Delivered by: Salma Khaled, Jessica Gamal, Youssef Etman

DWH Project delivery

1-State the modelling process

We are going to follow Kimbal's approach in designing our DWH for the following reasons:

1- The data integration requirement is focused on individual business areas (Marketing, Finance, Customer care)
2- It occupies less space compared to Inmon.
3- It's quicker and cheaper to set up compared to Inmon.
4- Requires less ETL work, therefore has faster loading.

In Kimbal's approach we follow the bottom-up methodology where ***we will first model 3 independent data marts, one for each department (Marketing, Finance & customer care) & then we will join them together to construct our DWH***.

We have 5 business processes namely:

Business Process	Granularity
Flight activity	Daily
Flyer miles redemption	Detailed (Per transaction)
Flyer promotions responsivity	Daily
Reservation	Daily
Customer care handling	Detailed (Per report)

2-Logical data models

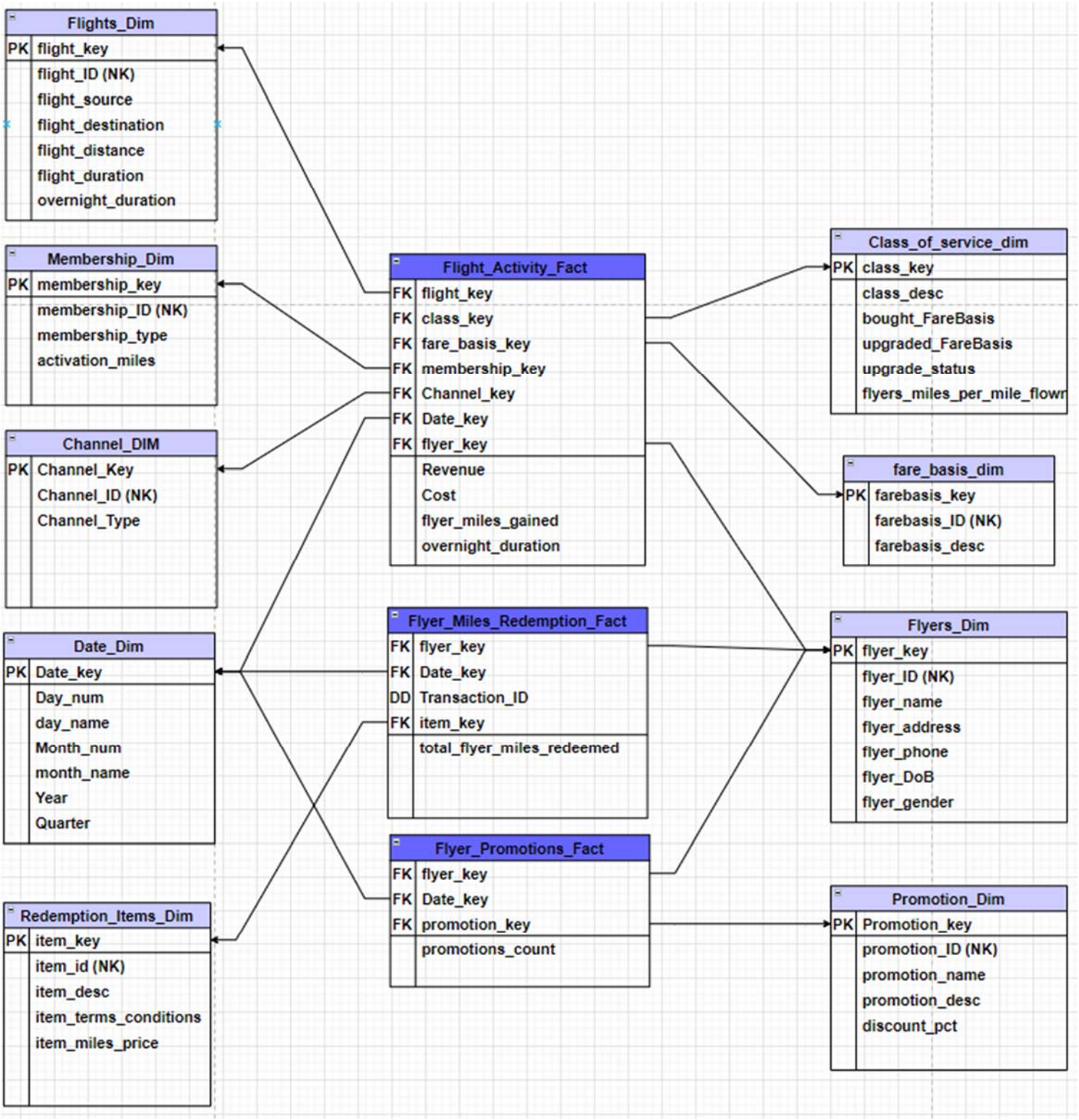
2-1-Marketing data mart

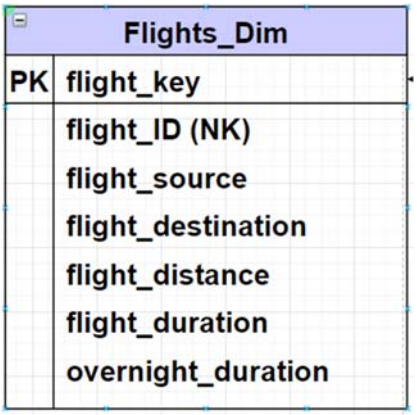
The marketing team is interested in modelling 3 different business processes:

First is the flight activity which measures the number of flights and the flyer miles gained per flight, per flyer, per fare basis, per program membership (will be explained later).
It also measures the total overnight duration spent in the airport (as a transit) per flyer.

Second is the flyer miles redemption activity which measures how many flyer miles are redeemed per flyer, per redemption item.

Third is the flyer promotions responsivity which measures how many promotions have been received per flyer, per promotion type.



