Rakamin Academy KALBE Nutritionals

Final Project
Kalbe Nutritionals - DE

Utus Karta Sanggam

Agenda

1 Introduction

2 Challenge & Answer

3 Summary

4 My Profile

5 Quote

Rakamin Academy KALBE Nutritionals

Introduction

In implementing this Project Based Internship program, we get an overview of how a Data Engineer works at Kalbe Nutritionals.

We will also learn how to solve problems and work on projects that are in accordance with Kalbe Nutritionals company activities.

With the challenges given in this final task as a representative of the real project so that we can be ready to handle it.

Rakamin Academy

KALBE Nutritionals

Challenge & Answer

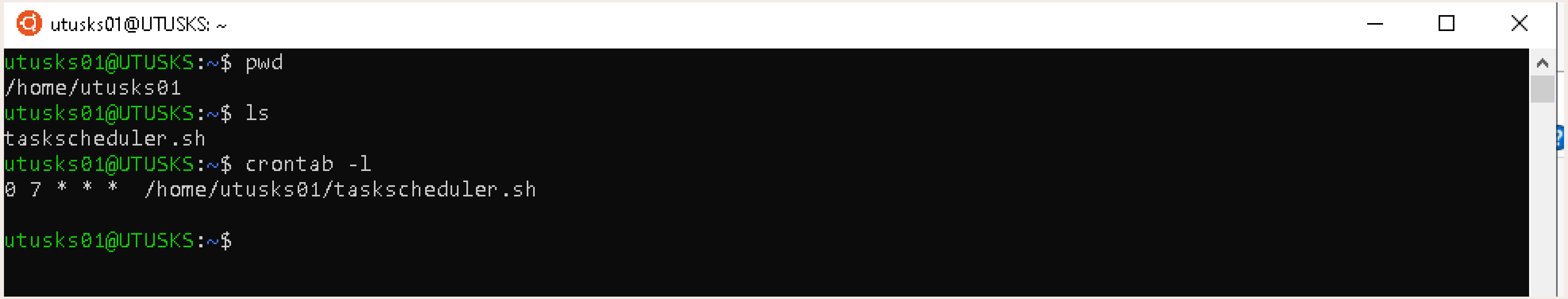
Challenge 1

- **To check whether directory exist inside a given path**

```
path=/hdfs/data/data1
name_of_directory=data1
dir=$path/$name_of_directory
filename_excel=Dataset_Final_Project_KN-DE.xlsx
source_dir=/local/data/market
if [ -d $dir ]; then
 echo “There is ${name_of_directory } Directory Exists!”
else
 echo “${name_of_directory} Directory Not Exists!”
 mkdir $dir
 exit 1
fi
```

Challenge 1 (Continued)

- **Final Step:**
  - **Create a crontab syntax to run the script at 07:00 AM Daily**

