Project: E-commerce Data Modeling Challenge

Business Context: We operate an e-commerce platform selling a wide range of products. We need to build a data warehouse to track sales performance, inventory levels, and product information. Data comes from multiple sources and has different update frequencies. Data Description You are provided with the following data points:

Product Information (updated daily):

product_id (unique identifier)
product_name
category (e.g., Electronics, Apparel)
supplier_id
supplier_name
unit_price (the price can change over time)

Sales Transactions (streamed in real-time):

order_id
product_id
customer_id
order_timestamp
quantity_sold
sale_price_per_unit
shipping_address (city, state, zip code)

Inventory Levels (snapshot taken every hour):

product_id
warehouse_id
stock_quantity
snapshot_timestamp

Requirements

Design a dimensional data warehouse model that addresses the following:

Data Model Design:

Create a star or snowflake schema with fact and dimension tables to store this data efficiently.
Your model must handle changes in product prices over time (Slowly Changing Dimensions).
The design must accommodate both real-time sales data and hourly inventory snapshots.

Schema Definition:

Define the tables with appropriate primary keys, foreign keys, data types, and constraints.

Data Processing Considerations:

Explain how your model supports analyzing historical sales with the product prices that were active at the time of sale.
Describe how to handle the different granularities of the sales (transactional) and inventory (hourly snapshot) data.

Deliverables

- A complete Entity-Relationship Diagram (ERD) illustrating your proposed data model.
- SQL DDL statements for creating all tables, keys, and indexes.

A written explanation detailing:

- The reasoning behind your modeling choices (e.g., why you chose a specific SCD type).
- The trade-offs you considered.
- How your model enables key business queries, such as "What was the total revenue by product category last month?" and "What is the current inventory level for our top 10 selling products?"
- Your recommended indexing strategy to optimize query performance.