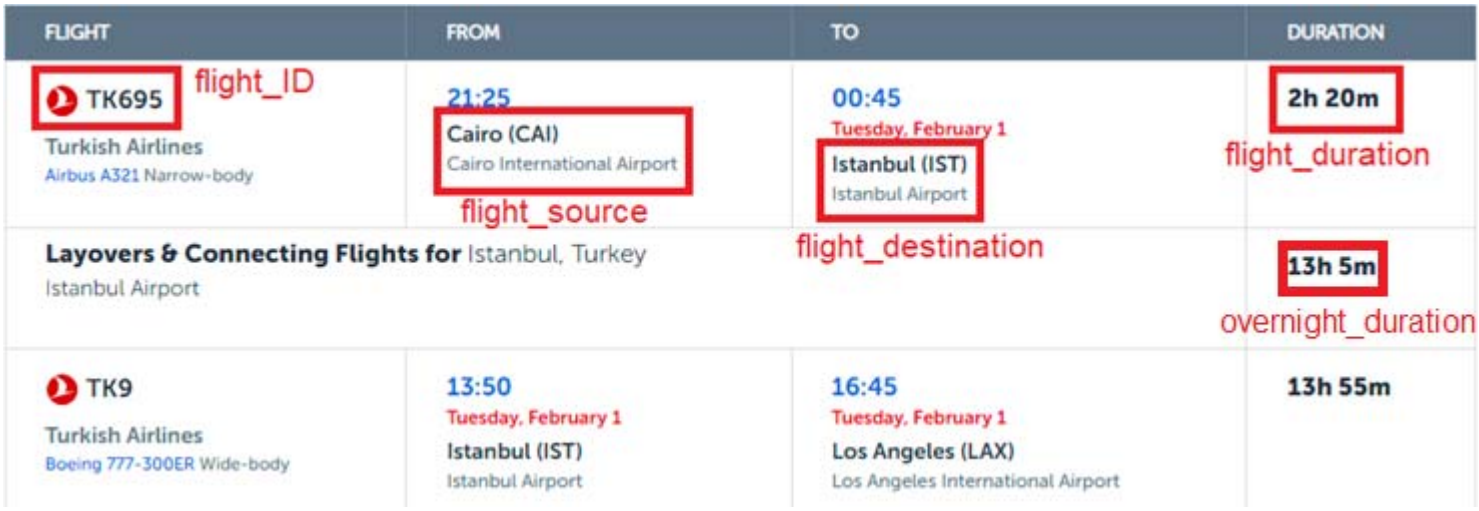


Explaining the dimensions:

Looking at the operational source system we find these attributes:

- Flight_ID
- Flight_source
- Flight_destination
- Flight_duration
- Overnight_duration

We also added a final attribute which is the flight_distance which helps in calculating how many flyer miles our frequent flyers earn per flight. We believe this attribute is also available in the source systems, and if not it can be easily googled and added.



Source: https://www.turkishairlines.com/

Flyers_Dim	
PK	flyer_key
	flyer_ID (NK)
	flyer_name
	flyer_address
	flyer_phone
	flyer_DoB
	flyer_gender

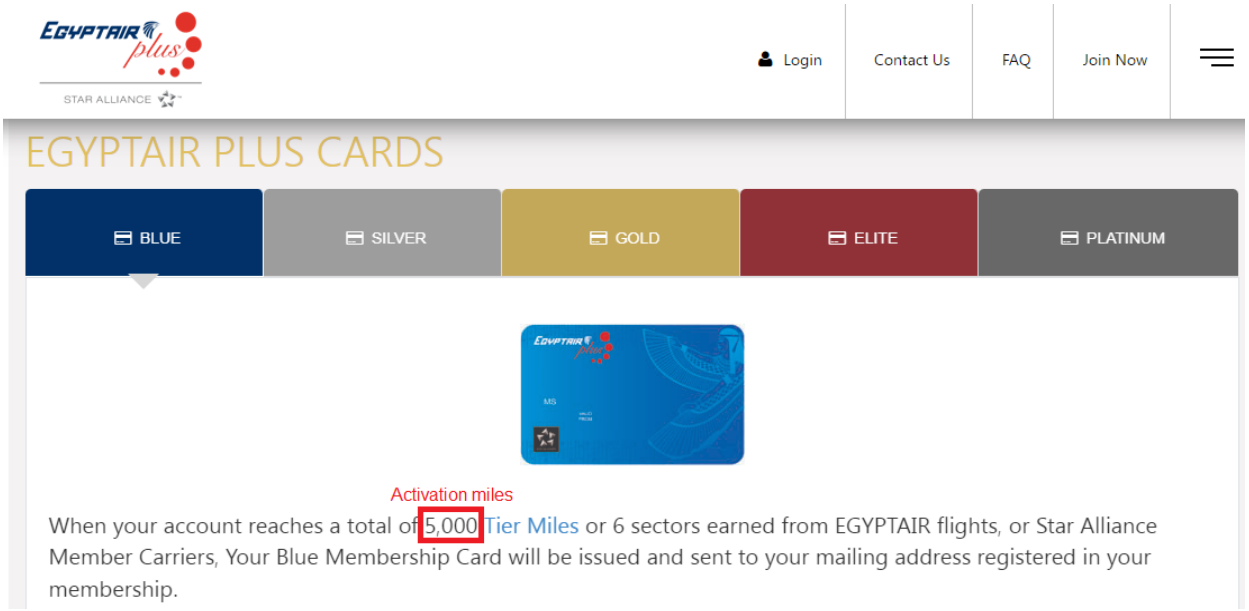
- These attributes are found in the source systems that oversee flyers registration.
- The flyer_ID here will be the flyer's passport number.

Membership_Dim	
PK	membership_key
	membership_ID (NK)
	membership_type
	activation_miles

This dimension is a small lookup table that enables our airline company to keep track of it's frequent or loyal flyers. All Flyers start as normal flyers "white status", and the more total miles they fly with our airline company the better their status becomes therefore becoming frequent flyers.

The following table demonstrates the idea:

Status	*Activation miles*	*Status*
White	0	Non frequent
Gold	50,000	Frequent
Platinum	300,000	Frequent
Titanium	600,000	Frequent

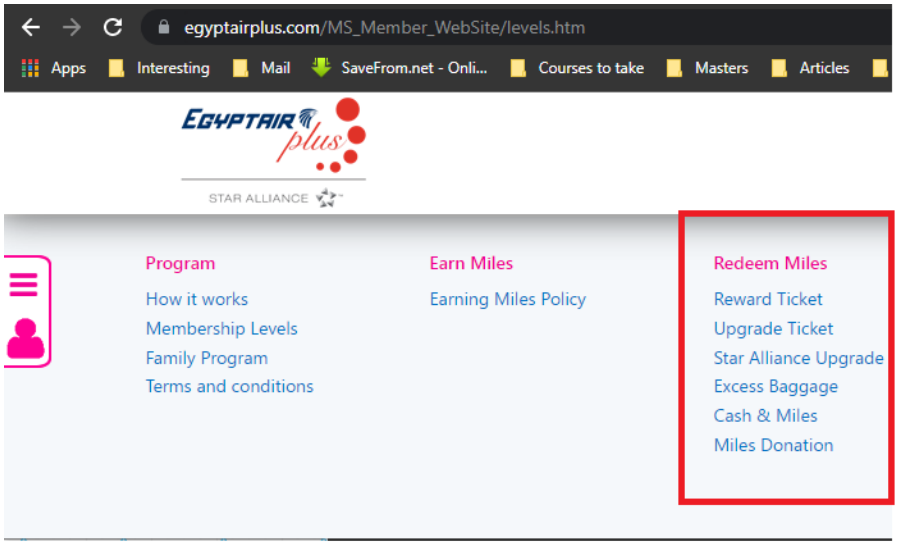


Source: https://www.egyptairplus.com/MS_Member_WebSite/levels.htm

---

This table holds the items or services that can be redeemed as a reward using the frequent flyer miles, examples include:

- Get Excess baggage
- Upgrade your fare basis
- Donate to charities



Instead of the class_of_service dimension we could have had two dimensions namely:

1- Four rows in the class dimension table to indicate first, business, premium economy, and economy classes.
2- The upgrade indicator dimension also would have just three rows in it, corresponding to upgrade, downgrade, or no class change.

