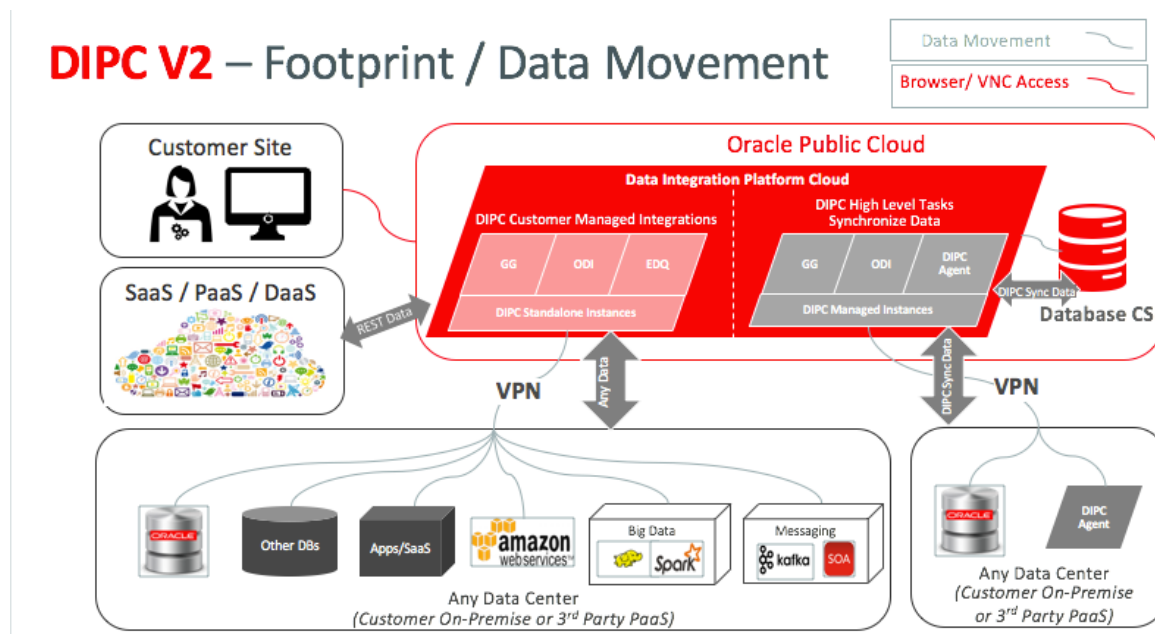# Data Integration Platform Cloud
Hands On Lab

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: Setup DIPC Connections

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

CDB User   : C##GGSRC
Source Schema/User : DIPC_SRC
OLTP_Schema/User   : DIPC_TGT
Passwords   : welcome1
Server   : DIPC
Port   : 1521
Services   :
                                        orclcdb.localdomain (CDB Connection)
                                        orclpdb1.localdomain (PDB Connections)

The next steps will walk you through how to setup each.

