ITI DATA MANAGEMENT TRACK

INTAKE 44

AIRLINE _DWH CASE STUDY

Presented by :
Abdelrahman Ahmed

Ahmed Samy

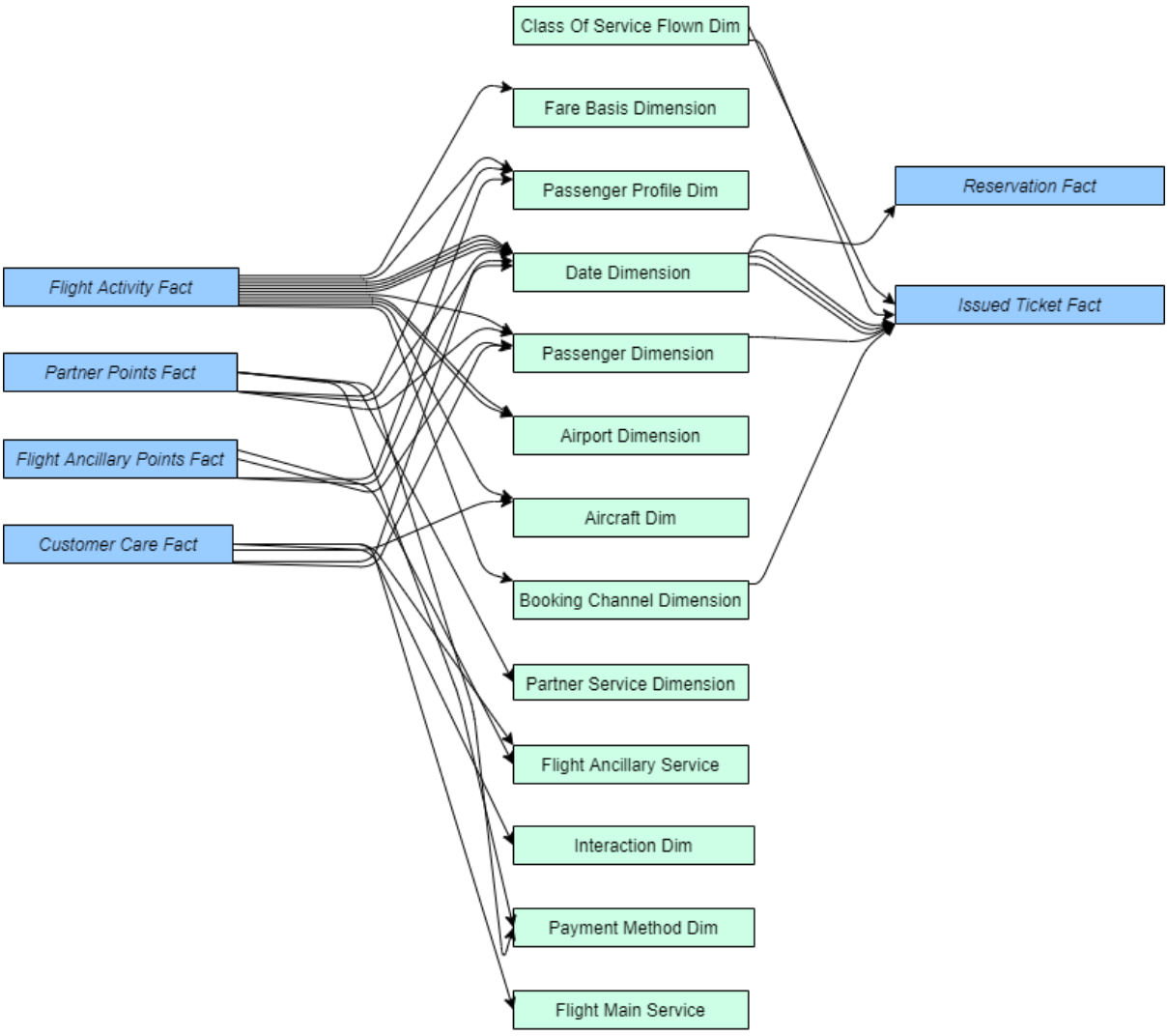
Manar Ayman

Rahaf Mohamad

Contents

INTRODUCTION

In an era of rapid transformation, the airline industry faces a pressing need for agile strategies and data-driven solutions. Our project aims to provide actionable insights to a major airline company, leveraging rigorous analysis of flight activity, reservation processes, and customer interactions. By aligning our analysis with the company's objectives, we aim to unlock opportunities for enhanced operational efficiency and customer satisfaction. Through a meticulous review of passenger behaviour, revenue streams, and service utilization, we seek to empower executive management with strategic recommendations for sustainable growth. Our approach integrates advanced analytics with a deep understanding of industry dynamics, ensuring relevance and impact in our findings. As we present our insights, we invite the company to embrace data-driven decision-making as a cornerstone of its strategic vision. Together, we can navigate the complexities of the aviation landscape and chart a course towards continued success and innovation.



STEPS OF DESIGN

We built our model based on Kimball model, Taking "The Data Warehouse Toolkit - The Definitive Guide to Dimensional Modelling " Chapter-12 specifically our reference for this case study.

INVESTIGATING THE BUS MATRIX

-Measures &Grain(Ticket) to be investigated
- Is it (accumulated or transaction without numeric measures.)
-attributes ask CHTGPT

To be investigated

	Date	Time	Airport	Passenger	Booking Channel	Class of Service	Fare Basis	Aircraft	Communication Profile	Transaction ID #
Reservations	X	X	X	X	X	X	X	X		Conf #
Issued Tickets	X	X	X	X	X	X	X	X		Conf # Ticket #
Unearned Revenue & Availability	X	X	X			X		X		
Flight Activity	X	X	X	X	X	X	X	X		Conf # Ticket #
Frequent Flyer Account Credits	X		X	X	X	X	X			Conf # Ticket #
Customer Care Interactions	X	X	X	X					X	Case # Ticket #
Frequent Flyer Communications	X	X	X	X					X	
Maintenance Work Orders	X	X	X					X		Work Order #
Crew Scheduling	X	X	X			X		X		

1: Subset of bus matrix row for an airline.

