Part B: Data Models for Starbucks

4. Starbucks Data Warehouse Schema

The Starbucks Data Warehouse Schema in the attached image represents a star schema designed to centralize and integrate data from different operational systems into a unified structure. This schema is optimized for analytical queries, reporting, and business intelligence. The star schema revolves around a central Fact Table connected to multiple Dimension Tables, allowing efficient querying and data aggregation.

Components of the Data Warehouse Schema

1. FactSales Transaction Table:

- **Role:** This is the central fact table containing transactional data, such as sales, linked to dimensional tables for deeper insights.
- **Key Attributes:**
    - **OrderID:** Unique identifier for each transaction.
    - **CustomerKey (FK):** Links to the DimCustomer Table.
    - **DateKey (FK):** Links to the DimDate Table.
    - **ProductKey (FK):** Links to the DimProduct Table.
    - **EmployeeKey (FK):** Links to the DimEmployee Table.
    - **SupplierKey (FK):** Links to the DimSupplier Table.
    - **Quantity, UnitPrice, and TotalPrice:** Store metrics for analytics, such as sales volume and revenue.
- **Purpose:** This table allows tracking of sales data, such as which products were sold, by whom, to whom, and when, along with associated metrics.

2. DimEmployee Table:

- **Role:** Stores employee-related information for analysis, such as performance or contribution to sales.
- **Key Attributes:**
    - **EmployeeKey (FK):** Unique identifier linking the employee to the fact table.
    - **EmployeeID (PK):** Unique ID for employees in the source operational system.
    - **FirstName, LastName, Position, HireDate, Salary:** Detailed attributes of employees.
- **Purpose:** Provides insight into employee performance, sales impact, and workforce analysis.

3. DimCustomer Table:

- **Role:** Maintains detailed customer information for understanding purchasing behavior and customer segmentation.
- **Key Attributes:**
  - **CustomerKey (FK):** Unique identifier linking the customer to the fact table.
  - **CustomerID (PK):** Identifier for customers in the source system.
  - **FirstName, LastName, Email, PhoneNumber, Address:** Detailed attributes for customer profiling.
- **Purpose:** Enables segmentation, targeted marketing, and analysis of customer loyalty.

4. DimProduct Table:

- **Role:** Contains detailed information about the products sold.
- **Key Attributes:**
  - **ProductKey (FK):** Unique identifier linking the product to the fact table.
  - **ProductID (PK):** Product identifier from the operational system.
  - **ProductName, ProductDescription:** Details about each product.
- **Purpose:** Supports product-level analytics, such as identifying top-selling products and performance of product categories.

5. DimSupplier Table:

- **Role:** Stores supplier information to track procurement and supply chain performance.
- **Key Attributes:**
  - **SupplierKey (FK):** Unique identifier linking the supplier to the fact table.
  - **SupplierID (PK):** Identifier for suppliers in the source system.
  - **SupplierName, ContactNumber, Email:** Details about suppliers.
- **Purpose:** Enables analysis of supplier performance, sourcing efficiency, and cost optimization.

6. DimDate Table:

- **Role:** A time dimension table that provides a detailed calendar for temporal analysis.
- **Key Attributes:**
  - **DateKey (FK):** Links to the fact table for time-based queries.
  - **FullDate, Year, Quarter, Month, Day, DayOfWeek:** Attributes for date-based aggregations.
- **Purpose:** Facilitates time-series analysis, such as sales trends over time and seasonality.

Why This Schema is Crucial for Starbucks

1. **Centralized Data Analysis:** The star schema consolidates data from disparate systems (Employee, Customer, and Supplier Management Systems), creating a single source of truth for analytics.
2. **Improved Decision-Making:** Enables Starbucks to make data-driven decisions by analyzing customer behavior, employee performance, and product trends in one unified structure.
3. **Scalable and Flexible:** The schema is optimized for querying large datasets, allowing Starbucks to scale as data volumes grow without compromising performance.
4. **Efficiency in Querying:** The fact-dimension relationships allow quick aggregation of key metrics, such as total sales, average revenue per customer, or product-specific performance.

