

# Final Project Kalbe Nutritionals - DE

Utus Karta Sanggam

## Data Engineering



# Agenda



**Introduction**



**Challenge &  
Answer**



**Summary**



**My Profile**



**Quote**

# 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.



# Challenge & Answer



# Challenge 1 and 2

Create a shell/bash script to check whether directory exists inside a given path.

- Variables:

- path=/hdfs/data/data1 name\_of\_directory=data1

- Conditions:

- If directory exists inside the path:

- Echo "There is [Directory Name] Directory Exists!"

- If not:
    - Echo "[Directory Name] Directory Not Exists!"
    - Create a directory inside the path.

- Final Step:

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

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

- Variables:

- filename\_excel=daily\_market\_price.xlsx

- source\_dir=/local/data/market

- target\_dir=Refer to Question Number 1 Path

- Conditions:

- Copy file from source directory into target directory.

- Create a log file inside the same path with "File Moved Successfully" as a log content if success.

# Challenge 1 and 2 (Continued)

```
directory_check.ps1 > ...
1 param(
2
3     [string]$path = "hdfs/",
4
5     [string]$name_of_directory = 'data',
6
7     [string]$filename_excel = "source/Dataset_Final_Project_KN-DE.xlsx"
8
9 )
10
11 # Check if the directory exist
12 $directoryPath = Join-Path $path /$name_of_directory
13
14 if (Test-Path $directoryPath -PathType Container) {
15
16     Write-Host "$($name_of_directory) Directory Exists!"
17 }
18
19 else {
20     Write-Host "$($name_of_directory) Directory Not Exists! Please create the directory...."
21
22     # Create the directory
23     New-Item -ItemType Directory -Path $directoryPath | Out-Null
24 }
25
26 #Final Step : create cront task schedule to run the script daily at 07:00 AM
27
28 $trigger = New-ScheduledTaskTrigger -Daily -At 7am
29
30 $action = New-ScheduledTaskAction -Execute "PowerShell.exe" -Argument "-ExecutionPolicy Bypass -File ""C:\Users\Ajiel Sinatra\Documents\Internship Data Engineer Kalbe Nutritional\Minggu 4> .\directory_check.ps1
31
32 Register-ScheduledTask -TaskName "CheckDirectoryTask" -Trigger $trigger -Action $action -User "root" -RunLevel Highest
33
34 Write-Host "Scheduled task 'CheckDirectoryTask' created cront task schedule to run the script daily at 07:00 AM"
```

PROBLEMS OUTPUT DEBUG CONSOLE **TERMINAL** PORTS

```
PS C:\Users\Ajiel Sinatra\Documents\Internship Data Engineer Kalbe Nutritional\Minggu 4> .\directory_check.ps1
data Directory Exists!
Scheduled task 'CheckDirectoryTask' created cront task schedule to run the script daily at 07:00 AM
PS C:\Users\Ajiel Sinatra\Documents\Internship Data Engineer Kalbe Nutritional\Minggu 4>
```

# Challenge 3

## Complete Syntax to insert data from Python to MySQL.

```
python_to_mysql_conneccion.py > ...
2 import mysql.connector
3
4 # Connection setup configuration and testing
5 con = mysql.connector.connect(user='root',password='12345',port='3306',host='localhost',database='kalbe')
6 if con.is_connected():
7     print('Connected')
8 # Create Object cursor
9 cursor = con.cursor()
10
11 # Add Value to Inventory Table of KALBE Database
12 insert_sql = ("INSERT INTO inventory(item_code,item_name,item_price,item_total) VALUES ('2346','Fitbar',15000,200)")
13 #values = ('2346','Fitbar',15000,200)
14
15 try:
16     #Execute to insert
17     cursor.execute(insert_sql)
18     con.commit()
19     print('Insert into Inventory Table has succeeded.....!!!')
20
21 except:
22     #Rollback if any issues
23     con.rollback()
24     raise
25
26 finally:
27     #Close connection
28     con.close()
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS Python Debug Console + - □ ✕ ... ^

Connected  
Insert into Inventory Table has succeeded.....!!!  
PS C:\Users\Ajiel Sinatra\Documents\Internship Data Engineer Kalbe Nutritional\Minggu 4> |

## Challenge 4-a

Convert this instruction into SQL Query Language.

- o Create a database with 'KALBE' as the name.constraints.

```
CREATE DATABASE KALBE;
```

- o Use a database digunakan untuk MySQL.

```
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 varchar(5) NOT NULL, item_name varchar(50) NOT NULL, item_price int NOT NULL DEFAULT 0, item_total int NOT NULL DEFAULT 0, PRIMARY KEY (item_code), UNIQUE KEY item_unique (item_code) )ENGINE = InnoDB;
```

## 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, item_total
FROM Inventory
WHERE item_total = (SELECT max_total_item FROM
(SELECT max(item_total) as max_total_item FROM inventory) as max);
```

- Update the Item\_price of the output of question bullet

```
WITH item_total_max AS ( SELECT MAX(item_total) as max_total FROM Inventory)
UPDATE Inventory
JOIN item_total_max ON Inventory.item_total = item_total_max.max_total
set Inventory.item_price = '100000';
```

## 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'
```

- Delete the Item\_name that has the lowest number of Item\_total.