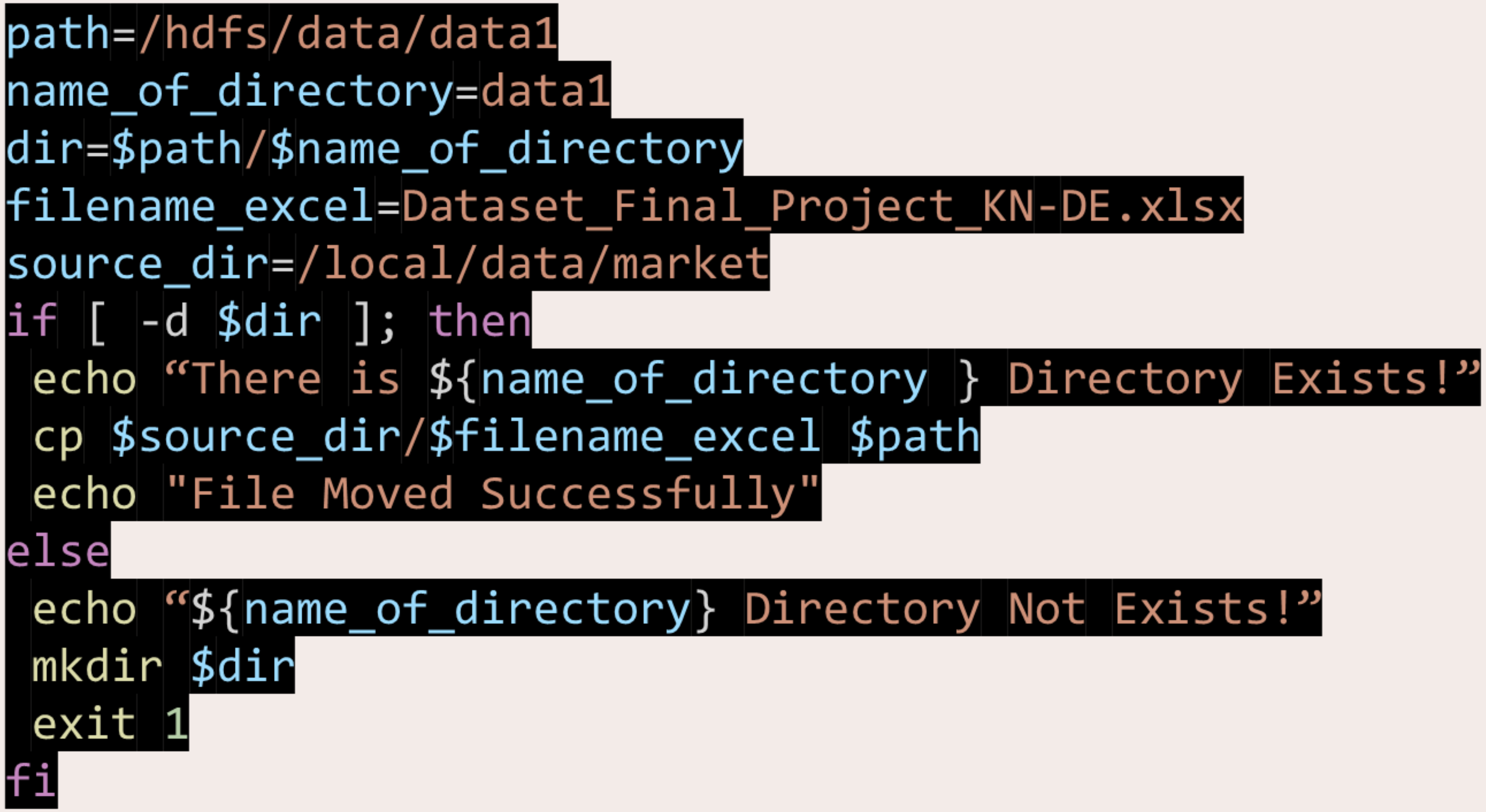


```
utusks01@UTUSKS:~$ pwd
/home/utusks01
utusks01@UTUSKS:~$ ls
taskscheduler.sh
utusks01@UTUSKS:~$ crontab -l
0 7 * * *  /home/utusks01/taskscheduler.sh

utusks01@UTUSKS:~$
```

Challenge 2

Using the question number 1 script, add another condition if directory exists inside the path.



```
path=/hdfs/data/data1
name_of_directory=data1
dir=$path/$name_of_directory
filename_excel=Dataset_Final_Project_KN-DE.xlsx
source_dir=/local/data/market
if [ -d $dir ]; then
 echo “There is ${name_of_directory } Directory Exists!”
 cp $source_dir/$filename_excel $path
 echo "File Moved Successfully"
else
 echo “${name_of_directory} Directory Not Exists!”
 mkdir $dir
 exit 1
fi
```

Challenge 3

Complete Syntax to insert data from Python to MySQL.

```
# import library mysql connector
import mysql.connector

# Connection setup configuration and testing
con = 
mysql.connector.connect(user='root',password=‘*****',port='3306',host='localhost
',database='kalbe’)

if con.is_connected(): print('Connected’)

# Create Object cursor
cursor = con.cursor()

# Add Value to Inventory Table of KALBE Database
insert_sql = ("INSERT INTO inventory(item_code,item_name,item_price,item_total)
VALUES ('2346','Fitbar',15000,200)")
#values = ('2346','Fitbar',15000,200)
```

Challenge 3 (Continued)

Complete Syntax to insert data from Python to MySQL.

```
try:
    #Execute SQL Command to insert
    cursor.execute(insert_sql)

    #Execute commit connection
    con.commit()
    print('Insert into Inventory Table has succeeded.....!!!’)

except:
    #Rollback if any issues
    con.rollback()

finally:
    #Close connection
    cursor.close()
    con.close()
```

Challenge 4-a

Convert this instruction into SQL Query Language.

- Create a database with ‘KALBE’ as the name.constraints.

