

GIST Engineer

You should send your solution in a written form – document via email to your contact person. Use a format of your own choice, so that you can clearly explain your answers. Please make sure to show us your best and send the solutions on time.

One of typical responsibilities of a GIST Engineer is to export data from various systems into centralised storage. In this assignment, your job is to run an export from one of Bloomreach's product APIs into a data warehouse.

Communication with the Bloomreach Engagement web application is documented [here](#).

Parameters you may find handfull:

- Project token: 962d1b9a-481b-11ea-9329-6a9665f4e6e6
- Basic auth username:
po5sj576o13t3n9m72fhc2d5r0fzd1l4c9w2kkgkd1l1kislqvf8016n1g1dvh
- Basic auth password:
pgpn186dbc7gpzl8gj1g0aiflu5cj6hu7xy0lnejfc4invuwu92tt3316w5kupgo
- Google oauth json:
Supplied along this exercise as a standalone file

Context

ETL jobs run in Kubernetes environment where each code release equals to new image creation. You are imitating exporting/ updating data to/ in production, so your operations should ideally be guaranteed to be atomic. Additionally, your script is part of a larger ecosystem of different ETL jobs. You can create whatever design in Google Cloud Platform (GCP) you see fitting for this purpose.

Tasks

1. Write a script in python which exports all customers in chunks of MAXIMUM 20 lines per insert from the above mentioned Bloomreach Engagement project to a GCP bigquery table under dataset **<<your first name>>_<<your last name>>_coding_exercise**. For creating infrastructure in GCP, please use terraform. Use supplied system account credentials when authenticating with GCP APIs in both cases.

2. Your colleague finds out that the parameter 'number of employees' in the data set is not up-to-date. The requirement is to increase 'number of employees' by 20% for all companies in the dataset. Write a script in python which updates 'number of employees' accordingly, and upserts the data to bigquery only for companies with non-null number of employees.

In your reply to this assignment, please attach the script file and shortly explain your architectural choices.