

Final Project Kalbe Nutritional - 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



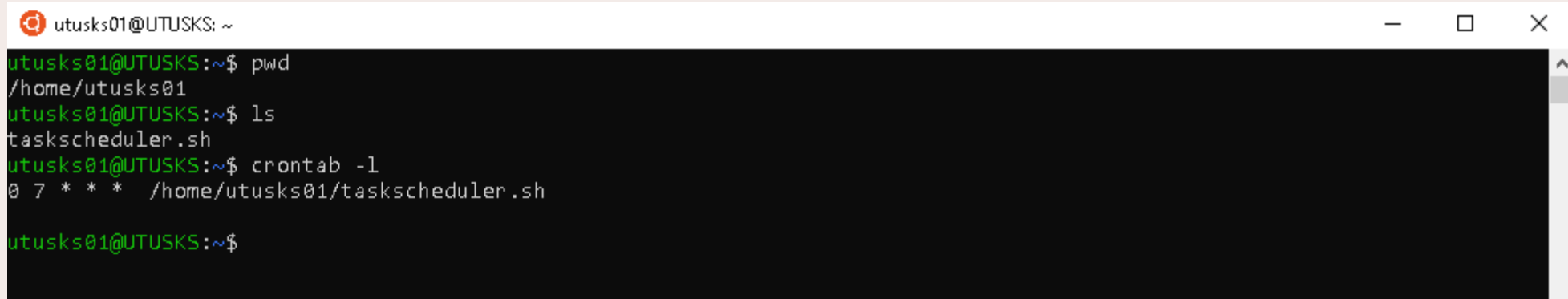
Challenge 1

- **To check whether directory exist inside a given path (Linux/Ubuntu)**

```
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: (Linux/Ubuntu)**
 - **Create a crontab syntax to run the script at 07:00 AM Daily**

A terminal window titled 'utusks01@UTUSKS: ~' with standard window controls. The terminal shows a sequence of commands and their outputs: 'pwd' returns '/home/utusks01', 'ls' lists 'taskscheduler.sh', and 'crontab -l' adds a new entry '0 7 * * * /home/utusks01/taskscheduler.sh' to the crontab. The prompt returns to '~\$' after each command.

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

Challenge 1 (Continued)

- **To check whether directory exist inside a given path (PowerShell)**

```
param (  
[string]$path_source = "local/data/market",  
[string]$path_target = "hdfs/data",  
[string]$name_of_directory = "data1"  
)  
  
# Check if the directory exist  
$directoryPath = Join-Path $path_target /$name_of_directory  
  
if (Test-Path $directoryPath -PathType Container) {  
    Write-Host "$($name_of_directory) Directory Exists!"  
}  
else {  
    Write-Host "$($name_of_directory) Directory Not Exists! Please create directory."  
  
    # Create a new folder in the target destination path.  
    New-Item -Path "$path_target\$name_of_directory" -ItemType "Directory"  
    Write-Host "Folder '$name_of_directory' created successfully in '$path_target'."  
}
```

Challenge 1 (Continued)

- **Final Step: (PowerShell)**

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

```
# Get the user account name.  
$username = "UTUSKS\Utus Karta Sanggam"
```

```
# Replace "password" with the actual password for the user.  
$password = "*****"
```

```
# Check if the user account name is null or empty.  
if ($username -eq $null -or $username -eq "") {  
    # Write an error message to the console.  
    Write-Error "The user account name cannot be null or empty."  
}
```

```
# Exit the script.  
Exit-Script  
}
```

```
# Create the scheduled task.  
$trigger = New-ScheduledTaskTrigger -Daily -At 7am  
$action = New-ScheduledTaskAction -Execute "PowerShell.exe" -Argument "-ExecutionPolicy Bypass -File  
""C:\source_path\taskscheduler.ps1"""  
Register-ScheduledTask -TaskName "CheckDirectoryTask" -Trigger $trigger -Action $action -User  
$username -Password $password -RunLevel Highest
```


Challenge 2

Using the question number 1 script, add another condition if directory exists inside the path.
(Linux/Ubuntu)

```
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 2 (Continued)

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

```
)
param (
    [string]$path_source = "local/data/market",
    [string]$path_target = "hdfs/data",
    [string]$name_of_directory = "data1"
)

$filename_excel = Dataset_Final_Project_KN-DE.xlsx

# Check if the directory exist
$directoryPath = Join-Path $path_target /$name_of_directory

if (Test-Path $directoryPath -PathType Container) {
    Write-Host "$($name_of_directory) Directory Exists!"
    # Use the copy-Item cmdlet
    Copy-Item -Path $path_source\$filename_excel -Destination $directoryPath -Recurse
    Write-Host "File has moved Successfully"
}
else {
    Write-Host "$($name_of_directory) Directory Not Exists! Please create directory."

    # Create a new folder in the target destination path.
    New-Item -Path "$path_target\$name_of_directory" -ItemType "Directory"
    Write-Host "Folder '$name_of_directory' created successfully in '$path_target'."
}
```

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-(i)

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-(ii)

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-(iii)

Insert below data into the table:

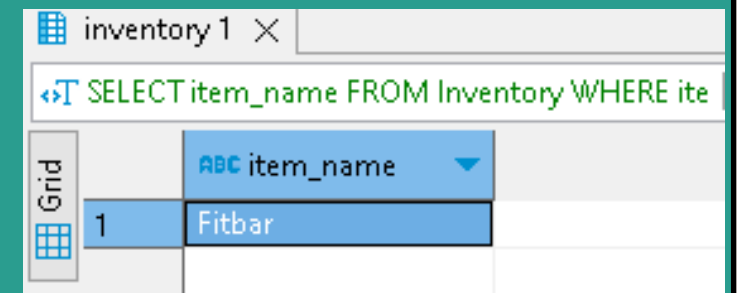
Item_code	Item_name	Item_price	Item_total
2341	Promag Tablet	3000	100
2342	Hydro Coco 250ML	7000	20
2343	Nutrive Benecol 100ML	20000	30
2344	Blackmores Vit C 500Mg	95000	45
2345	Entrasol Gold 370G	90000	120

```
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-(iv)

- 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);
```



The screenshot shows a database query editor window titled 'inventory 1'. The query entered is 'SELECT item_name FROM Inventory WHERE ite'. Below the query, there is a 'Grid' view showing the results. The first row is highlighted, showing 'Fitbar' as the item_name.

