

Data Engineering-Final Project

Part-2

Creating database, schema and checking if it exists.

```
1  --Create a Database for the Term Project
2  CREATE OR REPLACE DATABASE DE_Final;
3  --Create a Schema within the above DB for the Term Project
4  CREATE OR REPLACE SCHEMA DE_Final.STAGE_DATA;
5  -- Checking for schema if not exist.
6  CREATE SCHEMA IF NOT EXISTS STAGE_DATA;
7  --so the schema exists.
8  --Now we will be making the dimension tables,
```

Results

Chart

	status
1	STAGE_DATA already exists, statement succeeded.

-Create 3 Dimension Tables in your Stage Schema in Snowflake.

- Customer_dim

```
10  CREATE TABLE Customer_Dim (
11      CustomerID INT PRIMARY KEY,
12      FirstName VARCHAR(255),
13      LastName VARCHAR(255),
14      Email VARCHAR(255),
15      Phone VARCHAR(20),
16      CustAddress VARCHAR(255),
17      CustState VARCHAR(50),
18      CustCity VARCHAR(50),
19      CustZip VARCHAR(10));
20
21  --creating product dimension table
```

Results

Chart

	status
1	Table CUSTOMER_DIM successfully created.

- Product_Dim-

```

21      --creating product dimension table.
22      CREATE TABLE Product_Dim (
23          ProductID INT PRIMARY KEY,
24          Brand VARCHAR(255),
25          Name VARCHAR(255),
26          Description TEXT,
27          Price DECIMAL(10, 2),
28          MSRP DECIMAL(10, 2),
29          Cost DECIMAL(10, 2)
30      );
31

```

Results Chart

status	
1	Table PRODUCT_DIM successfully created.

- Order_dim-

```

43      --Creating order dimension table.
44      CREATE TABLE OrderDetail_Dim (
45          OrderDetailID INT PRIMARY KEY,
46          ProductID INT,
47          OrderID INT,
48          FOREIGN KEY (ProductID) REFERENCES Product_Dim(ProductID),
49          FOREIGN KEY (OrderID) REFERENCES Orders_Fact(OrderID)
50      );
51
52
53      --reading the Customer_Dim table data loaded into the snowflake

```

Results Chart

status	
	Table ORDERDETAIL_DIM successfully created.


- Creating Fact Table-

```

32      --creating order fact table.
33      CREATE TABLE Orders_Fact (
34          OrderID INT PRIMARY KEY,
35          OrderAmt DECIMAL(10, 2),
36          OrderQty INT,
37          CustomerID INT,
38          OrderDT TIMESTAMP,
39          FOREIGN KEY (CustomerID) REFERENCES Customer_Dim(CustomerID)
40      );
41
42
43      --Creating order dimension table.

```

Results Chart



Object 'ORDERS_FACT' already exists.

Performing Data Cleaning and transformation.-

Reading data from all the tables

- Customer table

```
52
53 --reading the Customer_Dim table data loaded into the snowflake
54 | select * from Customer_Dim;
55
56 --reading the table data loaded into the snowflake
57 | select * from Product_Dim;
58
59 --reading the table data loaded into the snowflake
```

Results Chart

	CUSTOMERID	FIRSTNAME	LASTNAME	EMAIL	PHONE	CUSTADDRESS	CUSTSTATE
1	1	Penelope	Keddie	pkeddie0@behance.net	205-151-7125	3 8th Crossing	Alabama
2	2	Trip	Malyj	tmalyj1@oaic.gov.au	404-614-8569	13 Mcguire Road	Georgia
3	3	Jake	Gorger	jgorger2@springer.com	208-562-9998	68153 Maryland Center	Idaho
4	4	Johnath	Game	jgame3@shareasale.com	410-276-1726	7994 Killdeer Parkway	Maryland
5	5	Annela	Nattriss	anattriss4@chinustat.com	513-792-4285	02 Maryland Hill	Ohio

- Product_Dim

```
56 --reading the table data loaded into the snowflake
57 | select * from Product_Dim;
58
59 --reading the table data loaded into the snowflake
```

