

# PROJECT REPORT

CAMP BI V3 Batch 1 Group 4

## INDEX

1. Project Description

2. Unit Test Report

a.	Creation of database
b.	Creation of schemas
c	Creation of table as per dataset
d.	Creation of Integration Object
e.	Creation of external stage for loading the data structure
f.	Creation of snowpipe for autoingesting of data from s3 bucket
g.	Creation of stream on the given table
h.	SCD 2 operation on the consumer table

**Supervised by: Pritam Gorain**

### **Group Members:**

- ❖ Aryan Kaundal
- ❖ Kajol Gupta
- ❖ Saurabh Pratap Singh
- ❖ Shivam Gupta
- ❖ Shubham Dixit
- ❖ Vrishti Saxena

## ■ **PROJECT DESCRIPTION**

The project involves data ingestion and analysis from public datahub Kaggle [link](#).

- Steps involved in performing the data ingestion:
  - a. Loading data to external stage
  - b. Ingesting data into the landing schema
  - c. Ingesting data into the consumer table
  - d. Perform analysis on the given dataset

## **UNIT TEST REPORT**

1. Created a database named SF\_PROJECT;

The screenshot shows a SQL query editor interface. The query is as follows:

```
DEMO_DB.PUBLIC ▾  
1 | create database sf_project;  
2 |  
3 | create schema itr_rds;  
4 |  
5 | create schema itr_rds_landing;  
6 |  
7 | create schema itr_dis;  
8 |  
9 | create table superstore (  
10 |     row_id int,  
11 |     order_id varchar,  
12 |     order_date date,  
13 |     ship_date date,  
14 |     ship_mode string,  
15 |     customer_id varchar,  
16 |     customer_name string,  
17 |     segment string,  
18 |     city string);
```

Below the query editor, there is a navigation bar with tabs: Objects, Query, Results, and Chart. The Results tab is active, showing a table with the following data:

	status
1	Database SF_PROJECT successfully created.

On the right side, there is a Query Details panel showing the following information:

- Query duration: 87ms
- Rows: 1
- status: 100% filled

At the bottom right, there is a watermark that says "Activate Windows. Go to Settings to activate Windows."

## 2. Created three schemas named ITR\_RDS, ITR\_RDS\_LANDING and ITR\_DIS

```
SF_PROJECT.ITR_RDS_LANDING ▾  
  
1 create database sf_project;  
2  
3 create schema itr_rds;  
4  
5 create schema itr_rds_landing;  
6  
7 create schema itr_dis;  
8  
9 show schemas;  
10  
11  
12 create table superstore (  
13     row_id int,  
14     order_id varchar...
```

Objects Query Results Chart

	created_on	name	is_default	is_current	database_name	owner
1	2022-03-24 03:30:38.139 -0700	INFORMATION_SCHEMA	N	N	SF_PROJECT	
2	2022-03-24 02:40:40.699 -0700	ITR_DIS	N	N	SF_PROJECT	ACCOUNTA
3	2022-03-24 02:40:23.550 -0700	ITR_RDS	N	N	SF_PROJECT	ACCOUNTA
4	2022-03-24 02:40:33.384 -0700	ITR_RDS_LANDING	N	Y	SF_PROJECT	ACCOUNTA
5	2022-03-24 02:40:15.910 -0700	PUBLIC	N	N	SF_PROJECT	ACCOUNTA

Query Details

Query duration 208ms

Rows 5

created\_on 100% filled

name 100% filled

Activate Windows  
Go to Settings to activate Windows.