Measures &Grain to be investigated

Investigate
-Track redemption and earning of points
- Reason of transaction instead of separate dimension
-mapping between miles and points

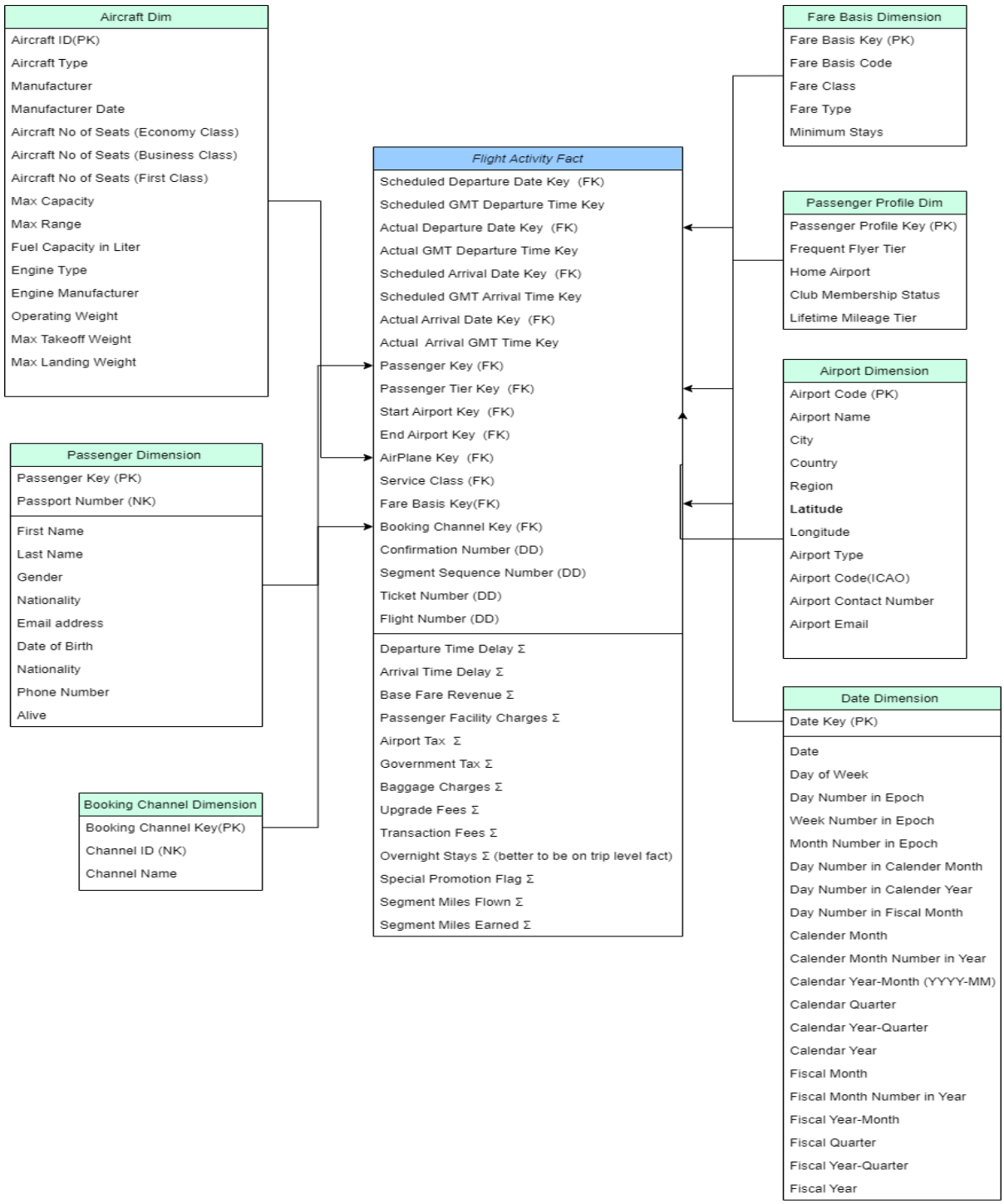
We started our work by investigating the bus matrix stated in this chapter to decide which business processes will fit our project requirements and what is the dimensions that intersect with it and here are our findings:

