**Technical Test**

Business Intelligence Data Engineer

# Overview

Before any love story begins, we need to make sure we know each other and meet each other expectations. This is the part where we want to get to know you better.   
  
Since our team builds data products on a daily basis, we want to make sure that you do too. If you successfully pass this test, we will be more assured that you are able to explore, mine, and wrangle the massive datasets that GO-JEK has access to in order to build highly impactful data products.  
  
So, let’s start building NOW!

# What are your responsibilities?

* Design, build, launch and maintain extremely efficient & reliable large-scale batch and real-time data pipelines with data processing frameworks like Dataflow, Flink, Spark and Kafka, which would enable teams to consume and understand data faster
* Run extensive and intensive POCs on data processing frameworks, technologies, and platforms
* Maintain and improve data quality through testing, tooling and continuously evaluation of performance

# Required qualifications

* Strong understanding of data streaming concepts
* Experience with building production-grade data processing tools such as BigQuery, Kafka, Pub/Sub, Dataflow, Flink
* Experience in writing codes in Java, Python, SQL
* Passionate about crafting clean code and story-like documentation
* Knowledgeable about system design, data structures, and algorithms
* Strong familiarity about data modeling, data access, and data storage techniques, like Data Lake and Data Warehousing concepts

# Preferred qualifications

* Understand the value of collaboration within teams and be very customer oriented
* Interested in being the glue between engineering and analysis

**Technical Test**

Business Intelligence Data engineer

# Instructions

1. There are 6 tasks in this technical test
2. For each task, we expect below outputs:
3. SQL Files
4. Source Code
5. Python Script
6. Docker Packages ( Dockerfile and source code )
7. Any notes application
8. Any notes application

# 

# Question 1

# For analysis, build a summary table to show how many customers use multiple GO-JEK services in a daily basis, along with the combination of services used:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| order\_date | no\_of\_service | total\_customer | detail | | order\_payment |
| order\_type | total\_customer\_per\_order\_type |
| 2018-04-03 | 1 | 20 | CAR | 5 | CASH |
| RIDE | 4 |
| FOOD | 1 |
| SEND | 10 |
| 2018-04-02 | 3 | 15 | FOOD,CAR,RIDE | 10 | CASH |
| RIDE,SEND,TIX | 5 |

Several rules for this task:

1. Take only order\_status = “Completed”
2. No repetition on details of order\_type
   1. E.g. same combinations like RIDE, CAR, SEND and CAR, RIDE, SEND are unacceptable, it needs to be unique
3. Group for each order\_payment and its combination
   1. I.e. aggregations by CASH, GOPAY, CASH&GOPAY (ALL)
4. Used date in Jakarta timezone

You will use BigQuery for this task, please follow these steps to access the data source:

1. Join this group below by clicking the [link](https://groups.google.com/forum/#!forum/bi-dwhdev/join)
2. After your join request is approved, please go to this [link](https://bigquery.cloud.google.com/table/bi-dwhdev-01:source.daily_order?tab=preview)

# Question 2

# Based on your SQL Query from task Question 1, please create a job to run the SQL in daily basis with any programming language and fulfill these conditions:

1. By default it will run query with D-1 date as start date and today as end date
2. It can receive variables sent when executing the script which contains start date and end date for backfill purposes

# Question 3

# Given a CSV file on [this link](https://goo.gl/1Vhc7n) as data source, please create a python script to reformat the data to JSON files with output like [this](https://goo.gl/FMkVkg)

# Question 4

# Create a Docker File for both scripts from task 2 and 3 with below conditions:

1. Using Unix / Linux based image
2. Include all requirements or dependencies to run the script

* Please provide documentation on how to run the application

# Question 5

# Please spare 30 minutes to watch below videos about how [GO-JEK Journey to Next Generation Data](https://youtu.be/qZz5bEeALJ0?t=41m44s)

# If you have a chance to build the architecture with us, what technology or process that you will replace or add to Next Gen Data, and share us your reasons

# Question 6 Here at GO-JEK we have 10 core values that we follow:

1) IT'S NOT ABOUT YOU

• Puts the company above themselves

• Obsesses about customer problems, not personal problems

• Has a purpose beyond personal success

2) STAND UP FOR WHAT YOU BELIEVE IN

• Says what they mean

• Has courage to disagree

• Has a strong moral compass

3) COLLABORATE WITH COMPASSION

• Is a pleasure to work with

• Supports others in areas beyond their scope

• Considers how their actions affects others

4) BE FAST AND FEARLESS

• Takes calculated risk

• Values failure as much as success

• Has strong sense of urgency

5) EARN YOUR TITLE

• Walks the talk

• Gets down in the trenches

• Trusts their team

6) BECOME A SCIENTIST

• Follows the numbers intensively

• Finds solutions in unexpected places

• Learns independently and shares knowledge

7) ALWAYS BE PREPARED

• Does their homework

• Plans out every scenario

• Takes action to reduce risk

8) CRITICISM IS A GIFT

• Actively seeks feedback from others

• Gives helpful feedback to others unprompted

• Genuinely acts upon feedback given

9) COMMUNICATE WITH PURPOSE

• States objectives in every interaction

• Aligns early and consistently

• Focuses on what is actionable

10) SHOOT FOR GREATNESS

• Goes the extra mile

• Thinks big

• Loves to challenge themselves

1. Please choose three Core Values from the ten values above that you think is your strength, share us one specific example where you showed it
2. Please choose three Core Values from the ten values above that you think is your weakness, share us one specific example where you showed it
3. Please order the ten Core Values based on your priority if you work at GO-JEK and why