**Relational Database Analytics Platform for Ecommerce**

**Purpose & Motivation**

Ensure atomicity for sales transactions

Ability to run analytics for business insights and decision making

Specifically to e-commerce: product performance, understand trends, segment customers, customer behavior

Production ready features

Scalable, secure, fault tolerance, high performance

**Dataset:**

https://www.kaggle.com/datasets/abdul0haadi/e-commerce-sales-dataset-csv?resource=download

[E-commerce sales dataset .csv](https://www.kaggle.com/datasets/abdul0haadi/e-commerce-sales-dataset-csv?resource=download)

**Database design:**

Schemas – star, snowflake

Normalization 1NF, 2NF, 3NF

**Data Cleaning:**

Format dates? Handle missing values, pre-processing for proper date formats

Drop columns?

**Schema Design**

**Columns in dataset:**

Row ID, Order ID, Order Date, Ship Date, Ship Mode,

Customer ID, Customer Name, Segment, Country, City, State, Postal Code, Region,

Product ID, Category, Sub-Category, Product Name,

Sales, Quantity, Discount, Profit

**Tables**

Customers

Orders

Products

**ER Diagram Features**

One-to-many relationship: customer -> many orders

One-to-many relationship: product -> many orders

One-to-many relationship: order -> product

**3. Project Organization**

**Who leads each phase?**

**Database design**

Using MySQL Workbench - Database design for loading data, design schema, clean/transform data, normalization, ER diagram, queries

**Backend (Owen)**

Launch in a Virtual Private Cloud (VPC)

**OLTP**

Export the schema and data to be imported into AWS Aurora MySQL

This will be our cloud OLTP database

Aurora can handle real-time transactions like new orders and updating customer data

\*API Gateway could be added for data importing from the internet

**OLAP**

ETL? Schema type snowflake vs star

Amazon Redshift data warehouse

**Visualization & Deployment**

From Redshift -> Quicksight for dashboard analytics

WEB UI dashboard with Quicksight

**Technology stack:**

We will be using a relational database because the dataset is organized in such a way that it can be normalized and queried. This will lead to better efficiency when running queries, efficient reads and writes.

MySQL Workbench for loading data, design schema, clean/transform data, normalization, ER diagram, queries

AWS Aurora – high throughput, high volume relational database engine

AWS Redshift – data warehouse

AWS Quicksight – Dashboard analytics

Internally (IAM-authenticated users), Externally (via embedded dashboards in apps or websites

**Timeline:**

Schema design

Query/function development

Interface or script prototyping (Note for Option 2: you must design a script that can be replicated

Integration & testing

Demo preparation

**Collaboration**

GitHub

Final\_project.mwb file for SQL script

**IAM Setup for a 3-Person AWS Project Team**

1. Create IAM Users for Each Team Member
2. Create IAM Groups Based on Roles
3. Attach IAM Policies to Groups