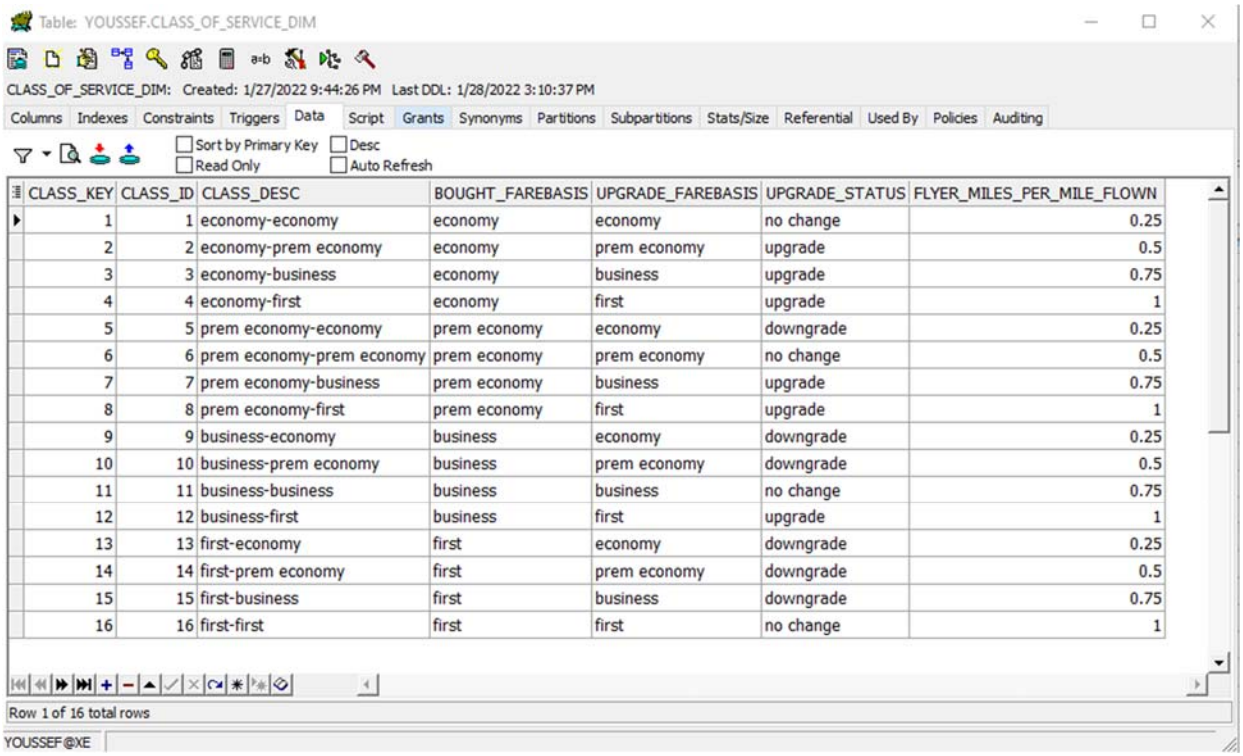
Because the row counts are so small, you can elect instead to combine the dimensions into a single class of service dimension as follows.

Class of Service Key	Class Purchased	Class Flown	Purchased-Flown Group	Class Change Indicator
1	Economy	Economy	Economy-Economy	No Class Change
2	Economy	Prem Economy	Economy-Prem Economy	Upgrade
3	Economy	Business	Economy-Business	Upgrade
4	Economy	First	Economy-First	Upgrade
5	Prem Economy	Economy	Prem Economy-Economy	Downgrade
6	Prem Economy	Prem Economy	Prem Economy-Prem Economy	No Class Change
7	Prem Economy	Business	Prem Economy-Business	Upgrade
8	Prem Economy	First	Prem Economy-First	Upgrade
9	Business	Economy	Business-Economy	Downgrade
10	Business	Prem Economy	Business-Prem Economy	Downgrade
11	Business	Business	Business-Business	No Class Change
12	Business	First	Business-First	Upgrade
13	First	Economy	First-Economy	Downgrade
14	First	Prem Economy	First-Prem Economy	Downgrade
15	First	Business	First-Business	Downgrade
16	First	First	First-First	No Class Change

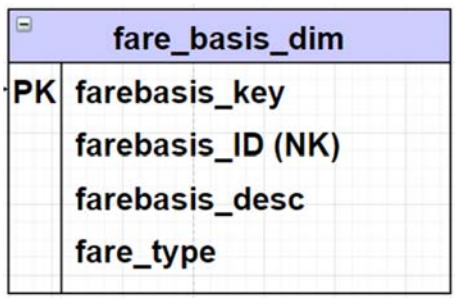
Source: The data warehouse toolkit-3rd ed-Ralph Kimball

This simplifies the queries required to analyze activity based on whether there was an upgrade, downgrade or no change in the fare basis.
We added to this table the “flyers_miles_per_mile_flown” column which indicates how many flyer miles rewards are earned for every one mile flown, this depends on the upgraded_farebasis (class flown) column

Snapshot of our filled class_of_service_dim table:



CLASS_KEY	CLASS_ID	CLASS_DESC	BOUGHT_FAREBASIS	UPGRADE_FAREBASIS	UPGRADE_STATUS	FLYER_MILES_PER_MILE_FLOWN
1	1	economy-economy	economy	economy	no change	0.25
2	2	economy-prem economy	economy	prem economy	upgrade	0.5
3	3	economy-business	economy	business	upgrade	0.75
4	4	economy-first	economy	first	upgrade	1
5	5	prem economy-economy	prem economy	economy	downgrade	0.25
6	6	prem economy-prem economy	prem economy	prem economy	no change	0.5
7	7	prem economy-business	prem economy	business	upgrade	0.75
8	8	prem economy-first	prem economy	first	upgrade	1
9	9	business-economy	business	economy	downgrade	0.25
10	10	business-prem economy	business	prem economy	downgrade	0.5
11	11	business-business	business	business	no change	0.75
12	12	business-first	business	first	upgrade	1
13	13	first-economy	first	economy	downgrade	0.25
14	14	first-prem economy	first	prem economy	downgrade	0.5
15	15	first-business	first	business	downgrade	0.75
16	16	first-first	first	first	no change	1

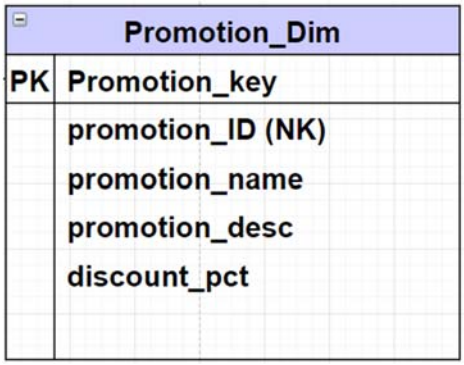


This Dimension holds the classes fare basis and determines whether the ticket has a discount on it or is full paid.
The “fare_type” column has one of two values: full fare or upgraded

Code	Description
F	First Class
P	Premium, usually first class
A	First class, discounted
R	First class suites (A380)
J	Business class, full fare
C	Business class, full fare
D	Business class, discounted
I	Business class, discounted
Z	Business class, discounted
Y	Coach, full fare, upgradeable
Q	Coach, discounted, non-upgradeable
B, M	Coach, discounted, upgradeable, restricted dates
O, T	Coach, no upgrade, no refunds
W, X	Coach, charters, frequent flyer award
S	Coach, extended economy

Source: http://craignow.blogspot.com/2010/08/airline-fare-basis-codes-explained.html

