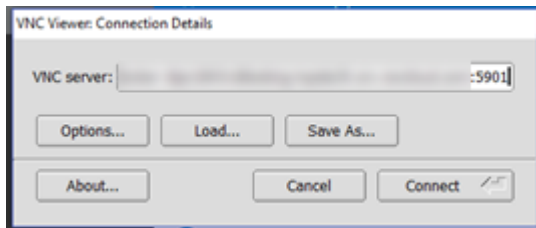
There will also be tasked that will be performed within the VM as the Data Integration Platform Cloud administrator.

### Task 0: Preparation Steps

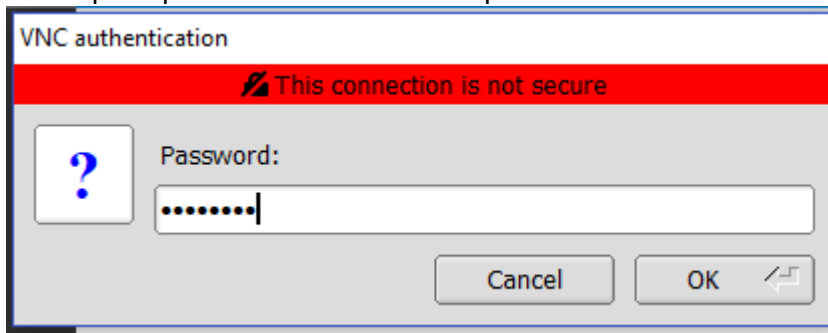
*In these steps you will clean up and setup the environment for this exercise*

## Data Integration Platform Cloud: Hands-on Lab

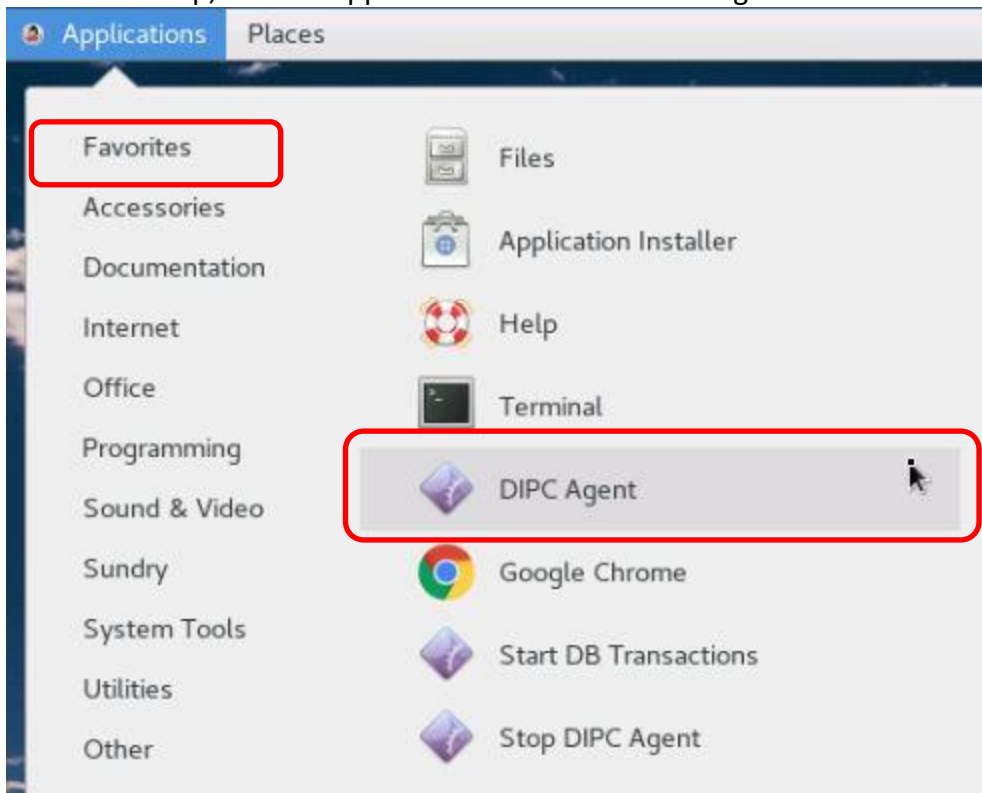
1. Using your favorite VNC Client enter the URL (<ravello hostname>:5901) and click Connect



2. When prompted enter welcome1 as password



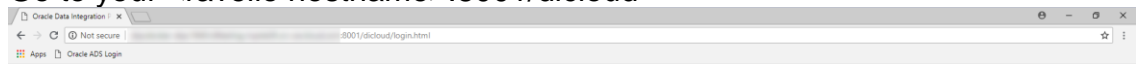
3. On the Desktop, click on Applications then select DIPC Agent under Favorites



4. 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**
5. Now you can log into Data Integration Platform Cloud

## Data Integration Platform Cloud: Hands-on Lab

- a. On your laptop, open Chrome (Firefox works too)
- b. Go to your <ravello hostname>:8001/dicloud

