

DATA CHALLENGE

TASK 1 - DATA EXTRACTION & LOADING

We want to load the attached data sources into 2 separate tables in RDBMS. Please provide the necessary code to do so.

Please consider the following while designing your solution you can also descope as required.

1. Data Volume
2. Data Quality
3. Scalability
4. Ease of deployment

track_events.json

weather.json

TASK 2 - ARCHITECTURE

Imagine we are receiving data as shown above from different sources from multiple event streams. We want to capture those events, write the data into a datalake and ingest certain data into a datawarehouse for data modelling to be used for BI/Reporting tools.

Please design an architecture that would allow us to deal with the above scenario and propose technologies you would use (benefits/drawbacks of different tools would be helpful).

Present your results in a Github Repository and share the results back with

<https://github.com/ccbrandenburg>

<https://github.com/igavrylenko>

<https://github.com/vikramsinghchandel>