## 3. Created a table named SUPERTORE as per the data set

```
SF_PROJECT.ITR_RDS_LANDING ▾  
  
27 sales double,  
28 quantity int,  
29 discount double,  
30 profit double  
31 );  
32  
33 desc table superstore;  
34
```

Objects Query Results Chart

	name	type	kind	null?	default	primary key	unique key	che
1	ROW_ID	NUMBER(38,0)	COLUMN	Y	null	N	N	null
2	ORDER_ID	VARCHAR(16777216)	COLUMN	Y	null	N	N	null
3	ORDER_DATE	DATE	COLUMN	Y	null	N	N	null
4	SHIP_DATE	DATE	COLUMN	Y	null	N	N	null
5	SHIP_MODE	VARCHAR(16777216)	COLUMN	Y	null	N	N	null
6	CUSTOMER_ID	VARCHAR(16777216)	COLUMN	Y	null	N	N	null
7	CUSTOMER_NAME	VARCHAR(16777216)	COLUMN	Y	null	N	N	null
8	SEGMENT	VARCHAR(16777216)	COLUMN	Y	null	N	N	null
9	CITY	VARCHAR(16777216)	COLUMN	Y	null	N	N	null
10	COUNTRY	VARCHAR(16777216)	COLUMN	Y	null	N	N	null
11	STATE	VARCHAR(16777216)	COLUMN	Y	null	N	N	null
12	POSTAL_CODE	NUMBER(38,0)	COLUMN	Y	null	N	N	null
13	REGION	VARCHAR(16777216)	COLUMN	Y	null	N	N	null

Query Details

Query duration 42ms

Rows 21

name 100% filled

type 100% filled

Activate Windows  
Go to Settings to activate Windows.

## 4. Created Integration object named SUPER\_STORE

```
SF_PROJECT.ITR_RDS_LANDING ▾  
  
31 );  
32  
33 desc table superstore;  
34  
35 create or replace storage integration superstore_int  
36 type = external_stage  
37 storage_provider = s3  
38 enabled = true  
39 storage_aws_role_arn = 'arn:aws:iam::700674097429:role/flatbucket4_policy_role'  
40 storage_allowed_locations = ('s3://flatbucket4/');  
41  
42 DESC INTEGRATION superstore_int;  
43
```

Objects Query Results Chart

	property	property_type	property_value	...	property_d
1	ENABLED	Boolean	true		false
2	STORAGE_PROVIDER	String	S3		
3	STORAGE_ALLOWED_LOCATIONS	List	s3://flatbucket4/		[]
4	STORAGE_BLOCKED_LOCATIONS	List			[]
5	STORAGE_AWS_IAM_USER_ARN	String	arn:aws:iam::122191154513:user/hygv-s-insa5198		
6	STORAGE_AWS_ROLE_ARN	String	arn:aws:iam::700674097429:role/flatbucket4_policy_role		
7	STORAGE_AWS_EXTERNAL_ID	String	GT45806_SFCRole=3_CNGUAPIW5d12F/D3W4g14ZMZ4TY=		
8	COMMENT	String			

Query Details ...

Query duration 199ms

Rows 8

property Aa

100% filled

property\_type Aa

String 5

List 2

Boolean 1

## 5. Created external stage named SUPERSTORE\_STAGE for loading data structures

```
SF_PROJECT.ITR_RDS_LANDING ▾  
  
33 desc table superstore;  
34  
35 create or replace storage integration superstore_int  
36 type = external_stage  
37 storage_provider = s3  
38 enabled = true  
39 storage_aws_role_arn = 'arn:aws:iam::700674097429:role/flatbucket4_policy_role'  
40 storage_allowed_locations = ('s3://flatbucket4/');  
41  
42 DESC INTEGRATION superstore_int;  
43  
44 create or replace stage sf_project.itr_rds_landing.superstore_stage  
45 storage_integration = superstore_int  
46 url = 's3://flatbucket4/'  
47 file_format = (type = csv field_delimiter=',' skip_header = 1 null_if = ('NULL','null') empty_field_as_null = true  
48 field_optionally_enclosed_by='');  
49 list @sf_project.itr_rds_landing.superstore_stage;  
50
```

Objects Query Results Chart

	name	size	md5	last_modified
1	s3://flatbucket4/Sample - Superstore.csv	2,287,806	b3066905e9eb4d477453d473dbf20c02	Tue, 22 Mar 2022 08:06:12 GMT
2	s3://flatbucket4/disk1.csv	24,400	687c8a7896d603b6c87d5932db7672be	Tue, 22 Mar 2022 10:17:23 GMT
3	s3://flatbucket4/superstore.csv	69,463	97b1f8dda821eb1355cfcfa9aafbe223	Tue, 22 Mar 2022 09:35:38 GMT
4	s3://flatbucket4/test1.csv	23,139	e85b97976f637b0a4a56ed6ee0792513b	Thu, 24 Mar 2022 09:25:39 GMT

Query Details ...