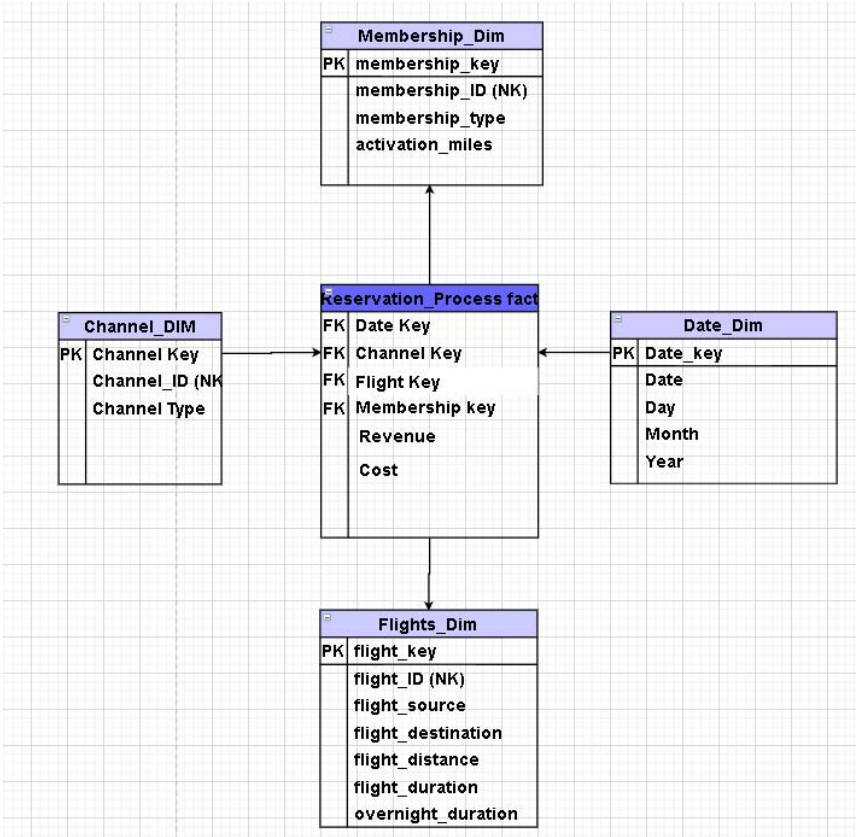
---



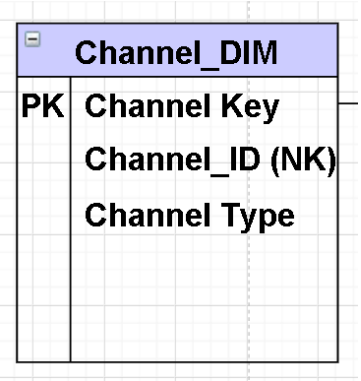
This dimension is filled by extracting and loading data from CMS (Campaign management system). It determines how many promotions has each flyer received whether this promotion was redeemed or not.
To know if a promotion was redeemed we can check the fare_type from the fare basis to see if the ticket had a discount or not.

2-2-Finance data mart

The finance team will be interested in analyzing the company profit which will be measured bt the subtraction of the total revenues earned and total cost paid.

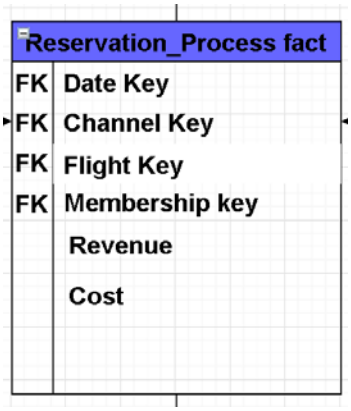


Explaining the dimensions:



This dimension is a lookup table that enables our airline company to keep track of its reservation processes and the type each reservation takes. This dimension contains different channels the flyer can use in his reservation and here is an example of these channels (airline website, directly going to the airport, travel agency, airline application, ……)

Explaining the fact:



This fact contains 2 measures (revenue and cost) that come from the source system which enables me to measure the company total profit by subtracting these 2 measures

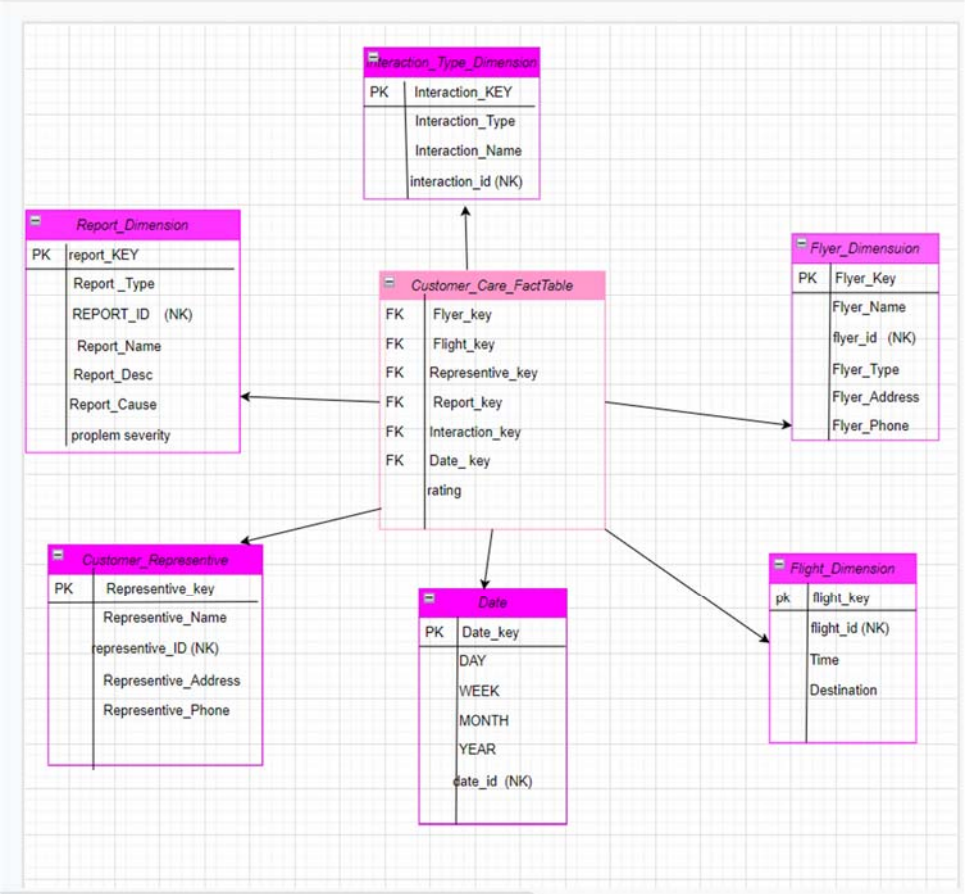
- Profit can be measured per channel (application, directly….) or per membership (gold, platinum……..) or per flight.

2-3 customer care data mart

The customer care team manage people's affairs takes care of their questions, and takes feedback from them and handle customer inquiries, complaints and keep their feedback for business enhancements.

