

Description

Using **Java 11**, **Spring Boot 2** and **Maven** implement a backend service running in **Cloud Run** and create (configure) a data pipeline in GCP that uses this service for automatic processing of **Avro** files with embedded schema that are being uploaded to a **Cloud Storage** bucket.

Whenever a new file is uploaded to the bucket, it is automatically processed by the backend service. This service parses each file and writes the data from the file into two different tables in **BigQuery**. One table contains all the fields from the files and another table contains only a subset of fields (only non-optional fields according to the Avro schema).

Avro file schema example:

```
{ "namespace": "example.gcp",  
  "type": "record",  
  "name": "Client",  
  "fields": [  
    { "name": "id", "type": "long" },  
    { "name": "name", "type": "string" },  
    { "name": "phone", "type": [ "string", "null" ] },  
    { "name": "address", "type": [ "string", "null" ] }  
  ]  
}
```

Requirements

- Push code and any supplementary documentation (if any) to a **GitHub** repository.
- Adhere to the best practices of software development (e.g. SOLID, type safety, code readability, testability, ...).

- A candidate is expected to be able to present the solution, explain its different parts and their configuration, and demonstrate it in real-time using its GCP account and the GCP Free Tier:

<https://cloud.google.com/free>

Optional requirements (tasks)

- enlarge the schema with more fields of different types.
- write data into different BigQuery tables in parallel;
- gracefully handle all possible errors and exceptions in code including when using GCP Java SDKs;
- investigate and present possible options for productionizing the solution from the point of view of scalability, availability, and performance.