DEFINING BUSINESS PROCESSES

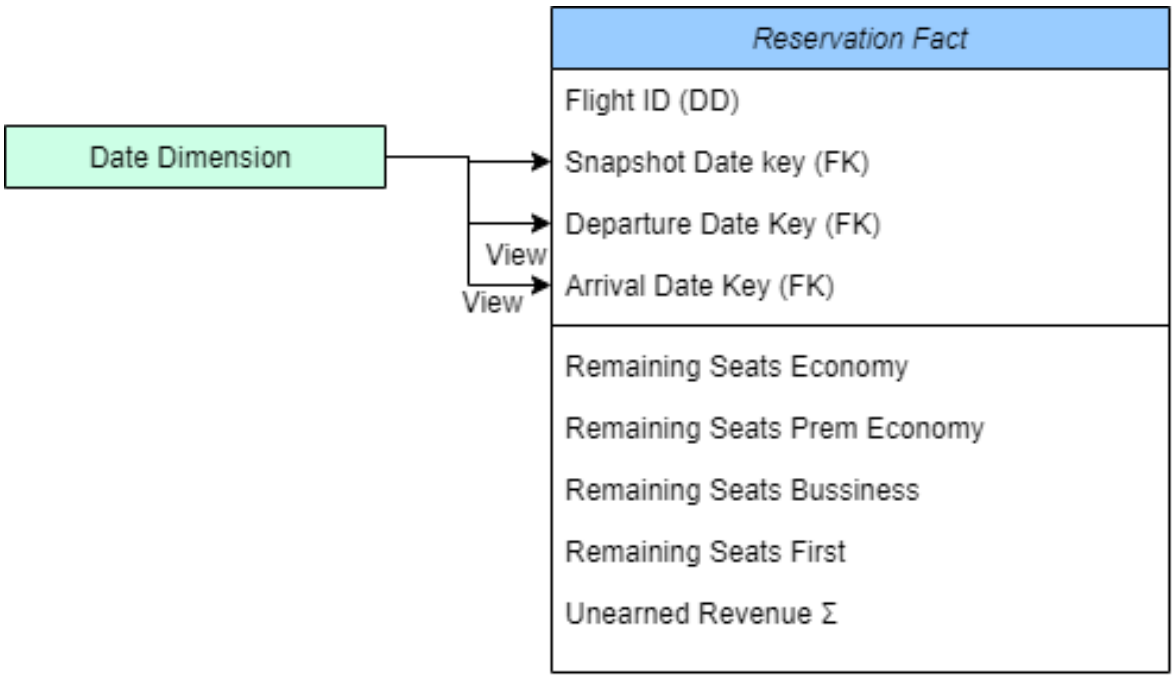
1- Flight Activity.
2- Reservations.
3- Points of flights' ancillary services (spending/earning) tracking for frequent flyers.
4- Points of Partners' services (spending/earning) tracking for frequent flyers.
5- Customer Care Interactions.
6- Ticket issuance (Bonus) .

DEFINING FACTS AND GRAIN

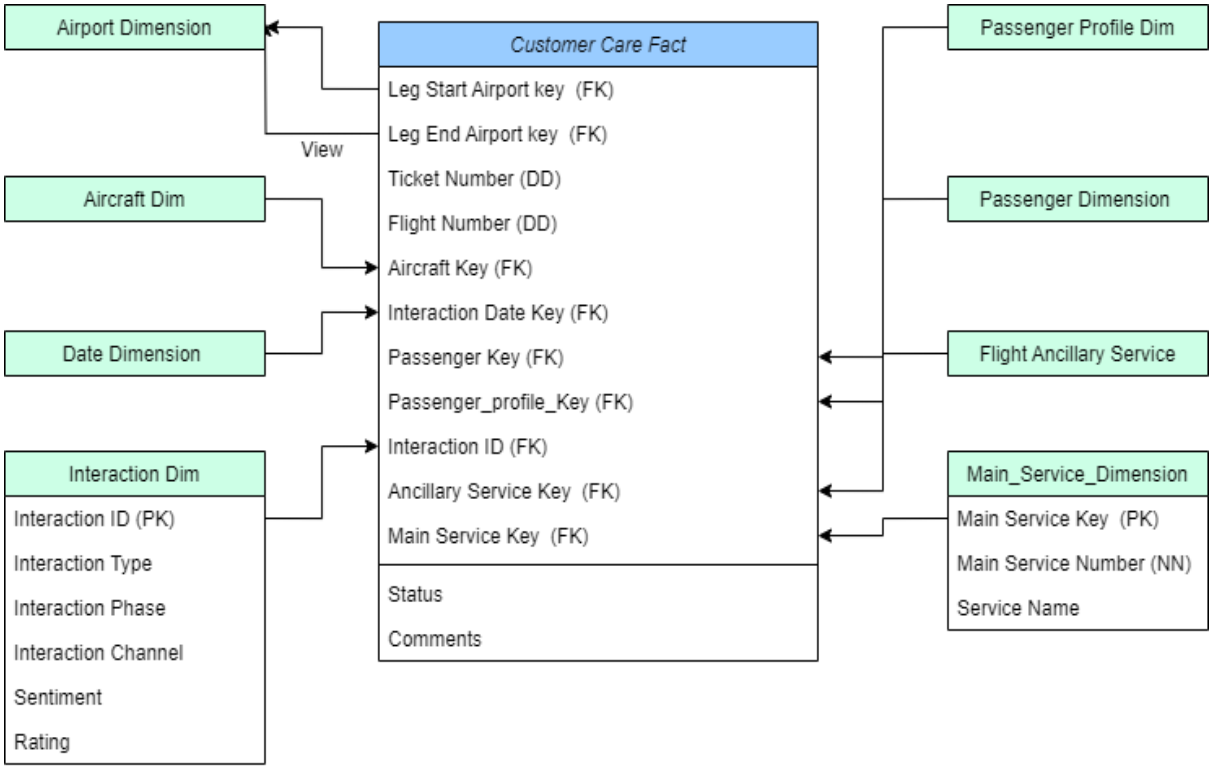
Flight Activity Fact: On **segment** level to provide a better usability for both marketing and BI teams, however. *it could be chosen on trip level but on average trip consists of 3 segments so 3X performance enhancements is considered small for the Kimball to lose some needed details*.



