**Airline Company Data Warehouse**

**TEAM 1**

* Mohamed AlGhaly
* Salma Ahmed
* Ahmed Ali

**Project Documentation**

**STEPS**

1. Understanding the business.
   * Conduct research on how Major Airline Companies operate.
2. Requirements Gathering.
   * Understand Business Needs.
   * Identify Data Sources.
3. Analyzing Source Systems’ Data.
4. Defining Business KPIs & DWH Objectives.
5. Defining Business Processes.
6. Defining Granularity for each business process.
7. Choosing the Optimal DWH Architecture.
8. Deciding the technology stack.
9. Defining Dimensions and Fact Tables.
   * We need also to define the type of each (Fact table, Dimension Table, Measure)
10. Data Warehouse Modeling.
    * Creating the Logical Model (The Schema).
11. Creating the Physical Model.
12. Generating & Populating data.
    * In this step we will simulate data sources’ data by generating data using a Python script.
13. Data Integration.
    * Moving data from source Systems to the DWH.
14. Indexing & Partitioning.
15. Data Analysis.
    * For step 7: we will work on a traditional DBMS (Oracle DBMS) on a local device.
    * For step 11: we don’t have a data source, so we will generate some realistic data to populate the DWH.

PS:

* The mentioned steps show a more **generalized** DWH Modeling **approach**, some of the steps may not be applicable in our case.
* We have excluded some crucial steps like choosing the technology stack which includes hardware, database software (e.g., Oracle, SQL Server, Snowflake) and choosing between On-Premises and Cloud.
* We have also excluded other steps like choosing the DWH provider, estimating DWH size, defining sparsity of the data, and determining hardware specifications.
* We act as the business owners and we don’t have Data Sources to refer to, so any assumptions made will be from our point of view and we will try to include as detailed description as possible.

**STEP1**

Understanding the business

**Assumptions about the business:**

Frequent Flyers Program (FFP)

**Core Objective:** Encourage repeated bookings by rewarding customers with points that can be redeemed for various benefits.

* Passengers can earn points from various channels, not only direct airlines, but also indirect airline partners such as cobranded credit cards, supermarkets, car rental companies, and international hotels and resorts. The points can be either redeemed for flights, flight upgrades or additional services such as extra baggage and airport lounge access.
* Airlines often run **promotions** that offer additional points for flying certain routes, during specific periods, or for using certain services.
* As stated on the business requirements (**the first deliverable should focus on the flight activity**), we are only considered with the flight activity, so we will ignore point earned by using our partners’ services and we will only focus on point earned by traveling with us!
* **Redeeming Points:**
  + **Flight Tickets:** The most common use of accumulated points is to redeem them for free or discounted flights.
  + **Class Upgrades:** Points can often be used to upgrade to a higher class of service, such as from economy to business class.

**STEP2**

Defining KPIs & DWH Objectives

DWH Objectives:

The objective of this project is to assist the executive management to analyze their current business processes and expand the company by discovering new opportunities.

We can achieve this by creating a robust and structured Data Warehouse that consolidates data from various sources, facilitating analytical reporting and business intelligence.

Business KPIs & Analysis Scope:

* Enhancing the overall business performance.
* Analyzing the behavior of company’s frequent flyers.
* Measuring the performance of our marketing team.
* Measuring the performance of our Loyalty Program.
* Analyzing the company’s profit.
* Which customers use our services more frequently.
* Analyzing our revenues in each country.
* Defining our main source of revenue.
* Which customer segmentation (gold, platinum, titanium) is most valuable.
* Our most popular booking channels.
* Improving customer satisfaction.
* What are the booking patterns of frequent flyers, and what types of fare classes do they typically book.
* How do customer demographics, such as age or income level, impact travel behavior and preferences.
* Ensure a seamless operation of our loyalty program.