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 weblogic/welcome1
        The following page should appear –
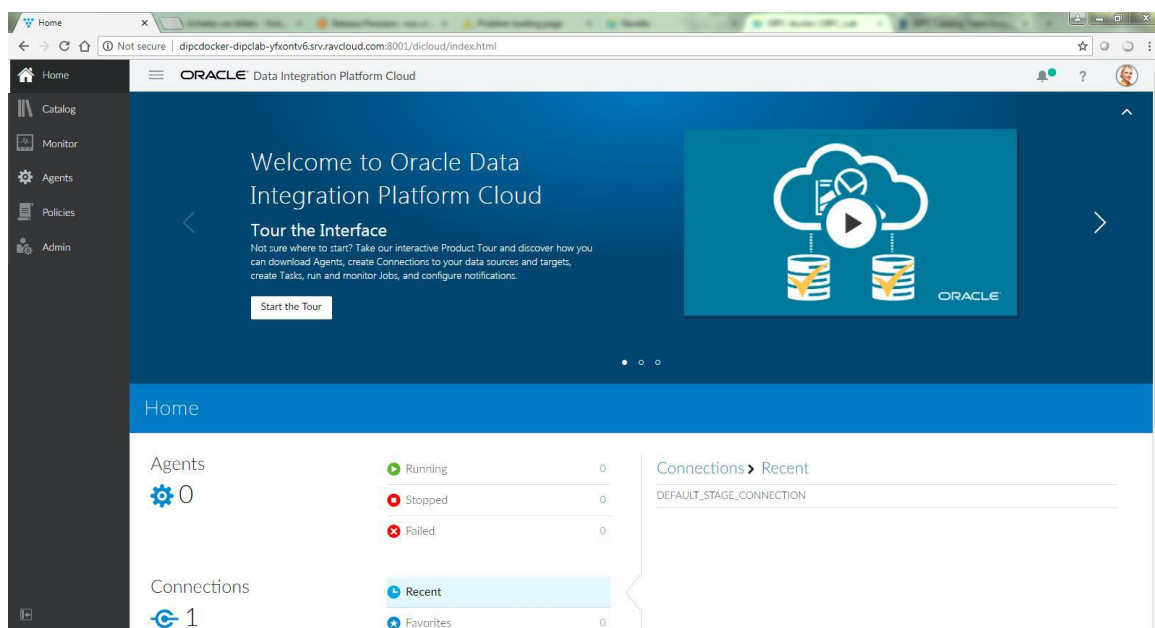
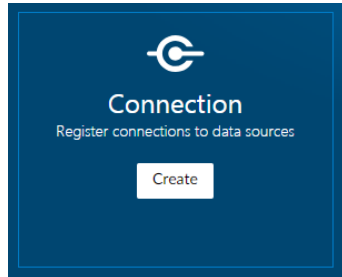2. Click Home and click on the right arrow in the carousel

3. Create Oracle CDB Connection
   a. Click Create under Connection

   b. Enter the following information:
      i. Name: Oracle CDB
      ii. Agent – DIPC.localdomain:7005
      iii. Type - Oracle CDB – selecting Oracle will expand the Connection Settings
      iv. Hostname: DIPC
      v. Port: 1521
      vi. Username: C##GGSRC
      vii. Password: welcome1
      viii. Service Name: orclcdb.localdomain

      c.   Review Connections settings.

      d.   Click Test Connection then Save

4. Create Sync Source Connection
   a. In Catalog, click Create > Connection
   b. Enter the following information
      i. Name: Sync Source
      ii. Description – Sales OLTP Source Data
      iii. Agent – DIPC.localdomain:7005
      iv. Type Oracle – selecting Oracle will expand the Connection Settings
      v. Hostname: DIPC
      vi. Port: 1521
      vii. Username: DIPC_SRC
      viii. Password: welcome1
      ix. Service Name: dics12c
      x. Schema: DIPC_SRC
      xi. CDB Connection: ORACLE_CDB

Connection	×

Identifier *   SYNC_SOURCE

Description   Sales OLTP Source Data

Agent   DIPC.localdomain:7005 ▾

Type *   Oracle ▾

Connection Settings

Hostname *   DIPC

Port *   1521

Username *   DIPC_SRC

Password *   ••••••••

Service *   ● Service Name   *   orclpdb1.localdomain

        ○ SID

Schema Name   DIPC_SRC [Default] ▾

CDB Connection   ORACLE_CDB ▾

Test Connection   Cancel   Save

      c.   Review Connections settings.

      d.   Click Test Connection then Save

5. Create Sync Target Connection
    a. In Catalog, click Create > Connection
    b. Enter the following information
        i. Name: Sync Target
        ii. Description – Target Schema
        iii. Agent – DIPC.localdomain:7005
        iv. Type Oracle – selecting Oracle will expand the Connection Settings
        v. Hostname: DIPC
        vi. Port: 1521
        vii. Username: DIPC_TGT
        viii. Password: welcome1
        ix. Service Name: dics12c
        x. Schema: DIPC_SRC
        xi. CDB Connection: ORACLE_CDB

Connection	✕
Identifier *	SYNC_TARGET
Description	Target Schema
Agent	DIPC.localdomain:7005 ▾
Type *	Oracle ▾
Connection Settings	
Hostname *	DIPC
Port *	1521
Username *	DIPC_TGT
Password *	••••••••
Service *	◉ Service Name * orclpdb1.localdomain
	○ SID
Schema Name	DIPC_TGT [Default] ▾
CDB Connection	ORACLE_CDB ▾

