Data Engineer Questions

1) You are given 3 files
   - regusers.tsv : Userids, Birth date and gender of the registered Users
   - urlmap.tsv : Links -> Product mappings
   - web_log.tsv : Raw website log

   Tasks
   ➔ Write a script to ingest the three tables into a Postgres database
   ➔ Join the tables on whichever features you think are important to an analytical study (you need not use all the features) and export the joined table to a NoSQL database of your choice
   ➔ Write a scheduling script in a tool of your choice to run this join daily and write/append it to the NoSQL DB you chose above
   ➔ Explain your logic, assumptions, schema and how you would scale this architecture as the volume grows

2) At Voicemod, our data comes from 2 sources:
   - Web
   - Desktop Application

   1) Theoretically orchestrate a pipeline from these two sources to a database (cloud) on which the analytical team can run ad-hoc SQL queries
   2) We have multiple verticals or departments and need to have separate tables that are relevant to the KPIs they are tracking. How would you change the previous pipeline/architecture so that relevant data for each department goes to their local data mart which is connected to a BI solution?
   3) If we store data from these 2 sources in a data lake, would the pipeline change?