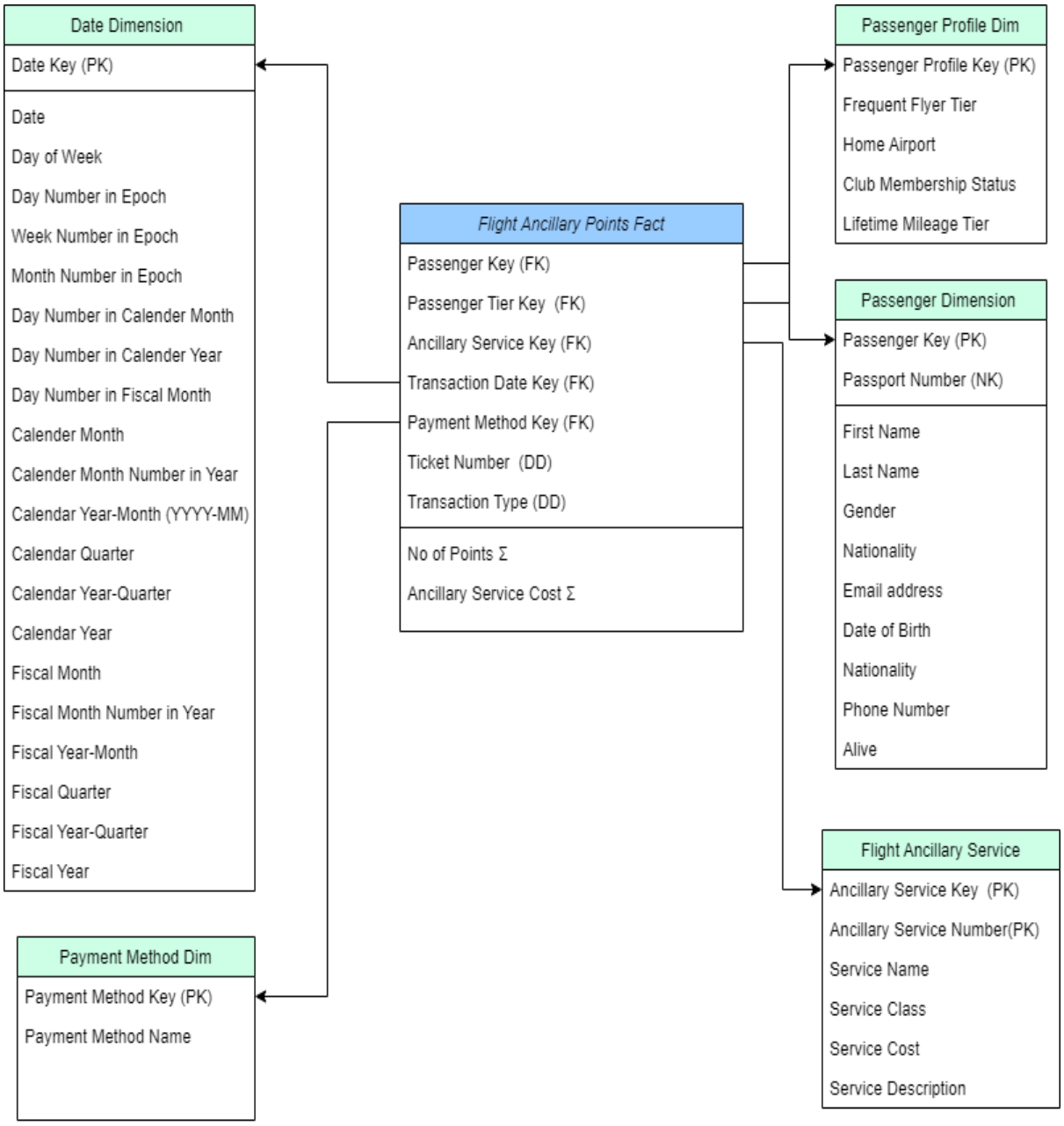
Reservations Fact: It's a daily snapshot fact table, on **Flight** level as the purpose of this fact to allow Business users to track the unearned revenues for every day.



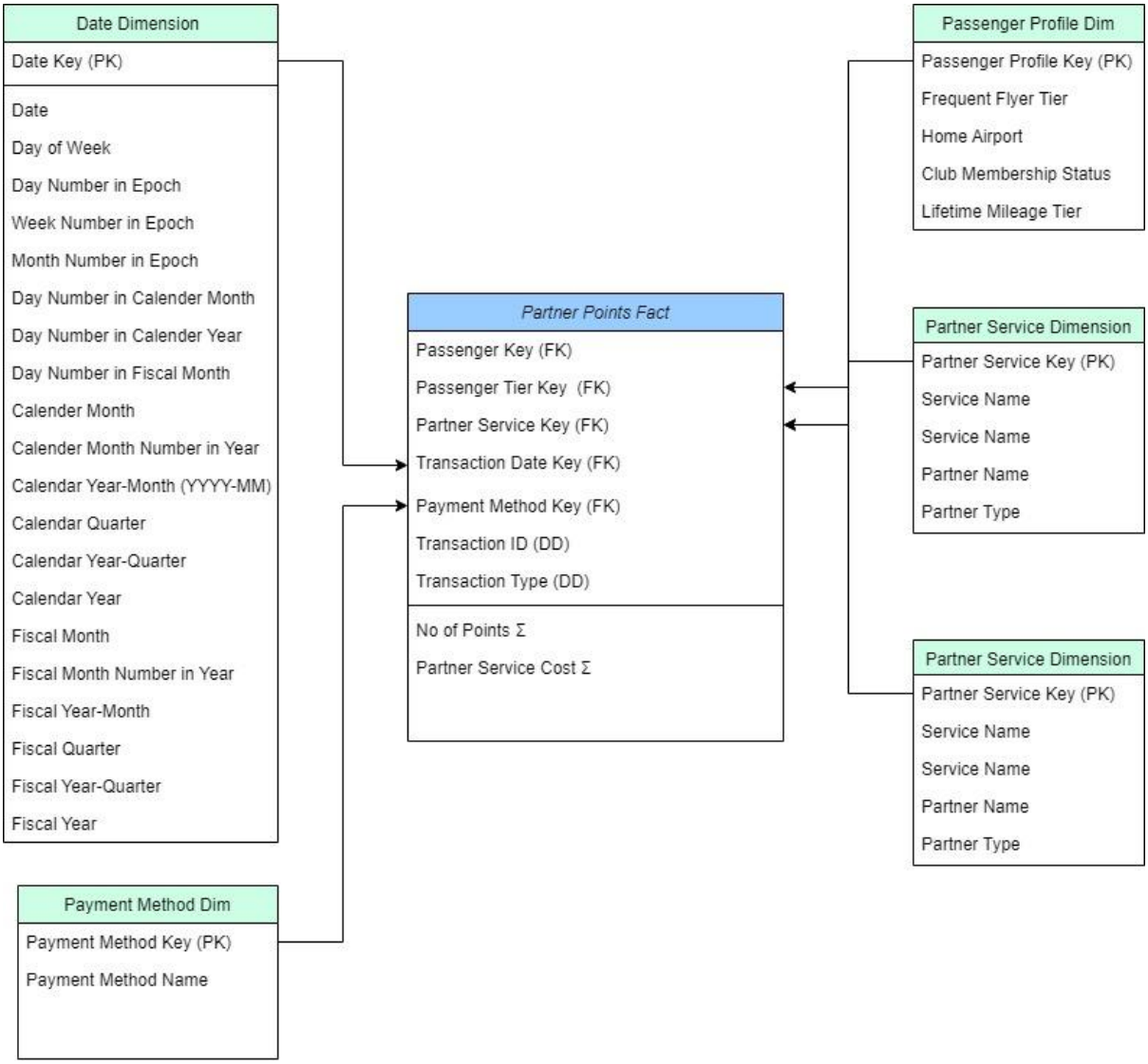
Customer Care Interactions: Transactional table on Leg level to allow tracing interaction on the lowest level of grain like (Complains, Feedbacks, Inquires).



