DW Project Documentation

Team Members

Abdulrahman Hatem
Abdullah Galal
Najat Khalaf
Nada Nady

First of all, we will define The
Business Process

We are modeling a DW model for a major airline to assist the executive management to analyze their current business processes and expand the company by discovering new opportunities, and this will be approached through some points we take care of:

- ## Flight Activity

  ✓ In this process will analyze flight activity including data on routes, passenger numbers. The report would also include an analysis of trends and patterns in flight activity, as well as recommendations for optimizing the airline's operations. such as number of passenger, empty seats, number of crew , flight ID, aircraft ID, airport ID , captain id, flight attendant.

- ## Marketing Analysis

  ✓ Analyze what flights the company's frequent flyers take.
  ✓ What fare basis they pay, 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 and what proportion of these frequent flyers have gold, platinum or titanium status.

- ## Reservations

  ✓ Mainly focus on Profit for particular (Customers ID, Channel ID, Flight ID, FBC ID, Payment Method , Date ID)

- ## Customer Care

  This process focus on
  ✓ Handling customer inquiries, complaints and feedbacks
  ✓ Interaction type and problem severity

Grains

1. Flight Fact

Per flight (Aircraft ID, route, source airport, destination airport, departure time, arrival time, captain, and flight)

2. Marketing Analysis

Per each customer , from the perspective of the marketing team , they are focusing on customer "lowest level", is he upgrade , uses the promotions and discounts , how often he travels, at the end the marketing team can take action.

3. Reservation

Here we use it for each Ticket, which contain (customer ID, flight ID, aircraft ID , route ID , from a certain airport at a given departure time, to an arrival airport at a specific arrival time, fare class type, channel ID , and a payment method)

4. Customer Care

For the level of each complain we can track the action, feedback, problem severity also for specific customer, on a specific flight, channel, payment method, with a specific crew members, and exact date.

Facts & Measurements

1.Flights Fact

Using this fact, we can keep track of each flight and gather information that will enable us to identify which flights are most popular, also which (individual aircraft, route, source airport, destination airport, departure time, arrival time, captain, and flight attendant)

1- Measurements:

- Number of passengers.
- Number of empty seats.
- Number of the crew.

2- Dimensions:

- Crew Member
- Aircraft
- Airport
- Flight

Facts & Measurements

2. Marketing Analysis Fact

We can say that it is factless fact, no measurements have been calculated, we only use it to help Marketing team to analyze the frequent customers, are they using promotions, miles bonus?

1.Measurements:

- No measurements

2.Dimensions:

- Flight date
- Promotion
- Flyer miles
- Upgrade
- Customer
- Flight

Facts & Measurements

3.Reservation Fact

Using this fact, we can extract which customers brings the most value, which date "season"/flight/airplane/airport makes the best profit, which reservation channel and which payment method delivers the most money.

1.Measurements:

- Price.
- Discounts.
- Overnight stand.

2.Dimensions:

- Customer
- Payment method
- Channel
- Flight date
- Fare class
- Flight

Facts & Measurements

4.Customer Care Fact

Using this fact, we keep up with information about customer satisfaction and suggests solutions to enhance business success by attending to customers' requirements and categorize the problems.

1.Measurements:

- Severity type.

2.Dimensions:

- Customer
- Flight
- Payment method
- Channel
- Flight date
- Action
- Crew members

Dimensions

1- Crew Member

- Data about all the employees in the company.

2- Customer

- Data about the passengers.

3- Date

- Calendar dimension for any DW Model.

4- Channel

- The channels are the source for tickets' reservation.

5- Payment Method

- The different methods of payment.

6- Fare Class

- The tickets classes (Economy, Business, etc).

7- Flyer Miles

- Holding miles' number of the flight.

Dimensions

8- Promotion
- Data about the discounts that the company offer.

9- Upgrade
- Data about types of upgrades the company offer.

10- Aircraft
- Data about the airplanes the company owns.

11- Airport
- Data about the airports that the aircrafts use.

12- Flight
- Data about the specific times for each flight.

13- Route
- Data about the details for each flight.