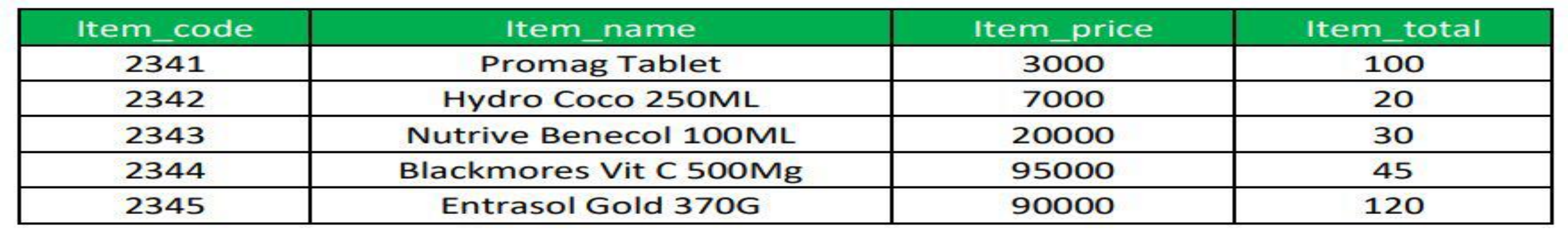
```sql
CREATE DATABASE KALBE;
```

- Use a database digunakan untuk MySQL.

```sql
USE KALBE;
```

Challenge 4-b

Inside the database, create a table with the name ‘Inventory’, with columns Item_code, Item_name, Item_price, and Item_total. Choose its own best data type and the length of it according to best practice. Choose one unique column as a primary key and decide columns

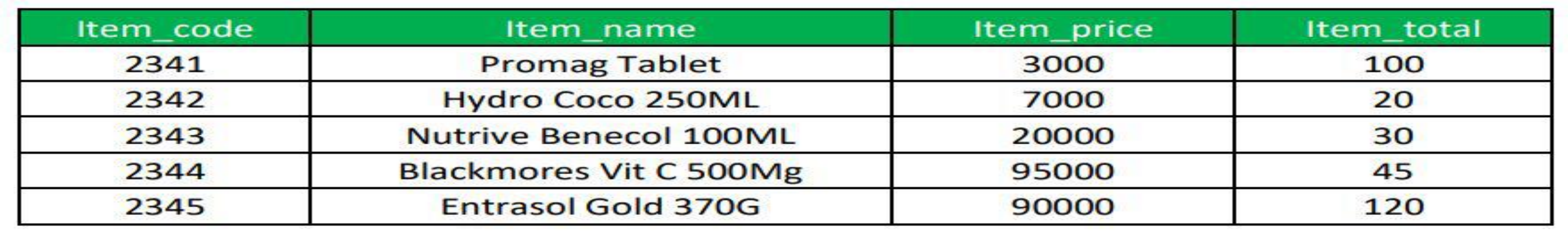


Item_code	Item_name	Item_price	Item_total
2341	Promag Tablet	3000	100
2342	Hydro Coco 250ML	7000	20
2343	Nutrive Benecol 100ML	20000	30
2344	Blackmores Vit C 500Mg	95000	45
2345	Entrasol Gold 370G	90000	120

```
CREATE TABLE Inventory (
item_code INT PRIMARY KEY,
item_name VARCHAR(100),
item_price DECIMAL(10,2),
item_total INT);
```

Challenge 4-c

Insert below data into the table:



Item_code	Item_name	Item_price	Item_total
2341	Promag Tablet	3000	100
2342	Hydro Coco 250ML	7000	20
2343	Nutrive Benecol 100ML	20000	30
2344	Blackmores Vit C 500Mg	95000	45
2345	Entrasol Gold 370G	90000	120

```
INSERT INTO Inventory (item_code, item_name, item_price, item_total)
VALUES ('2341','Promag Tablet',3000,100), ('2342','Hydro Coco 250ML',7000,20),
('2343','Nutrive Benecol 100ML',20000,30), ('2344','Blackmores Vit C
500Mg',95000,45), ('2345','Entrasol Gold 370G',90000,120);
```

Challenge 4-d

- Show Item_name that has the highest number in Item_total. :

```
SELECT item_name
FROM Inventory
WHERE item_total = (SELECT MAX(item_total)
FROM Inventory);
```

- Update the Item_price of the output of question bullet

```
UPDATE Inventory
SET item_price = ‘77500’
WHERE item_total = (SELECT MAX(item_total)
FROM Inventory);
```

Challenge 4-e

- What will happen if we insert another Item_name with Item_code of 2343 into the table?