Test Connection   Cancel   Save

c. Review Connections settings.
d. Click Test Connection then Save

6. Review Catalog after saving –
7. Data Entities are harvested and profiled at Connection creation, their popularity is also calculated by reviewing the DB query logs
   <u>Note</u>: This process may take some time (5 minutes or so), the Catalog will show a message when new updates are available You can refresh the page manually to see the new Data Entities
8. Click an entity – SRC_ORDERS You can browse the Catalog pages to find it or you can use the Search bar (search for SRC_ for example)

Catego...	Name	Type	Last Updated	Popularity	
	TRG_REGION	Oracle Table	3/20/2018, 1:44:58 PM		★ ≡
	TRG_EDW	Oracle Table	3/20/2018, 1:44:58 PM		★ ≡
	SRC_ORDER_LINES	Oracle Table	3/20/2018, 1:44:58 PM		★ ≡
	SRC_AGE_GROUP	Oracle Table	3/20/2018, 1:44:57 PM		★ ≡
	SRC_ORDERS	Oracle Table	3/20/2018, 1:44:57 PM		★ ≡
	TRG_CUSTOMER	Oracle Table	3/20/2018, 1:44:56 PM		★ ≡

‹ ▤ SRC_ORDERS

Summary	Attributes	Data	History

Name	SRC_ORDERS
Identifier	SRC_ORDERS
Resource Name	SRC_ORDERS
Description	
Type	Table
Namespace	DIPC_SRC
Contact	
Popularity	
Tags	Add Tags...

Notice the Popularity score calculated for SRC_ORDERS, 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

9. Click on the Attributes tab

Key	Attribute	Data Type	Sample Values
🔑	ORDER_ID	NUMBER (10)	1,2,3,4,5
	STATUS	VARCHAR2 (3)	COM,CLO
	CUST_ID	NUMBER (10)	103,104,1002,105,1001
	ORDER_DATE	DATE (7)	1999-02-12 00:00:00.0,1990-03-23 00:00:00.0,2001-01-11 00:00:00.0,1998-05-10 00:00:00.0,1991-04-26 00:00:00.0
	CUSTOMER	VARCHAR2 (35)	

10. Click on an Attribute, ORDER_DATE for example

Key	Attribute	Data Type	Sample Values		Profile for ORDER_DATE	×
♀.	ORDER_ID	NUMBER (10)	1,2,3,4,5		Total Rows	81
	STATUS	VARCHAR2 (3)	COM,CLO		Rows with Data	100.00%
					Rows with No Data	0.00%
	CUST_ID	NUMBER (10)	103,104,1002,105,1001		Distinct Values	67
	ORDER_DATE	DATE (7)	1999-02-12 00:00:00.0,1990-03-23 00:00:00.0,2001-01-11 00:00:00.0,1998-05-10 00:00:00.0,1991-04-26 00:00:00.0		Duplicate Values	14
					Minimum Value	18-FEB-90
	CUSTOMER	VARCHAR2 (35)			Maximum Value	15-SEP-01
					Average Length	8
					Density	1.49%

11. Notice the Profiling statistics appearing in the right-hand side drawer
12. Click Data tab

ORDER_ID		STATUS	CUST_ID	ORDER_DATE	CUSTOMER
1		COM	1001	2001-01-11 00:00:00.0	
2		CLO	1002	1999-02-12 00:00:00.0	
3		CLO	103	1990-03-23 00:00:00.0	
4		CLO	104	1991-04-26 00:00:00.0	
5		COM	105	1998-05-10 00:00:00.0	
6		CLO	106	1998-06-23 00:00:00.0	

