

Data Warehouse

By: Takdir Zulhaq Dessiaming



Overview

▶ Introduction	01
▶ Data Flow dan ERD	02
▶ Staging	03
▶ Mapping to DWH	04
▶ Target Output	05
▶ Dashboard	06



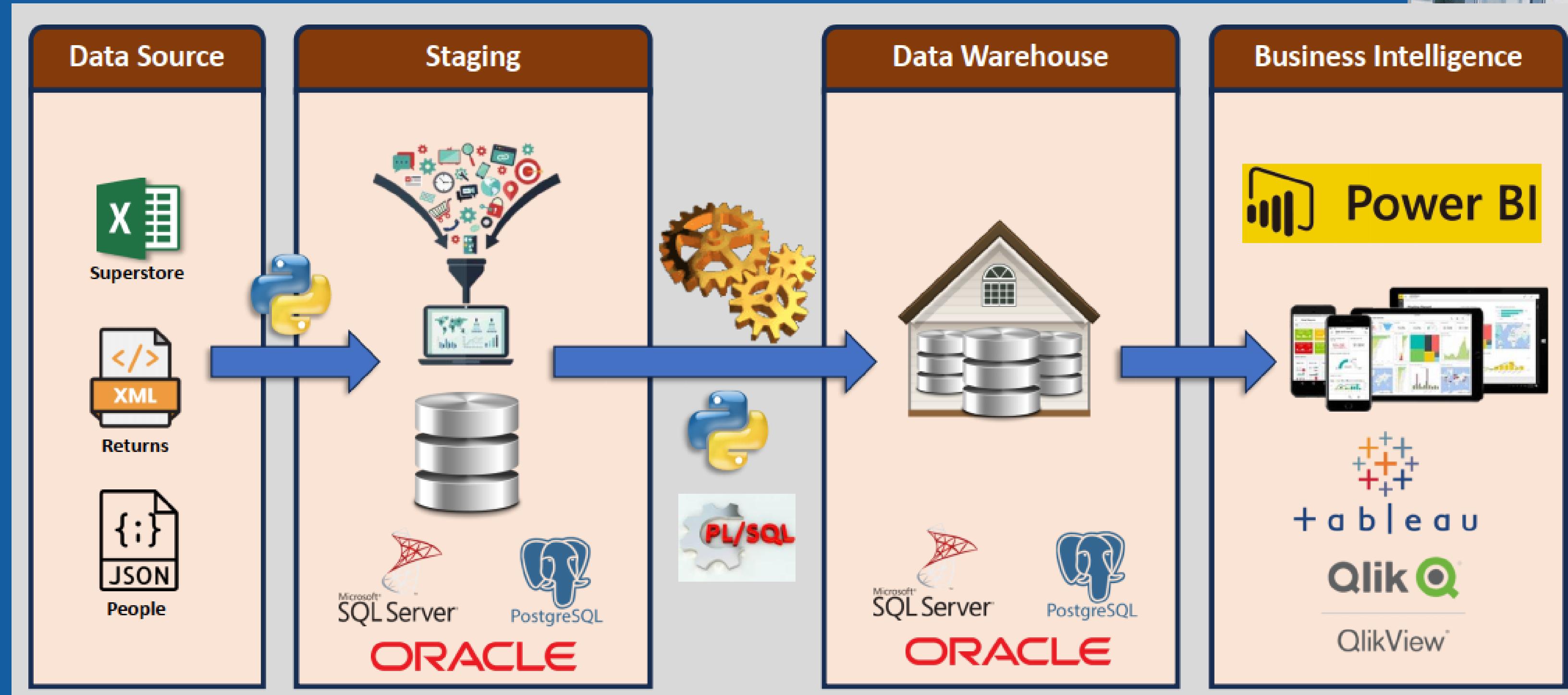


1. Introduction

Project ini bertujuan untuk membuat sistem Data Warehouse dengan menerapkan konsep ETL, dan menggunakan kombinasi bahasa pemrograman Python dan SQL.