Schema

Aircraft
aircraft_id(pk)
model
type
manfucture
capacity

Airport
Airport_id(PK)
Name
City
Country

Flight_Activity FACT
ID(PK)
flight_id(FK)
aircraft_id(FK)
SRCairport_id(FK)
dest_airport_id(FK)
captain_id(FK)
flight_attend_id(FK)
no_passenger
empty_seats
count_crew

Crew_member
Member_id (PK)
FName
LName
Gender
Job_Type
Hire_Date
DOB
Country
Email
Phone

Route
route_id(PK)
source_city
dest_city
source_country
dest_country
distance

Flight
Flight_id(PK)
Arrival_Time
Depature_Time

Customer_service FACT
CSer_ID(PK)
Action_id(FK)
Customer_Id(FK)
Date_Id(FK)
flight_id(FK)
PM_ID(FK)
channel_id(FK)
Member_id(FK)
problem_severity

Action
Action_id(PK)
Name
Type
description

Channel
channel_id(PK)
Name
Description
Status

Flight_Date
Date_Id(PK)
Month
Day
Year
Minute
Hour
Week

Payment_Mehtod
Payment_id(PK)
Name
Flight_id(FK)

Reservation FACT
Resrv_Id(PK)
Customer_Id(FK)
PM_ID(FK)
channel_id(FK)
flight_id(FK)
Fare_class_id(FK)
Date_Id(FK)
seat_number
Base_price
Discount
Price
overnight_stand

fare_class
Fare_class_id(PK)
Name
description

Customer
Customer_Id(PK)
FName
Lname
Geder
City
Country
DoB
Email
phone
Status

Marketing_analysis FACT
UP_ID(PK)
Date_Id(FK)
Customer_Id(FK)
Promotion_id(FK)
FM_ID earned(FK)
FM_ID redeemed(FK)
flight_id(FK)
MA_Id(FK)

promotion
Promotion_id(PK)
Name
Start_date
END_Date

Flyer Miles
FM_ID(PK)
Miles

Upgrade
UP_ID(PK)
Name
Start_Date
End_Date

- ## Why we used Star schema?

We find that the star schema is suitable for our case study and perform better in implementation, after searching also we find that star schema is better for many reasons like:

- Streamlined query execution: Star schema modeling enables us to streamline and improve query execution. Queries may be run effectively and rapidly without the use of intricate joins or subqueries since the fact table serves as the foundation of the schema and is linked to the dimension tables by foreign keys.

- Star schema modelling is very scalable and has a great data handling capacity. The schema can handle massive volumes of data without affecting query performance or data analysis by dividing the data into smaller, more manageable tables.

- Better data analysis: Star schema modelling offers a clear and simple method for analyzing data from several aspects. By traversing through the dimension tables, analysts can quickly drill down into the data, enabling more intricate analysis and insights.

Indexes

For Indexes, We used

1. Cluster index on each PK columns " By default on Toad".
2. B-tree Index on FK " idx_flight_activity_flight_id ON hr.flight_activity(flight_id);"
3. Unique Index on E-mail column.