Results Chart

PRODUCTID	BRAND	NAME	DESCRIPTION	PRICE	MSRP	COST
1	TechWave	SpeedBoost Wireless Charging Pad	Fast - efficient - wireless power.	39.65	44.41	14.67
2	TechWave	PixelGrip Gaming Mouse	Precision control for gamers	26.55	29.74	9.82
3	TechWave	SoundSync Bluetooth Earbuds	Seamless audio freedom anytime	69.99	78.39	25.9
4	TechWave	UltraView HD Webcam	Crystal-clear video conferencing	32.42	36.31	12

Order_Detail_Dim

```
59 --reading the table data loaded into the snowflake
60 | select * from OrderDetail_Dim;
61
62 --reading the order fact table data loaded into the snowflake
63 | select * from Orders_Fact;
64
65
66
```

Results Chart

ORDERDETAILID	PRODUCTID	ORDERID
1	213	16
2	60	6
3	100	10

- Fact Table

```

62 --reading the order fact table data loaded into the snowflake
63 select * from Orders_Fact;
64
65
66

```

	ORDERID	ORDERAMT	ORDERQTY	CUSTOMERID	ORDERDT
1	1	543.5	3	20	2022-07-14 15:43:22.000
2	2	138.62	4	156	2022-09-25 08:25:31.000

- Checking for Null values in Customer_Dim.

```

67 --Checking for missing values in customer_dim
68 -- Customer_Dim
69 SELECT
70     COUNT(*) AS total_rows,
71     COUNT(CustomerID) AS customer_id_count,
72     COUNT(FirstName) AS first_name_count,
73     COUNT(LASTNAME) AS last_name_count,
74     COUNT(EMAIL) AS email_count,
75     COUNT(PHONE) AS Phone_count,
76     COUNT(CUSTADDRESS) AS Address_count,
77     COUNT(CUSTSTATE) AS custstate_count,
78     COUNT(custcity) AS city_count,
79     COUNT(custzip) AS zip_count,
80 FROM Customer_Dim;
81

```

TOTAL_ROWS	CUSTOMER_ID_COUNT	FIRST_NAME_COUNT	LAST_NAME_COUNT	EMAIL_COUNT	PHONE_COUNT	AI
200	200	200	200	200	191	

Only the customer_dim table has null values in 5 columns. We will be doing more analysis and data cleaning on customer_dim. --sales fact table, Orderdetail table, productdim does not have any null values

- I will Change the Null values in phone to NA.

```

83 update customer_dim
84 set phone = 'NA'
85 WHERE phone is null;
86

```

number of rows updated
9

- Checking for Duplicates in Customer Dim.

```

87 select distinct(phone) from customer_dim ;
88 select * from customer_dim
89 where CUSTADDRESS is null
90 and custstate is null
91 and custcity is null
92 and custzip is null ;
93
94

```

	CUSTOMERID	FIRSTNAME	LASTNAME	EMAIL	PHONE	CUSTADDRESS	CUSTSTATE	CUSTCITY
1	100	Tyler	Erasmus	terasmus2r@cbslocal.com	727-425-1723	null	null	null

Deleting the Duplicates in Customer.

```

96 DELETE FROM customer_dim WHERE customerid = '100';
97
98

```

	number of rows deleted
1	1

- Checking if the row is deleted. Query Produced no result.

```

99 -- checking if the customer_id is deleted
00 select * from customer_dim
01 where customerid = '100';
02
03 --we see that certain states have abbreviations in custstate column, this needs to be changed to full r

```

CUSTOMERID	FIRSTNAME	LASTNAME	EMAIL	PHONE	CUSTADDRESS	CUSTSTATE	CUSTCITY
------------	-----------	----------	-------	-------	-------------	-----------	----------

Query produced no results

- Changing the state abbreviations to full names-

```

--updating state abbreviations to full name, there are 3 states which needs to be updated
update customer_dim
set custstate = 'Texas'
WHERE custstate = 'TX';

--updating for AZ to Arizona
update customer_dim
set custstate = 'Arizona'
WHERE custstate = 'AZ';

update customer_dim
set custstate = 'California'
WHERE custstate = 'CA';

```

- Quality Check-

```

155 --looking for duplicates on customer_dim
156 -- Customer_Dim
157 SELECT
158     CustomerID, FirstName, LastName, Email, Phone, CustAddress, CustState, CustCity, CustZip,
159     COUNT(*)
160 FROM Customer_Dim
161 GROUP BY CustomerID, FirstName, LastName, Email, Phone, CustAddress, CustState, CustCity, CustZip
162 HAVING COUNT(*) > 1;
163

