**Data Engineer Assessment /**

**1 - Build an online RDBMS-database (technology depends on you)**

• DB name must be ***de\_assessment***.

• Min. 5 tables and every table must include 50 rows of data.

• DB type must be retail business (supermarket, automotive sales, or online retail sales).

• FK, Constraints, and Normalization rules should be considered

• id and created\_at are mandatory columns that each table must contain.

**2 - ETL with Apache-Airflow (technology in your local system)**

• Build an ETL, which converts every data in ***db\_assessment*** to a CSV file and then transfers it to GCP Bucket on a daily basis. \*\*\*

• The bucket name must be ***de\_assessment\_bucket***.

• ETL process must be Incremental, e.g. yesterday we had 50 data and we transferred these data to buckets, today we have 10 more new data. And the ETL process must transfer these data to the bucket, not the whole data.

\*\*\* Table for everyday activity MUST be in a CSV file.

**3 - ETL with GCP services (technology in your GCP account)**

• We have a bucket that is called ***de\_assessment\_bucket***. Now, we have to convert and transfer this data to Big Query to make an analysis.

• Every CSV file must be parsed and the parsed data must be added to the tables.

• BQ tables must at least include a ***wild\_cards*** table and a partition table.

• ETL process must be Incremental. For example; yesterday we transferred 5 CSV files from the bucket to the BQ tables. Today we’ll transfer 2 more new CSV files. And the ETL process must transfer these CSV files to the tables, not whole CSV files.

**Deadline:** Friday 26th February

create table branches (nu int, subeKod int, subeAdi varchar(50), il varchar(50), ilce varchar(50), semt varchar(50), calisanSayisi int, kapasite int, urunSayisi int, created\_id date, primary key (subeKod));

create table products(nu int, urunKod int, urunAdi varchar(200), listeFiyatı float, kdvOran float, marka varchar(50), kategorri varchar(50), mensei varchar(50), created\_at date, primary key (urunKod));

create table customers (id int, isim varchar(50), gender varchar(5), dogumTarihi date, il varchar(50), ilce varchar(50), telefonNu varchar(50), adres varchar(200), primary key(id));

create table orders (nu int, kullaniciKod int, faturaAdresi varchar(200), siparisTarih timestamp, toplam float, faturaKod int,subeKodu int, primary key(faturaKod), foreign key (kullaniciKod) references customers(id), foreign key (subeKodu) references branches(subeKod));

create table orders\_details (nu int, faturaKod int, urunKod int, fiyat float, kdv float, adet int, primary key(nu), foreign key(faturaKod) references orders(faturaKod), foreign key(urunKod) references products(urunKod));

gcloud sql export csv database gs://okans\_database/okan.csv --database=de\_assessment --query="select isim from customers"

./cloud\_sql\_proxy -instances="amiable-vent-305512:europe-west1:database"=tcp:8080 -credential\_file="credit.json" &