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

Catalog   Data Entities ▾

SRC_ORDER| 🔍

SRC_ORDERS
SRC_ORDER_LINES

Cate
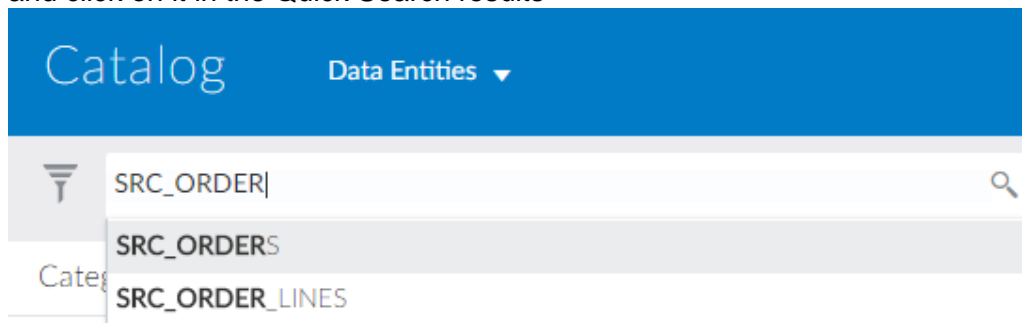
14. Click Attributes tab

Key	Attribute	Data Type	Sample Values
♀.	ORDER_ID	NUMBER (10)	1,2,3,4
♀.	LORDER_ID	NUMBER (10)	1,2,3
	PRODUCT_ID	NUMBER (10)	1,2,3,4,5
	QTY	NUMBER (10)	1689,1477,1829,549,1163
	AMOUNT	NUMBER (10, 2)	59292,146223,136809,20934,172840.5

15. Click on a column, for example column – QTY
16. Notice the Profiling statistics

Key	Attribute	Data Type	Sample Values		Profile for QTY	×
♀.	ORDER_ID	NUMBER (10)	1,2,3,4		Total Rows	236
♀.	LORDER_ID	NUMBER (10)	1,2,3		Rows with Data	100.00%
					Rows with No Data	0.00%
	PRODUCT_ID	NUMBER (10)	1,2,3,4,5		Distinct Values	182
	QTY	NUMBER (10)	1689,1477,1829,549,1163		Duplicate Values	54
					Minimum Value	1
	AMOUNT	NUMBER (10, 2)	59292,146223,136809,20934,172840.5		Maximum Value	5500
					Average Length	4
					Density	0.55%

17. Click Data tab
18. 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. To start this client open a terminal window in the Ravello Console. If needed press Enter to see the Login window, re-enter the DIPC user password (welcome1) and press Enter or click Unlock
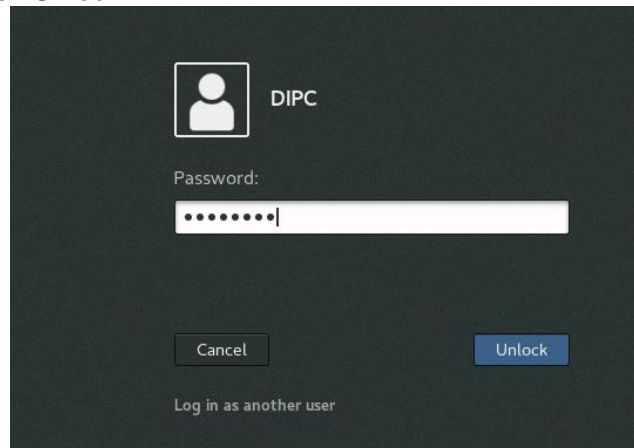
3. Open a Terminal
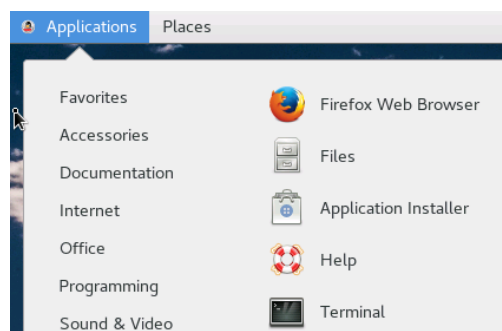
4. From the home directory execute ./startDIPCDemoClient.sh