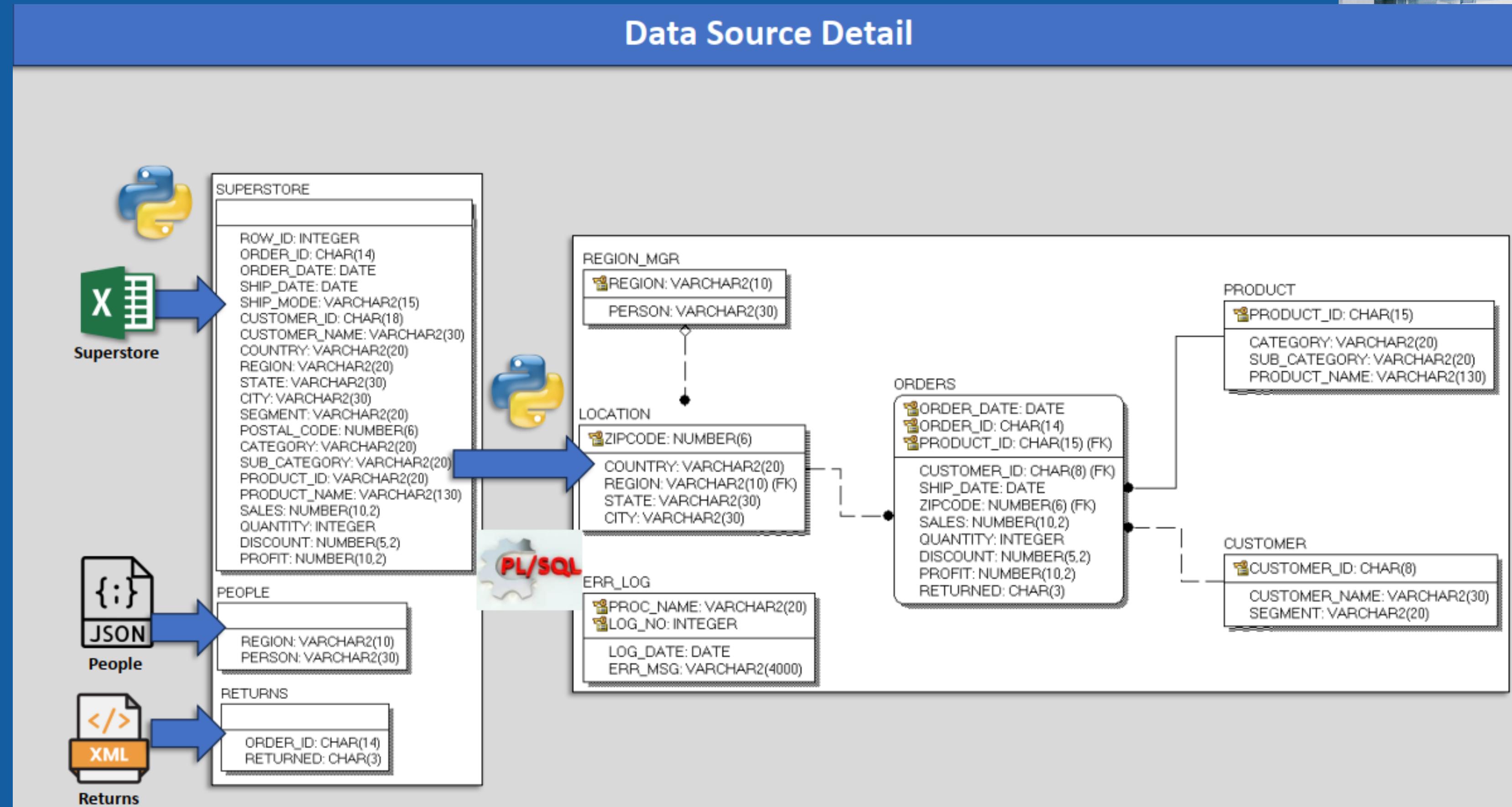
2. Data Flow

Data flow pada test ini adalah memasukkan source data (Excel, XML, dan JSON) ke dalam database yang kita anggap sebagai staging/tempat sementara. Lalu dari staging tersebut dilakukan proses ETL untuk menghasilkan model data yang diinginkan dan mengirimkan hasil model data tersebut ke database yang kita anggap sebagai Data Warehouse. Setelah itu akan disconnect ke PowerBI agar visualisasinya terotomatisasi