Flight Ancillary Points Fact: On level of ticket level to trace points spending and earing through the flights of the flight ancillary services like (Extra baggage, accessing lounge, extra meal, etc.)



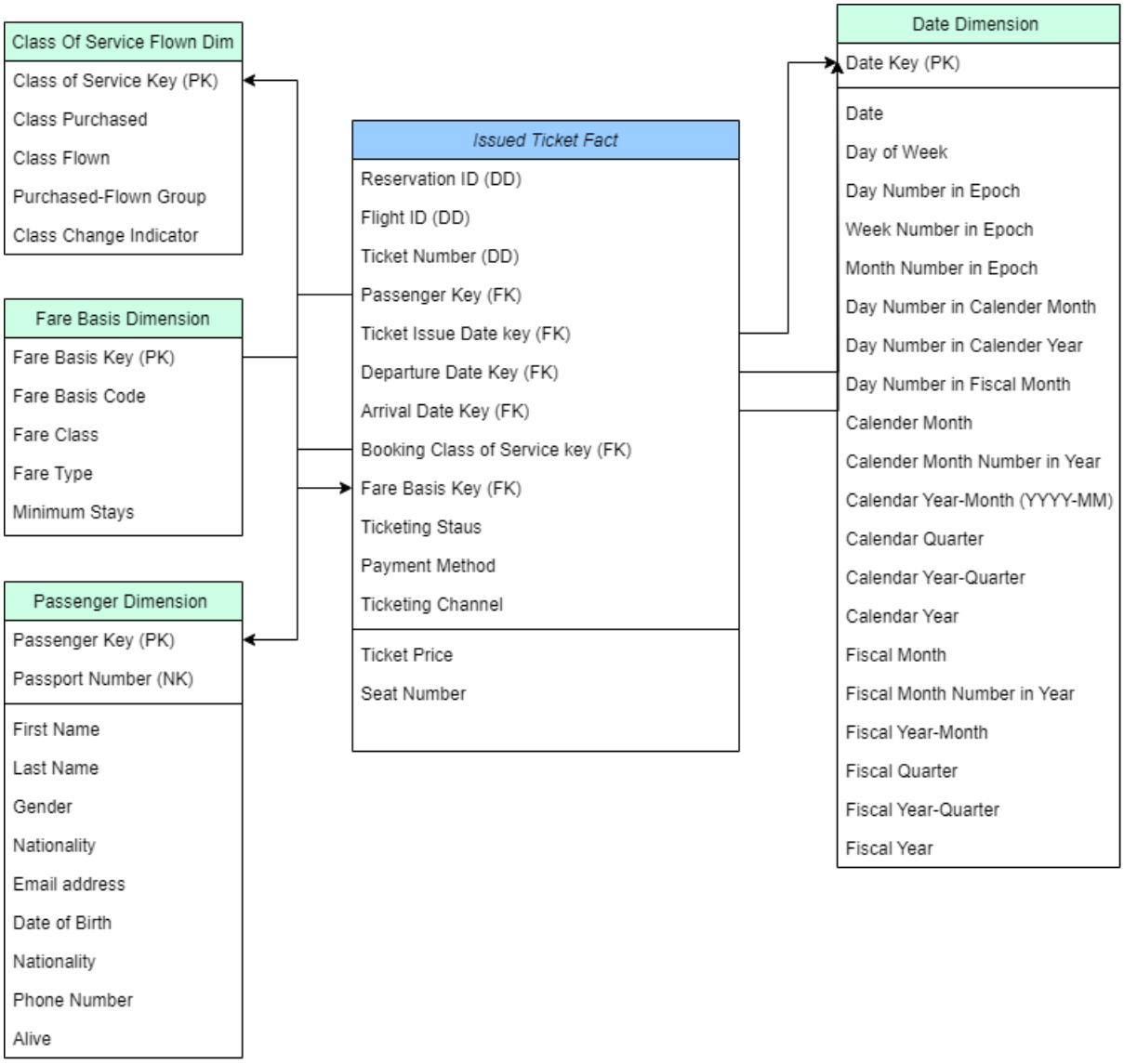
Partner Points Fact: On transaction_id level, and *this is separated from the previous fact to allow tracing the services on the flight independent of partners' services like (Hotels, Transportations, Retail, etc…)*



Issued Ticket Fact Table:

Transactional table capturing individual ticket issuance events at the granularity of each ticket issued, providing insights into passenger ticketing activities, such as reservation IDs, flight details, ticket issuance dates, and payment methods.

It enables analysis of ticket issuance trends, passenger booking behaviors, popular flight routes, and revenue generation patterns. This data can be used to optimize ticketing processes, understand passenger preferences, and enhance revenue management strategies.



