

Black Friday Sale Analysis for Walmart: Real-Time Sales and Inventory Management

Background

Black Friday, the biggest shopping day of the year, presents a unique opportunity for retailers like Walmart to boost sales, attract new customers, and clear inventory. Managing the surge in transaction volumes, understanding customer preferences, and optimizing inventory in real time are critical challenges that require sophisticated data solutions.

Objective

Develop a real-time data warehousing solution that ingests, processes, and visualizes sales data from Walmart's Black Friday event. This solution should provide insights into sales performance, customer behavior, and inventory levels to enable data-driven decision-making for inventory management, marketing strategies, and customer experience enhancement.

Data Sources **(Write mock scripts to generate such kind of data on scale with random values)**

Real-time streams:

- **Stream 1: Sales Transactions Stream**

Fields: transaction_id, product_id, timestamp, quantity, unit_price, store_id

Sample data:

```
{ "transaction_id": "T1001", "product_id": "P501", "timestamp": "2023-11-24 00:01:00", "quantity": 2, "unit_price": 299.99, "store_id": "W001" }
{ "transaction_id": "T1002", "product_id": "P502", "timestamp": "2023-11-24 00:03:00", "quantity": 1, "unit_price": 99.99, "store_id": "W002" }
```

- **Stream 2: Inventory Updates Stream**

Fields: product_id, timestamp, quantity_change, store_id

Sample data:

```
{ "product_id": "P501", "timestamp": "2023-11-23 23:59:00", "quantity_change": -2, "store_id": "W001" }
{ "product_id": "P502", "timestamp": "2023-11-24 00:02:00", "quantity_change": -1, "store_id": "W002" }
```

Preloaded Dimensional Data:

- **Products**

Fields: product_id, name, category, price, supplier_id

Sample data:

```
{ "product_id": "P501", "name": "Electronics Gadget", "category": "Electronics", "price": 299.99, "supplier_id": "S101" }
```

- **Stores**

Fields: store_id, location, size, manager

Sample data:

```
{ "store_id": "W001", "location": "New York, NY", "size": 25000, "manager": "John Doe" }
```

Requirements

- **Data Warehousing:**
 - Design and implement a data warehouse to ingest real-time sales and inventory data.
 - Ensure the data warehouse supports high-velocity data ingestion and provides capabilities for advanced analytical queries.
- **Visualization and Dashboards:**
 - Develop dashboards for each key metrics
- **Derive Key Metrics**
 - Identify hourly, daily, and weekly sales trends during the Black Friday event. Highlight periods of peak sales activity.
 - Analyze sales trends by geographic region or store location to understand regional preferences and demand patterns.
 - Determine the top-selling products and categories based on quantity sold and revenue generated. Highlight which products are exceeding sales expectations and which are underperforming.
 - Monitor real-time inventory levels and identify products that are at risk of stockouts. Provide alerts for items that require immediate replenishment to avoid lost sales.
 - Calculate the sell-through rate (the ratio of units sold to units received) for each product to evaluate inventory turnover and effectiveness in meeting customer demand.
 - Calculate total revenue generated in real-time, segmented by product category, store, and overall for the Walmart chain.
 - Identify high-revenue stores and regions
 - Rank stores by sales volume and revenue to identify top performers

Tech Stack:

Anything - Any open source, Any paid tool, Any cloud solution, Any programming language, free to use anything

Evaluation Criteria

- Completeness of the Pipeline: End-to-end functionality from source to destination data ingestion
- Data Processing Efficiency: Speed and precision of calculations.
- Dashboard Usability: Clarity, real-time capabilities, and user experience.
- Scalability: Ability to handle vast data volumes and scale with increased data rates.
- Documentation: Clear documentation of architecture, data flows, and tools/components used.

Must do things for project submission

- Complete code should be checked in into Github along with readme file, architecture diagram, installation steps and step by step process to execute the project

- Teams need to create a 4-5 minutes of working Video demo of their project, need to explain the architecture and working of complete project with real time data
- No project submission will be accepted if above 2 points are not taken care