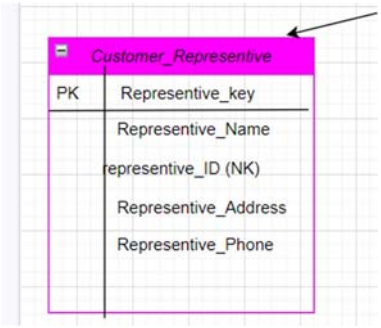
My assumptions level of gurnality per report to handle their questions each day receive people feedback and inquiries about trip or anything within or after trip, assume customer representative team receive all inquiries and complaints of customers for each trip or any problem in system so my assumptions we have 1 fact table and 6-dimension table

And measure quality per representative or per any one of dimension measure it.



Explaining the dimensions:

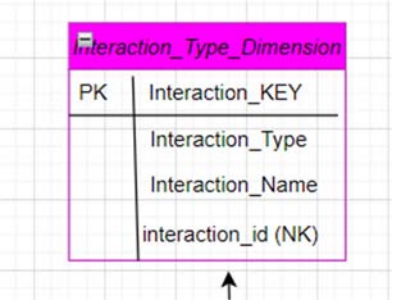
1) customer_ representative dimension: assume we have representative employee and he handle and solve problem and care about customer feedback, so I suppose we have dimension called customer representative get info of it from sources like name, id, address, and phone



---

2) interaction type dimension: it determine type of customer contact with our company, they have many types like via email, phone, chatbot ,application, customer care office

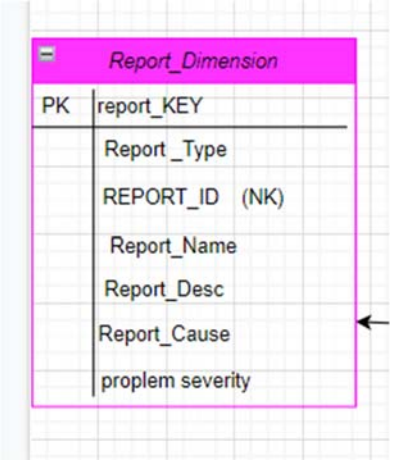
In this dimension table: attributes: interaction, interaction name, interaction type gets all from source system



---

3) report _dimension: in this dimension I measure complaints, feedback and inquires of customer to enhance services, I suppose three in one dimension because it shares same measures

We search on website of EGYPTAIR, and we found care team and we can ask or give feedback or complaints about something from form in website as we do dimension flyer contain information about customer and on which flight on dimension flight then on report dimension, we make we contain report type consist of feedback or inquires or complaints



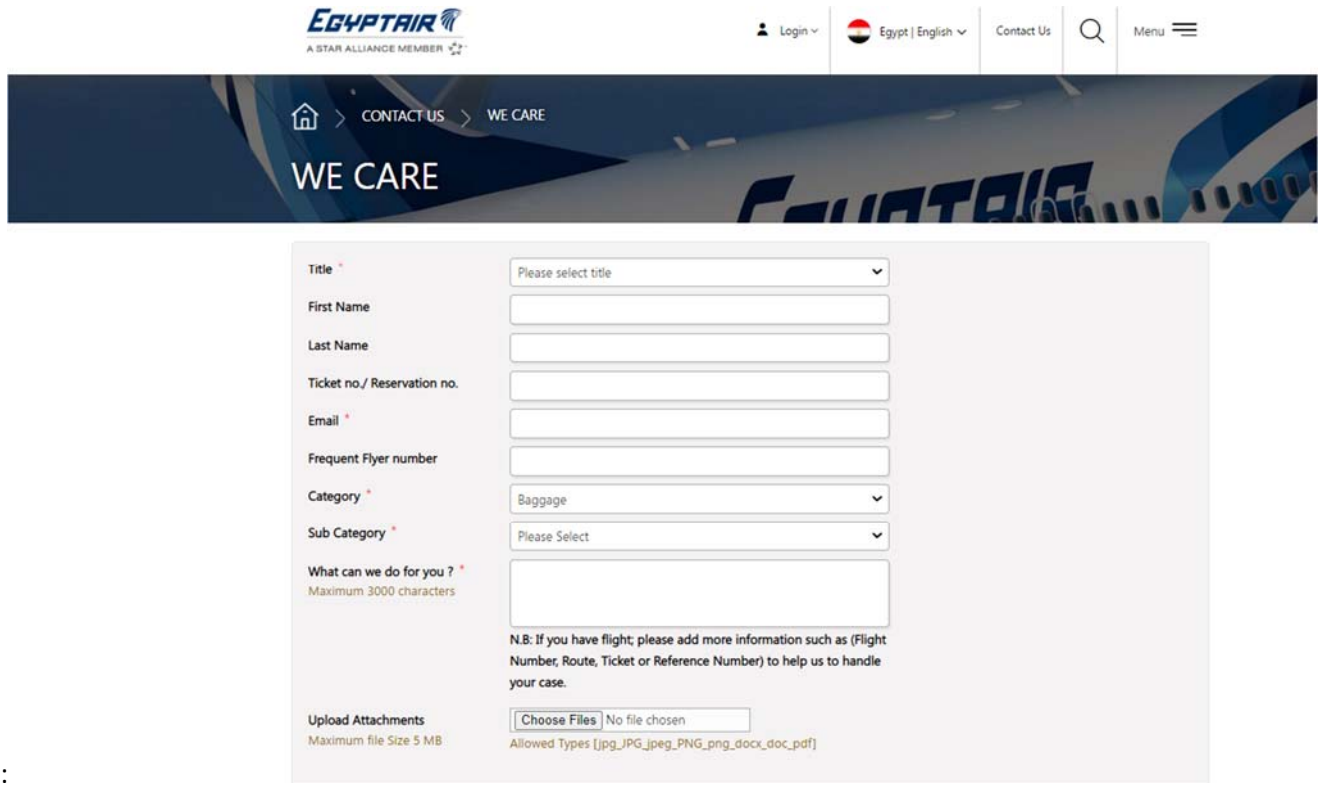
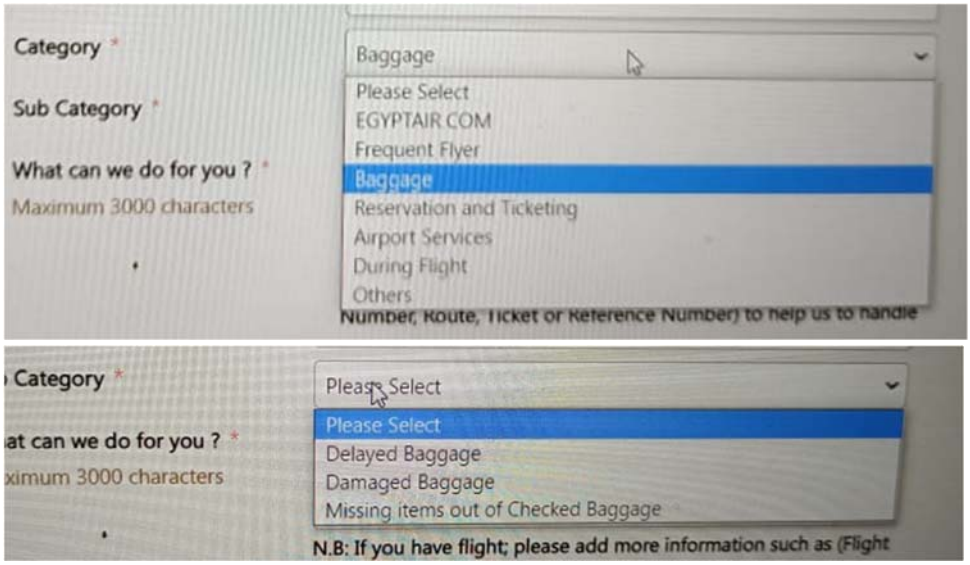