Relationships in the Schema

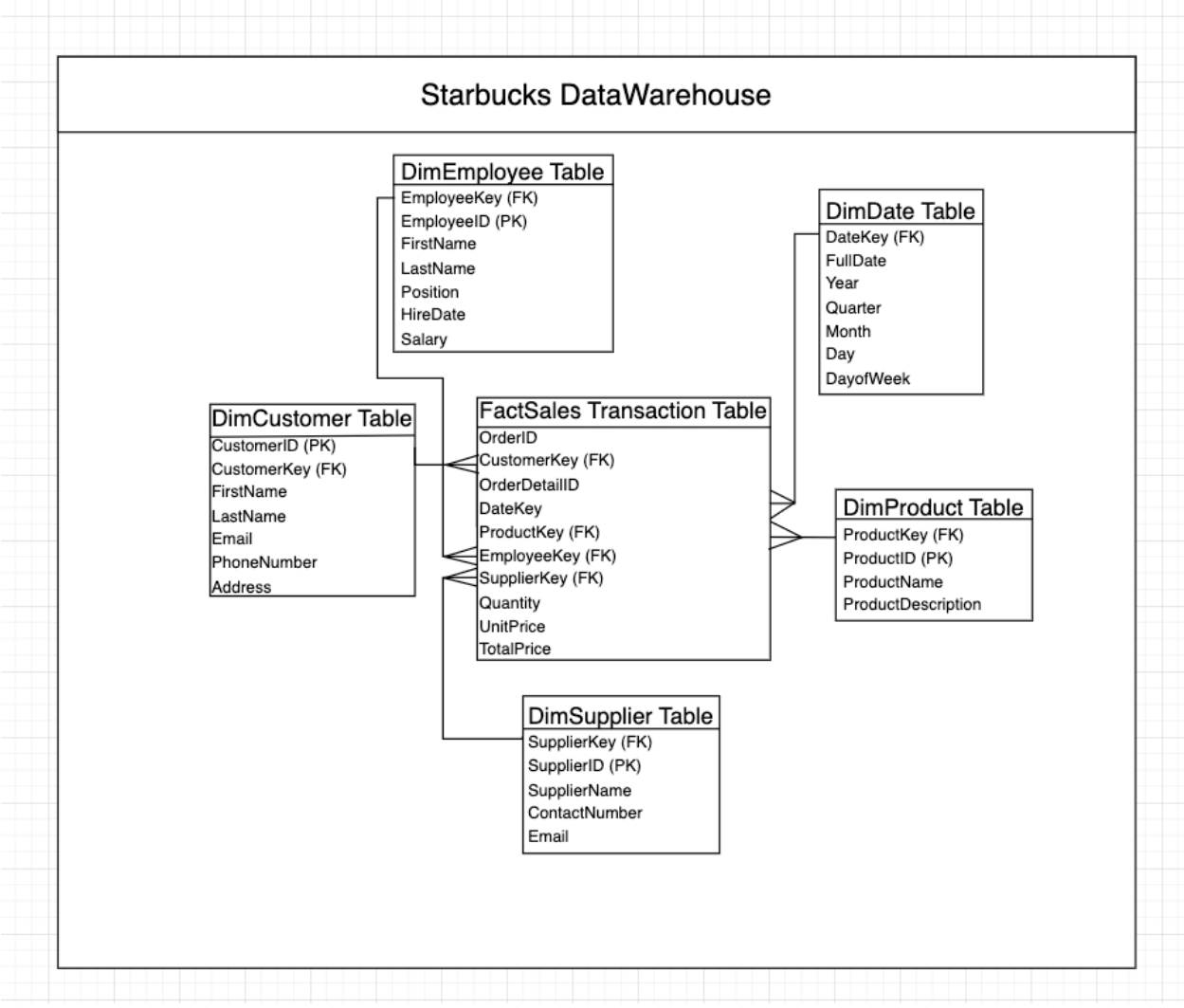
1. **Fact Table to Dimension Tables:**
    - The FactSales Transaction Table serves as the central hub and is connected to all dimension tables via foreign keys (CustomerKey, EmployeeKey, ProductKey, SupplierKey, and DateKey).
    - Each foreign key points to a unique record in the corresponding dimension table, allowing detailed drill-down analysis.
2. **One-to-Many Relationships:**
    - Each dimension table has a one-to-many relationship with the fact table. For example:
        - One employee (DimEmployee Table) can be linked to multiple transactions in the FactSales Transaction Table.
        - One product (DimProduct Table) can appear in multiple sales transactions.