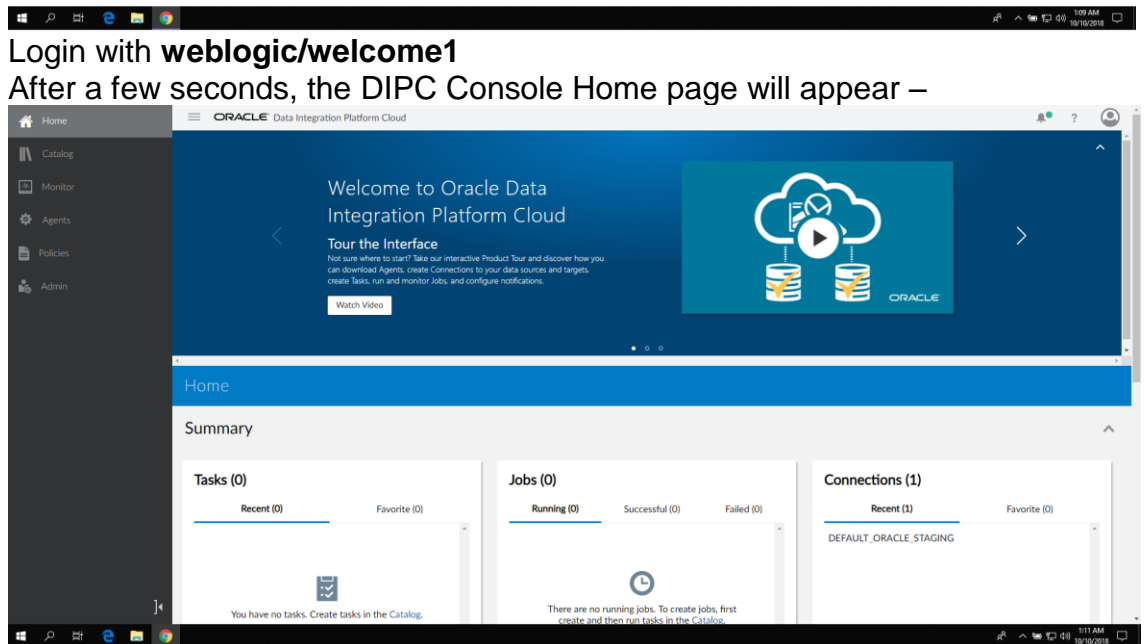


ORACLE Data Integration Platform Cloud

Username \*

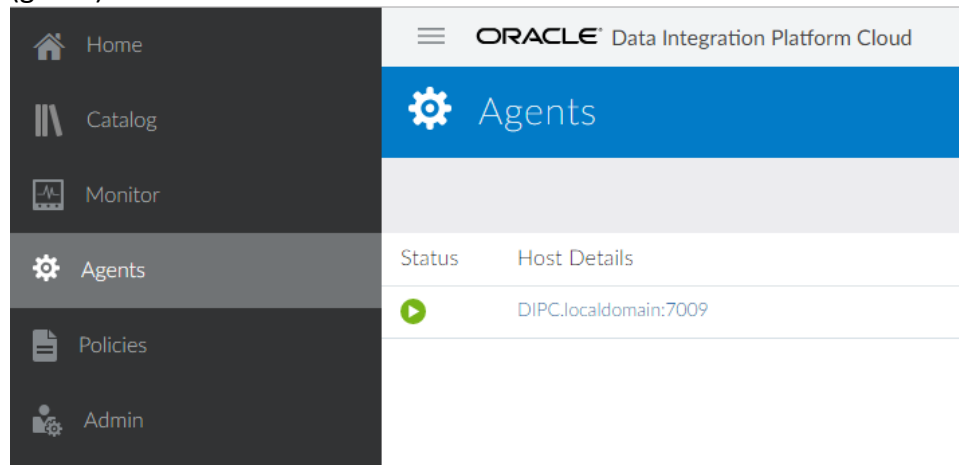
Password \*

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



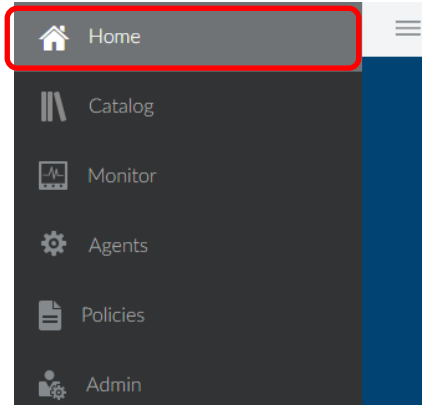
6. Click on Agents to display the list of Agents available in this environment. The Agent (DIPC.localdomain:7009) we just started will be displayed and should be in Running