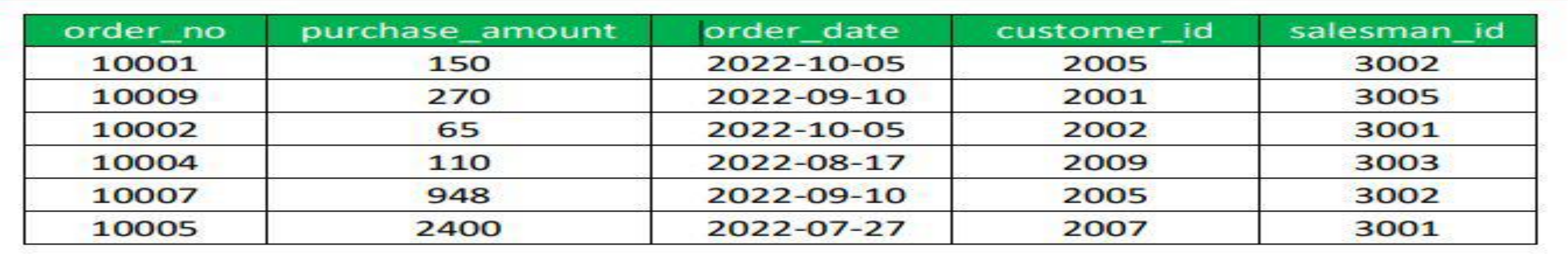
```
INSERT INTO Inventory (item_code, item_name, item_price, item_total)
values ('2343','Fitbar',3000,100);
-->>> SQL Error [1062] [23000]: Duplicate entry '2343' for key 'inventory.PRIMARY’
It results in a primary key constraint violation due to primary key column
must have unique value. The insertion will fail.
```

- Delete the Item_name that has the lowest number of Item_total.

```
DELETE FROM Inventory
WHERE item_total = (SELECT MIN(item_total)
FROM inventory);
```

Challenge 5

Create a Query to display all customer orders where purchase amount is less than 100 or exclude those orders which order date is on or greater than 25 Aug 2022 and customer id is above 2001. Sample table: customer_orders

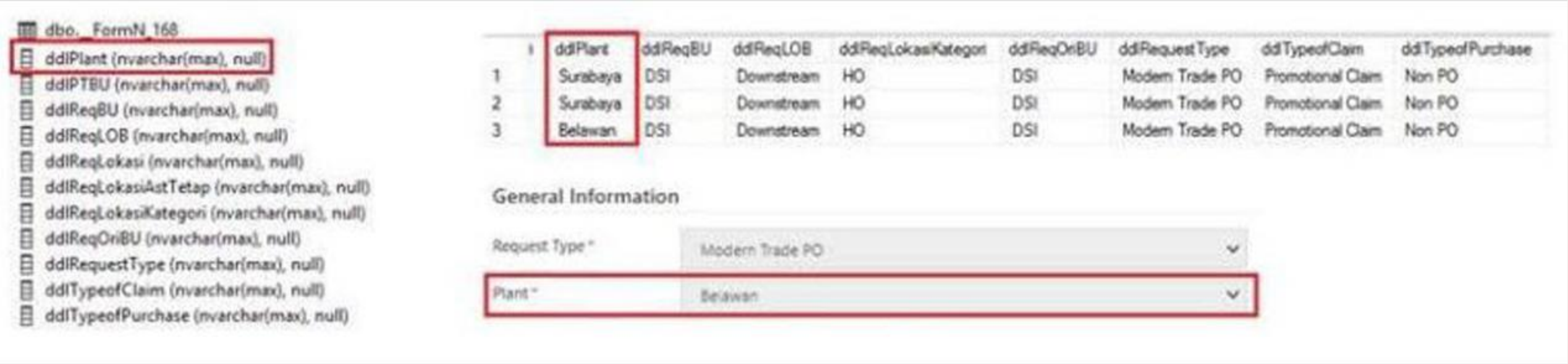


order_no	purchase_amount	order_date	customer_id	salesman_id
10001	150	2022-10-05	2005	3002
10009	270	2022-09-10	2001	3005
10002	65	2022-10-05	2002	3001
10004	110	2022-08-17	2009	3003
10007	948	2022-09-10	2005	3002
10005	2400	2022-07-27	2007	3001

```
SELECT *
FROM customer_orders as co
WHERE (purchase_amount < 100 OR (order_date < '2022-08-25’ AND customer_id <=
'2001'));
```

