

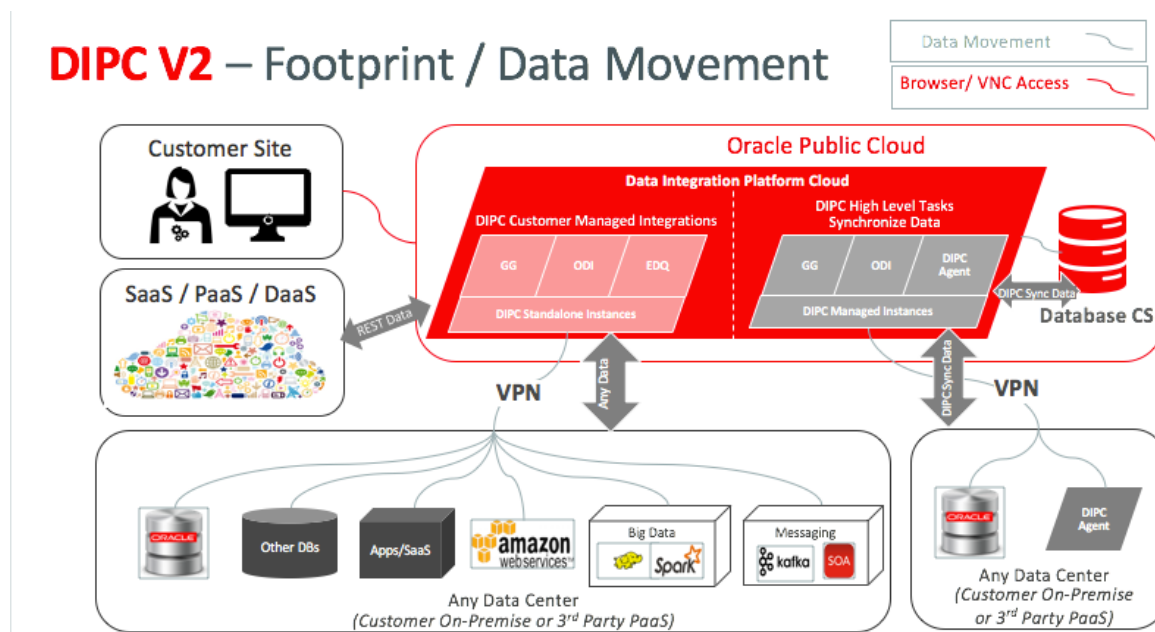


Data Integration Platform Cloud Hands On Lab



Architecture Overview

This Hands-on lab demonstrates the power and simplicity of the Data Integration Platform. For this Hands-on lab data will be sourced from an OLTP system made up of a 3NF schema, data will be extracted and synchronized in real time to an operational reporting schema, this data will then be mini batched to an analytical data warehouse. The goal of this lab to demonstrate how DIPC removes the complexity in creating a real-time operation reporting schema as well as a near real-time analytical data warehouse, enabling even novice users to perform the creation, initial load and real time synchronization of the operational reporting schema in just a couple of clicks. Historically these types of data integration tasks required expert ETL and DBA resources as well as a detailed understanding of the source schema. Built from the ground up on proven technologies, DIPC provides enterprise data integration and data governance features for today's enterprise.