ERD

Berikut adalah relasi antar tabel yang menjadi dasar pada pembuatan Data Warehouse



3. Staging

People

	ABC Person	ABC Region
1	Anna Andreadi	West
2	Chuck Magee	East
3	Kelly Williams	Central
4	Cassandra Brandom	South

Returns

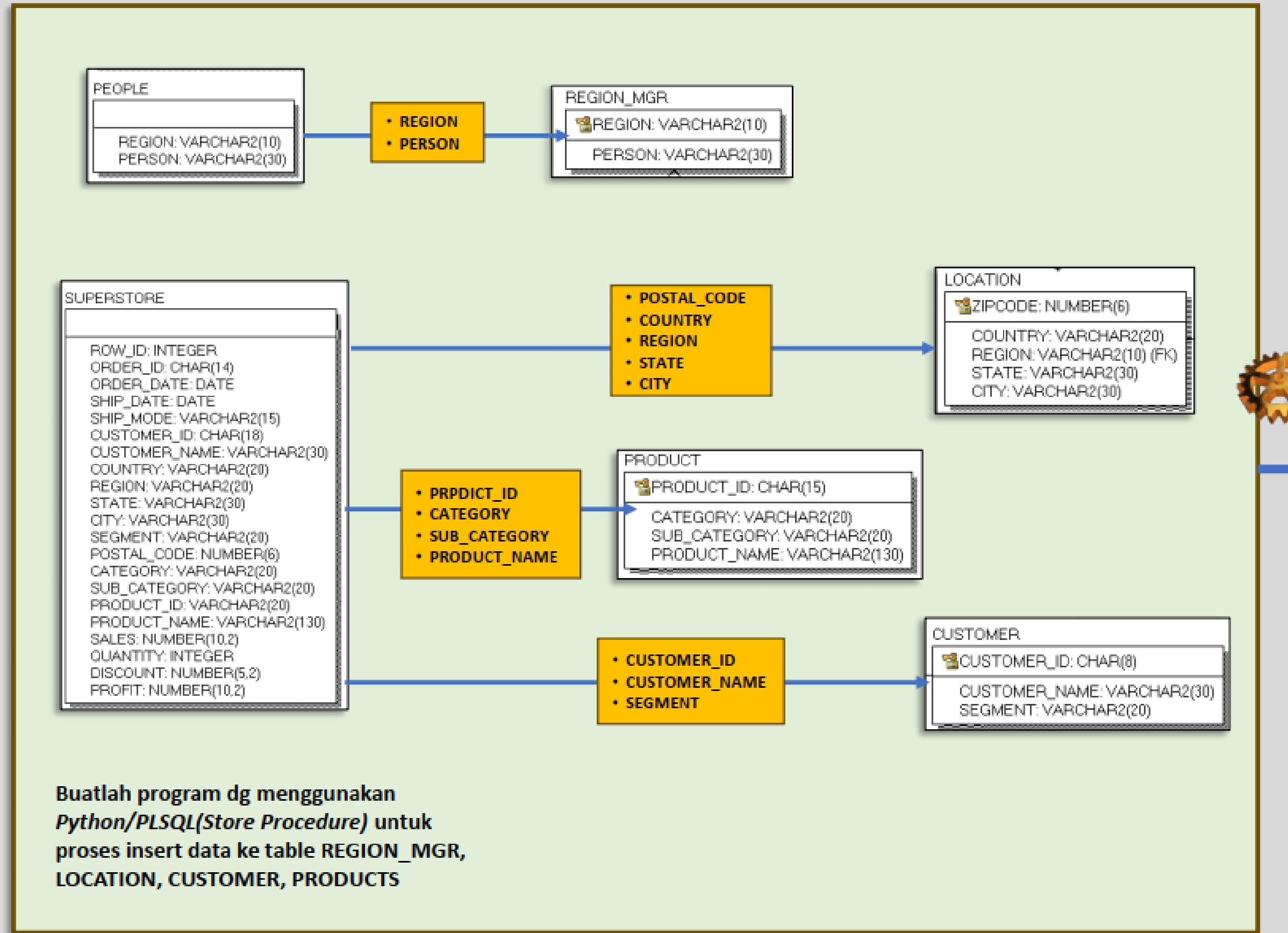
	ABC Order_ID	ABC Returned
1	CA-2014-153822	Yes
2	CA-2014-129707	Yes
3	CA-2011-152345	Yes
4	CA-2012-156440	Yes
5	US-2014-155999	Yes
6	CA-2011-157924	Yes