Query duration 2.2s

Rows 4

name Aa

100% filled

## 6. Created a snowpipe named SUPERSTORE\_PIPE for autoingesting the data from S3 bucket – flatbucket4

```
SF_PROJECT.ITR_RDS_LANDING ▾  
  
41  
42 DESC INTEGRATION superstore_int;  
43  
44 create or replace stage sf_project.itr_rds_landing.superstore_stage  
45     storage_integration = superstore_int  
46     url = 's3://flatbucket4'  
47     file_format = (type = csv field_delimiter=',' skip_header = 1 null_if = ('NULL','null') empty_field_as_null = true  
         field_optionally_enclosed_by='');  
48  
49 list @sf_project.itr_rds_landing.superstore_stage;  
50  
51 create or replace pipe sf_project.itr_rds_landing.superstore_pipe auto_ingest=true as  
52     copy into sf_project.itr_rds_landing.superstore  
53     from @sf_project.itr_rds_landing.superstore_stage;  
54  
55 show pipes;  
56  
57 alter pipe sf_project.itr_rds_landing.superstore_pipe refresh;  
58
```

Objects Query Results Chart

	File	Status
1	Sample - Superstore.csv	SENT
2	disk1.csv	SENT
3	superstore.csv	SENT
4	test1.csv	SENT

Query Details ...  
Query duration 2.2s  
Rows 4

File 100% filled  
Go to 100% filled to activate Windows.

## 7. Created stream SUPERSTORE\_STREAM on table SUPERSTORE

```
SF_PROJECT.ITR_RDS_LANDING ▾  
  
44 create or replace stage sf_project.itr_rds_landing.superstore_stage  
45     storage_integration = superstore_int  
46     url = 's3://flatbucket4'  
47     file_format = (type = csv field_delimiter=',' skip_header = 1 null_if = ('NULL','null') empty_field_as_null = true  
         field_optionally_enclosed_by='');  
48  
49 list @sf_project.itr_rds_landing.superstore_stage;  
50  
51 create or replace pipe sf_project.itr_rds_landing.superstore_pipe auto_ingest=true as  
52     copy into sf_project.itr_rds_landing.superstore  
53     from @sf_project.itr_rds_landing.superstore_stage;  
54  
55 show pipes;  
56  
57 alter pipe sf_project.itr_rds_landing.superstore_pipe refresh;  
58  
59 select * from sf_project.itr_rds_landing.superstore;  
60  
61 create or replace stream superstore_stream on table superstore;  
62
```

Objects Query Results Chart

	status
1	Stream SUPERSTORE_STREAM successfully created.

Query Details ...  
Query duration 215ms  
Rows 1

status 100% filled  
Go to status 100% filled to activate Windows.

8. Performed SCD 2 operation on consumer table SUPERSTORE\_TARGET as per changes that happen in the source table SUPERSTORE

```
SF_PROJECT.ITR_RDS_LANDING
,s.quantity ,s.discount,s.profit,metadata$action,0,CURRENT_TIMESTAMP() );
111
112 ALTER TASK superstore_task RESUME;
113
114 update superstore set region='northeast' where row_id=1;
115 delete from superstore where row_id =2;
116 update superstore set region='southwest' where row_id=2;
117
118 select * from superstore_target;
119 select * from superstore_stream;
120
121 select row_id,order_id,region,stream_type,rec_version from superstore_target;
122
123
```

	ROW_ID	ORDER_ID	REGION	STREAM_TYPE	...	REC_VERSION
1	1	CA-2016-152156	northeast	INSERT		0
2	1	CA-2016-152156	South	INSERT		-1
3	2	CA-2016-152156	South	DELETE		9,999
4	3	CA-2016-138688	West	INSERT		0
5	4	US-2015-108966	South	INSERT		0
6	5	US-2015-108966	South	INSERT		0
7	6	CA-2014-115812	West	INSERT		0