DIMENSIONS

Date Dimension: Stores various date-related attributes like day, month, year, and quarter. Used for time-based analysis and reporting, facilitating queries related to passenger activities, flight schedules, and revenue trends over time, Note that its recommended to implement the date by creating views to reference different type of date keys like (departure and arrival) .

Interaction Dimension: Captures details about different types of interactions with passengers, including inquiries, complaints, and feedback. Enables analysis of customer sentiment, satisfaction levels, and service quality across various communication channels.

Ancillary Service Dimension: Contains information about additional services offered to passengers, such as meals, Wi-Fi, and extra luggage. Facilitates analysis of ancillary service utilization, revenue generation, and passenger preferences.

Main Service Dimension: Represents primary services provided by the airline, such as booking, ticketing, and baggage handling. Supports analysis of service usage, performance, and customer satisfaction.

Passenger Dimension: Stores demographic and contact information about passengers, including names, genders, and contact details. Facilitates customer segmentation, loyalty program management, and personalized marketing initiatives.

Passenger Profile Dimension: There are no specific dimension for frequent flyers , all passengers considered a frequent flyers but the one who has a memberships in the club the airline company will track his point spending and earning , Actually this dimension is the one that specifying the frequent flyer tier of the passenger when he made the flight, Captures details about passengers' frequent flyer tiers, club memberships, and lifetime mileage tiers. Enables segmentation of passengers based on loyalty status, travel patterns, and mileage accrual.

Fare Basis Dimension: Contains information about fare types and restrictions associated with ticket pricing. Supports analysis of fare structures, pricing strategies, and revenue management.

Class of Service Flown Dimension: Describes the class of service purchased and flown by passengers, such as economy, business, or first class. Facilitates analysis of seat occupancy, revenue distribution, and service differentiation, Note that this dimensions track the upgrades and downgrades by providing all possible combination between all the classes purchased a flown with if its upgraded or downgraded.

Payment Method Dimension: Stores details about payment methods used for transactions, including credit cards, debit cards, and bank transfers. Supports analysis of payment trends, transaction volumes, and revenue channels.

Partner Service Dimension: Contains information about services offered by partner organizations, such as hotels, transportation companies, and retail outlets, and the partners themselves. Enables analysis of partnership effectiveness, service utilization, and revenue sharing.

Booking Channel Dimension: Represents different channels through which bookings are made, including online platforms, call centres, and travel agencies. Facilitates analysis of booking channel performance, customer preferences, and distribution costs.

Aircraft Dimension: Stores details about aircraft, including type, manufacturer, seating capacity, and technical specifications. Supports analysis of fleet management, operational efficiency, and maintenance scheduling.

Airport Dimension: Contains information about airports, including codes, names, locations, and contact details. Facilitates analysis of flight routes, passenger traffic, and airport performance metrics, Here the same idea of views as date dimension to be implemented for (start and end airports).

DEFINING MEASURES

Flight Activity Fact Analysis:

Required:

- Flights taken by frequent flyers.
- Fare basis paid.
- Frequency of upgrades
- Earning and redemption of frequent flyer miles
- Response to special fare promotions.
- Duration of overnight stays.

Additional:

- Revenue breakdown by fare type
- Analysis of flight delays by severity
- Distribution of flights by aircraft type
- Trend analysis of passenger facility charges
- Seasonal variation in flight demand

Reservation Fact Analysis:

Required:

- Analysis of company profit through reservation process
- Distribution of reservations by channel

Additional:

- Trend analysis of booking cancellations
- Seasonal variation in booking patterns
- Analysis of booking changes and their impact on profit

Customer Care Fact Analysis:

Required:

- Handling of customer inquiries, complaints, and feedback
- Analysis of interaction type and problem severity

Additional:

- Trend analysis of customer care response times
- Distribution of interaction types by channel
- Analysis of customer satisfaction scores

Flight Ancillary Points Fact Analysis:

Required:

- Earning and redemption of frequent flyer miles
- Analysis of ancillary service usage

Additional:

- Revenue breakdown by ancillary service
- Trend analysis of ancillary service redemption
- Comparison of ancillary service usage by passenger profile

Partner Points Fact Analysis:

Required:

- Earning and redemption of partner points
- Analysis of partner service usage

Additional:

- Revenue breakdown by partner service
- Trend analysis of partner service redemption
- Comparison of partner service usage by passenger profile

Issued Ticket Fact Analysis:

Required:

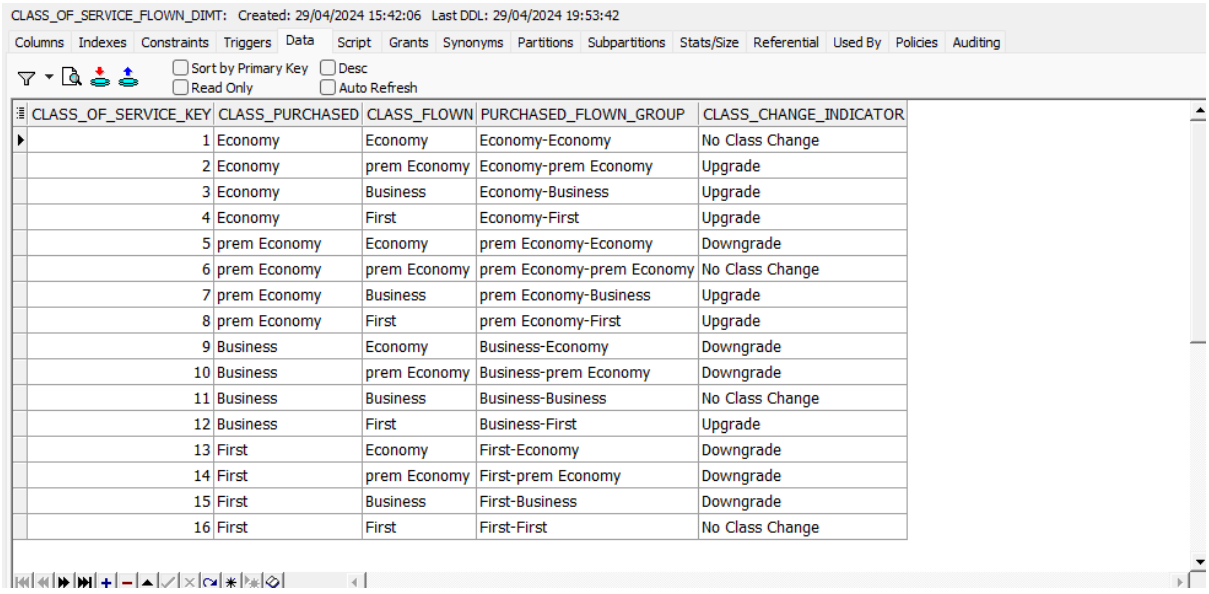
- Insight into ticket issuance trends and patterns
- Analysis of passenger booking behaviours.
- Understanding revenue generation from ticket sales