ü
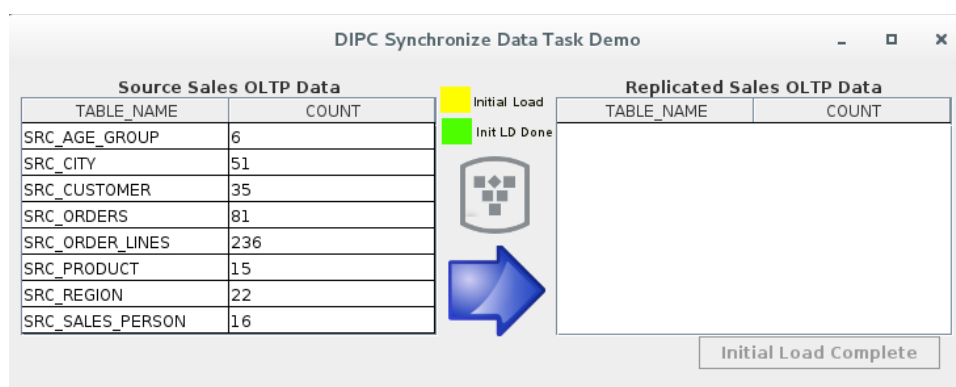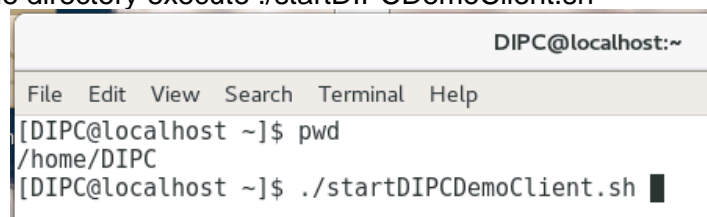
5. Once the Synchronize Data task is saved and executed this client will be used to visually monitory the Replicated Schema, the tables and their row counts
6. Within the DIPC App Click Home
7. Click Create under Synchronize Data
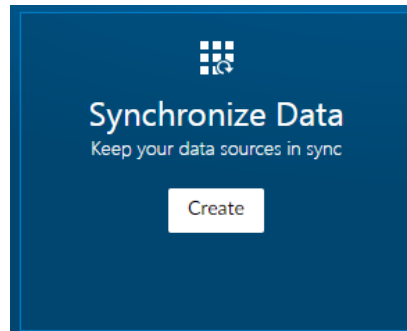
8. Name your task –Sync Sales Data
9. Description: Sync Schemas - DIPC_SRC to DIPC_TGT

10. Select your source connection and schema
    a. Connection: Sync Source
    b. Schema: DIPC_SRC
11. Select your target connection and schema
    a. Connection: Sync Target
    b. Schema: DIPC_TGT
12. 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

Note: If you run into any issues when trying to select a Connection refresh the page manually. The Schemas may take some time to appear as well, this is expected.

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

14. The Configure Entities screen helps you create Include or Exclude rules to define precisely which database objects will be moved over 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

15. Leave the Configure Entities settings as-is and click Cancel to go back to the previous screen
16. To better view the Demo Client as well as the DIPC monitoring, try to setup your screen as follows with the Demo Client at the top and the DIPC App below or have the two side by side. It is recommended to use two screens if possible.
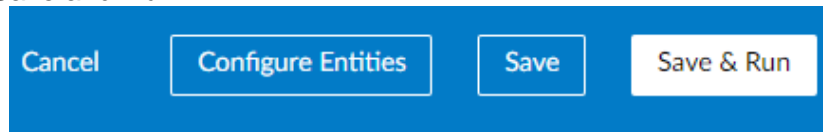