Superstore

	123 Row_ID	ABC Order_ID	⌚ Order_Date	⌚ Ship_Date	ABC Ship_Mode	ABC Customer_ID	ABC Customer_Name	ABC Segment	ABC Country	ABC City	ABC State	123 Postal_Code
1	1	CA-2013-152156	2013-11-09	2013-11-12	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42,420
2	2	CA-2013-152156	2013-11-09	2013-11-12	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42,420
3	3	CA-2013-138688	2013-06-13	2013-06-17	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California	90,030
4	4	US-2012-108966	2012-10-11	2012-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33,311
5	5	US-2012-108966	2012-10-11	2012-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33,311
6	6	CA-2011-115812	2011-06-09	2011-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90,030
7	7	CA-2011-115812	2011-06-09	2011-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90,030
8	8	CA-2011-115812	2011-06-09	2011-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90,030
9	9	CA-2011-115812	2011-06-09	2011-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90,030
10	10	CA-2011-115812	2011-06-09	2011-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90,030

ABC Region	ABC Product_ID	ABC Category	ABC Sub_Category	ABC Product_Name	123 Sales	123 Quantity	123 Discount	123 Profit
South	FUR-CH-10000454	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stackir	731.94	3	0	219.58
West	OFF-LA-10000240	Office Supplies	Labels	Self-Adhesive Address Labels for Type	14.62	2	0	6.87
South	FUR-TA-10000577	Furniture	Tables	Bretford CR4500 Series Slim Rectangul	957.58	5	0.45	-383.03
South	OFF-ST-10000760	Office Supplies	Storage	Eldon Fold 'N Roll Cart System	22.37	2	0.2	2.52
West	FUR-FU-10001487	Furniture	Furnishings	Eldon Expressions Wood and Plastic D	48.86	7	0	14.17
West	OFF-AR-10002833	Office Supplies	Art	Newell 322	7.28	4	0	1.97
West	TEC-PH-10002275	Technology	Phones	Mitel 5320 IP Phone VoIP phone	907.15	6	0.2	90.72

4. Mapping to DWH



Error Handling

SEMUA ERROR PADA SAAT PROSES INSERT KE TABLE ORDERS DITULIS KE TABLE ERR_LOG

PROC_NAME: VARCHAR2(20)
LOG_NO: INTEGER
LOG_DATE: DATE
ERR_MSG: VARCHAR2(4000)