Query Details

Query duration 128ms

Rows 100

ROW\_ID 123

1 99

ORDER\_ID

## ▪ **DATA ANALYSIS ON THE GIVEN DATASET**

--1. Calculate the number of Orders those with Ship Mode as 'Second Class'

SELECT COUNT(\*), SHIP\_MODE FROM superstore WHERE SHIP\_MODE= 'Second Class' GROUP BY SHIP\_MODE;

```
SF_PROJECT.ITR_RDS_LANDING
1 SELECT COUNT(*), SHIP_MODE FROM superstore WHERE SHIP_MODE= 'Second Class' GROUP BY SHIP_MODE;
2
3
4
```

	COUNT(*)	SHIP_MODE
1	19	Second Class

Query Details

Query duration 40ms

Rows 1

COUNT(\*) 123



## --2. List down the most valuable customers Country wise

SELECT COUNTRY, CUSTOMER\_NAME, COUNT(\*) FROM SUPERSTORE GROUP BY COUNTRY, CUSTOMER\_NAME ORDER BY COUNTRY, COUNT(\*) DESC;

SF\_PROJECT.ITR\_RDS\_LANDING ▾

```
1 SELECT COUNT(*), SHIP_MODE FROM superstore WHERE SHIP_MODE= 'Second Class' GROUP BY SHIP_MODE;
2
3 SELECT COUNTRY, CUSTOMER_NAME, COUNT(*) FROM SUPERSTORE GROUP BY COUNTRY, CUSTOMER_NAME ORDER BY COUNTRY, COUNT(*) DESC;
4
5 SELECT SUM(SALES) FROM SUPERSTORE WHERE CATEGORY = 'Furniture';
6
7
```

Objects Query Results Chart

	COUNTRY	CUSTOMER_NAME	COUNT(*)
1	Chicago	Christopher Schild	1
2	Chicago	Paul Stevenson	1
3	Columbia	Patrick O'Donnell	1
4	Concord	Andrew Allen	1
5	Decatur	Stewart Carmichael	2
6	Dover	Lena Hernandez	2
7	Durham	Julie Creighton	1
8	Eagan	Odella Nelson	2
9	Fort Lauderdale	Sean O'Donnell	2
10	Fort Worth	Harold Pawlan	2
11	Fremont	Ken Black	2
12	Gilbert	Brendan Sweed	2
13	Henderson	Claire Gute	1

Query Details ...

Query duration 32ms

Rows 49

COUNTRY Aa

Los Angeles 6

Houston 5

New York City 3

+ 31 more

CUSTOMER\_NAME Aa

Patrick O'Donnell 2

Joel Eaton 2

Tracy Blumstein 2

+ 43 more

## --3. Total Sales for Category 'Furniture'

SELECT SUM(SALES) FROM SUPERSTORE WHERE CATEGORY = 'Furniture';

SF\_PROJECT.ITR\_RDS\_LANDING ▾

```
1 SELECT COUNT(*), SHIP_MODE FROM superstore WHERE SHIP_MODE= 'Second Class' GROUP BY SHIP_MODE;
2
3 SELECT COUNTRY, CUSTOMER_NAME, COUNT(*) FROM SUPERSTORE GROUP BY COUNTRY, CUSTOMER_NAME ORDER BY COUNTRY, COUNT(*) DESC;
4
5 SELECT SUM(SALES) FROM SUPERSTORE WHERE CATEGORY = 'Furniture';
6
7 SELECT CATEGORY, SUM(PROFIT) FROM SUPERSTORE GROUP BY CATEGORY ORDER BY SUM(PROFIT) DESC;
```

Objects Query Results Chart

	CATEGORY	SUM(PROFIT)
1	Technology	785.6685
2	Office Supplies	741.1687
3	Furniture	-1,870.9283

Query Details ...

