British Airway DWH design

**Brief**   
This System is especially design to assist the management of the famous airline company **British** **Airway** to make better data-driven decisions and to discover new opportunities to expand their business. They have specific business requirements that we’ll elaborate and fulfill later on in this documentation.

**Why -Three Layer Architecture Approach**

We’ve choose **3-Layer Architecture approach** as it initially covers complete business areas and combines data according to current needs in a subsequent step. In contrast, the 2-layer Architecture Approach for Large-Scale Data Warehouse Systems defines DW based on actual users’ requirements. This leads to higher complexity and higher volume of extracted data. More conceptual work is necessary, so that duration and costs of the implementation are affected.

In traditional 2-Layer architectures, data are usually not available in a reusable format, and must be reloaded from sources. In our approach, the propagation layer holds data in a format that enables rebuilding into the overlying levels. Further, data have to be remodeled for new business requirements, Business often requires data that are not included in the initial concept of DW’s design. Even in case such data are part of already connected sources, they are not included in data extraction in traditional architectures. Hence, the data flow from source to DW must be enhanced and all data must be loaded again. In **the 3-layered architecture**, we extract all potentially relevant data into the DW when a new source is extracted.

At the same time data-vault won’t be the best fit as the data flow coming from source systems aren’t that fast that require a real-time or a near-real time analysis it would be a waste of money and resources, As a previous day data won’t have this much of an effect of our case, it’s like buying a Ferrari and ride it in Cairo’s roads.

**The 3NF Model**

**Customer** (**Cust\_Id**, first\_name,last\_name, phone, passport\_id , Nationality)

**Customer**\_**rank( Cust\_Id, actual miles** , Home\_airport)

**Rank**( **Actual\_miles**, gained\_miles\_rank\_type)

note : knowing that if the rank has a vaule (titanium,gold , platinum) then it's a frequent flyer if not it's a regular customer

**Reservation**(**Ticket\_num** ,Customer\_id, , Flight\_id, Class, Payment\_Channel, Origina\_fair, Actual\_fair,Payment\_Method, promotion\_id)

**Promotion**( **Promotion\_id**, promotion\_type,promotion\_percentage)

**FareBase(Ticket\_num,** base\_faire , No\_of\_ways, changeability, direct\_ticketing,baggaes\_fees, tax\_amount )

**Flight**( flight\_id ,Aircraft\_id, Destination\_airport, Arrival\_airport, sceduled\_departure\_fdate, actual\_departure\_fdate , sceduled\_arrival\_fdate, actual\_arrical\_fdate,

sceduled\_arrival\_ftime, actual\_arrival\_ftime, scheduale\_arrival\_ftime, actual\_arrival\_ftime,

**Segment**( **Sgement\_Id,flight\_id,** Segment\_origin\_airPort,Segment\_destination\_airPort, Segment\_milesFlown, sceduled\_departure\_sdate, actual\_departure\_sdate , sceduled\_arrival\_sdate, actual\_arrical\_sdate,

sceduled\_arrival\_stime, actual\_arrival\_stime, scheduale\_arrival\_stime, actual\_arrival\_stime)  
  
**Leg** (**leg\_id**, **segment\_id**, leg\_depart\_time, leg\_arrival\_time )

**AirCraft**(**Aircraft\_id**, Aircraft\_model, Aircraft\_specs , Aircrafr\_capacity(No\_of\_seats) ,Aircraft\_maintHist, Traveled\_miles )

**AirPort**(**AirPort\_id,** Name, Country , city , AirCraft\_id )

**AirCraft**/**AirPort( AirCraft\_id** , **AirPort\_id** ,Segment\_id, flag(Origin/Destination) )

**Interaction**(**Inter\_ID**, Channel, Desc , Status, Timing , flight\_id , feedback , problem\_sev)

InteractCustomer (**inter\_id,** cust\_id, agent\_id)

Agent (**ssn** , name , phone , Numberofteractions , working\_hr )

**First Business Process**

* Frequent Flyers management system to analyze the frequent flyer behavior and monitor his activity to ensure him the best experience.