Figure1 :

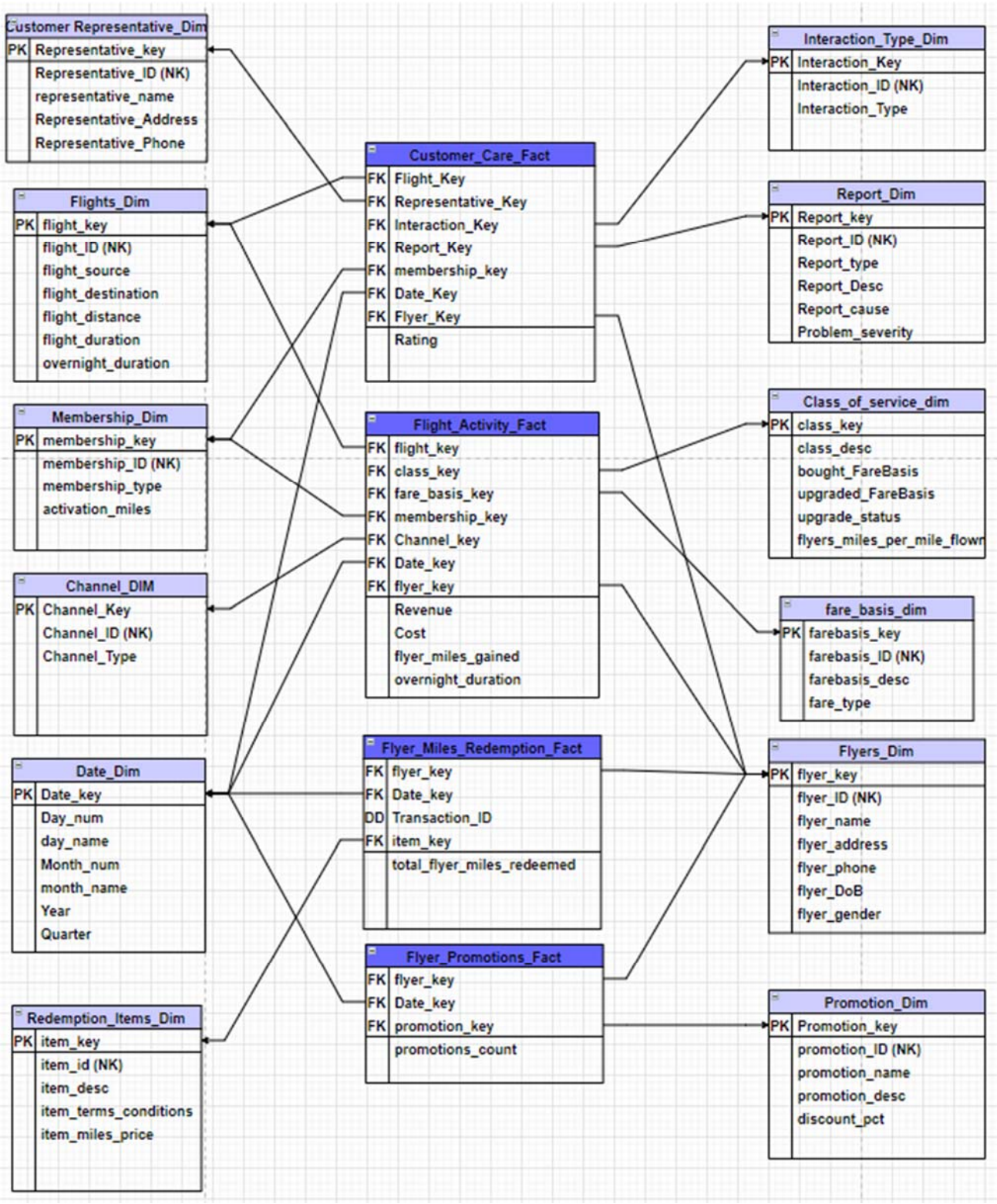
And also report dimension contain report cause attributes and report desc I take this idea also from EGYPTAIR website category contain types of report description and in sub category contain cause of it so I do this 2 attributes in my dimension as we see in figure 2



And we simulate that in our logical and physical design

2-4- Integrating data marts

After constructing the 3 independent data marts we integrate them all together to construct our DWH. We have a lot of common dimensions that can be shared



3-Indexes we used:

A Bitmap index is the index we used the most.it creates a matrix of all possible values of a field with every row of that field. In this matrix, a value of 1 or 0 is stored, the value is a 1 if the value of this database field = the value of the matrix column.

i.e.: Let's take the membership_dim as our example, possible values are only white, gold, platinum, titanium.

Let's say there are 4 flyers, Youssef, Salma, Jessica, Malk.

The Bitmap index will look like this:

white, gold, platinum, titanium.

Youssef : 1 | 0 | 0 | 0

Salma : 0 | 1 | 0 | 0

Jessica : 0 | 0 | 1 | 0

Malk : 0 | 0 | 0 | 1

That is a Bitmap index. When the index gets activated (i.e. you want all Platinum flyers), the database engine will only look into the Platinum column of the membership and return rows that = 1. This is much faster compared to a B-tree whereas you have to traverse the tree, find the node, compare all letters of the membership field to realize yes indeed Jessica is a platinum flyer.