Types of Indexes in Oracle
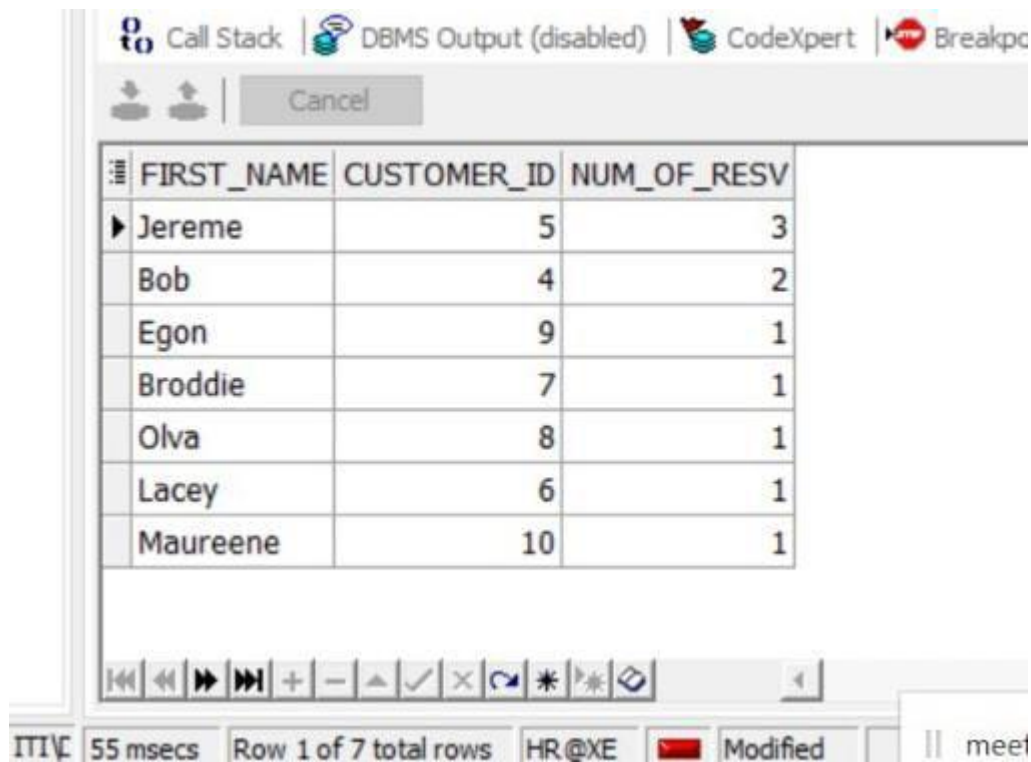
Oracle defines two types of indexes: the B-Tree (Balanced Tree) Index and the Bitmap Index.

B-tree	Bitmap
• Index-organized tables An index-organized table differs from a heap-organized because the data is itself the index.	• Function-based indexes This type of index includes columns that are either transformed by a function, such as the UPPER function, or included in an expression. B-tree or bitmap indexes can be function-based.
• Reverse key indexes In this type of index, the bytes of the index key are reversed, for example, 103 is stored as 301. The reversal of bytes spreads out inserts into the index over many blocks.	• Application domain indexes A user creates this type of index for data in an application-specific domain. The physical index need not use a traditional index structure and can be stored either in the Oracle database as tables or externally as a file.
• UNIQUE Descending indexes This type of index stores data on a particular column or columns in descending order. • B-tree cluster indexes	

Sample of SQL queries

1-Which customers use our services more frequent ?

```sql
select * from(
select   first_name ,r.customer_id, count(r.customer_id) num_of_resv
from hr.reservation r,hr.customer c
where r.customer_id=c.customer_id
group by first_name,r.customer_id
  )order by num_of_resv desc;
```
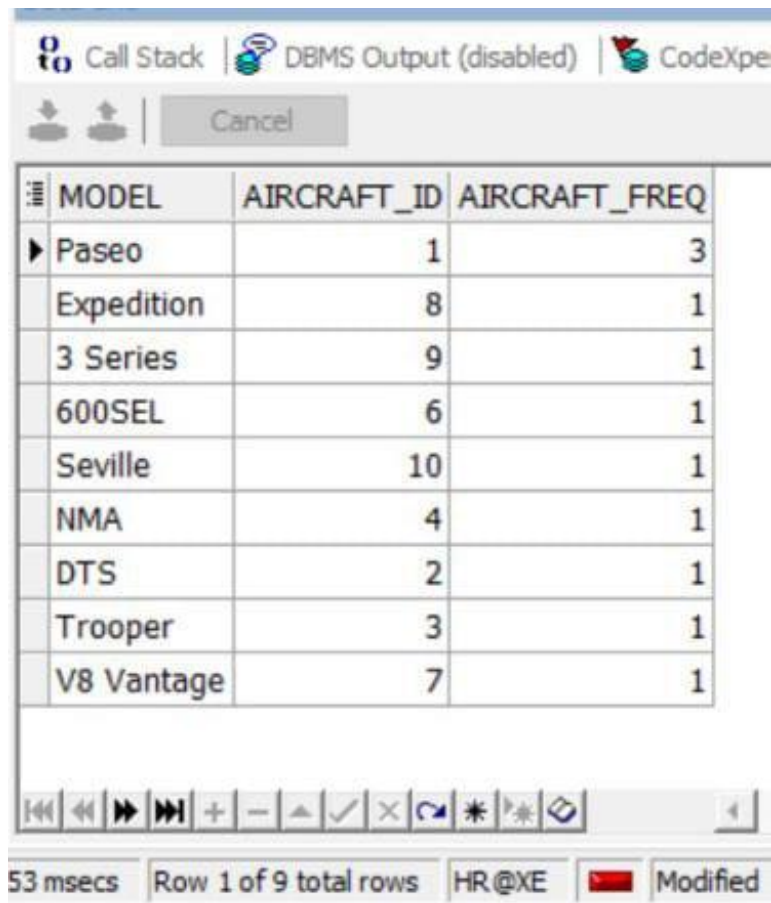
FIRST_NAME	CUSTOMER_ID	NUM_OF_RESV
Jereme	5	3
Bob	4	2
Egon	9	1
Broddie	7	1
Olva	8	1
Lacey	6	1
Maureene	10	1