Benefits of a Star Schema Over Independent OLTP Systems

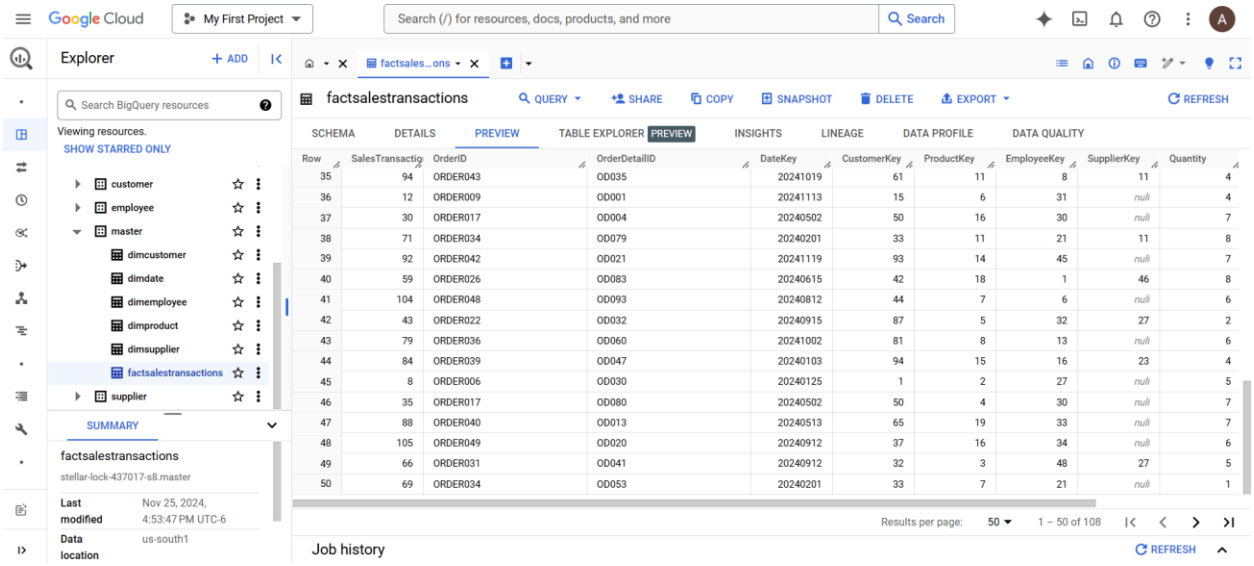
1. **Integrated Data Analysis:** Combines customer, employee, product, and supplier data into a single schema, eliminating silos and enabling holistic analytics.
2. **Better Query Performance:** Star schemas are optimized for read-heavy operations, making them ideal for business intelligence queries and reporting.
3. **Simplified Maintenance:** Centralized schema simplifies database maintenance compared to managing separate OLTP systems for employees, customers, and suppliers.
4. **Enhanced Scalability:** As Starbucks' data grows, the star schema can easily adapt, allowing the addition of more dimensions or facts without disrupting the existing structure.

Benefits of the Data Warehouse for Starbucks

1. **Customer Insights:** Enables Starbucks to analyze customer purchase patterns and create targeted marketing campaigns, enhancing customer loyalty and driving revenue growth.
2. **Operational Efficiency:** Provides insights into supplier performance and employee contributions, helping optimize the supply chain and improve workforce management.
3. **Product Performance Analysis:** Identifies best-selling products, underperforming items, and opportunities for introducing new products.
4. **Time-Based Analytics:** Allows Starbucks to track trends over time, such as seasonal demand variations or annual growth rates.



BIGQUERY:


Google Cloud
My First Project
Search (/) for resources, docs, products, and more
Search
Explorer
ADD
Search BigQuery resources
Viewing resources.
SHOW STARRED ONLY
customer
employee
master
dimcustomer
dimdate
dimemployee
dimproduct
dimsupplier
factsalestransactions
supplier
SUMMARY
factsalestransactions
stellar-lock-437017-s8.master
Last modified
Nov 25, 2024, 4:53:47 PM UTC-6
Data location
us-south1
factsales...ons
factsalestransactions
QUERY
SHARE
COPY
SNAPSHOT
DELETE
EXPORT
REFRESH
SCHEMA
DETAILS
PREVIEW
TABLE EXPLORER
PREVIEW
INSIGHTS
LINEAGE
DATA PROFILE
DATA QUALITY
Row SalesTransactio OrderID OrderDetailID DateKey CustomerKey ProductKey EmployeeKey SupplierKey Quantity
35 94 ORDER043 OD035 20241019 61 11 8 11 4
36 12 ORDER009 OD001 20241113 15 6 31 null 4
37 30 ORDER017 OD004 20240502 50 16 30 null 7
38 71 ORDER034 OD079 20240201 33 11 21 11 8
39 92 ORDER042 OD021 20241119 93 14 45 null 7
40 59 ORDER026 OD083 20240615 42 18 1 46 8
41 104 ORDER048 OD093 20240812 44 7 6 null 6
42 43 ORDER022 OD032 20240915 87 5 32 27 2
43 79 ORDER036 OD060 20241002 81 8 13 null 6
44 84 ORDER039 OD047 20240103 94 15 16 23 4
45 8 ORDER006 OD030 20240125 1 2 27 null 5
46 35 ORDER017 OD080 20240502 50 4 30 null 7
47 88 ORDER040 OD013 20240513 65 19 33 null 7
48 105 ORDER049 OD020 20240912 37 16 34 null 6
49 66 ORDER031 OD041 20240912 32 3 48 27 5
50 69 ORDER034 OD053 20240201 33 7 21 null 1
Results per page: 50
1 – 50 of 108
Job history
REFRESH