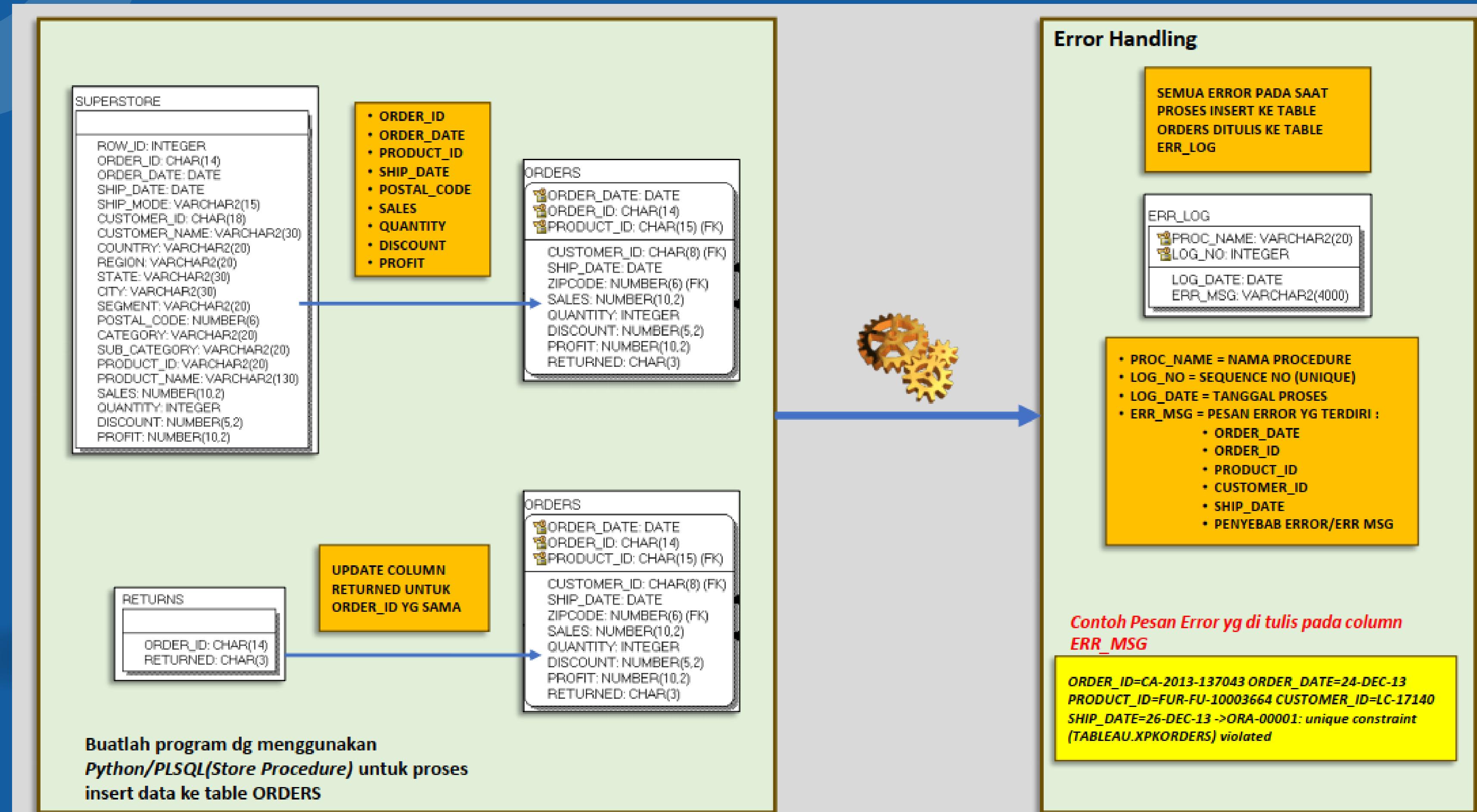
- PROC_NAME = NAMA PROCEDURE
- LOG_NO = SEQUENCE NO (UNIQUE)
- LOG_DATE = TANGGAL PROSES
- ERR_MSG = PESAN ERROR YG TERDIRI :

