Information Technology Institute Data Management Track Data Warehouse Project (Intake 44/2024)

Data Model Design Report

1. Overview of the Project

The goal of this project is to build a data warehouse for an airline to analyze its operations and improve customer satisfaction. We want to give executives a way to track flight activity, reservations, flight revenue, and segment-level details. By collecting both operational data and customer-centric information, the airline can see the full picture of its business.

2. Data Model Design

The data warehouse employs a dimensional structure with a star schema design to support analytical processing and business insights. This design was chosen for several key advantages, including scalability, query performance, and ease of understanding. It also aligns with our goal to maintain a logical separation of data, while still ensuring strong connections between related information. Moreover, it provides the flexibility needed to allow for future expansion as business requirements evolve.

3. Data Model Components

Our data warehouse design encompasses multiple fact tables, each serving a distinct role in capturing and analyzing key aspects of airline operations. Here's an overview of the fact tables and how they contribute to business insights.

Flight Activity Trip

The Flight Activity Trip fact table provides detailed information about individual flight segments and related metrics. It includes data on origin and destination airports, booking channels, passenger profiles, and various financial elements like fare revenue, taxes, and fees.

Reservations Fact

The Reservations Fact table captures comprehensive reservation details, including passenger data, origin and destination airports, aircraft types, and class of service. It provides a detailed view of how reservations are made and the financial implications.

Flight Revenue Fact

The Flight Revenue Fact table focuses on the revenue aspects of individual flights, including unearned revenue and the number of remaining seats by class. It helps the airline track revenue generation and seat capacity utilization.

Segment Level Flight Activity Fact

The Segment Level Flight Activity Fact table provides granular insights into flight segments, capturing segment-specific information such as departure and arrival times, passenger data, and various financial metrics.

Frequent Flyer Points

The Frequent Flyer Points fact table tracks the activities of the airline's loyalty program. It helps monitor points accumulation, redemptions, and total points earned by frequent flyers.

Customer Interaction

The Customer Interaction fact table captures various interactions between the airline and its customers, including inquiries, complaints, and feedback. This table is crucial for analyzing customer service quality and understanding passenger satisfaction.

Passenger Tier Data (Factless)

The Passenger Tier Data factless fact table tracks changes in passenger tiers and related activities. This table is used to analyze passenger tier upgrades and other tier-related interactions.

Bringing It All Together

These fact tables collectively provide a robust framework for analyzing the airline's operations and customer-centric data.

4. Indexing and Partitioning in Data Warehousing

Partitioning is like organizing your closet. When you know where everything is, you can grab what you need without tearing the place apart. In our data warehouse project, partitioning serves the same purpose—it breaks down large tables into smaller, more manageable chunks, making data easier to find and queries faster to run.

Partitioning in Our Data Warehouse

We use partitioning to boost performance, scalability, and maintainability. By creating smaller data segments, we make the data warehouse more efficient and easier to work with. Here's how we approach partitioning:

Range Partitioning

Range partitioning splits data based on a range of values, like dates or numbers. It’s a great way to organize data by time, making it easy to focus on specific periods without scanning the whole table. In our data warehouse, range partitioning helps us manage large tables by:

Date: By creating partitions for specific dates or periods (like weeks, months, or years), we can quickly retrieve data from a particular timeframe. This is crucial for analyzing trends, especially when looking at flight activity or reservation patterns.

Time: We can partition by specific times of day, which is useful when analyzing busy periods like peak travel hours or off-peak times.

Places: Sometimes, it's helpful to create partitions based on locations, like airports or regions. This allows us to run queries focused on specific geographic areas without scanning unrelated data.

List Partitioning : This method organizes data based on a predefined list of values. For example, we might partition by specific flight routes or airline classes, making it easier to analyze data for those specific categories.

Indexing in Our Data Warehouse

Indexes are like a table of contents—they help you find things quickly. Different types of indexes serve different purposes, depending on the kind of data and queries we're working with. Here's how we approach indexing:

Primary Index

Primary indexes are created on the primary key of a table, ensuring each record has a unique identifier. This makes it easy to look up records quickly. In our project, primary indexes are essential for:

Fact Tables: We implement primary indexes on composite keys, like the `Passenger Tier Data` table, to ensure each record has a unique combination of keys. This makes it easier to join tables and retrieve specific data points.

Passenger Tier Data: In this factless fact table, the primary index is on a composite key consisting of Passenger Key, Date Key, Time Key, Passenger Profile Key, and Tier Change Key. This composite key reflects the unique combination of attributes that define an entry in the fact table, allowing for fast lookups and efficient joins with related tables.

Tracking Passenger Tier Changes: The composite primary index on Passenger Tier Data allows the data warehouse to track changes in passenger tiers efficiently, enabling insights into tier upgrade patterns and loyalty program effectiveness.

Frequent Flyer Points: The primary index here is on the Passenger Key, Passenger Profile Key, Date Key, Time Key, and Points Activity Key. This combination represents the unique identifier for each frequent flyer points record, aiding in query performance when analyzing frequent flyer data.

Analysis of Frequent Flyer Behavior: With primary indexes on Frequent Flyer Points, queries that examine frequent flyer activities are executed more quickly, enabling efficient analysis of points accumulation and redemption.

Dimension Tables: Primary indexes on dimension tables like `Points Transaction Profile` ensure unique identification and speed up joins with fact tables.

Points Transaction Profile: The primary index on Points Transaction Profile Key ensures unique identification for each profile record. This index facilitates fast joins with related fact tables, such as Frequent Flyer Points

Join Operations in Analytical Queries: Primary indexes on dimension tables, such as Points Transaction Profile, speed up join operations, allowing for more complex queries regarding how they earn and redeem their frequent flyer miles.

Bitmap Index

Bitmap indexes are great for columns with a limited number of unique values (low cardinality). They use bitmaps to represent data, making them efficient for complex queries. In our data warehouse, bitmap indexes are used for:

Frequent Flyer Status: Since frequent flyer status has a limited set of possible values (like Blue, Silver, Gold), a bitmap index allows for quick filtering and analysis.

Date and Time: Bitmap indexes on dates and times help speed up time-based queries, allowing us to quickly identify records within a specific timeframe.

Why These Choices Matter

By using a combination of partitioning and indexing techniques, we ensure our data warehouse is fast, scalable, and easy to maintain. These strategies not only boost query performance but also make it easier to expand the data warehouse as the airline grows and data volumes increase. It's all about keeping things organized and accessible, so we can quickly find the insights we need to drive better business decisions.

AirOps Data Warehouse

By:

Marina Fawzy

Hussien Khaled

Radwa Esmaiel