OR

17. Run Synchronize Data Task
    a. Click Save and Run

    b. A notification will appear that the job was saved.
    c. A new DIPC Job will be created to executed the task.
       A notification will appear in the notification bar as below:
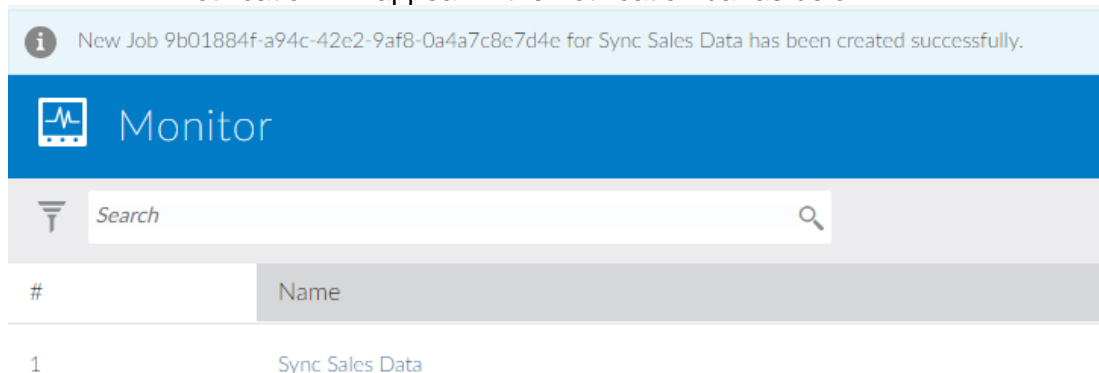
    d. The job will automatically appear within the Jobs page.
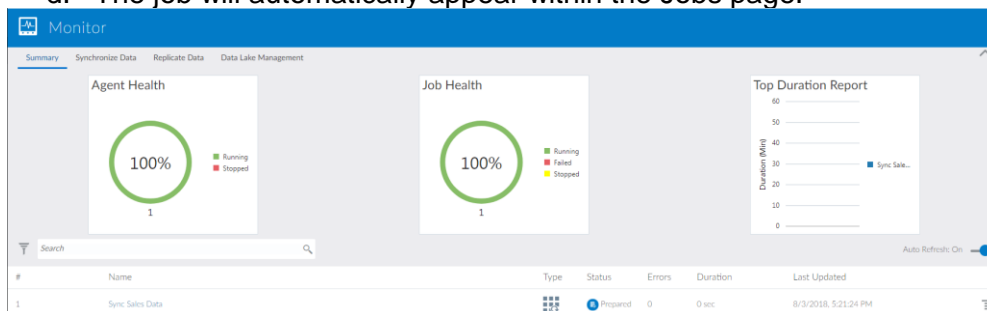
    e. Click on the Job to see the Job Details
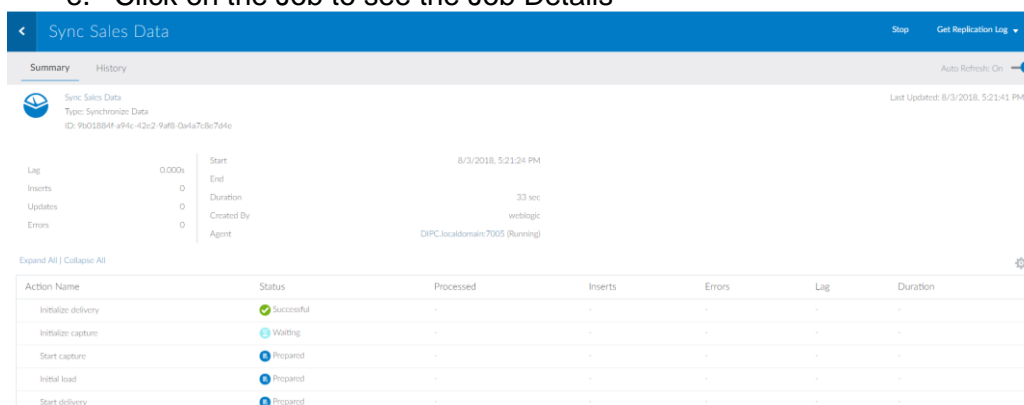
    f. Auto-refresh is on, statuses will be updated frequently
    g. As the job executes, the Initial Load process is created in ODI while DIPC configures OGG for the Source Capture and Target Delivery

h. As this job executes, the Replicated Sales OLTP Source Data table will be updated in the Demo Client. 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 for the Replicated Sales OLTP Data side to show anything. This is normal.

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

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

k. 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 7 minutes or more)

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

▲ Initial load	✅ Successful	557	557	0	-	181s
▲ Initial load_PROC	✅ Successful	557	557	0	-	181s
Procedure:Initial load_PROC:DROP DBLINK	⚠️ Warning	0	0	0	-	0s
Procedure:Initial load_PROC:CREATE DBLINK	✅ Successful	0	0	0	-	0s
Procedure:Initial load_PROC:SOURCE_ODI_VARIABLE	✅ Successful	0	0	0	-	0s
Procedure:Initial load_PROC:DBLINK_DATAPUMP	✅ Successful	0	0	0	-	19s
Procedure:Initial load_PROC:DBLINK_DATAPUMP_ASYNC_METRICS	✅ Successful	557	557	0	-	162s
Procedure:Initial load_PROC:TARGET_ODI_VARIABLE	✅ Successful	0	0	0	-	0s
Procedure:Initial load_PROC:DROP DBLINK	✅ Successful	0	0	0	-	0s

m. 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

Lag	4.649s	Start	
Inserts	462	End	
Updates	0	Duration	
Errors	0	Created By	
Warnings	0	Agent	