Row 1 of 7 total rows   HR@XE   Modified

2-What are the most popular routes/aircrafts ?

```
select * from (
select model,air.aircraft_id ,count(air.aircraft_id) aircraft_freq
from hr.flight_activity f,hr.aircraft air
where air.aircraft_id = f.aircraft_id
group by air.aircraft_id,air.model
)
order by aircraft_freq desc;
```

MODEL	AIRCRAFT_ID	AIRCRAFT_FREQ
Paseo	1	3
Expedition	8	1
3 Series	9	1
600SEL	6	1
Seville	10	1
NMA	4	1
DTS	2	1
Trooper	3	1
V8 Vantage	7	1

53 msecs   Row 1 of 9 total rows   HR @XE   Modified

3-What are the most popular destcountry?

```
select * from (
select r.DEST_COUNTRY  ,r.route_id ,count(f.route_id) as freq_country
from hr.flight f,hr.route r
where f.route_id=r.route_id
group by r.route_id   ,r.DEST_COUNTRY
)
order by freq_country desc;
```

DEST_COUNTRY	ROUTE_ID	FREQ_COUNTRY
Philippines	7	4
Czech Republic	5	1
Brazil	9	1
Indonesia	10	1
France	3	1
Philippines	6	1
China	4	1
Russia	1	1
China	2	1

ITI\[44 msecs   Row 1 of 9 total rows   HR@XE   Modified

4-Which booking channel is most rewarding?

```sql
select ch.name,ch.channel_id ,count(r.channel_id) freq_channel
from hr.reservation r ,hr.channel ch
where r.channel_id=ch.channel_id
group by ch.CHANNEL_ID   ,ch.name
order by freq_channel desc;
```
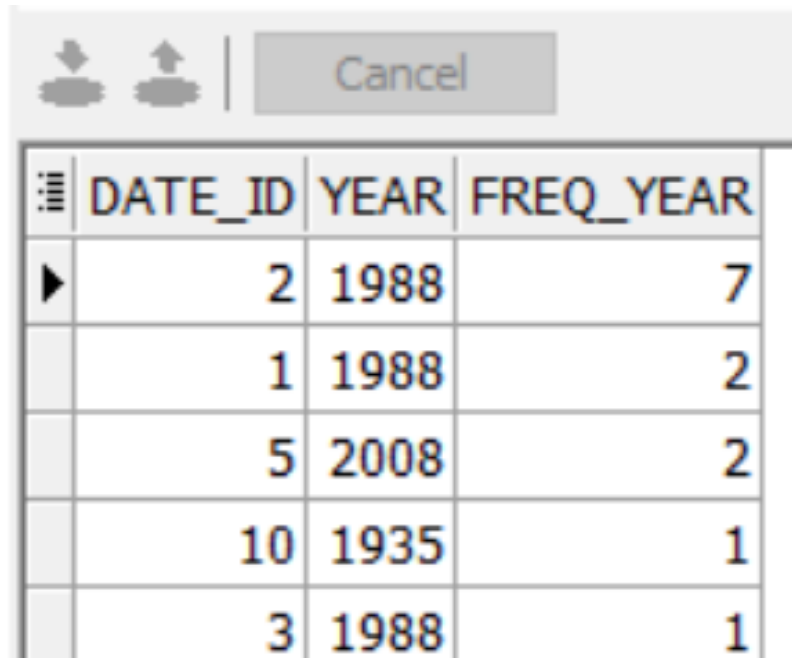
NAME	CHANNEL_ID	FREQ_CHANNEL
mobile app	7	7
text	1	4
online	9	1