(green) status

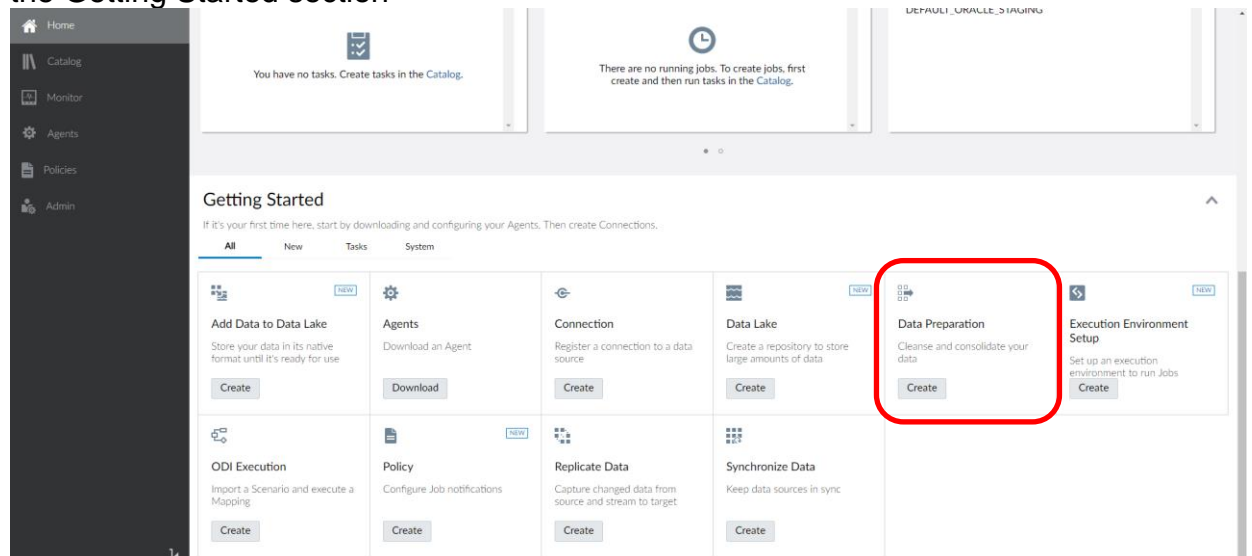


## Task 1: Prepare External Website Data


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


<  Create Data Preparation Task Cancel Save & Transform


General Information

Name \*

Identifier \*


Description


Source Configuration 

Connection \*  


Schema \*

Data Entity \*  Select...


Target Configuration 

Connection \*  

Schema \*

Data Entity \*  

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

4. For Source Configuration click the add new connection button 


### General Information

Name \*

Identifier \*

Description

### Source Configuration

Connection \*  

Schema \*

Data Entity \*  Select...



Connection

General Information

Name \* LocalFileStore

Identifier \* LOCALFILESTORE

Description Local File Store

Agent \* DIPC.localdomain:7009

Type \* File

Connection Settings

Directory \* /tmp

Test Connection Cancel Save

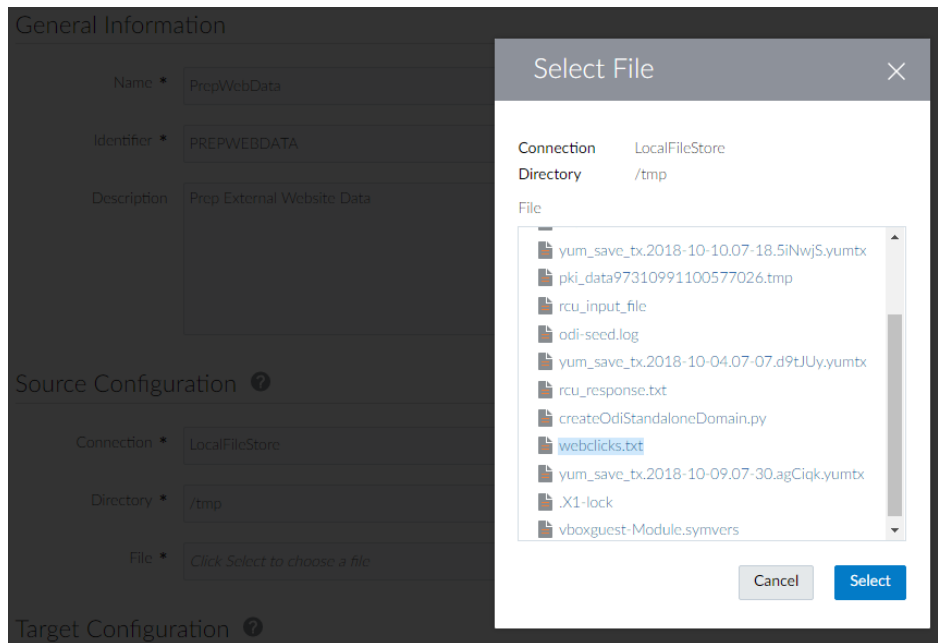
- a. Name: LocalFileStore
  - b. Identifier: use default
  - c. Description: Local File Store
  - d. Agent: DIPC.localdomain:7009
  - e. Type: File
  - f. Directory: /tmp
5. Review Connections settings.
  6. Click Test Connection – Review for Errors.
    - a. Click Save if successful
  7. Now click on Select... next to File, scroll down in the list and pick webclicks.txt then click Select

Source Configuration ?

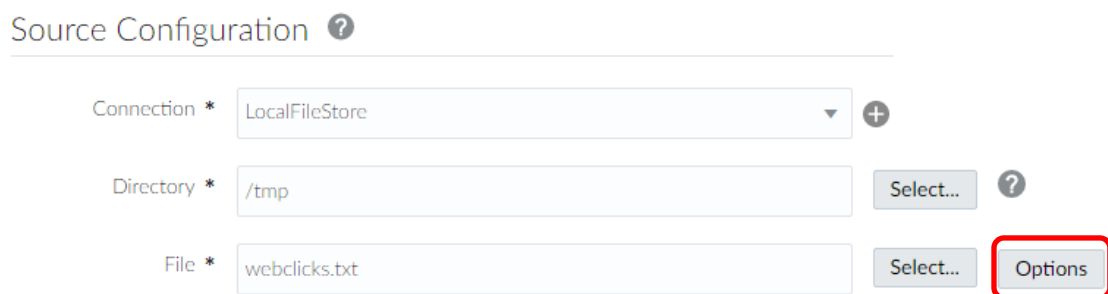
Connection \* LocalFileStore +

Directory \* /tmp Select... ?