Expand All | Collapse All

Information Notification ✕

i Procedure:Initial load_PROC:DBLINK_DATAPUMP

```
declare
l_dp_job_name varchar2(400);

--local functions/procedures
FUNCTION add_quotes(p_str varchar2)
return varchar2 is begin
return '"' || p_str || '"';
end add_quotes;

function import_dblink_job_asynch
(
param_src_schema in varchar2 default 'SCOTT',
param_src_table_filter_expr in varchar2 default null,
param_src_table_exclude_expr in varchar2 default null,
```

Done

Action Name			
Initialize capture			
Initialize delivery			
Start capture			
▲ Initial load			
▲ Initial load_PROC			
Procedure:Initial load_PROC:DROP DBLINK	⚠️ Warning	0	0
Procedure:Initial load_PROC:CREATE DBLINK	✅ Successful	0	0
Procedure:Initial load_PROC:ODI_VARIABLE	✅ Successful	0	0
Procedure:Initial load_PROC:DBLINK_DATAPUMP	✅ Successful	0	0
Procedure:Initial load_PROC:DBLINK_DATAPUMP_METRICS	✅ Successful	462	462
Procedure:Initial load_PROC:DROP DBLINK	✅ Successful	0	0

Task 3: Validate DIPC data Synchronization Job

At this point DIPC has created and orchestrated the initial load and the data synchronization process for the OLTP source and operational data store. A new DIPC job is created and executed and can be monitored with the Jobs view.

1. You can also monitor the process by monitoring the DIPC Agent terminal window. You could also monitor the DIPC Managed Server, however there is a lot of detailed information here
2. Click Jobs and refresh until the newly created DIPC Job shows up.
   a. You should see 5 Actions
3. The DIPC agent is orchestrating this process. Additional details can be seen in the GG logs as well as within ODI Studio.
4. Before doing anything else monitor GG and watch for extracts and replicats to be created and started and RUNNING
   a. Open Terminal on DIPC host machine
   b. cd /home/DIPC
   c. Execute ./GGINFO_ALL.sh
   d. Ensure both Extract and Replicat are running

```
[DIPC@localhost ~]$ ./GGINFO_all.sh

Oracle GoldenGate Command Interpreter for Oracle
Version 12.2.0.1.170221 OGGCORE_12.2.0.1.00GGBP_PLATFORMS_170123.1033_FBO
Linux, x64, 64bit (optimized), Oracle 12c on Jan 23 2017 23:22:14
Operating system character set identified as UTF-8.

Copyright (C) 1995, 2017, Oracle and/or its affiliates. All rights reserved.



GGSCI (localhost.localdomain) 1>
Program     Status      Group       Lag at Chkpt  Time Since Chkpt

MANAGER     RUNNING
JAGENT      STOPPED
EXTRACT     RUNNING     EXT0000     00:00:02      00:00:09
REPLICAT    RUNNING     REPL0000    00:00:00      00:00:06


GGSCI (localhost.localdomain) 2> [DIPC@localhost ~]$
```

5. Once all rows are green in the Demo Client, proceed to next step

Task 4: Monitor Source Inserts/Updates/Deletes

1. Now that the initial load is complete and the extract/replicat 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 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. Once the synchronization is done, the target Sales DW schema is updated by a scenario that was created in the Standalone ODI Instance included with DIPC. DIPC and standalone ODI will work hand in hand to implement the end-to-end data flow.

Note: We will be using an ODI Execution Task in the next Lab to kick off the execution of that scenario

Note: If the window is too small through the Ravello Console you can also use VNC to log into the VM. Use your favorite VNC client and enter <Ravello Instance Hostname>:5901 as the URL to connect to.
When prompted enter the password: #DIPCR0CKS#

Note: If you are unable to connect to VNC, you can start vncserver manually on the host using the vncserver command.

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.
9. Both rows should show as yellow
10. To perform a delete click last row and click on the icon. Notice this record will be delete from the replicated schema as well.

11. [Optional] Go to the Jobs page and click on the menu and Stop to stop the Capture and Delivery processes.

OR
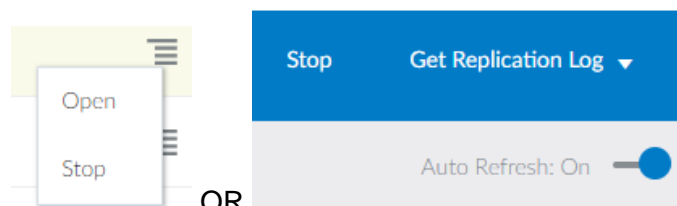
Summary:

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.