Case Study: Data-Driven Event Growth & Optimization

Company Context:

You've joined EventX, a fast-growing Series A startup that provides an event planning platform helping individuals and businesses organize events, sell tickets, and engage attendees. The company has seen rapid growth, but leadership is struggling with fragmented data and unclear performance metrics.

Currently, EventX collects data from:
Web and app interactions (e.g., page views, clicks, searches)
Ticket sales & purchases (e.g., event types, pricing, discount usage)
Marketing campaigns (e.g., email open rates, ad performance)
Customer feedback (e.g., survey responses, complaints)
In-progress and completed events (e.g., expected revenue, booked vendors, incoming and outgoing payments)

As the **Senior Analytics Engineer**, your goal is to build scalable, efficient data infrastructure and ensure accurate, accessible insights for decision-making.

Part 1: Data Modeling & Pipeline Design

Scenario:

EventX currently stores raw event data in multiple third-party sources (Google Analytics, Stripe, HubSpot, PostgreSQL, Equals, Monday.com, and Google Sheets). The company wants to consolidate this into a **modern data warehouse** and improve reporting reliability.

Tasks:

1. Data Modeling:
Design a high-level schema for a data warehouse to support analytics and reporting.
Define key fact and dimension tables to track revenue, customer behavior, and event performance.
Explain how you'd ensure data consistency and integrity.

2. ETL/ELT Pipeline Strategy:
Propose an **ETL/ELT pipeline** using modern tools (e.g., dbt, Airflow, Snowflake, BigQuery).
Outline how you'd handle incremental updates and late-arriving data.
Discuss how you would maintain **fault-tolerance** in this pipeline as EventX grows.

Part 2: Business Impact Analysis

Scenario:

The CEO wants insights into ticket sales trends to improve marketing efforts and increase revenue. The marketing team is unsure how discounts and promotions affect sales.

Tasks:

1. SQL Challenge:

Revenue Analysis by Event Category

- Write a query to calculate the total revenue per event category.
- Use window functions to calculate the cumulative revenue for each event category over time (from earliest to latest event).
- Calculate the percentage change in total revenue month-over-month for each event category.

Discount Impact on Sales

- Write a query to calculate the percentage of tickets sold at a discount per event category.
- Use window functions to find the moving average of the discount percentage (over the last 3 months) for each event category.
- Calculate the total revenue lost due to discounts for each event category (assuming the difference between the original price and the discounted price).

Customer Segment Analysis

- Write a query to calculate the average revenue per customer segment.
- Use window functions to calculate the rank of each customer segment by average revenue (rank segments in descending order of average revenue).
- Additionally, use a window function to calculate the total revenue per customer segment while considering how customer segments change over time.

Performance Optimization

- Suggest optimizations for running the above queries efficiently, particularly focusing on how to utilize indexes and partitioning
- Discuss the trade-offs between performance and data accuracy when working with large datasets

2. KPI Recommendations:

Identify three key metrics to track ticket sales effectiveness.
Propose a dashboard structure for leadership to monitor these insights.