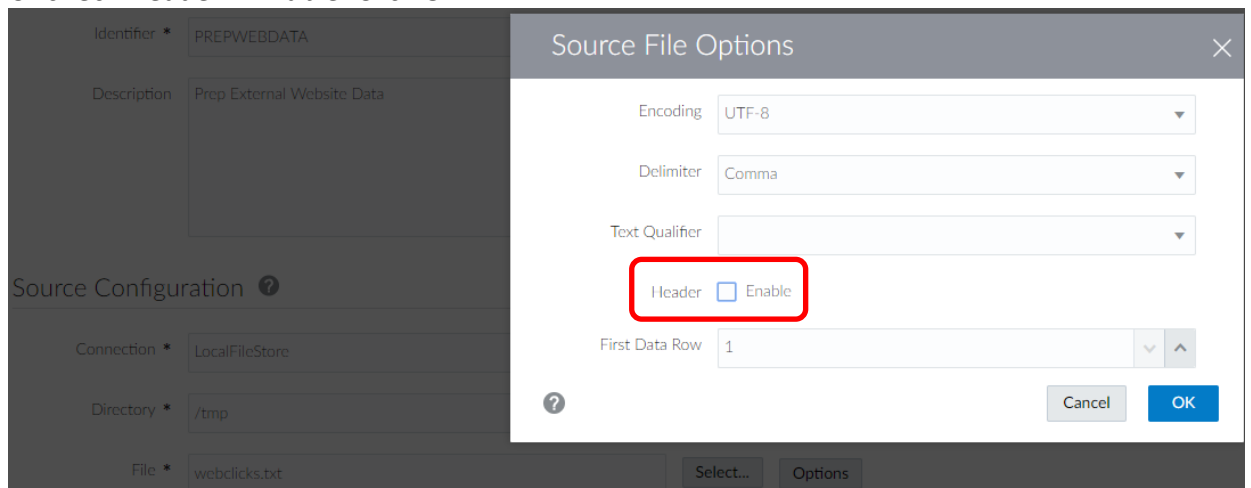
File \* webclicks.txt Select... Options



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 (if needed otherwise simply select Sync Target)

### Target Configuration ?

Connection \*  

Schema \*

Data Entity \*  ?

a. Enter the following information

- Name: Sync Target
- Identifier: use default
- Description: Target Schema
- Agent: DIPCLocaldomain:7009
- Type: Oracle Database
- Hostname: DIPCL
- Port: 1521
- Username: DIPCL\_TGT
- Password: welcome1
- Service Name: orclpdb.us.oracle.com
- Schema: DIPCL\_TGT
- CDB Connection: leave blank

Connection

Name \*

Identifier \*

Description

Agent \*

Type \*

Connection Settings

Hostname \*

Port \*

Username \*

Password \*

Service \* ☒ Service Name \*   
☐ SID

Schema Name

CDB Connection

?

Test Connection

Cancel

Save

- 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



### General Information

Name \*

Identifier \*

Description

### Source Configuration ?

Connection \*  +

Directory \*  Select... ?

File \*  Select... Options

### Target Configuration ?

Connection \*  +

Schema \*

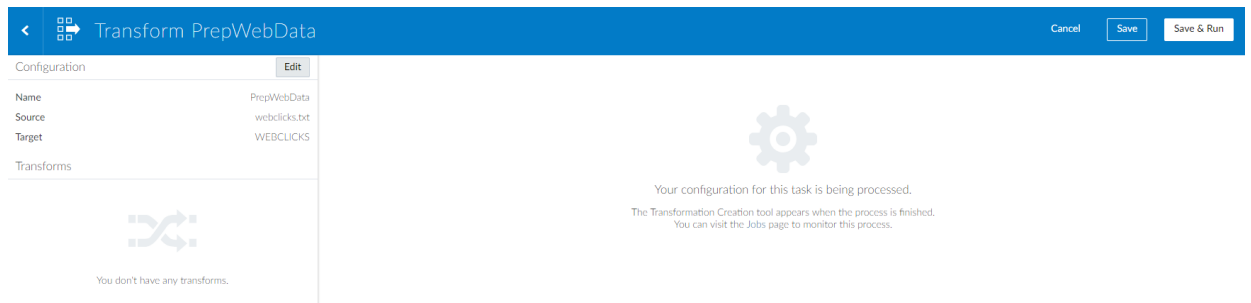
Data Entity \*  ?