5-Which year have more reservations?

```
select fd.date_id,year,count(fd.date_id) freq_year
from hr.flight_date fd,  hr.reservation r
where r.date_id=fd.date_id
group by fd.date_id ,fd.year
order by freq_year desc;
```
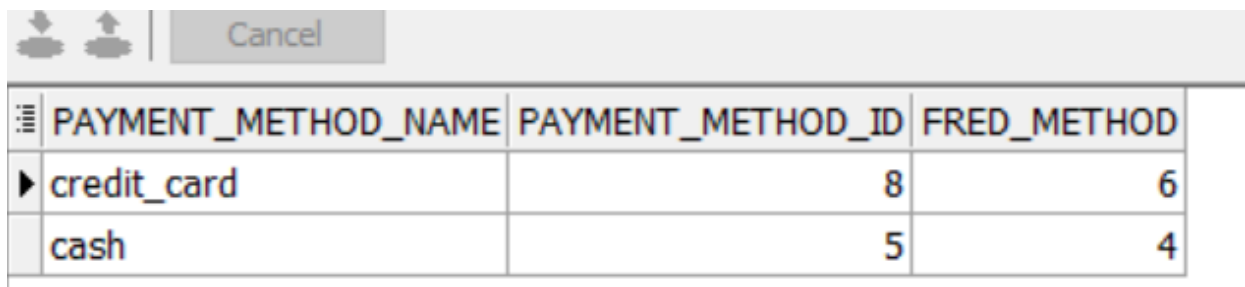
DATE_ID	YEAR	FREQ_YEAR
2	1988	7
1	1988	2
5	2008	2
10	1935	1
3	1988	1

6-Which payment method most popular ?
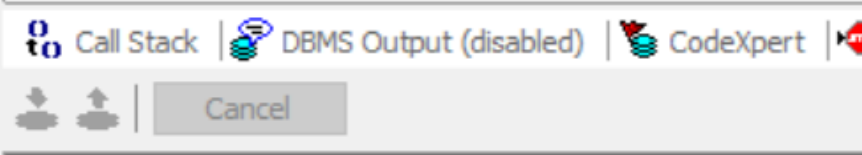
```sql
select pm. PAYMENT_METHOD_NAME,pm.PAYMENT_METHOD_id,count(r.PM_ID) fred_method
from hr.reservation r,hr.payment_method pm
where r.pm_id=pm.PAYMENT_METHOD_id
group by pm.PAYMENT_METHOD_id,pm. PAYMENT_METHOD_NAME
order by fred_method desc;
```

PAYMENT_METHOD_NAME	PAYMENT_METHOD_ID	FRED_METHOD
credit_card	8	6
cash	5	4

7-How much miles they earn ?

```
select first_name, m.customer_id,sum(miles) earn_miles
from hr.marketing_analysis M ,hr.Flyer_Miles F, hr.customer c
where M.Flyer_miles_earned_id = f.Flyer_miles_id and
m.customer_id =c.customer_id
group by m.customer_id,first_name
order by earn_miles desc;
```
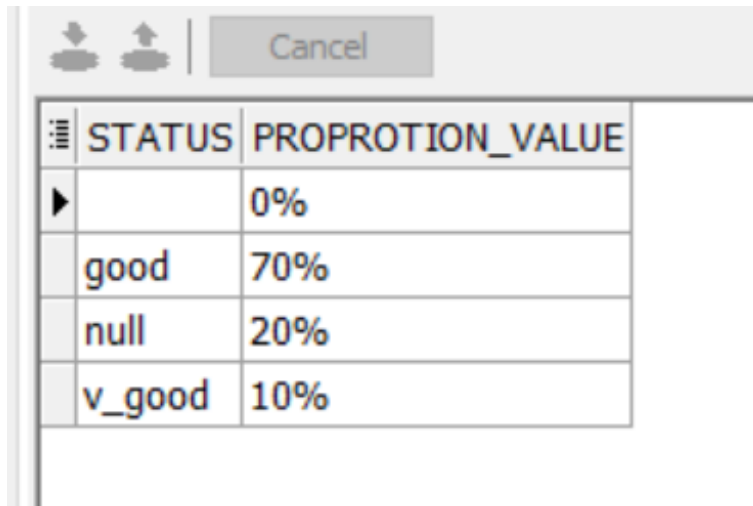
FIRST_NAME	CUSTOMER_ID	EARN_MILES
Carlita	3	3164
Bob	4	3156
Maureene	10	3156
Olva	8	3156
Lacey	6	3156
	2	3156
Broddie	7	2475