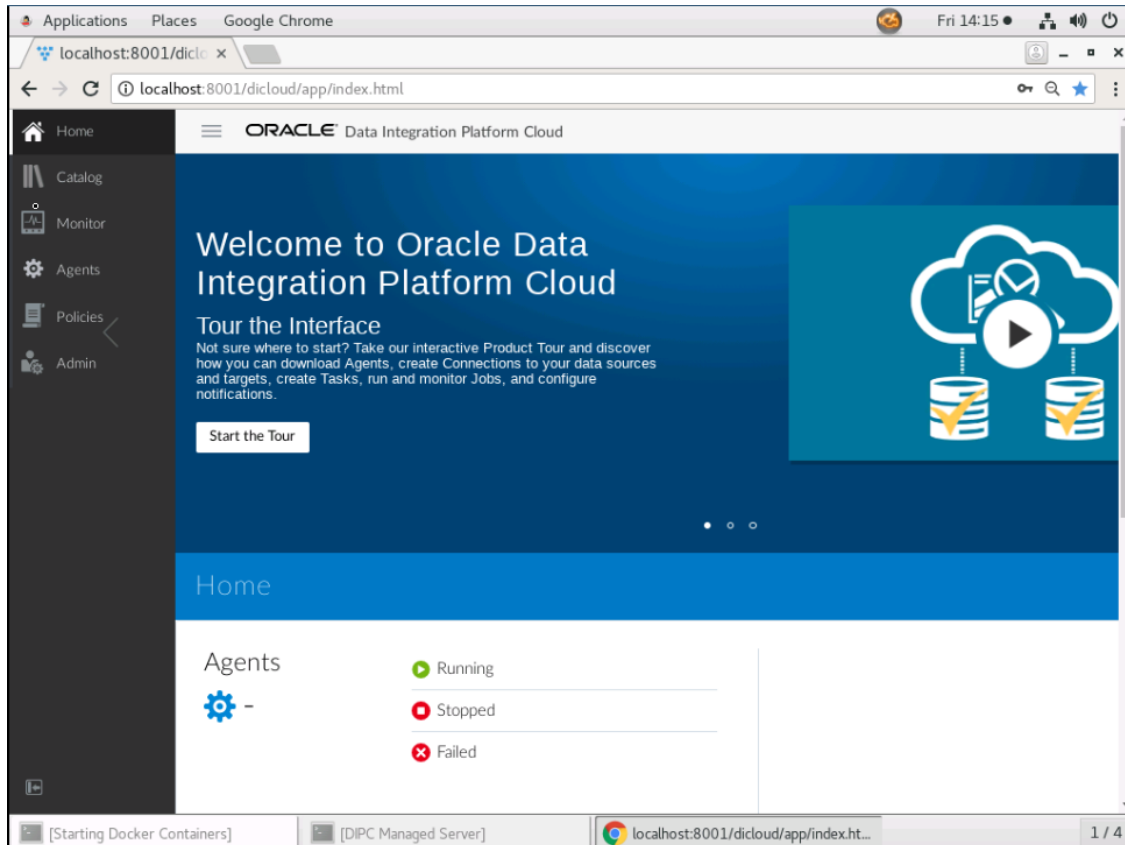




Task 1: Prep External Website Data

1. Log into Data Integration Platform Cloud
 - a. From your local laptop/desktop start Chrome (you can also continue to use Chrome on the Ravello instance)
 - b. Go to your <Ravello Instance hostname>:8001/dicloud/login.html
 - c. Login with weblogicwelcome1

The following page should appear –



2. Click Home
3. Click Home and click on the right arrow in the carousel
4. Select Create under Data Preparation
5. On the Create Data Preparation Task enter the following information
 - a. Name: PrepWebData
 - b. Description – Prep External Website Data
6. For Source Configuration click the add new connection button
 - a. Name: LocalFileStore
 - b. Description – Local File Store
 - c. Agent – leave blank
 - d. Type: File
 - e. Directory: /tmp
7. Review Connections settings.
8. Click Test Connection – Review for Errors.
 - a. Click Save
9. Now click on Select... next to File, pick webclicks.txt then click Select



10. Next click on Options and uncheck Header > Enable. Click OK
11. In Target Configuration, select Sync Target if it already exists or click on Create button to create it.

NOTE: This connection may have already have been setup in a previous HOL.

