Data Engineering Technical Assessment (Senior Data Engineer)

Dear candidate,

In this assessment, you are asked to design an End-2-End system that should fulfill the requirements listed below. In addition, there is a short code assignment.

Assignments:

1. **Pipeline design**:

   Requirements:

   - The assignment is to develop a data processing pipeline that will run different data processing "tasks" (a "task" is a code that can be run) on raw data. This system should process the tasks and store the processes' results in a database or a data lake.
   - The system should be run on demand and process different predefined "tasks".
   - One task can invoke other tasks.
   - The system also produces telemetry on its runs (stores/updates the status of the runs).
   - **Pipeline monitoring**: Each execution should be monitored and store its **state** and **results**. This should be ready to be queried at any time.
   - You may use known orchestration services/frameworks.

   a. Provide a diagram of the system (components/flows)
   b. Provide the data model (including db. schema. Which database will you use? Why?)
   c. An example of a query syntax to retrieve specific telemetry of a task that was run

2. **Coding skills assessment** (not related to the above)
   a. You are asked to write a process that checks for the arrival of new files and loads them into their corresponding tables in the DB.
   b. There are two types of files in the folder (See apendinx below):
      i. Objects_detection
         1. The format of the file name - objects_detection_[timestamp].json
         2. These files will hold streaming detection events that are sent from Mobileye's cars.
         3. Each file can hole one or more events.
      ii. Vehicles_status
         1. The format of the file name - vehicles_status_[timestamp].json
         2. This files will holde the latest status of each vehicle.
   c. You can choose whatever DB you wish to hold the received information, but you should take care of common queries that users can perform based on that information. You should also provide the code that configure the DB (creates table/defines scheme and so).

General guidelines:

- Make sure you understand the assignment before you start.
- You should consider within your design: code reuse, flexibility for changes, the ability to run in the cloud, reslliance and the ability to handle at scale.

- As for the coding assignment, please add a readme with clear execution instructions.
- Pay attention to performance (CPU, memory usage, execution time)
- Feel free to ask questions during the assignment (call our rep that gave you the assignment) per need.
- Don’t spend more than 4-5 Hours 😊 You might add assumptions and mention what would you do differently if you had more to work on it.

Once you have finished you should submit the artifacts you created (designs and code) – please upload them to the web and send us the link.

If during your work you are using any assumptions, please list them within the design document or comment in your code – wherever makes more sense.

Good luck!

Appendix – Event Examples

objects_detection_events file content sample:


```
"objects_detection_events": [
     {
          "vehicle_id": "ebab5f787798416fb2b8afc1340d7a4e",
          "detection_time": "2022-06-05T21:02:34.546Z",
          "detections": [
               {"object_type": "pedestrians",
               "object_value": 3},
               {"object_type": "cars",
               "object_value": 2},
               {"object_type": "signs",
               "object_value": 3},
          ]
          "object_type": "pedestrians",
          "object_value": 3,
     },
     {
          "vehicle_id": "ebab5f787798416fb2b8afc1340d7a4e",
          "detection_time": "2022-06-05T21:05:20.590Z",
          "detections": [
               {"object_type": "cars",
               "object_value": 4},
          ]
     },
     {
          "vehicle_id": "ebab5f787798416fb2b8afc1340d7a4e",
          "detection_time": "2022-06-05T21:11:35.567Z",
          "detections": [
               {"object_type": "trucks",
               "object_value": 5},
               {"object_type": "obstacles",
               "object_value": 2},

          ]
     }

]
```


vehicle_status file content sample:


```
"vehicle_status": [

     {

          "vehicle_id": "ebab5f787798416fb2b8afc1340d7a4e",
          "report_time": "2022-05-05T22:02:34.546Z",
          "status": "driving",
     },

     {
```

```
        "vehicle_id": "ebae3f787798416fb2b8afc1340d7a6d",
        "report_time": "2022-05-06T00:02:34.546Z",
        "status": "accident",
    },
    {
        "vehicle_id": "qbae3f787798416fb2b8afc1340ddf19",
        "report_time": "2022-05-09T00:02:34.546Z",
        "status ": "parking"
    }
]
```