Challenge 6

Please explain what is wrong with this picture and give the best solution for this case.



dbo._FormN_168
- ddlPlant (nvarchar(max), null)
- ddlPTBU (nvarchar(max), null)
- ddlReqBU (nvarchar(max), null)
- ddlReqLOB (nvarchar(max), null)
- ddlReqLokasi (nvarchar(max), null)
- ddlReqLokasiAstTetap (nvarchar(max), null)
- ddlReqLokasiKategori (nvarchar(max), null)
- ddlReqOriBU (nvarchar(max), null)
- ddlRequestType (nvarchar(max), null)
- ddlTypeofClaim (nvarchar(max), null)
- ddlTypeofPurchase (nvarchar(max), null)

	ddlPlant	ddlReqBU	ddlReqLOB	ddlReqLokasiKategori	ddlReqOriBU	ddlRequestType	ddlTypeofClaim	ddlTypeofPurchase
1	Surabaya	DSI	Downstream	HO	DSI	Modern Trade PO	Promotional Claim	Non PO
2	Surabaya	DSI	Downstream	HO	DSI	Modern Trade PO	Promotional Claim	Non PO
3	Belawan	DSI	Downstream	HO	DSI	Modern Trade PO	Promotional Claim	Non PO

General Information

Request Type *: Modern Trade PO

Plant *: Belawan

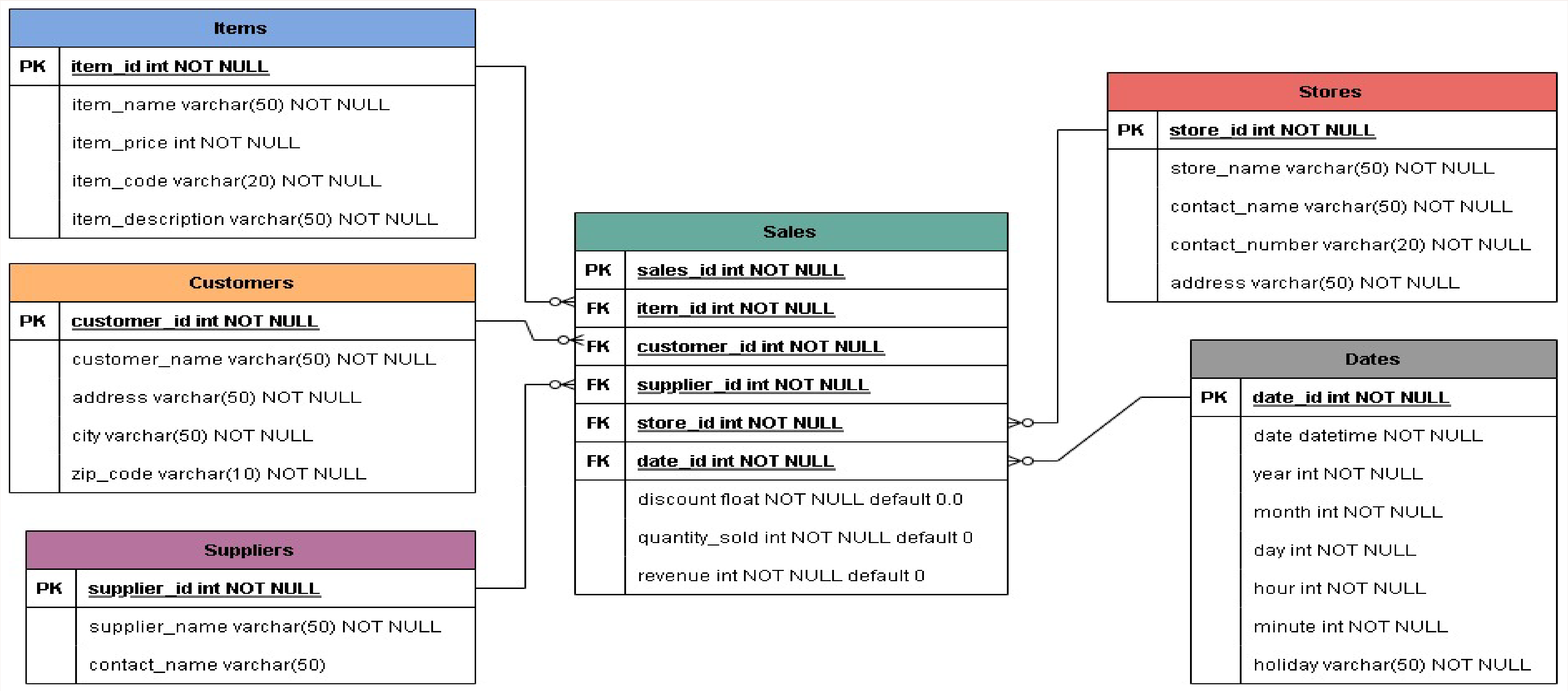
Based on General Information with filtering Requested Type : Modern Trade PO and Plant : Belawan, **on column ddPlant should be Belawan only but also we see Surabaya (Duplicate).** It’s wrong and the best solution for this case :