Additional:

- Identification of popular flight routes and times
- Optimization of ticketing processes based on booking channel performance

SAMPLE FROM DATA POPULATION

CLASS OF SERVICE FLOWN DIMENSION



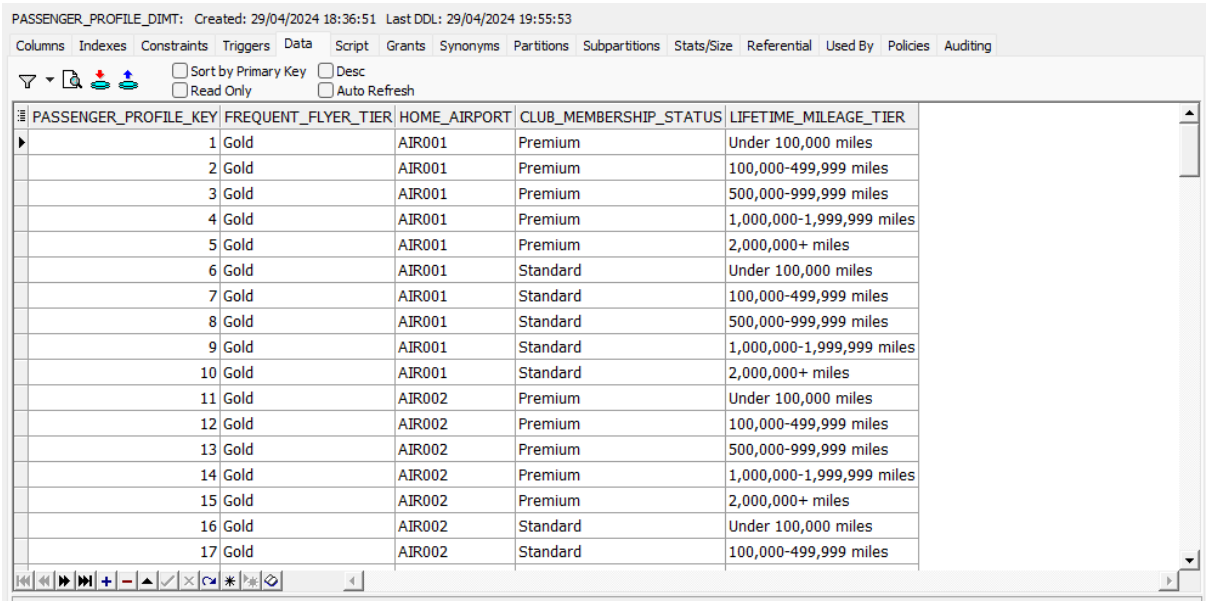
CLASS_OF_SERVICE_FLOWN_DIMT: Created: 29/04/2024 15:42:06 Last DDL: 29/04/2024 19:53:42

Columns | Indexes | Constraints | Triggers | Data | Script | Grants | Synonyms | Partitions | Subpartitions | Stats/Size | Referential | Used By | Policies | Auditing

Sort by Primary Key | Desc | Read Only | Auto Refresh

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

PASSENGER PROFILE DIMENSION



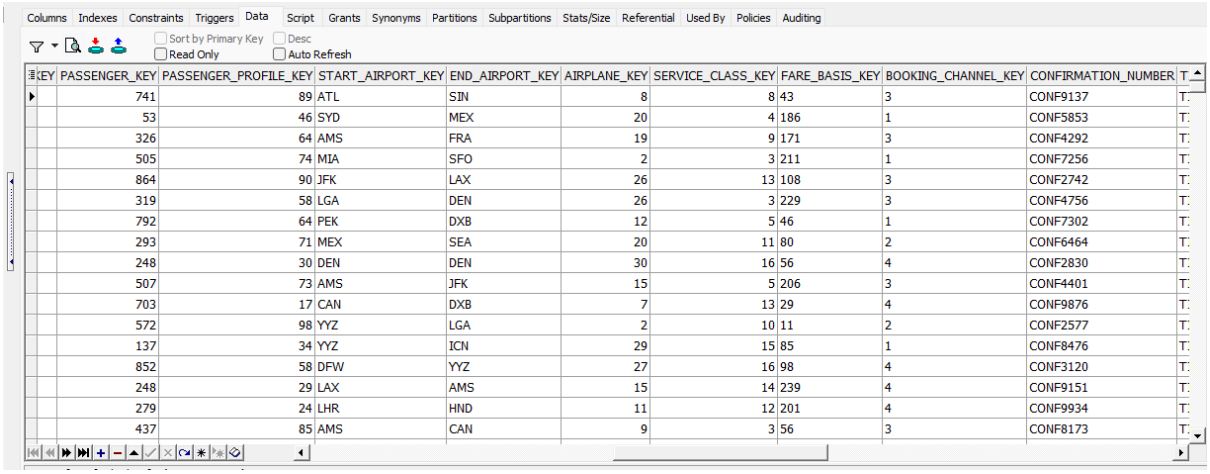
PASSENGER_PROFILE_DIMT: Created: 29/04/2024 18:36:51 Last DDL: 29/04/2024 19:55:53