 - ORDER_DATE
 - ORDER_ID
 - PRODUCT_ID
 - CUSTOMER_ID
 - SHIP_DATE
 - PENYEBAB ERROR/ERR MSG

Contoh Pesan Error yg di tulis pada column ERR_MSG

```
ORDER_ID=CA-2013-137043 ORDER_DATE=24-DEC-13
PRODUCT_ID=FUR-FU-10003664 CUSTOMER_ID=LC-17140
SHIP_DATE=26-DEC-13 ->ORA-00001: unique constraint
(TABLEAU.XPKORDERS) violated
```

Mapping to DWH



5. Target Output

Customer

	ABC CUSTOMER_ID	ABC CUSTOMER_NAME	ABC SEGMENT
1	AA-10315	Alex Avila	Consumer
2	AA-10375	Allen Armold	Consumer
3	AA-10480	Andrew Allen	Consumer
4	AA-10645	Anna Andreadi	Consumer
5	AB-10015	Aaron Bergman	Consumer
6	AB-10060	Adam Bellavance	Home Office
7	AB-10105	Adrian Barton	Consumer

Location

	123 ZIPCODE	ABC COUNTRY	ABC REGION	ABC STATE	ABC CITY
1	1,040	United States	East	Massachusetts	Holyoke
2	1,453	United States	East	Massachusetts	Leominster
3	1,752	United States	East	Massachusetts	Marlborough
4	1,810	United States	East	Massachusetts	Andover
5	1,841	United States	East	Massachusetts	Lawrence
6	1,852	United States	East	Massachusetts	Lowell
7	1,915	United States	East	Massachusetts	Beverly

Product

	ABC PRODUCT_ID	ABC CATEGORY	ABC SUB_CATEGORY	ABC PRODUCT_NAME
1	FUR-BO-10000112	Furniture	Bookcases	Bush Birmingham Collection Bookcase, Dark Cherry
2	FUR-BO-10000330	Furniture	Bookcases	Sauder Camden County Barrister Bookcase, Planked
3	FUR-BO-10000362	Furniture	Bookcases	Sauder Inglewood Library Bookcases
4	FUR-BO-10000468	Furniture	Bookcases	O'Sullivan 2-Shelf Heavy-Duty Bookcases
5	FUR-BO-10000711	Furniture	Bookcases	Hon Metal Bookcases, Gray
6	FUR-BO-10000780	Furniture	Bookcases	O'Sullivan Plantations 2-Door Library in Landvery Oak
7	FUR-BO-10001337	Furniture	Bookcases	O'Sullivan Living Dimensions 2-Shelf Bookcases

Region Mgr

	ABC Person	ABC Region
1	Anna Andreadi	West
2	Chuck Magee	East
3	Kelly Williams	Central
4	Cassandra Bran	South

5. Target Output

Orders

	ORDER_DATE	ORDER_ID	PRODUCT_ID	CUSTOMER_ID	SHIP_DATE	ZIPCODE	SALES	QUANTITY	DISCOUNT	PROFIT	RETURNED
1	2011-09-07	CA-2011-100006	TEC-PH-10002075	DK-13375	2011-09-13	10,024	377.97	3	0	109.61	No
2	2011-07-08	CA-2011-100090	FUR-TA-10003715	EB-13705	2011-07-12	94,122	502.49	3	0.2	-87.94	No
3	2011-03-14	CA-2011-100293	OFF-PA-10000176	NF-18475	2011-03-18	32,216	91.06	6	0.2	31.87	No
4	2011-01-29	CA-2011-100328	OFF-BI-10000343	JC-15340	2011-02-04	10,024	3.93	1	0.2	1.33	No
5	2011-04-08	CA-2011-100363	OFF-FA-10000611	JM-15655	2011-04-15	85,301	2.37	2	0.2	0.83	No
6	2011-05-25	CA-2011-100391	OFF-PA-10001471	BW-11065	2011-05-29	10,035	14.62	2	0	6.73	No
7	2011-04-18	CA-2011-100678	OFF-AR-10001868	KM-16720	2011-04-22	77,095	2.69	2	0.2	1.01	No