11. Click Save & Transform

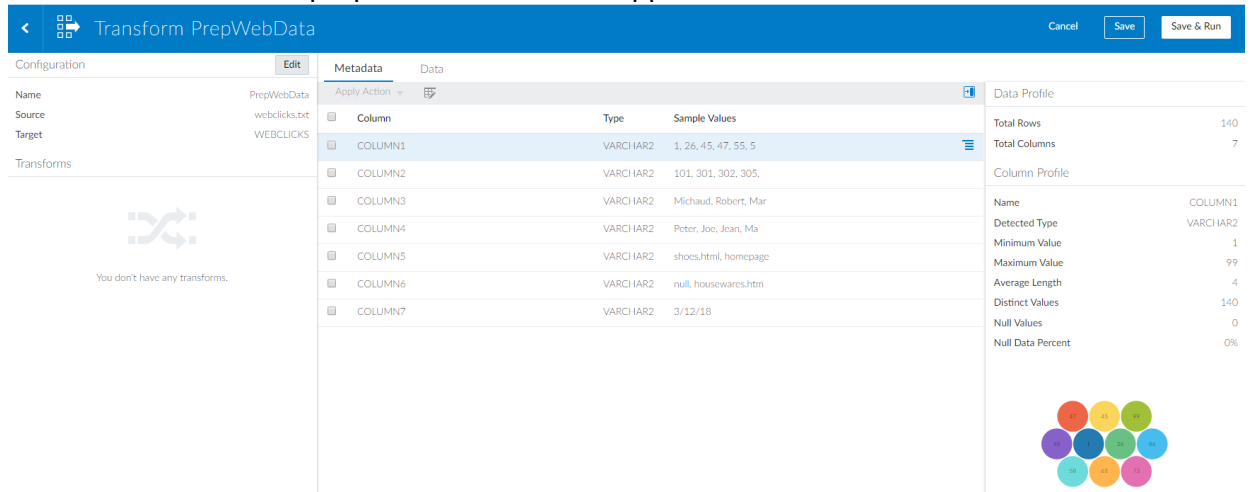


12. The following screen will appear as the file is being parsed and profiled

## Data Integration Platform Cloud: Hands-on Lab

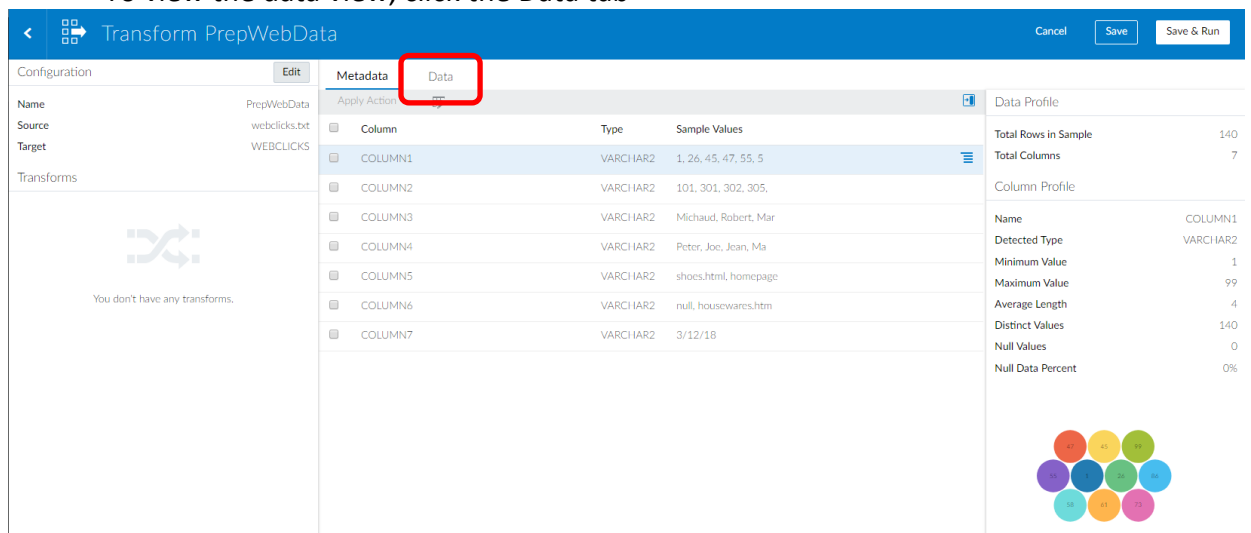


### 13. Once finished the data preparation screen will appear

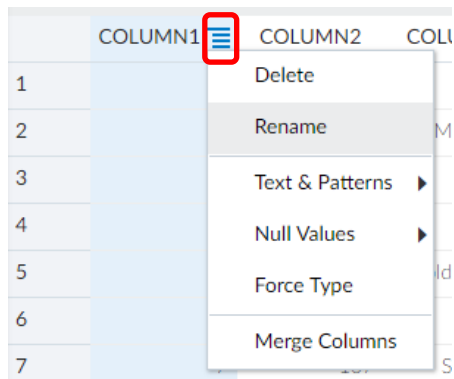


### 14. Begin prepping/transforming the file

- 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.
- There are also two views in which data can be prepared and transformed. To view the data view, click the Data tab