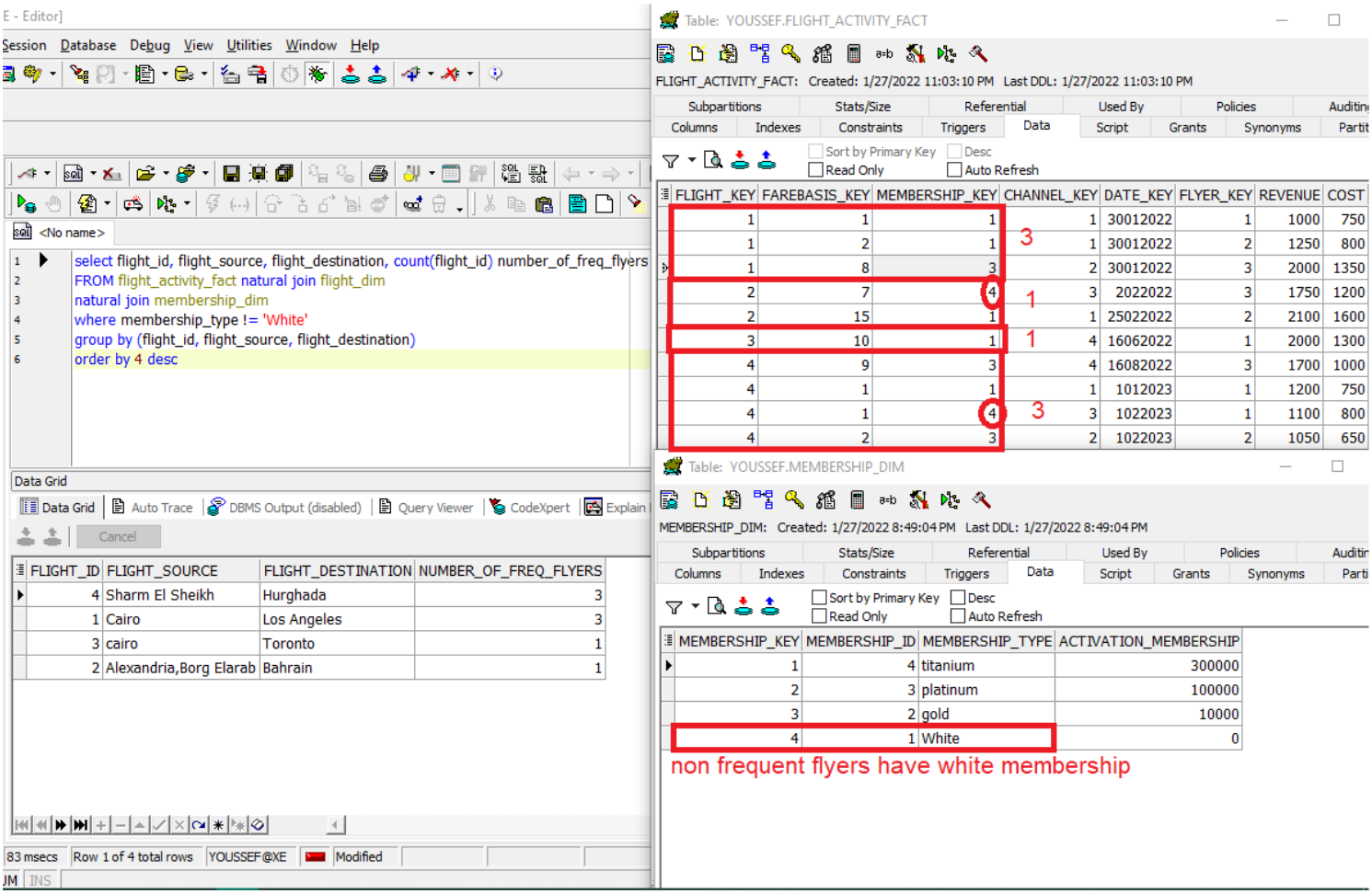
On a table with one million rows, a column with 10,000 distinct values is a candidate for a bitmap index. A bitmap index on this column can outperform a B-tree index, particularly when this column is often queried in conjunction with other indexed columns.

In ad hoc queries and similar situations, bitmap indexes can dramatically improve query performance. `AND` and `OR` conditions in the `WHERE` clause of a query can be resolved quickly by performing the corresponding Boolean operations directly on the bitmaps before converting the resulting bitmap to rowids. If the resulting number of rows is small, the query can be answered quickly without resorting to a full table scan.

4-SQL queries:

1-What flights the company's frequent flyers take?

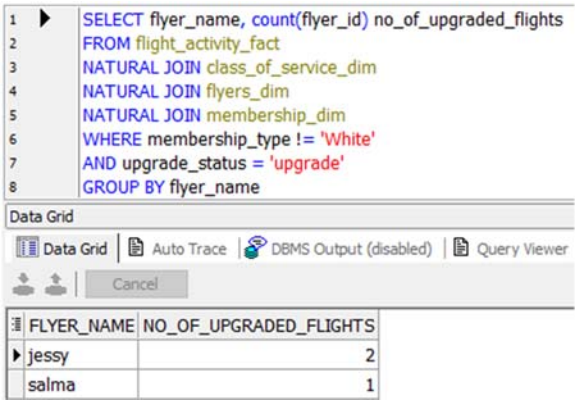
```
select flight_id, flight_source, flight_destination, count(flight_id) number_of_freq_flyers
FROM flight_activity_fact natural join flight_dim
natural join membership_dim
where membership_type != 'White'
group by (flight_id, flight_source, flight_destination)
order by 4 desc
```



2- How frequent do frequent flyers upgrade their fare basis?

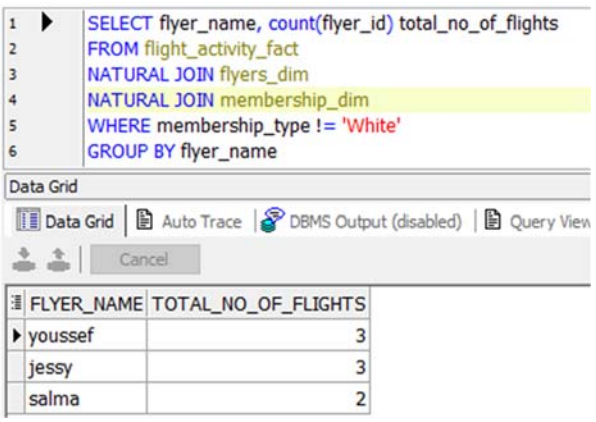
To solve this we first get the no of upgraded flights for each frequent flyer

```
SELECT flyer_name, count(flyer_id) no_of_upgraded_flights
FROM flight_activity_fact
NATURAL JOIN class_of_service_dim
NATURAL JOIN flyers_dim
NATURAL JOIN membership_dim
WHERE membership_type != 'White'
AND upgrade_status = 'upgrade'
GROUP BY flyer_name
```



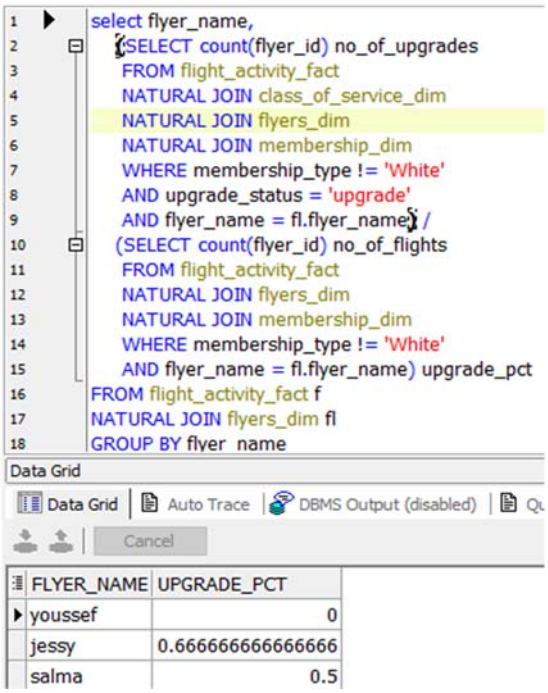
Then we get the total no of flights for each frequent flyer

```
SELECT flyer_name, count(flyer_id) total_no_of_flights
FROM flight_activity_fact
NATURAL JOIN flyers_dim
NATURAL JOIN membership_dim
WHERE membership_type != 'White'
GROUP BY flyer_name
```



Then we divide those by each other to get the frequency (percentage) of upgrades for each frequent flyer