- a. Optional: Enter the following information
 - **Name:** Sync Target
 - **Description** – Target Schema
 - **Agent** – DIPC.localdomain:7005
 - Type Oracle – selecting Oracle will expand the Connection Settings
 - **Hostname:** DIPC
 - **Port:** 1521
 - **Username:** DIPC_TGT
 - **Password:** welcome1
 - **Service Name:** orclpdb1.localdomain
 - **Schema:** DIPC_TGT
 - **CDB Connection:** ORACLE_CDB

- b. Optional: Click Test Connection, Review for errors
- c. Optional: Click Save
12. Enter WEBCLICKS next to Data Entity under Target Configuration
 - a. Connection: DIPC_TGT
 - b. Schema: DIPC_TGT
 - c. Data Entity: webclicks



< Create Data Preparation Task

General Information

Name * PrepWebData

Identifier * PREPWEBDATA

Description Prep External Website Data

Source Configuration ?

Connection * LocalFileStore

Directory * /tmp

File * webclicks.txt

Select... ?

Select... Options

Target Configuration ?

Connection * Sync Target

Schema * DIPC_TGT

Data Entity * WEBCLICKS

13. Click Save and Transform

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

< Transform PrepWebData

Cancel Save Save & Run

Configuration Edit

Name PrepWebData

Source webclicks.txt

Target WEBCLICKS

Transforms

You don't have any transforms.

Your configuration for this task is being processed.
The Transformation Creation tool appears when the process is finished.
You can visit the Jobs page to monitor this process.

15. Once finished the data preparation screen will appear

Data Integration Platform Cloud: Hands On Lab



Transform PrepWebData Cancel Save Save & Run

Configuration Edit **Metadata** **Data**

Name: PrepWebData
Source: webclicks.txt
Target: WEBCLICKS

Transforms: You don't have any transforms.

Column	Type	Sample Values
COLUMN1	VARCHAR2	1, 26, 45, 47, 55, 5
COLUMN2	VARCHAR2	101, 301, 302, 305,
COLUMN3	VARCHAR2	Michaud, Robert, Mar
COLUMN4	VARCHAR2	Peter, Joe, Jean, Ma
COLUMN5	VARCHAR2	shoes.html, homepage
COLUMN6	VARCHAR2	null, housewares.htm
COLUMN7	VARCHAR2	3/12/18

Data Profile

Total Rows: 140
Total Columns: 7

Column Profile

Name: COLUMN1
Detected Type: VARCHAR2
Minimum Value: 1
Maximum Value: 99
Average Length: 4
Distinct Values: 140
Null Values: 0
Null Data Percent: 0%

16. 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 prepped and transformed. To view the data view click the data tab

Configuration Edit **Metadata** **Data**

Name: Prep External Website Dat
Source: webclicks.txt
Target: webclicks

Transforms: You don't have any transforms.

Column	Type	Sample Values
COLUMN1	VARCHAR2	1, 26, 45, 47, 55, 5
COLUMN2	VARCHAR2	101, 301, 302, 305,
COLUMN3	VARCHAR2	Michaud, Robert, Mar
COLUMN4	VARCHAR2	Peter, Joe, Jean, Ma
COLUMN5	VARCHAR2	homepage.html, Shoes
COLUMN6	VARCHAR2	null, housewares.htm
COLUMN7	VARCHAR2	3/12/18

Data Profile

Total Rows: 140
Total Columns: 7

Column Profile

Name: COLUMN1
Detected Type: VARCHAR2
Minimum Value: 1
Maximum Value: 99
Average Length: 4
Rows with No Data: 0
Distinct Values: 140

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

Configuration Edit **Metadata** **Data**

Name: PrepWebData
Source: webclicks.txt
Target: WEBCLICKS

Transforms: You don't have any transforms.

