Answer :

1. Query :



with

--1. Reformat order\_time into Jakarta Timezone and trunct the time

--2. Distinct the data by fields order date, customer\_no and order\_type,

-- the purpose is to see how many order\_type that a customer use in one order\_date regardless he order\_payment

--3. filter order\_status = 'Completed'

distdata as (

select distinct DATE(order\_time, "Asia/Jakarta") order\_date,

customer\_no,

order\_type

from `bi-dwhdev-01.source.daily\_order`

where order\_status = 'Completed'

order by 1,2,3)

-- Select count of customer for each order\_type in a order\_date

-- in this query, we know how many order\_type a customer us in a day

, cust as (

select order\_date, customer\_no, count(\*) cnt from distdata

group by order\_date, customer\_no)

-- Only change the order\_time into order\_date and use Jakarta timezone

, thedata as (

select DATE(order\_time, "Asia/Jakarta") order\_date,

customer\_no,

order\_type,

order\_payment

from `bi-dwhdev-01.source.daily\_order`

where order\_status = 'Completed'

order by 1,2,3)

-- Query to show aggregate data of customer, in this query we can see what order\_types are each customer use

-- and what are the order payment

,pre\_data as(

select thedata.order\_date,

thedata.customer\_no,

count(thedata.customer\_no)cnt\_cust,

cust.cnt,

FORMAT("%T", ARRAY\_AGG(distinct(thedata.order\_type) order by (thedata.order\_type)) ) order\_type,

FORMAT("%T", ARRAY\_AGG(distinct(thedata.order\_payment) order by (thedata.order\_payment)) ) order\_payment

from thedata

inner join cust

on thedata.order\_date = cust.order\_date

and thedata.customer\_no = cust.customer\_no

group by thedata.order\_date, thedata.customer\_no, cust.cnt

order by 1,2)

-- aggregate the data

,data\_sum as (

select order\_date,

no\_of\_service,

sum(total\_customer\_per\_order\_type) over (partition by order\_date, no\_of\_service) total\_customer,

order\_type,

total\_customer\_per\_order\_type,

order\_payment

from (

select

order\_date,

cnt as no\_of\_service,

replace(replace(replace(order\_type,'[',''),']',''),'"','')order\_type,

sum(cnt\_cust) total\_customer\_per\_order\_type,

replace(replace(replace(order\_payment,'[',''),']',''),'"','') order\_payment

from pre\_data

group by order\_date, cnt, order\_type, order\_payment

)

order by 1,2,4)

select

order\_date,

no\_of\_service,

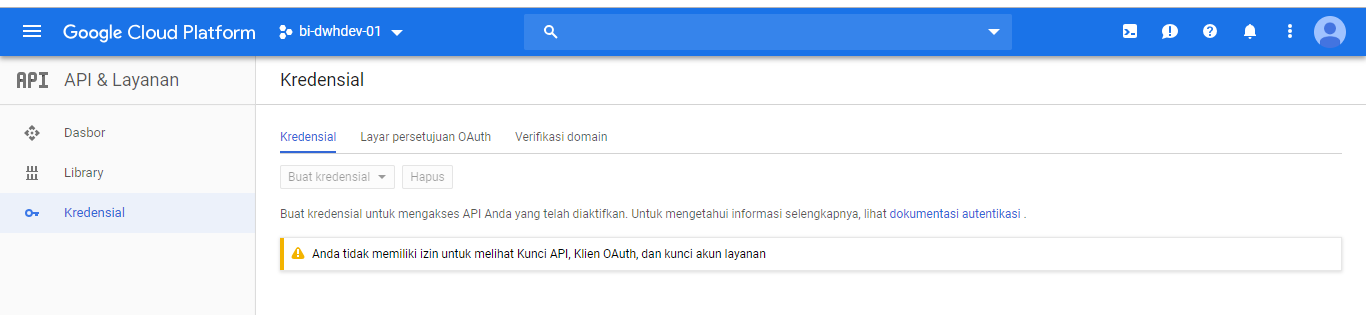
total\_customer,

order\_type,

total\_customer\_per\_order\_type,

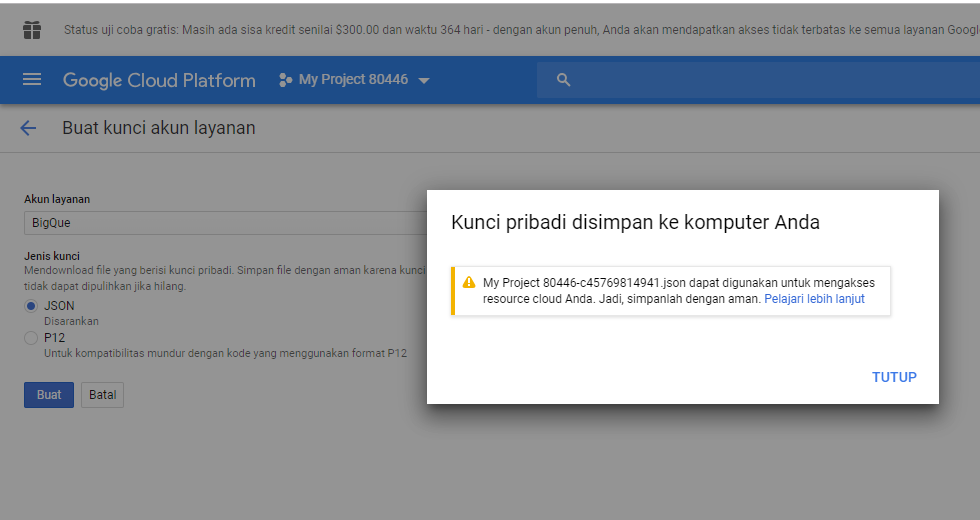
order\_payment from data\_sum

1. In Order to create BigQuery Job, I found that we need to have a permission to access data from client



When I try to get that permission, there is a restriction that say I cannot see the key.

