

## **Objective : Build a Real Time Analytics Platform for E-commerce data**

We have three data sources ,

CRM Database ( Master Data Refresh )

Transactional Database (Insert / update )

Click Stream data ( Only insert )

### **Assumptions**

Transactional Data and Click stream data are both high traffic and volume.

The CRM database does not change in real time and is a slow changing data.

### **Data Ingestion Strategy**

The best approach to do a real time / near real time data ingestion would be to ingest data using **pub sub** model. This model provides the data at the destination with the least data latency. Create **Data flow** pipelines to subscribe to these topics and ingest data into BQ.

Data flow **automatically scales up or down** resources and can process data in parallel based on workload and does not involve manual intervention. Data flow charges based on resources consumed not provisioned.

### **DataProcessing**

**Dataflow** : Utilize dataflow to perform real time data cleansing , transformation and enrichment.

**BigQuery** : BQ makes a good choice for this design as the use case is to utilise this for analytics. It can handle both batch and streaming effectively.

### **Cloud Storage**

Cloud storage may be used to store raw data where required and is very cost effective solution for storing large data for short duration.

### **Data Access :**

**Big Query Studio** may be used by users who wish to query the data.

IAM profiles with the right roles may be used to class various access for different persona and protect sensitive data

**Looker Studio** may be used to create reporting dashboards and visualisation of data