	COLUMN1	COLUMN2	COLUMN3	COLUMN4	COLUMN5	COLUMN6	COLUMN7
1	1	101	Brndt	Paul	homepage.html		
2	2	102	McCarthy	Robin	homepage.html	watches.ht	
3	3	103	Travis	Peter	homepage.html	watches.ht	
4	4	104	Larson	Joe	homepage.html		
5	5	105	Goldschmidt	Tony	homepage.html	ousewares.ht	
6	6	106	Baker	William	homepage.html	ousewares.ht	
7	7	107	Swenson	Jack	homepage.html	ousewares.ht	
8	8	201	Sartois	Joan	homepage.html		3/12/18
9	9	202	Michaud	Phillipe	homepage.html	Shoes.html	3/12/18
10	10	203	Robert	Christian	homepage.html	Shoes.html	3/12/18
11	11	204	Martin	Christine	homepage.html		3/12/18
12	12	205	Piaget	Luc	homepage.html	watches.html	3/12/18
13	13	206	Gentil	Michele	homepage.html	watches.html	3/12/18
14	14	207	Dupont	Marie-Charlotte	homepage.html		3/12/18
15	15	301	Edwards	Caroline	homepage.html		3/12/18
16	16	302	Marlow	Justin	homepage.html	Shoes.html	3/12/18

Data Profile

Total Rows: 140
Total Columns: 7

Column Profile

Name: COLUMN6
Detected Type: VARCHAR2
Minimum Value: Shoes.html
Maximum Value: watches.html
Average Length: 11
Distinct Values: 4
Null Values: 36
Null Data Percent: 25.7%

18. For Column 6 transform Shoes.html to shoes.html using the Text & Patterns > Replace transform

19. 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.

20. Add the following column transforms
 - a. Column 1: Rename WROWID
 - b. Column 2: Rename CUSTID
 - c. Column 5: Rename PAGENAME
 - d. Column 6: Rename PAGEREFERRER
 - e. Column 7: Rename VISITDATE
 - f. Merge COLUMN3 and COLUMN4 and create a new NAME column





Merge columns

Available Columns

WROWID
CUSTID
PAGENAME
PAGEREFERRER
VISITDATE

>
>>
<
<<

Selected Columns

COLUMN3
COLUMN4

Merge Column

Name *

Name|

Delimiter

Space

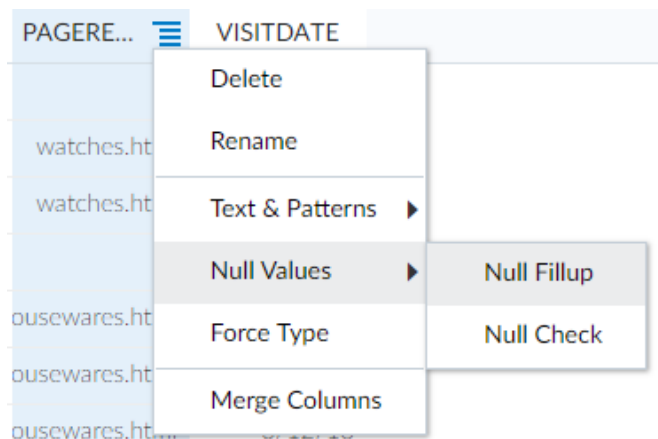
Prefix

Suffix

Cancel

Apply

- g. Delete columns 3 and 4
h. Column 6: Null Values > Null Fillup



Fill null values in column "PAGEREFERRER"

Value *

Unknown|

Cancel

Apply

21. Review Transforms and Data

Data Integration Platform Cloud: Hands On Lab



Transform PrepWebData [Cancel] [Save] [Save & Run]

Configuration | Edit | Metadata | Data | Data Profile

Name: PrepWebData
Source: webclicks.txt
Target: WEBCLICKS

Transforms

- Replace Shoes.html with shoes.html in COLUMN6
- Rename COLUMN1 to WROWID
- Rename COLUMN2 to CUSTID
- Rename COLUMN5 to PAGENAME
- Rename COLUMN6 to PAGEREFERER
- Rename COLUMN7 to VISITDATE
- Merge COLUMN3, COLUMN4 to Name
- Delete COLUMN3
- Delete COLUMN4
- Fill null with Unknown in PAGEREFERER