8-Proprotion of membership status passengers

```
select status, count(status) /(select count(status) from hr.customer )*100||'%' proprotion_value
from hr.customer
group by status;
```
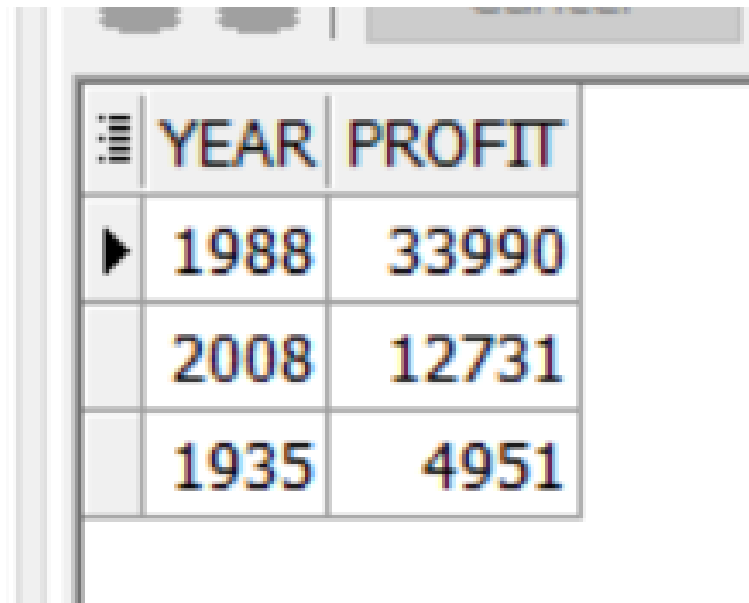
	STATUS	PROPROTION_VALUE
▶		0%
	good	70%
	null	20%
	v_good	10%

9-Profit per year

```
select year, sum(price) as profit
from hr.reservation r ,hr.flight_date  d
where r.date_id =d.date_id
group by year
order by profit desc;
```
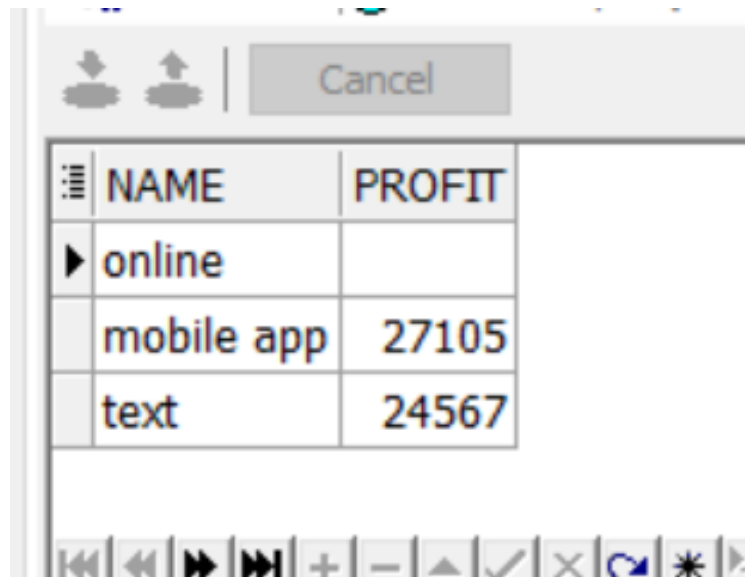
YEAR	PROFIT
1988	33990
2008	12731
1935	4951

10-Profit from every reservation channel

```
select name,sum(price) as profit
from hr.reservation r ,hr.channel ch
where r.channel_id=ch.channel_id
group by name
order by profit desc;
```

NAME	PROFIT
online	
mobile app	27105
text	24567