1. Filtering if Belawan only : where ddRequestType=“Modern Trade PO” and ddPlant=“Belawan”

2. Duplicate : ddPlant=“Surabaya” must be cleaned.

Challenge 7

Create a simple star schema for KALBE database consist of 1 Fact and 5 Dimensions using Physical Data Model Theory.



	Items
PK	item_id int NOT NULL
	item_name varchar(50) NOT NULL
	item_price int NOT NULL
	item_code varchar(20) NOT NULL
	item_description varchar(50) NOT NULL

	Customers
PK	customer_id int NOT NULL
	customer_name varchar(50) NOT NULL
	address varchar(50) NOT NULL
	city varchar(50) NOT NULL
	zip_code varchar(10) NOT NULL

	Suppliers
PK	supplier_id int NOT NULL
	supplier_name varchar(50) NOT NULL
	contact_name varchar(50)

	Sales
PK	sales_id int NOT NULL
FK	item_id int NOT NULL
FK	customer_id int NOT NULL
FK	supplier_id int NOT NULL
FK	store_id int NOT NULL
FK	date_id int NOT NULL
	discount float NOT NULL default 0.0
	quantity_sold int NOT NULL default 0
	revenue int NOT NULL default 0

	Stores
PK	store_id int NOT NULL
	store_name varchar(50) NOT NULL
	contact_name varchar(50) NOT NULL
	contact_number varchar(20) NOT NULL
	address varchar(50) NOT NULL

	Dates
PK	date_id int NOT NULL
	date datetime NOT NULL
	year int NOT NULL
	month int NOT NULL
	day int NOT NULL
	hour int NOT NULL
	minute int NOT NULL
	holiday varchar(50) NOT NULL

All Dimension (Items/Customers/Suppliers/Stores/dates) Tables have relationship with Fact Sales Table.

Summary

The challenges given in this final task divide on 3 (Three) main tasks as follow :

1. Create a shell/bash scripting and Cron scheduler.
   → Challenge 1 and 2.
2. Database Integration using SQL and Python.
   → Challenge 3.
3. Data Modeling (DDL, DML and Star Schema).
   → Challenge 4 – 7.

My Profile

Working Experience

Data Engineer Virtual Internship at PT. Kalbe Nutitionals. Sep-Oct, 2023

Data Analyst Virtual Internship at PT. Xeratic. June-August, 2023

Career Transition to learn more about Data Sept 2022 – Present :

- Data Science IDCamp 2023 Scholarship - period Sep 11 - Dec 25, 2023 by Dicoding.
- Alibaba Big Data Fundamentals DTS PROA 2023 Batch 3 period Sep 11 - Nov 15, 2023 by Alibaba Indonesia.
- Bootcamp Machine Learning & AI for Beginner - period July 1 - Aug 26, 2023 by DQLab.
- Data Engineer Course (Digital Talent Scholarship PROA 2023 Batch 1 period Mar-Apr, 2023) by DQLab.
- Bootcamp Data Analyst with SQL and Python period Jan-Mar, 2003 by DQLab
- Google Data Analytics Course (Digital Talent Scholarship/ DTS) PROA 2022-Batch 4 period Sep-Nov, 2022) by Cousera & Dicoding

Quote

“Business opportunities are like buses, there's always another one coming.”

Richard Branson

Thank you

Utus Karta Sanggam

utussanggams@gmail.com

Linkedin www.linkedin.com/in/utusks01