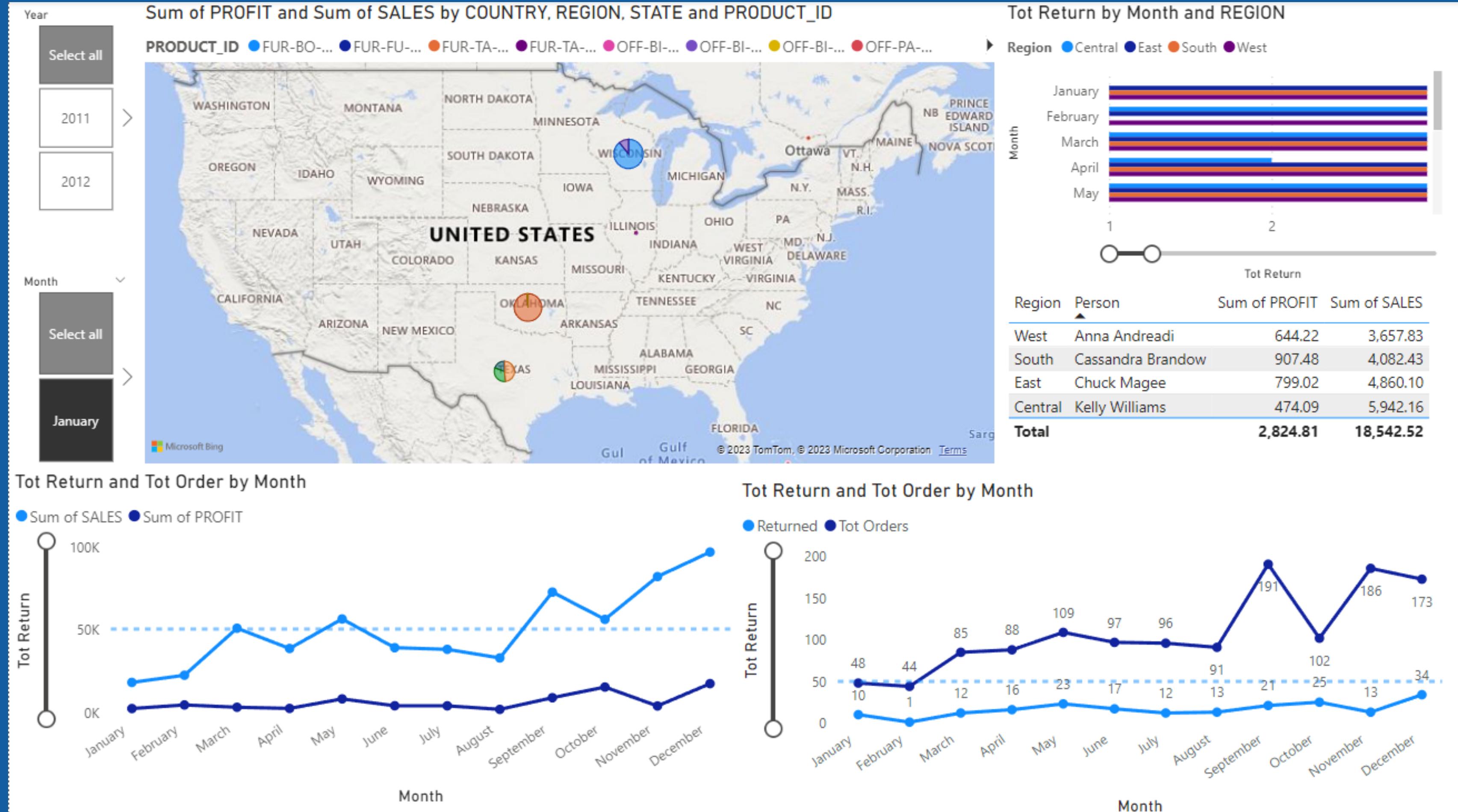
Error_Log

	PROC_NAME	LOG_NO	LOG_DATE	ERR_MSG
1	SYN_ORDERS	26,697	2023-12-16 10:49:46	ORDER_ID=CA-2013-129714 ORDER_DATE=2013-09-02
2	SYN_ORDERS	26,698	2023-12-16 10:49:46	ORDER_ID=US-2013-123750 ORDER_DATE=2013-04-16
3	SYN_ORDERS	26,699	2023-12-16 10:49:46	ORDER_ID=CA-2013-137043 ORDER_DATE=2013-12-24
4	SYN_ORDERS	26,700	2023-12-16 10:49:46	ORDER_ID=CA-2014-152912 ORDER_DATE=2014-11-10
5	SYN_ORDERS	26,701	2023-12-16 10:49:46	ORDER_ID=US-2011-150119 ORDER_DATE=2011-04-23
6	SYN_ORDERS	26,702	2023-12-16 10:49:46	ORDER_ID=CA-2012-103135 ORDER_DATE=2012-07-24
7	SYN_ORDERS	26,703	2023-12-16 10:49:46	ORDER_ID=CA-2014-118017 ORDER_DATE=2014-12-04
8	SYN_ORDERS	26,704	2023-12-16 10:49:46	ORDER_ID=CA-2013-140571 ORDER_DATE=2013-03-16
	ERR_MSG			
	ORDER_ID=CA-2013-129714 ORDER_DATE=2013-09-02 PRODUCT_ID=OFF-PA-10001970 SHIP_DATE=2013-09-04 : unique constraint (orders.XPKORDERS) violated			
	ORDER_ID=US-2013-123750 ORDER_DATE=2013-04-16 PRODUCT_ID=TEC-AC-10004659 SHIP_DATE=2013-04-22 : unique constraint (orders.XPKORDERS) violated			
	ORDER_ID=CA-2013-137043 ORDER_DATE=2013-12-24 PRODUCT_ID=FUR-FU-10003664 SHIP_DATE=2013-12-26 : unique constraint (orders.XPKORDERS) violated			