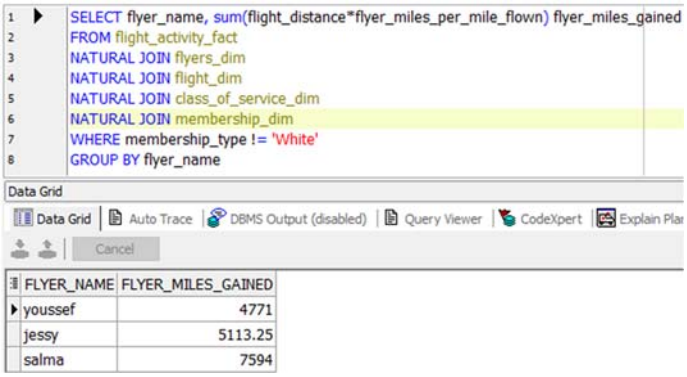
```
select flyer_name,
   (SELECT count(flyer_id) no_of_upgrades
    FROM flight_activity_fact
    NATURAL JOIN class_of_service_dim
    NATURAL JOIN flyers_dim
    NATURAL JOIN membership_dim
    WHERE membership_type != 'White'
    AND upgrade_status = 'upgrade'
    AND flyer_name = fl.flyer_name) /
   (SELECT count(flyer_id) no_of_flights
    FROM flight_activity_fact
    NATURAL JOIN flyers_dim
    NATURAL JOIN membership_dim
    WHERE membership_type != 'White'
    AND flyer_name = fl.flyer_name) upgrade_pct
FROM flight_activity_fact f
NATURAL JOIN flyers_dim fl
GROUP BY flyer_name
```



FLYER_NAME	UPGRADE_PCT
youssef	0
jessy	0.666666666666666
salma	0.5

3- How frequent flyers earn their frequent flyer miles?

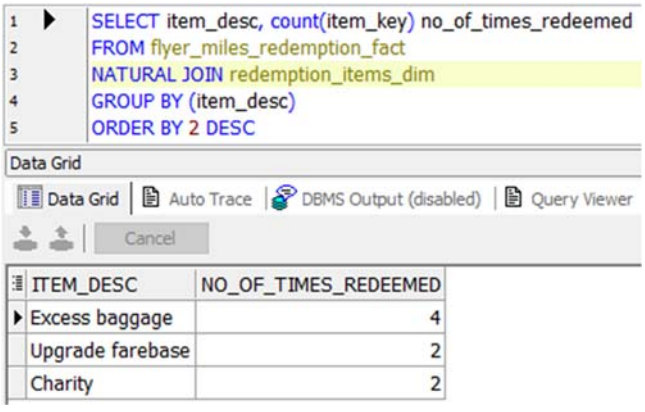
```
SELECT flyer_name, sum(flight_distance*flyer_miles_per_mile_flown) flyer_miles_gained
FROM flight_activity_fact
NATURAL JOIN flyers_dim
NATURAL JOIN flight_dim
NATURAL JOIN class_of_service_dim
NATURAL JOIN membership_dim
WHERE membership_type != 'White'
GROUP BY flyer_name
```



FLYER_NAME	FLYER_MILES_GAINED
youssef	4771
jessy	5113.25
salma	7594

4- How frequent flyers redeem their frequent flyer miles?

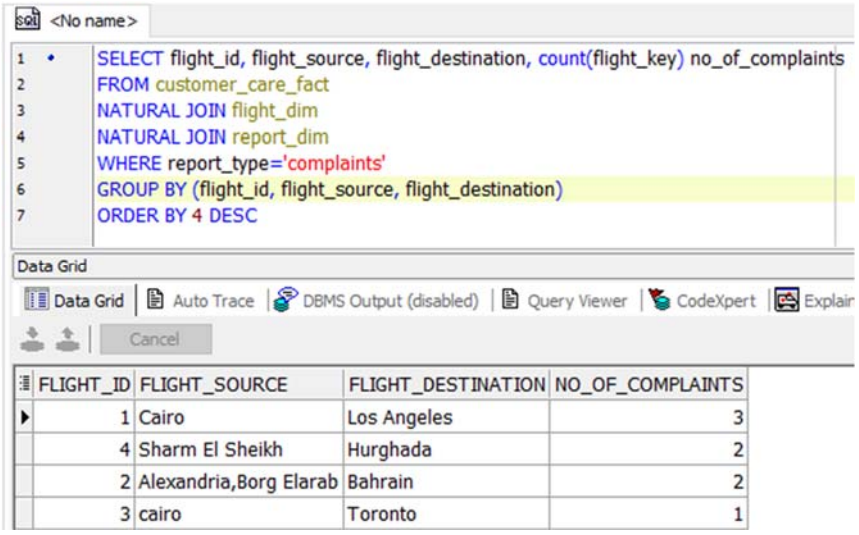
```
SELECT item_desc, count(item_key) no_of_times_redeemed
FROM flyer_miles_redemption_fact
NATURAL JOIN redemption_items_dim
GROUP BY (item_desc)
ORDER BY 2 DESC
```



ITEM_DESC	NO_OF_TIMES_REDEEMED
Excess baggage	4
Upgrade farebase	2
Charity	2

5- Which flights receive the most complaints?

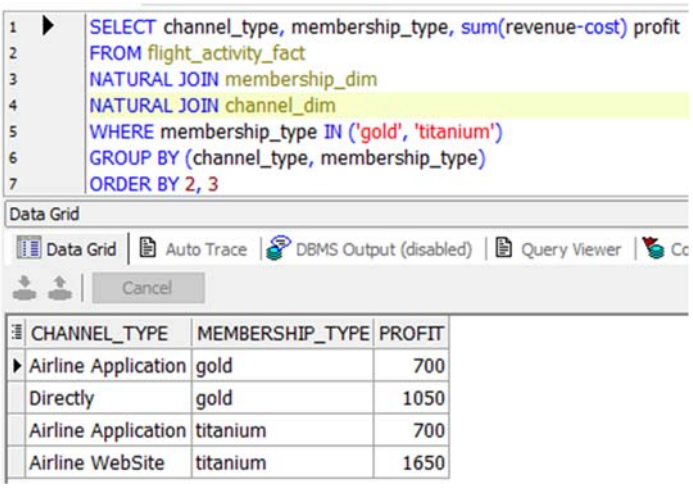
```
SELECT flight_id, flight_source, flight_destination, count(flight_key) no_of_complaints
FROM customer_care_fact
NATURAL JOIN flight_dim
NATURAL JOIN report_dim
WHERE report_type='complaints'
GROUP BY (flight_id, flight_source, flight_destination)
ORDER BY 4 DESC
```



FLIGHT_ID	FLIGHT_SOURCE	FLIGHT_DESTINATION	NO_OF_COMPLAINTS
1	Cairo	Los Angeles	3
4	Sharm El Sheikh	Hurghada	2
2	Alexandria,Borg Elarab	Bahrain	2
3	cairo	Toronto	1

6- What is the profit per channel for gold& titanium members?

```
SELECT channel_type, membership_type, sum(revenue-cost) profit
FROM flight_activity_fact
NATURAL JOIN membership_dim
NATURAL JOIN channel_dim
WHERE membership_type IN ('gold', 'titanium')
GROUP BY (channel_type, membership_type)
ORDER BY 2, 3
```



CHANNEL_TYPE	MEMBERSHIP_TYPE	PROFIT
Airline Application	gold	700
Directly	gold	1050
Airline Application	titanium	700
Airline WebSite	titanium	1650