```

Results Chart

CUSTOMERID	FIRSTNAME	LASTNAME	EMAIL	PHONE	CUSTADDRESS	CUSTSTATE	CUSTCITY
------------	-----------	----------	-------	-------	-------------	-----------	----------

```

164 --looking for duplicates on Order_fact_dim
165 -- Order_Fact Dim
166 SELECT
167     OrderDetailID, ProductID, OrderID,
168     COUNT(*)
169 FROM OrderDetail_Dim
170 GROUP BY OrderDetailID, ProductID, OrderID
171 HAVING COUNT(*) > 1;
172

```

Results Chart

ORDERDETAILID	PRODUCTID	ORDERID
---------------	-----------	---------

Query produced no results

```

--looking for duplicates on Product_dim
-- Product Dim
SELECT
    ProductID, Brand, Name, Description, Price, MSRP,
    Cost,
    COUNT(*)
FROM PRODUCT_DIM
GROUP BY ProductID, Brand, Name, Description, Price, MSRP,
    Cost
HAVING COUNT(*) > 1;

```

Results Chart

PRODUCTID	BRAND	NAME	DESCRIPTION	PRICE
-----------	-------	------	-------------	-------

Query produced no results

Creating the Prod_DE_Final schema-

185	--Creating Prod_DE_Final_schema.,
186	CREATE SCHEMA IF NOT EXISTS Prod_DE_Final;
187	USE SCHEMA Prod_DE_Final;
188	
189	
190	INSERT INTO DE_Final.Prod_DE_Final.Customer_Dim

	status
1	Statement executed successfully.

- Creating Customer Dim in Prod table

191	--Creating Customer_Fact_Table for Prod_DE_Final.
192	CREATE TABLE IF NOT EXISTS CustomerDim (
193	CustomerID INT PRIMARY KEY,
194	FirstName VARCHAR(255),
195	LastName VARCHAR(255),
196	Email VARCHAR(255),
197	Phone VARCHAR(20),
198	CustAddress VARCHAR(255),
199	CustState VARCHAR(50),
200	CustCity VARCHAR(50),
201	CustZip VARCHAR(10),
202	RecordInserted TIMESTAMP DEFAULT CURRENT_TIMESTAMP(),
203	RecordSource VARCHAR(255)
204);

	status
1	CUSTOMERDIM already exists, statement succeeded.

- Inserting data from staged to prod.

206	--Insertinmg data from staged to prod in customer dim
207	USE DATABASE DE_Final;
208	
209	INSERT INTO Prod_DE_Final.CustomerDim (CUSTOMERID, FIRSTNAME, LASTNAME, EMAIL, PHONE, CUSTADDRESS, CUSTSTATE, CUSTCITY, CUSTZIP
210)
211	SELECT CUSTOMERID, FIRSTNAME, LASTNAME, EMAIL, PHONE, CUSTADDRESS, CUSTCITY, CUSTSTATE, CUSTZIP
212	FROM STAGE_DATA.CUSTOMER_DIM;
213	

	number of rows inserted	Query Details
1	199	Query duration

- Creating order detail dim in prod.

```

213
214 --creating orderdetail dim for Prod_DE_Final.
215 CREATE OR REPLACE TABLE orderdetail_dim (
216 OrderDetailID number(19,0) not null ,
217 OrderID number(19,0) not null,
218 ProductID number(19,0) not null,
219 copy_date TIMESTAMP NOT NULL,
220 primary key (OrderDetailID)
221 );
222
223

```

Results Chart

status	
1	Table ORDERDETAIL_DIM successfully created.