```
DELETE FROM Inventory  
WHERE item_total = (SELECT min_item_total FROM (SELECT min(item_total)  
as min_item_total FROM inventory) as subquery);
```

## 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

ddPlant (nvarchar(max), null)

ddIPTBU (nvarchar(max), null)

ddReqBU (nvarchar(max), null)

ddReqLOB (nvarchar(max), null)

ddReqLokasi (nvarchar(max), null)

ddReqLokasiAstTetap (nvarchar(max), null)

ddReqLokasiKategori (nvarchar(max), null)

ddReqOnBU (nvarchar(max), null)

ddRequestType (nvarchar(max), null)

ddTypeofClaim (nvarchar(max), null)

ddTypeofPurchase (nvarchar(max), null)

	ddPlant	ddReqBU	ddReqLOB	ddReqLokasiKategori	ddReqOnBU	ddRequestType	ddTypeofClaim	ddTypeofPurchase
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

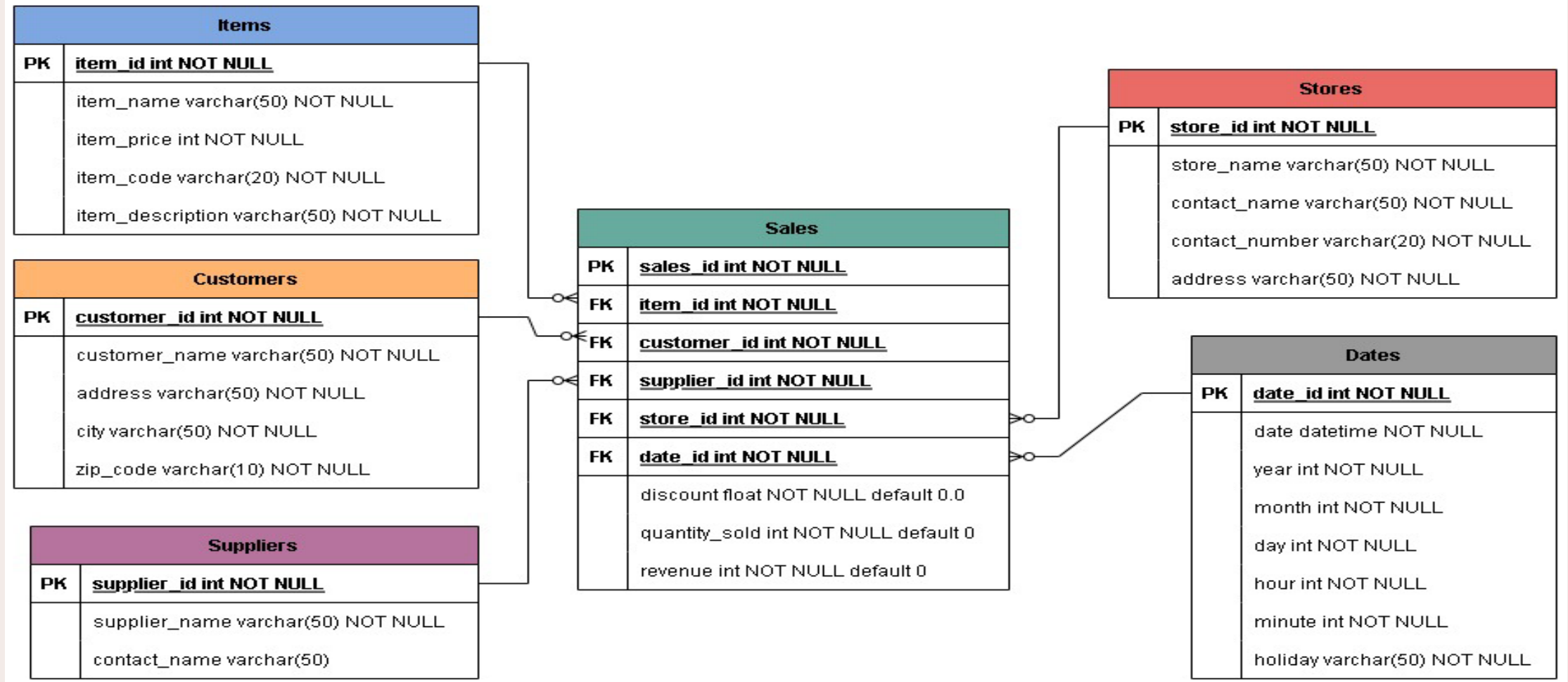
Based on General Information with filtering Requested Type : Modern Trade PO and Plant : Belawan, on column ddPlant should be Belawan (1 row only) but we see also Surabaya (3 rows).

It's wrong and I assume the query using where ddRequestType="Modern Trade PO" or ddPlant="Belawan" instead of where ddRequestType="Modern Trade PO" and ddPlant=" Belawan"



# Challenge 7

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



<https://app.diagrams.net/#G1xBZqExP-Xs2DAicMFDdB6nAeJNb7TgI7>



# 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 @ PT. Kalbe Nutitionals. Sep-Oct, 2023

Data Analyst Virtual Internship @ 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 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, 2023 by DQLab
- Google Data Analytics Course (Digital Talent Scholarship PROA 2022-Batch 4 period Sep-Nov, 2022) by Cousera & Dicoding

# Quote

“



”

# Thank you

Utus Karta Sanggam

utussanggams@gmail.com

[LinkedIn !\[\]\(4729e517bc6a7cd81c8025b9646574fb\_img.jpg\) www.linkedin.com/in/utusks01](https://www.linkedin.com/in/utusks01)