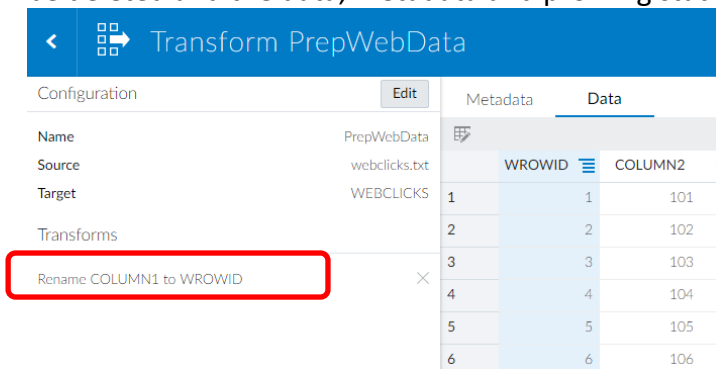


### 15. To transform data click the menu bar on the column



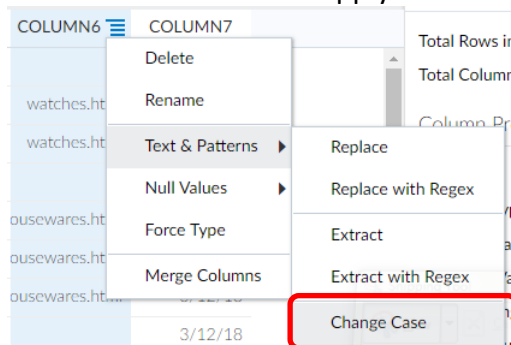
16. For Column 1 rename the column into WROWID

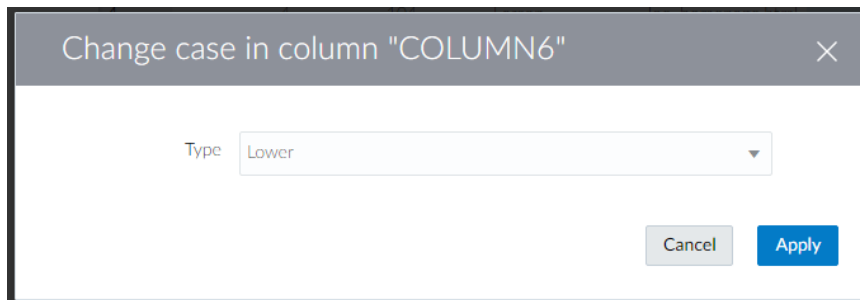
17. Click Apply. Clicking apply updates, the data, metadata and profiling statistics. Also note the transform is saved and displayed in the left-hand drawer. This transform can be deleted and the data, metadata and profiling statistics will be updated accordingly.



18. Add the following column transforms

- Column 2: Rename into CUSTID
- Column 5: Rename into PAGENAME
- Column 6: transform Shoes.html to shoes.html using Text & Patterns > Change Case to Lower and click Apply





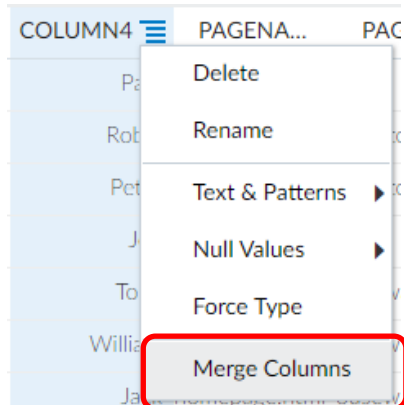
Change case in column "COLUMN6"

Type: Lower

Cancel Apply

Then rename COLUMN6 into PAGEREFFERRER

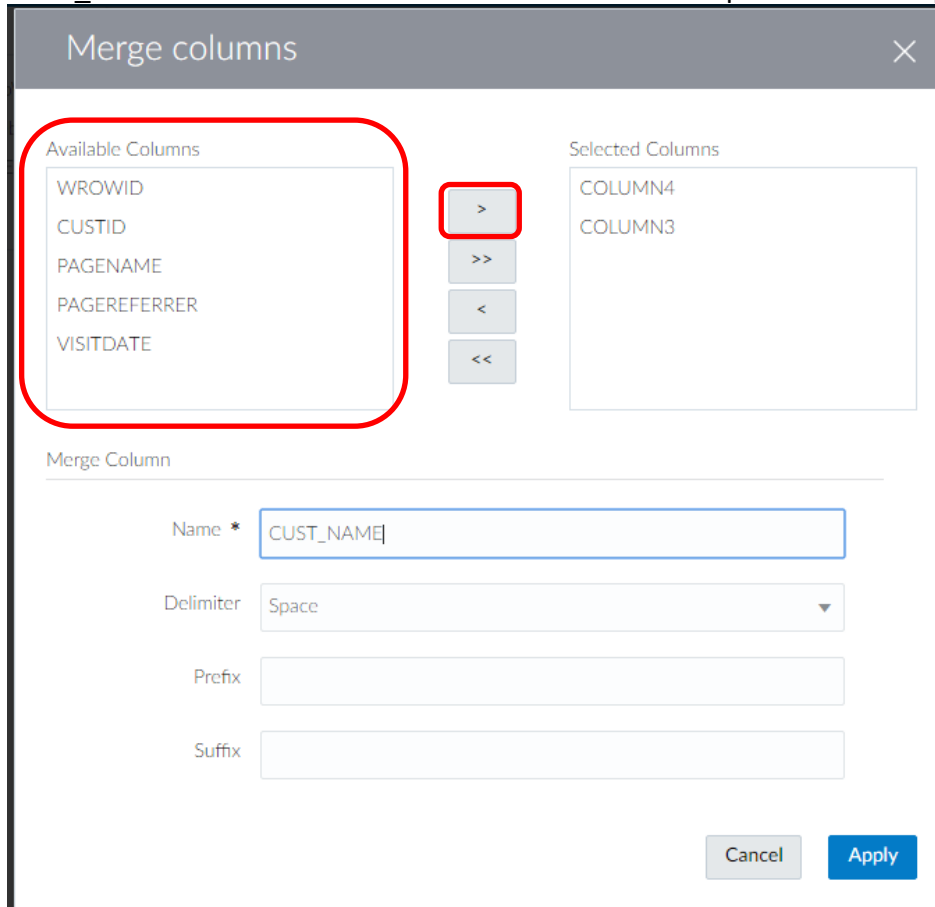
- d. Column 7: Rename into VISITDATE
  - e. Merge COLUMN3 and COLUMN4 to create a new CUST\_NAME column
- Click on COLUMN4 menu and select Merge Columns