	WROWID	CUSTID	Name	PAGENAME	PAGEREFERER	VISITDATE
1	1	101	Brendt Paul	homepage.html	Unknown	3/12/18
2	2	102	McCarthy Robin	homepage.html	watches.html	3/12/18
3	3	103	Travis Pctor	homepage.html	watches.html	3/12/18
4	4	104	Lanson Joe	homepage.html	Unknown	3/12/18
5	5	105	Goldschmidt Tony	homepage.html	ousewares.html	3/12/18
6	6	106	Baker William	homepage.html	ousewares.html	3/12/18
7	7	107	Swenson Jack	homepage.html	ousewares.html	3/12/18
8	8	201	Sartois Jean	homepage.html	Unknown	3/12/18
9	9	202	Michaud Philippe	homepage.html	shoes.html	3/12/18
10	10	203	Robert Christian	homepage.html	shoes.html	3/12/18
11	12	205	Plaget Luc	homepage.html	watches.html	3/12/18
12	13	206	Gorelli Michele	homepage.html	watches.html	3/12/18
13	14	207	Marie-	homepage.html	Unknown	3/12/18
14	15	301	Edwards Caroline	homepage.html	Unknown	3/12/18
15	16	302	Marlow Justin	homepage.html	shoes.html	3/12/18
16	17	303	Wilson John	homepage.html	shoes.html	3/12/18
17	18	304	McCartney George	homepage.html	Unknown	3/12/18
18	19	305	Kriegsin Haricett	homepage.html	Unknown	3/12/18

Data Profile

Total Rows: 140
Total Columns: 7

Column Profile

Name: PAGEREFERER
Detected Type: VARCHAR2
Minimum Value: Unknown
Maximum Value: watches.html
Average Length: 13
Distinct Values: 5
Null Values: 0
Null Data Percent: 0%

22. Click Save and Run

23. Review Job Execution

Monitor | Summary | Synchronize Data | Replicate Data | Data Lake Management

Agent Health: 100% (1 Running, 0 Failed, 0 Stopped)

Job Health: 67% (1 Running, 1 Failed, 2 Stopped)

Top Duration Report

#	Name	Type	Status	Errors	Duration	Last Updated
1	PrepWebData	PrepWebData	Running	0	12 sec	8/3/2018, 6:02:47 PM
2	PrepWebData	PrepWebData	Successful	0	7 sec	8/3/2018, 5:53:13 PM
3	Sync Sales Data	Sync Sales Data	Stopped	0	16 min 47 sec	8/3/2018, 5:38:11 PM

PrepWebData | Summary | History | Auto Refresh: Off

PrepWebData
Type: Prepare Data
ID: 1745548b-c772-40cd-97a1-ae0cf831f62

Lag: 0
Start: 8/3/2018, 6:02:37 PM
End: 8/3/2018, 6:03:48 PM
Duration: 1 min 12 sec
Errors: 0
Created By: weblogic

Action Name	Status	Processed	Inserts	Errors	Lag	Duration
INGEST_Process_webclicks.txt	Successful	-	-	-	-	-
TRANS_DATAPREP_Process_STAGE_030620181602231	Successful	140	140	0	-	0s
DEFAULT_PHYSICAL_STEP	Successful	140	140	0	-	0s
Prepare: drop database link to source:KIM Oracle to Oracle Pull (DB Link)	Warning	0	0	0	-	0s
drop table:KIM Data Prepare	Warning	0	0	0	-	0s
Create target table:KIM Data Prepare	Successful	0	0	0	-	0s
Create database link to source:KIM Oracle to Oracle Pull (DB Link)	Successful	0	0	0	-	0s
Insert new rows:KIM Data Prepare:Load WEBCLICKS	Successful	140	140	0	-	0s
Drop database link to source:KIM Oracle to Oracle Pull (DB Link)	Successful	0	0	0	-	0s

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.