**We have decided to use the state of the art Data Vault model because:**

- This approach is modular. meaning that the entire system is composed of smaller modules that can be incremented or removed without affecting the rest of the system.
- Its approach is closer to the business requirements, the analyst and data vault developer will need to understand the business very well to design and implement.
- It is similar to the 3 layer enterprise data warehouse approach. The data vault is extremely detailed at the finest possible grain with ALL data from source systems, and thus the data is always available in its smallest form whenever we need new information marts.
- It has 88% faster loading (ingestion) of data from the staging area's source tables. This is because the hubs, links, and satellites are all loaded in parallel, while in the 3 layer approach for example, the data needs to be inserted as a primary key in one table, before it can be used as a foreign key in the other tables.
- It uses a hash function which is 100% CPU and no disk, thus greatly reducing I/O operations, which are a bottleneck in any system. This is used to:
    1. Generate primary keys, rather than access a sequence object's CURRENTVAL from disk and increment one, or calculate the MAX() also from disk.
    2. Can support near-real time reporting by optimizing insertion, and updating the satellites by producing a hashdiff attribute out of the entire table by hashing the entire row together, and hashing rows from the source systems table, and comparing to the existing hashes. Rather than comparing column by column in what could be a table with 50+ columns.

---

**Business Process** : booking Reservations and Flight Activities

**Grain: per reservation**

**Dimensions**: promotion_dim , fare_basis_dim, passenger_dim,

Flight_dim, reservation_channel_dim, flight_dim, class_dim, aircraft_dim, interaction, surveys

**Facts** : reservation_fact , customer care _fact, Fact Flying , Fact_Overnight_Layover , Fact_Product_Sales .

—----------

# Details of each model component

## 1) date_dim

**It represents data about the Dates and it contains 4 attributes.**

- **Date_Key :is the surrogate key which identifies this table and it has Number datatype·**
- **Day: Shows the day and its datatype is varchar , EX: saturday, sunday…**
- **Month: shows the Month and it has varchar datatype, EX: November,**

  **Day of months: shows the day of the months and its datatype is number, Ex: 15 , 12**

- **Year: shows the year and it has number datatype, EX: 2025**

## 2) Promotion_dim

**It represents all data about the Promotions and it contains 7 attributes.**

- **promotion_key: is the surrogate key which identifies this table and it has Number datatype·**
- **Promotion_ID : is the business key which identifies each Promotion in the business side and it has Number datatype**
- **Promotion_code: Shows the promotion code and it has Varchar datatype, Ex: bbxxxu,**
- **Start_date: Shows Promotion Start Date and its datatype is Date, EX: 15/9/2023**
- **expire_date: Shows Promotion expire Date and its datatype is Date, EX: 15/10/2023**
- **Amount : represents the amount of discount and its datatype is Number, EX: 50 LE , 20 LE …**
- **Name : represent the name of the promotion and its datatype is varchar EX: black friday promotion**

## 3) Channel_dim

**It represents all the data about booking channels and it contains 3 attributes.**

- **Channel_Key is the surrogate key which identifies this table and it has Number datatype**
- **Channel_ID is the business key which identifies each Channel in the business side and it has Number datatype**

- **Channel_Type it describes the channel of the booking and its datatype is Varchar, EX: website channel**

### 4) fare_basis_dim

**It represents all the data about Fare basis that the airline company has and it contains 4 attributes.**

- **Fare_basis_Key is the surrogate key which identifies this table and it has Number datatype**
- **Fare_basis_id is the business key which identifies fare basis in the business side and it has Number datatype**
- **Fare_basis_Code identifies each Fare basis that the airline company uses and it has Varchar2 datatype, EX: ETF123 .**
- **Class_type: it shows the class type and it has Varchar datatype, EX: economy class, first class ….**
- **Fare_basis_Description Describes the meaning of this code and its datatype is Varchar , EX: full fare economy ticket**

### 5) Flights_dim:

**It represents all the data about all the flights and it contains 6 attributes.**

- **Flight_Key: is the surrogate key which identifies this table and it has Number datatype**
- **Flight_ID: is the business key which identifies Flights in the business side and it has Number datatype**
- **Departure_time: represents the time of the departure of each flight and its datatype is Date, EX: 1:20 am**

- **Departure_date: represents the date of the departure of each flight and its datatype is Date, EX: 15/3/2020**
- **·duration: represents minutes which the flight will take and its datatype is number, EX: 120 minutes**
- **Departure_airport: represent the airport name, which the flight will departure from and its datatype is varchar , EX: cairo airport**
- **Arrival_airport represents the airport name, which the flight will arrive to, And its datatype is varchar , EX: king fahd airport**
- **miles : represent the miles which the flight will took and its datatype is number**

## 6) Aircraft_dim:

**It represents all the data about all the Aircrafts and it contains 4 attributes.**

- **Aircraft_Key is the surrogate key which identifies this table and it has Number datatype, EX:1, 2, 3, …etc.**
- **Aircraft_ID is the business key which identifies Aircraft in the business side and it has Number datatype**
- **Aircraft_capacity represents the number of the seats in the aircraft And its datatype is number EX: 129 seats**
- **Aircraft_model represent the name of the aircraft and its datatype is varchar EX: Airco DH.4A**