COLUMN4

- Delete
- Rename
- Text & Patterns
- Null Values
- Force Type
- 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



Merge columns

Available Columns

- WROWID
- CUSTID
- PAGENAME
- PAGEREFFERRER
- VISITDATE

Selected Columns

- COLUMN4
- COLUMN3

Merge Column

Name \* CUST\_NAME

Delimiter Space

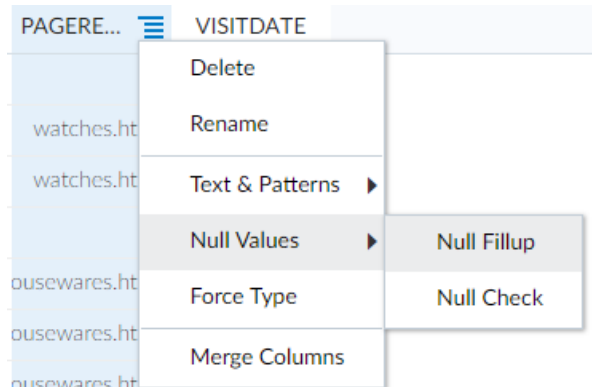
Prefix

Suffix

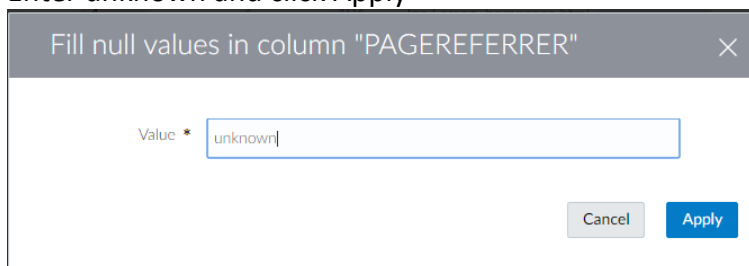
Cancel Apply

## Data Integration Platform Cloud: Hands-on Lab

- f. Delete COLUMN3 and COLUMN4 (click menu item and select Delete then Yes, Delete)
- g. PAGEREFFERRER: Click on menu then select Null Values > Null Fillup



Enter unknown and click Apply



### 19. Review Transforms and Data

The screenshot shows the 'Transform PrepWebData' interface. The 'Configuration' tab is active, showing a list of transforms. A red box highlights the following transforms: Rename COLUMN1 to WROWID, Rename COLUMN2 to CUSTID, Rename COLUMN5 to PAGENAME, Change case to lower in COLUMN6, Rename COLUMN6 to PAGEREFFERRER, Rename COLUMN7 to VISITDATE, Merge COLUMN4.COLUMN3 to CUST\_NAME, Delete COLUMN3, Delete COLUMN4, and Fill null with unknown in PAGEREFFERRER. The 'Data' tab shows a table with 20 rows of data. The 'Data Profile' tab on the right shows statistics for the 'PAGEREFFERRER' column.

WROWID	CUSTID	CUST_N...	PAGENA...	PAGERE...	VISITDATE
1	1	101	Paul Brendt	homepage.html	unknown
2	2	102	Robin McCarthy	homepage.html	watches.html
3	3	103	Peter Travis	homepage.html	watches.html
4	4	104	Joe Larson	homepage.html	unknown
5	5	105	Tony Goldschmidt	homepage.html	ousewares.html
6	6	106	William Baker	homepage.html	ousewares.html
7	7	107	Jack Swenson	homepage.html	ousewares.html
8	8	201	Jean Sartois	homepage.html	unknown
9	9	202	Philippe Michaud	homepage.html	shoes.html
10	10	203	Christian Robert	homepage.html	shoes.html
11	12	205	Luc Piaget	homepage.html	watches.html
12	13	206	Michele Gentil	homepage.html	watches.html
13	14	207	Chantale	homepage.html	unknown
14	15	301	Caroline Edwards	homepage.html	unknown
15	16	302	Justin Marlow	homepage.html	shoes.html
16	17	303	John Wilson	homepage.html	shoes.html
17	18	304	George McCartney	homepage.html	unknown
18	19	305	Harriett Keegan	homepage.html	unknown
19	20	306	Mary Jones	homepage.html	ousewares.html
20	21	307	Priscilla Hopkins	homepage.html	ousewares.html