Grid	item_name
1	Fitbar

- 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'
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'));
```

customer_orders 1 X						
SELECT * FROM customer_orders as co WHERE Enter a SQL expression to filter results (use Ctrl+Space)						
Grid	order_no	purchase_amount	order_date	customer_id	salesman_id	
1	10002	65	2022-10-05	2002	3001	

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)

ddIReqBU (nvarchar(max), null)

ddIReqLOB (nvarchar(max), null)

ddIReqLokasi (nvarchar(max), null)

ddIReqLokasiAstTetap (nvarchar(max), null)

ddIReqLokasiKategori (nvarchar(max), null)

ddIReqOnIBU (nvarchar(max), null)

ddRequestType (nvarchar(max), null)

ddITypeofClaim (nvarchar(max), null)

ddITypeofPurchase (nvarchar(max), null)

	ddPlant	ddIReqBU	ddIReqLOB	ddIReqLokasiKategori	ddIReqOnIBU	ddRequestType	ddITypeofClaim	ddITypeofPurchase
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 *

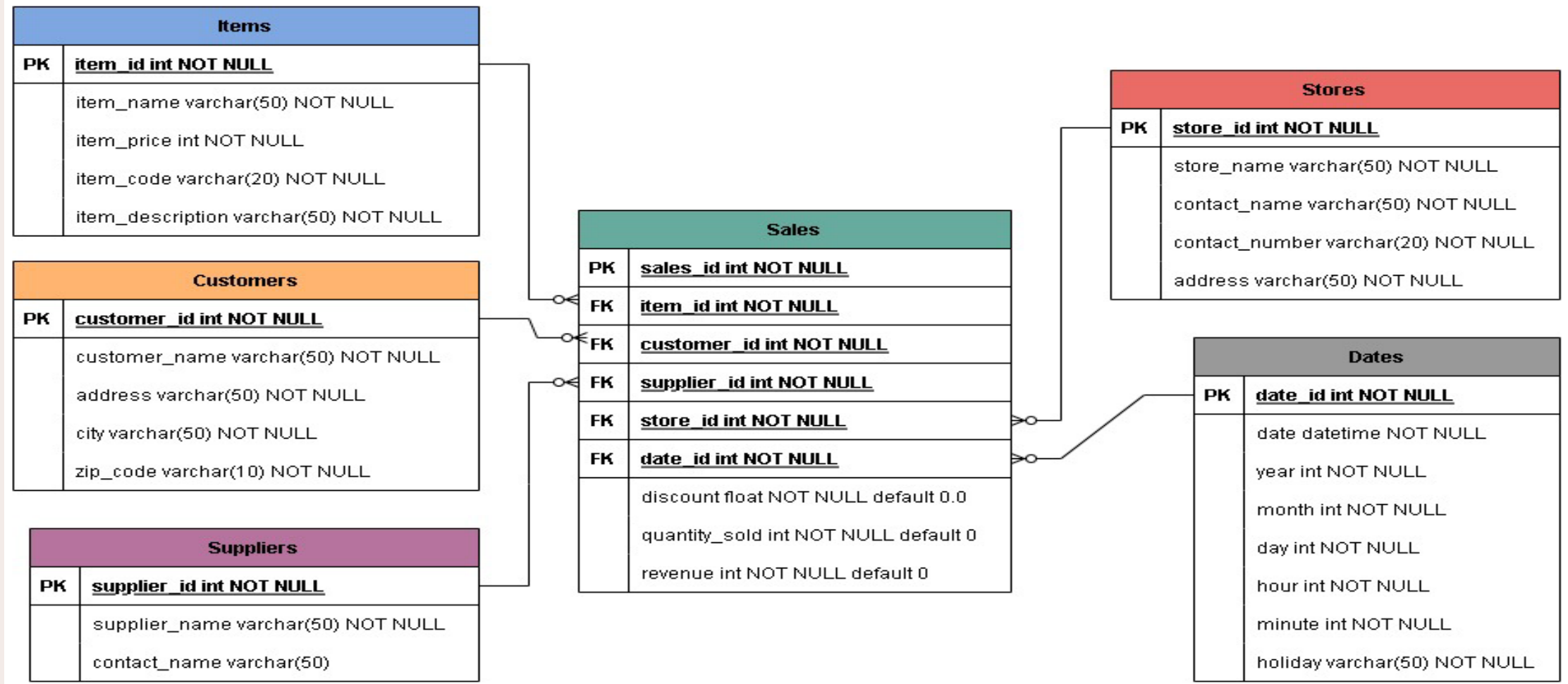
Belawan

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.



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, 2023 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

quotezany

”

Thank you

Utus Karta Sanggam

utussanggams@gmail.com

[LinkedIn !\[\]\(99f58673407353e96a019fbca558fd72_img.jpg\) www.linkedin.com/in/utusks01](https://www.linkedin.com/in/utusks01)