## 7) passenger_dim:

**It represents all the data about all the passengers in the company and it contains 4 attributes.**

- **passenger_Key: is the surrogate key which identifies this table and it has Number datatype, EX:1, 2, 3, …etc.**
- **National_ID : represent the national id of each passenger and it has varchar datatype**
- **Firist_name represents the name of the passenger And its datatype is varchar**
- **last_name represents the last name of the passenger And its datatype is varchar**
- **Address represents the address of the passenger And its datatype is varchar**
- **Phone_number: represents the phone number of the passenger And its datatype is varchar**
- **frequent_flyer_id: when passengers sign up or the frequent flyer program to get free miles when flying, they get a frequent_flyer_id. if they are not subscribed, this is NULL**
- **status: for frequent flyers, this is silver, gold, or titanium.**

**8) Reservation fact**

**It used to make analysis related to finance team which is mainly interested in analyzing the company profit**

**It has 14 attributes**

- **Date_key : which is reference to date_dimension to determine the date in which the reservation was made and the money was paid**
- **Promotion_key: which is reference to promotion_dimension to determine the promotion on this reservation**
- **Fare_basis_key: which is reference to fare_basis_dimension To determine the fare_basis_code and the class**
- **Channel_key : which is reference to the channel_dimension to determine the channel where the reservation done**
- **passenger_key : which is reference to the passenger_dimension to determine the passenger full data**
- **flight_key : which is reference to the flight_dimension to determine the flight full data**
- **payment_key : which is reference to the payment_dimension to determine the payment type**
- **aircraft_key : which is reference to the aircraft_dimension to determine the full data of the aircraft on this flight**
- **Ticket_number: which is degenerated dimension which determine the ticket number and its datatype is varchar**
- **Ticket_price : measurable fact, which determine the ticket price,**

  **And its datatype is number**

- **reservation _fees : measurable fact, which determine the fees of the reservation , And its datatype is number**
- **Tax : measurable fact, which determine the tax, and this tax is not included in the company revenue, it's for the government, And its datatype is number**
- **revenue : revenue is derived fact and will be calculated by this assumption:**

**If (the payment type is earned miles )**

**Then revenue = reservation_fees**

**Else{**

**if(reservation.promotion_key is not null && sysdate is less than promotion_dim.expire_date)**

**Then**

**revenue = reservation_fees +ticket_price - promotion_dim.amount**

**Else**

**revenue = reservation_fees +ticket_price**

**}**

### 9) Fact Flying

**This has data about the reserved passengers who actually took their flight. and has 8 attributes. the grain is per flight per customer.**

- **date_key (foreign key)**
- **passenger_key (foreign key)**
- **flight_key (foreign key)**
- **fare_basis_key (foreign key)**
- **payment_key (foreign key)**
- **promotion_key (foreign key)**
- **is_transit: a TRUE/FALSE value of whether this flight was direct or a transit flight.**
- **ticket_number (DD)**

### 10) Factless_PromotionFlight

**This table contains data about all the promotions on all the flights. regardless of whether or not any passengers (frequent or not) have responded to said promotions.**

**It has 2 attributes:**

- **promotion_key (foreign key)**
- **flight_key (foreign key)**

## 11)Fact_Overnight_Layover

**This table contains information about passengers who have stayed overnight in between transit flights**

- **date_key: (foreign key)**
- **bookin_time: the time the passenger has checked into the (airport) hotel**
- **passenger_key (foreign key)**
- **payment_key (foreign key)**
- **duration: length of stay in the hotel in hours**
- **price: depending on the payment_key, this could be either be dollars in cash, or it could be redeemed miles by frequent flyers.**

## 12) Dim_Products

**This contains the list of products sold by the airline company in its stores such as accessories, phone chargers, and etc. Those can be purchased either using cash, or redeemed using miles.**

- **product_key: integer surrogate key for products. (number)**
- **product_id: natural key for products in the airline source systems (varchar).**
- **product_name: the commercial name of the product (varchar)**
- **product_description (varchar)**

## 13) Fact_Product_Sales

**This contains the list of products purchased by passengers. its grain could either be monthly or daily, but will be chosen to be daily to match the grains of the other fact tables**

- **date_key: foreign key**
- **passenger_key: foreign key**
- **product_key: foreign key:**
- **payment_type: foreign key**
- **price: the price of the item in either miles or dollars (number)**

## 1- Customer Care :

**Measure feedback or problem severity from passengers according to interaction type whether it was before, within and after the trip.**

- **date_key (foreign key)**
- **passenger_key (foreign key)**
- **flight_key (foreign key)**
- **interaction_id (foreign key)**
- **survey _id (foreign key)**
- **Problem_severity (number value)**
- **feedback (number value)**

## 2- INTERACTION Table:

**It contains passengers' interactions before, within and after the trip .**

- **Interaction_id : to identifies each INTERACTION**
- **INTERACTION_type: (1- before, 2- within , 3- after)**

## 3- Surveys Table:

**It contains customer surveys whether it was about inquiries, complaints or feedback.**

- **survey _id (pk) : it represents a number for each category.**
- **Survey_category: ( inquiries, complaints , feedback)**
- **1 for feedback**
- **2 for inquiries**
- **3 for complaints.**
- **Customer_comment (text) : it contains anything the passenger writes on the Survey in order to keep their feedback for business analytics or reporting.**