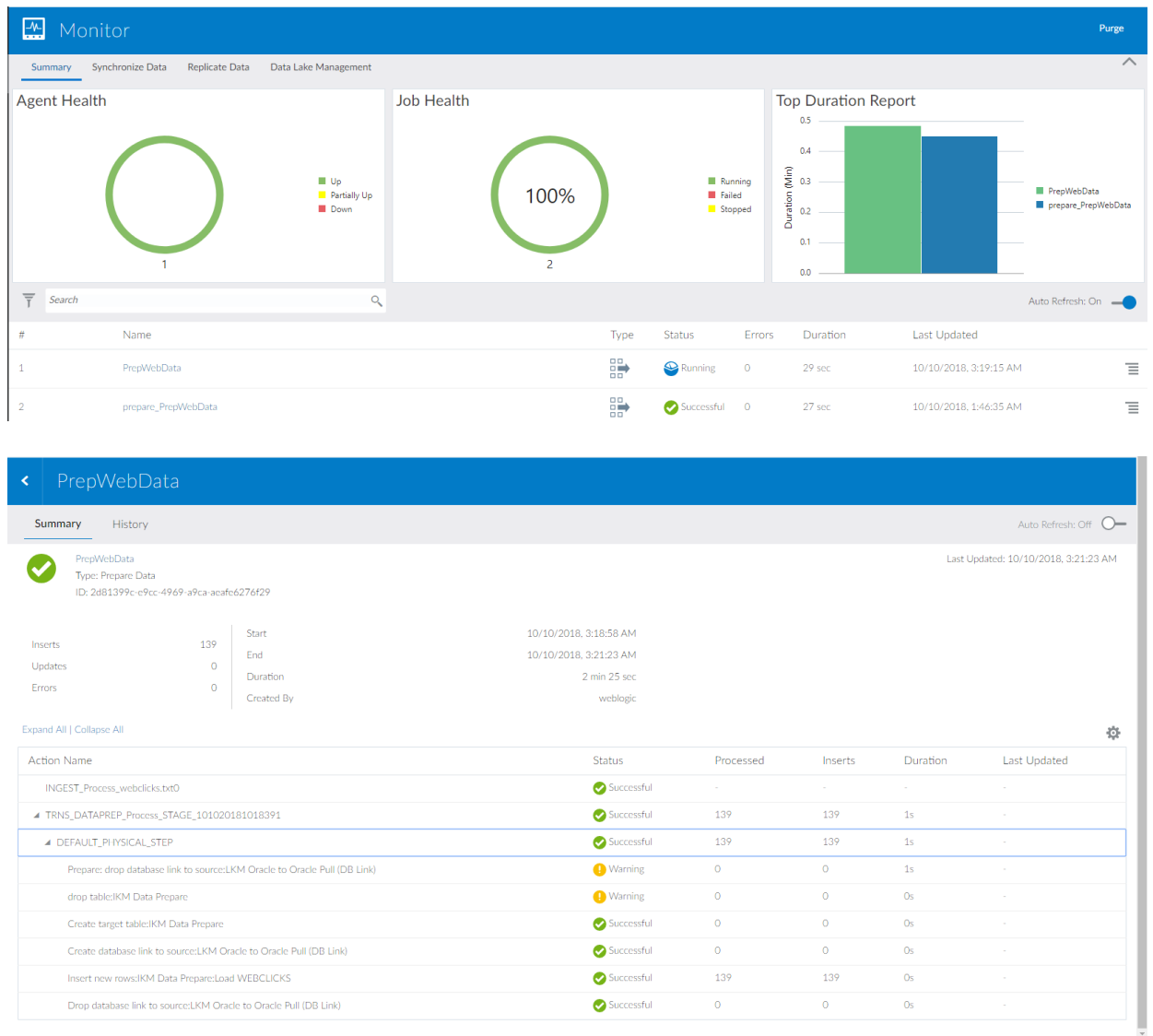
### 20. Click Save & Run



### 21. Review Job Execution

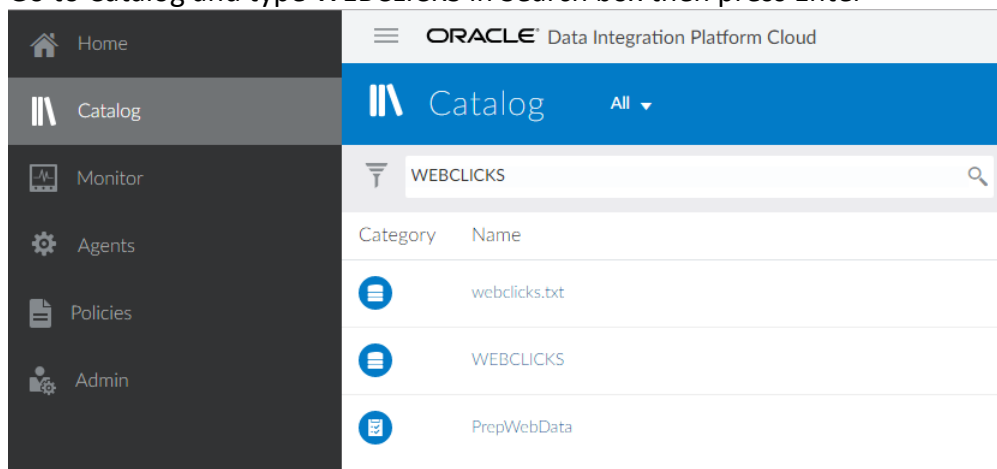


## Data Integration Platform Cloud: Hands-on Lab



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



## Data Integration Platform Cloud: Hands-on Lab

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

WEBCLICKS						Refresh Data Profile	Delete	Edit
Summary	Attributes	Data	History					
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		

## Summary

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