**STEP3**

Defining Business Processes

1. **Ticketing Transactions**:
   * This process captures transactional data related to bookings made by customers. It supports analyses of sales performance, customer booking behaviors, and revenue management.
   * This process is mainly concerned with the Finance Team.
   * The analysis scope is to analyze company’s revenues.
2. **Frequent Flyers**:
   * This process maintains detailed records of frequent flyer mileage transactions, supporting the management of the airline's loyalty program.
   * This process is mainly concerned with the Marketing Team.
   * The analysis scope is to analyze frequent flyer’s behavior.
3. **Customer Care**:
   * This process is mainly focused on achieving higher customer satisfaction.
   * It captures customers’ inquiries, complaints and keeps their feedback.
   * This process is mainly concerned with the Customer Support Team.
   * The analysis scope is to define any problems within the company and to ensure a better customer experience.
4. **Flights**:
   * This process records key operational data for each flight conducted by the airline. It supports performance analysis, operational planning, and decision-making processes related to flight operations.
   * It mainly focuses on our flight activity.
   * This process is mainly concerned with the Upper Management Team.
   * The analysis scope is to analyze the overall company’s performance.

**STEP4**

Defining Granularity

**Possible Granularities**

The following are the possible grains to consider across defined business processes.

* Transit:
  + The step from one airport to the other.
  + A flight from San Francisco to Minneapolis with a stop/transit in Denver will be mapped in 2 rows:
    - San Francisco to Denver and Denver to San Francisco.
* Flight:
  + The trip from journey to destination.
  + A flight from San Francisco to Minneapolis with a stop/transit in Denver will be mapped in a single row holding data about the source and destination only.
* Trip:
  + A return flight that takes you from Paris to London and takes you back from London to Paris is mapped in a single row.
* Daily, weekly, etc.

**Choosing Granularities for Each Fact Table**

1. Ticketing Transactions:

* We will work on the grain of each ticket booked by any customer.
* A customer books a flight from San Francisco to Minneapolis with a stop/transit in Denver will be mapped in a single row in this table.

1. Frequent Flyers:
   * The most detailed grain in this fact table will be of each flight booked by any customer **on the loyalty program.**
   * We will use this fact table to analyze frequent flyers’ behavior and their booking patterns in addition to how they redeem and earn flyer mileages.
2. Customer Care:
   * This most detailed grain is each action taken by a customer in our company, this action can be a complaint, feedback, or inquiry.
3. Flights:
   1. The most detailed grain is each segment flight organized by the company.

**STEP5**

Choosing DWH Architecture

* We have chosen a Typical Two-layer Architecture (Ralph Kimball’s) approach.
* It will be almost possible to build a 3rd normal form Enterprise Data Warehouse without having a solid understanding of data sources.

**STEP6**

Choosing Technology Stack

* We will implement the DWH on Oracle DBMS.

**STEP7**

Defining Dimensions & Facts

Step seven will be the capstone for the project which involves determining facts, dimensions, and measurements.

**FACTS & MEASUREMENTS:**

1. Ticketing Transactions ()
2. Frequent Flyers (Transactional Fact Table)
   1. What?
      1. This is a fact table to keep track of each flight booked by a frequent flyer, this fact table will be valuable for the marketing team to measure the benefits of our loyalty program, promotions, and other marketing campaigns.
      2. This fact table will also be used to analyze the behavior of our frequent flyers to help us build brand loyalty.
   2. Measurements:
      1. Points Redeemed (additive).
      2. Points Earned (additive).

Passengers can earn/redeem point from any other activities rather than booking a flight, but analyzing those activities are out of the analysis scope for this fact table, so we will only be concerned about points earned from booking a flight or points redeemed as a discount on a ticket!

PS: When we say points, we refer to flyer miles.

* 1. Dimensions:
     1. Passenger
     2. Class
     3. DATE
     4. AIRPORT
     5. Promotion
     6. Status
     7. Ticket
     8. Flight
     9. Fare Base

**STEP8**

DWH Modeling

**STEP9**

Physical Model

**STEP10**

Populating Data

**STEP11**

Creating Data Marts

**STEP12**

Indexing & Partitioning

**STEP12**

Gaining Insights