**System Key Performance Indicators**

* What flights the company’s frequent flyers take.
* How often they upgrade.
* How they earn and redeem their frequent flyer miles.
* Whether they respond to special fare promotions.
* How long their overnight stays are.
* What proportion of these frequent flyers have gold, platinum or titanium status.

**Granularity**

The **granularity** **level** will be a Frequent Flyer Ticket on a segment level which means that each row in the **Freqflyer\_segment\_fact** table will represent an actual reservation for a frequent flyer that has been committed.   
But this table with the segment level of granularity couldn’t answer all the business questions so we had to make another fact table **Leg\_fact** which has a granularity up to leg level.

Knowing that,  
A Flight for example The trip from England to America.  
A Segment if the same customer had a stop in France and changed his aircraft then continued to America it’s a different segment for the same flight.  
A leg, if the same customer had a stop in France and continued to America with the same Aircraft it’s a different leg with the same segment and flight.

**Assumptions and clarifications**

* Each row of the **Freqflyer\_segment\_fact** table represents a customer who actually bought a ticket and he’s already registered in source systems as a frequent flyer which also mean that he matches at least the Titanium category.
* To ensure that we follow the Object oriented method in DWH methodology we created one big dimension for a general customer that has a flag to identify the **frequent flyer** from the regular customer, with this approach we managed to provide different fact table with different granularity and business goals from the same dimension aka object.
* Titanium will be customers who travel more than 25000 miles and less than 35000 miles.  
  Gold will be customers who travel more than 35000 miles and less than 50000 miles.  
  Platinum will be customer who travel more than 50000 miles within 2 years maximum.
* To calculate The gained miles per segment for each frequent flyer in the fact table:  
  Gained Miles= (Actual Miles \* class\_type\_factor \* customer\_rank\_factor).

Class Type factor depends on Customer’s class (Economy, Business or First )  
Customer rank factor depends on which category this customer belongs to either gold, platinum or titanium.

* Gained Miles can be exchanged with free flight with the same mile numbers.
* Frequent flyers data is divided into 2 dimensions to keep monitoring the frequent changes without the hassles of unnecessary redundancy for example customer name, date of birth and nationality will be in frequent flyer profile dimension ­ -as they’re almost static attributes- On the other hand customer’s actual miles, gained miles and rank are on the frequent flyer dimension.

**Second Business Process**

* Reservation and profit analysis

**System Key Performance Indicators**

* How many flights are booked through website
* Overall sales though application

**Granularity**

Granularity level is up to sold ticket.

**Assumptions and clarifications**

* Assuming that the profit is calculated in an outside system and the finance department wants to get the overall sales
* Profit calculated By = Sum(price) from sales Fact table - Cost ( from ERP System)
* The customer can reserve a tickt through multiple channels (website , airport , travel agency or application) with different payment methods (cash/credit )

**Third Business Process**

* Analysis of customer care service

**System Key Performance Indicators**

* Providing customer care interaction monitoring either before, within and after the trip.
* keep the customer feedback about his inquiries, complaints to enhance his experience.
* Analysis should include interaction type and problem severity (if issue exists).

**Granularity**

The granularity level would be up to interaction in the **Customer\_inquires\_fact** whether it’s simple inquiry or complain.

**Assumptions and clarifications**

* Customer can make interaction through different channel (website,phone call , airport or application)
* Each interaction is handled by a customer service agent
* Interation types include (inquiry / complain )
* After each interation the customers gives feedback which is a score ranging from 1 to 5
* The common interaction have predefined degree of severity which is

0- simple inquiry

1- food complain

2- staff complain

4- missing luggage

5- delayed flight

If the customer interation doesn’t fall into any of these categories then it’s up to the agent to determine the severity degree of the issue

* Each interaction has a unique indentifier