Query duration 160ms

Rows 3

CATEGORY Aa

100% filled

SUM(PROFIT) 123

#### --4. Which Product provides the maximum profit

SELECT CATEGORY, SUM(PROFIT) FROM SUPERSTORE GROUP BY CATEGORY ORDER BY SUM(PROFIT) DESC;

SF\_PROJECT.ITR\_RDS\_LANDING ▾

```
1 SELECT COUNT(*), SHIP_MODE FROM superstore WHERE SHIP_MODE= 'Second Class' GROUP BY SHIP_MODE;
2
3 SELECT COUNTRY, CUSTOMER_NAME, COUNT(*) FROM SUPERSTORE GROUP BY COUNTRY, CUSTOMER_NAME ORDER BY COUNTRY, COUNT(*) DESC;
4
5 SELECT SUM(SALES) FROM SUPERSTORE WHERE CATEGORY = 'Furniture';
6
7
```

Objects Query Results Chart

	SUM(SALES)
1	10,351.8397

Query Details

Query duration 36ms

Rows 1

SUM(SALES) 123

100% filled

#### --5. Calculate the total profit made by each product category country wise

SELECT COUNTRY,CATEGORY, SUM(PROFIT) FROM SUPERSTORE GROUP BY CATEGORY, COUNTRY ORDER BY COUNTRY;

SF\_PROJECT.ITR\_RDS\_LANDING ▾

```
2
3 SELECT COUNTRY, CUSTOMER_NAME, COUNT(*) FROM SUPERSTORE GROUP BY COUNTRY, CUSTOMER_NAME ORDER BY COUNTRY, COUNT(*) DESC;
4
5 SELECT SUM(SALES) FROM SUPERSTORE WHERE CATEGORY = 'Furniture';
6
7 SELECT CATEGORY, SUM(PROFIT) FROM SUPERSTORE GROUP BY CATEGORY ORDER BY SUM(PROFIT) DESC;
8
9 SELECT COUNTRY,CATEGORY, SUM(PROFIT) FROM SUPERSTORE GROUP BY CATEGORY, COUNTRY ORDER BY COUNTRY;
```

Objects Query Results Chart

	COUNTRY	CATEGORY	SUM(PROFIT)
1	Chicago	Furniture	-15.2225
2	Chicago	Office Supplies	-48.9549
3	Columbia	Furniture	33.2156
4	Concord	Office Supplies	5.4432
5	Decatur	Office Supplies	64.2384
6	Dover	Technology	11.054
7	Durham	Office Supplies	62.8075
8	Eagan	Office Supplies	8.2062
9	Eagan	Technology	19.7714
10	Fort Lauderdale	Office Supplies	2.5164
11	Fort Lauderdale	Furniture	-383.031
12	Fort Worth	Office Supplies	-127.674
13	Fremont	Office Supplies	20.748

Query Details

Query duration 165ms

Rows 53

COUNTRY Aa

Houston 3

Los Angeles 3

New York City 3

+ 31 more

CATEGORY Aa

Office Supplies 28

Furniture 14

Technology 11



--6. Which region of United States have majority loss?

SELECT REGION, MIN(PROFIT) FROM SUPERSTORE GROUP BY REGION;

SF\_PROJECT.ITR\_RDS\_LANDING

4

5

6

7

8

9

10

11

SELECT SUM(SALES) FROM SUPERSTORE WHERE CATEGORY = 'Furniture';

SELECT CATEGORY, SUM(PROFIT) FROM SUPERSTORE GROUP BY CATEGORY ORDER BY SUM(PROFIT) DESC;

SELECT COUNTRY,CATEGORY, SUM(PROFIT) FROM SUPERSTORE GROUP BY CATEGORY, COUNTRY ORDER BY COUNTRY;

SELECT REGION, MIN(PROFIT) FROM SUPERSTORE GROUP BY REGION;

Objects

Query

Results

Chart

1

2

3

4

5

	REGION	...	MIN(PROFIT)
1	northeast		41.9136
2	West		-3.788
3	South		-383.031
4	Central		-147.963
5	East		-1,665.0522

Query Details

...

Query duration1.2s

Rows5

REGIONAa

100% filled

MIN(PROFIT)123

-1,665.0522

41.9136

Activate Windows  
Go to Settings to activate Windows.