- Inserting Data from staged to prod in order detail dim.

```

225
226 USE DATABASE DE_Final;
227
228 INSERT INTO Prod_DE_Final.orderdetail_dim (OrderDetailID,OrderID,ProductID
229 )
230 SELECT OrderDetailID, ProductID, OrderID
231 FROM STAGE_DATA.OrderDetail_Dim;
232
233
234 use database DE_Final;
235 USE SCHEMA Prod_DE_Final;

```

Results Chart

number of rows inserted		Query Details
1	300	Query duration

```

234
235 use database DE_Final;
236 USE SCHEMA Prod_DE_Final;
237
238 --Creating product dim table for Prod_DE_Final.
239 CREATE TABLE IF NOT EXISTS Product_Dim (
240 ProductID INT PRIMARY KEY,
241 Brand VARCHAR(255),
242 Name VARCHAR(255),
243 Description TEXT,
244 Price DECIMAL(10, 2),
245 MSRP DECIMAL(10, 2),
246 Cost DECIMAL(10, 2),
247 RecordInserted TIMESTAMP DEFAULT CURRENT_TIMESTAMP(),
248 RecordSource VARCHAR(255)
249 );
250 USE DATABASE DE_Final;

```

Results Chart

status	
1	PRODUCT_DIM already exists, statement succeeded.

-
- Creating Product Dim table and inserting data into it from staged.,


```

237 --Creating product dim table for Prod_DE_Final.
238 CREATE TABLE IF NOT EXISTS Product_Dim (
239     ProductID INT PRIMARY KEY,
240     Brand VARCHAR(255),
241     Name VARCHAR(255),
242     Description TEXT,
243     Price DECIMAL(10, 2),
244     MSRP DECIMAL(10, 2),
245     Cost DECIMAL(10, 2),
246     RecordInserted TIMESTAMP DEFAULT CURRENT_TIMESTAMP(),
247     RecordSource VARCHAR(255)
248 );
249 USE DATABASE DE_Final;
250
251 INSERT INTO Prod_DE_Final.Product_Dim (ProductID,Brand,Name,Description,Price,MSRP,Cost
252 )
253 SELECT ProductID,Brand,Name,Description,Price,MSRP,Cost
254 FROM STAGE_DATA.Product_Dim;
255

```

Results

Chart

PREVIEW

number of rows inserted

1	20
---	----

Query Details

Query duration 686ms

Rows 1

Query ID 01b385f2-0105-1033-0...

- Selecting data from customerdim in the prod schema.

```

275
276 use database de_final;
277 use schema prod_de_final;
278 select * from customerdim ;

```

Results

Chart

PREVIEW

	CUSTOMERID	FIRSTNAME	LASTNAME	EMAIL	P
1	1	Penelope	Keddie	pkeddie0@behance.net	2
2	2	Trip	Malyj	tmalyj1@oaic.gov.au	4
3	3	Jake	Gorger	jgorger2@springer.com	2
4	4	Johnath	Game	jgame3@shareasale.com	4
5	5	Angela	Nattriss	anattriss4@shinystat.com	5
6	8	Carlina	Shilliday	cshilliday7@shareasale.com	8
7	9	Maribel	Guile	mguile8@techcrunch.com	8
8	11	Tatiana	Bonner	tbonnera@php.net	7
9	12	Nobe	Folli	nfollib@columbia.edu	4
10	13	Wallis	Rushbv	wrushbvc@imareshack.us	7

Query Details

Query duration

Rows

Query ID 01b385f5-0105

CUSTOMERID

1

FIRSTNAME

100% filled

-Quality check for missing values, which shows no null values in Prod schema.

```

281 use database de_final;
282 use schema prod_de_final;
283 SELECT
284     COUNT(*) AS total_rows,
285     COUNT(CustomerID) AS customer_id_count,
286     COUNT(FirstName) AS first_name_count,
287     COUNT(LASTNAME) AS last_name_count,
288     COUNT(EMAIL) AS email_count,
289     COUNT(PHONE) AS Phone_count,
290     COUNT(CUSTADDRESS) AS Address_count,
291     COUNT(CUSTSTATE) AS custstate_count,
292     COUNT(custcity) AS city_count,

```

Results Chart

	TOTAL_ROWS	CUSTOMER_ID_COUNT	FIRST_NAME_COUNT	LAST_NAM
1	398	398	398	

Successful Connection to **Tableau**.

CUSTOMER_DIM (DE_ECOMMERCE_SALES.CUSTOMER_DIM) (STA...

CUSTOMER_DIM

Need more data
 Drag tables here to relate the

Name	Customerid	First Name	Last Name
CUSTOMER_DIM	CUSTOMER_DIM	CUSTOMER_DIM	CUSTOMER_D