So, I create my own project with data from the question, and I can create my own credentials



Instead of **bi-dwhdev-01**, the project name is **My Project 80446**. And I have the credentials to access the google bigquery from client.

I copy the table from **bi-dwhdev-01** to **My Project 80446 and** the table name become

`root-patrol-237412.order.daily\_ordr`

Not

`bi-dwhdev-01.source.daily\_order`

Python Script

+ SQL Query

Task Scheduller

Google Bigquery Data



Usage:

File Job.bat is a windows bat file that can be run with windows task scheduler. By default the script will run without argument and the sql query by default

Python Question2.py

1. Python Script



import csv

import json

f = open( 'driver\_registration.csv', 'r' )

reader = csv.DictReader( f, fieldnames = ('id', 'date\_created', 'date\_last\_modified','active\_date','name','phone','resign\_date'

,'resign\_reason','status','tipe','area','operator\_id','modified\_by','vehicle\_type'

,'helmet\_qty','jacket\_qty','vehicle\_brand','vehicle\_year','bike\_type'

,'first\_ride\_bonus\_awarded','is\_doc\_completed'))

out = json.dumps( [ row for row in reader ] )

jsonArray = json.loads(out)

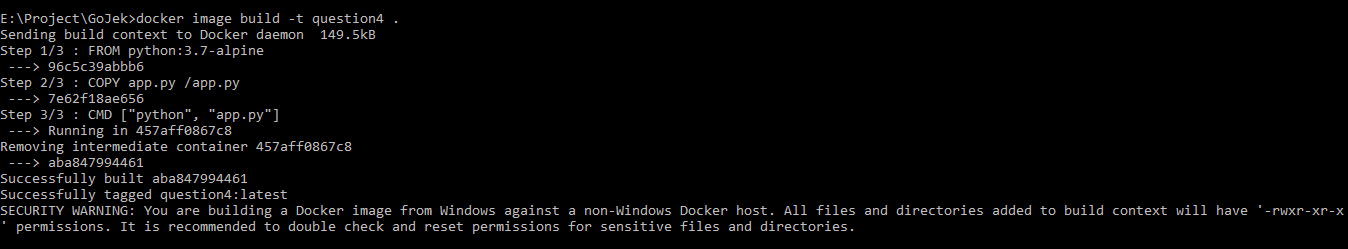
with open("Question 3.json", 'w') as f:

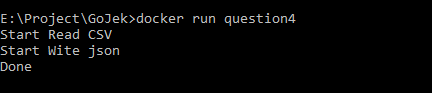
for jsonArrays in jsonArray :

f.write(str(jsonArrays)+ '\n')

1. Dockerfile, script and how to run it

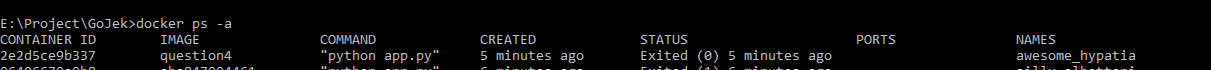


docker image build -t question4 .

docker run question4

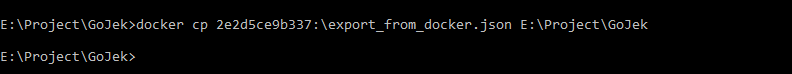
Check container id with

docker ps -a



Copy file from container to local directory using

docker cp 2e2d5ce9b337:\export\_from\_docker.json E:\Project\GoJek



1. If I have a chance to build an architecture at Gojek, I want to add Elastic Search / ELK to Data Warehouse at GoJek. Because Gojek has many data source name, table name, column name and text data (I assumed/imagine table that contains Go Food menu). It will be easier to search on indexed text data and using search engine on the data.
2. GoJek Core Value
   1. Strength

The three point that I think my strength are

* + 1. COLLABORATE WITH COMPASSION
    2. CRITICISM IS A GIFT
    3. COMMUNICATE WITH PURPOSE

Example: I like to discuss with other, discuss about how I solved a problem, discuss about how other people solved the same problem and I like to know what people think about it. I accept criticism as objective as possible.

* 1. Weakness
     1. SHOOT FOR GREATNESS
     2. BE FAST AND FEARLESS
     3. ALWAYS BE PREPARED

Example: I always curious about something, I like to challenge myself. When I interested in something, I am obsessed with it, searching, learning and doing everything about it. But at some point if I think its too complicated and I do not know

* 1. Order 10 Core
     1. IT'S NOT ABOUT YOU
     2. BE FAST AND FEARLESS
     3. ALWAYS BE PREPARED
     4. SHOOT FOR GREATNESS
     5. STAND UP FOR WHAT YOU BELIEVE IN
     6. COLLABORATE WITH COMPASSION
     7. COMMUNICATE WITH PURPOSE
     8. CRITICISM IS A GIFT
     9. EARN YOUR TITLE
     10. BECOME A SCIENTIST

I think the first priorities is to think it’s not about me. When doing something at workplace, its always about a team. When we make a mistake, of course we personally responsible about it, but when other people/customer see it, it become a team mistake. I think technology is at running state, its mean we must update our knowledge as fast as possible with current technology and decide whether it will be useful for our business or not. To gain the new technology, of course we can’t do it alone, so we need to communicate an discuss with other.