	ORDER_ID=CA-2014-152912 ORDER_DATE=2014-11-10 PRODUCT_ID=OFF-ST-10003208 SHIP_DATE=2014-11-13 : unique constraint (orders.XPKORDERS) violated			
	ORDER_ID=US-2011-150119 ORDER_DATE=2011-04-23 PRODUCT_ID=FUR-CH-10002965 SHIP_DATE=2011-04-27 : unique constraint (orders.XPKORDERS) violated			
	ORDER_ID=CA-2012-103135 ORDER_DATE=2012-07-24 PRODUCT_ID=OFF-BI-10000069 SHIP_DATE=2012-07-28 : unique constraint (orders.XPKORDERS) violated			
	ORDER_ID=CA-2014-118017 ORDER_DATE=2014-12-04 PRODUCT_ID=TEC-AC-10002006 SHIP_DATE=2014-12-07 : unique constraint (orders.XPKORDERS) violated			
	ORDER_ID=CA-2013-140571 ORDER_DATE=2013-03-16 PRODUCT_ID=OFF-PA-10001954 SHIP_DATE=2013-03-20 : unique constraint (orders.XPKORDERS) violated			

Tabel yang terbentuk ke Data Warehouse yaitu :

1. Region Manager, berisi tabel yang menunjukkan region/wilayah
2. Location, menunjukkan lokasi yang lebih rinci dari Region
3. Product, berisi data detail dari product
4. Customer, berisi data detail dari customer
5. Orders, tabel penjualan dan status pengembalian barang
6. Error Log, tabel yang berisi jika ada error saat menginput data ke data warehouse

Hasil Visualisasi



TERIMA KASIH

GITHUB