Columns | Indexes | Constraints | Triggers | Data | Script | Grants | Synonyms | Partitions | Subpartitions | Stats/Size | Referential | Used By | Policies | Auditing

Sort by Primary Key | Desc | Read Only | Auto Refresh

PASSENGER_PROFILE_KEY	FREQUENT_FLYER_TIER	HOME_AIRPORT	CLUB_MEMBERSHIP_STATUS	LIFETIME_MILEAGE_TIER
1	Gold	AIR001	Premium	Under 100,000 miles
2	Gold	AIR001	Premium	100,000-499,999 miles
3	Gold	AIR001	Premium	500,000-999,999 miles
4	Gold	AIR001	Premium	1,000,000-1,999,999 miles
5	Gold	AIR001	Premium	2,000,000+ miles
6	Gold	AIR001	Standard	Under 100,000 miles
7	Gold	AIR001	Standard	100,000-499,999 miles
8	Gold	AIR001	Standard	500,000-999,999 miles
9	Gold	AIR001	Standard	1,000,000-1,999,999 miles
10	Gold	AIR001	Standard	2,000,000+ miles
11	Gold	AIR002	Premium	Under 100,000 miles
12	Gold	AIR002	Premium	100,000-499,999 miles
13	Gold	AIR002	Premium	500,000-999,999 miles
14	Gold	AIR002	Premium	1,000,000-1,999,999 miles
15	Gold	AIR002	Premium	2,000,000+ miles
16	Gold	AIR002	Standard	Under 100,000 miles
17	Gold	AIR002	Standard	100,000-499,999 miles

FLIGHT ACTIVITY FACT ANALYSIS(NOT COMPLETE)



Columns | Indexes | Constraints | Triggers | Data | Script | Grants | Synonyms | Partitions | Subpartitions | Stats/Size | Referential | Used By | Policies | Auditing

Sort by Primary Key | Desc | Read Only | Auto Refresh

KEY	PASSENGER_KEY	PASSENGER_PROFILE_KEY	START_AIRPORT_KEY	END_AIRPORT_KEY	AIRPLANE_KEY	SERVICE_CLASS_KEY	FARE_BASIS_KEY	BOOKING_CHANNEL_KEY	CONFIRMATION_NUMBER	T
	741	89	ATL	SIN	8	8	43	3	CONF9137	T
	53	46	SYD	MEX	20	4	186	1	CONF5853	T
	326	64	AMS	FRA	19	9	171	3	CONF4292	T
	505	74	MIA	SFO	2	3	211	1	CONF7256	T
	864	90	JFK	LAX	26	13	108	3	CONF2742	T
	319	58	LGA	DEN	26	3	229	3	CONF4756	T
	792	64	PEK	DXB	12	5	46	1	CONF7302	T
	293	71	MEX	SEA	20	11	80	2	CONF6464	T
	248	30	DEN	DEN	30	16	56	4	CONF2830	T
	507	73	AMS	JFK	15	5	206	3	CONF4401	T
	703	17	CAN	DXB	7	13	29	4	CONF9876	T
	572	98	YYZ	LGA	2	10	11	2	CONF2577	T
	137	34	YYZ	ICN	29	15	85	1	CONF8476	T
	852	58	DFW	YYZ	27	16	98	4	CONF3120	T
	248	29	LAX	AMS	15	14	239	4	CONF9151	T
	279	24	LHR	HND	11	12	201	4	CONF9934	T
	437	85	AMS	CAN	9	3	56	3	CONF8173	T

BUSINESS QUERIES

```
-- Query 1: Retrieve passenger information along with their booking channel details and the total revenue generated from their bookings
SELECT p.First_Name, p.Last_Name, p.Email_Address, p.Phone_Number, bc.Channel_Name,
    COALESCE(SUM(ff.Base Fare Revenue + ff.Passenger_Facility_Charges + ff.Airport_Tax +
ff.Government_Tax), 0) AS Total_Revenue
FROM Passenger_DimensionT p
LEFT JOIN Flight_Activity_FactT ff ON p.Passenger_Key = ff.Passenger_Key
LEFT JOIN Booking_Channel_DimensionT bc ON ff.Booking_Channel_Key = bc.Booking_Channel_Key
GROUP BY p.First_Name, p.Last_Name, p.Email_Address, p.Phone_Number, bc.Channel_Name;
```

FIRST_NAME	LAST_NAME	EMAIL_ADDRESS	PHONE_NUMBER	CHANNEL_NAME	TOTAL_REVENUE
David	Miller	david_miller@example.com	123456789001		0
Amber	Fowler	amber_fowler@example.com	123456789004	website	822.06
Brent	Martin	brent_martin@example.com	123456789027	mail	962.45
Shirley	Williams	shirley_williams@example.com	123456789035	mail	496.04
Scott	Bailey	scott_bailey@example.com	123456789036	cash desk	377.71
Ashley	Foster	ashley_foster@example.com	123456789049	cash desk	658.19
Edward	Rogers	edward_rogers@example.com	123456789051	call	466.33
Raymond	Mann	raymond_mann@example.com	123456789052	mail	1115.81
Sandra	Smith	sandra_smith@example.com	123456789053		0
Tonya	Torres	tonya_torres@example.com	123456789076		0
Crystal	Foley	crystal_foley@example.com	123456789087	website	415.13

```
-- Query 2: Get details of flights including their departure and arrival airports along with scheduled departure and arrival dates, and the average delay in departure and arrival times
SELECT ff.Flight_Number, dd1.Datee AS Scheduled_Departure_Date, dd2.Datee AS Scheduled_Arrival_Date,
       da.Airport_Name AS Departure_Airport, aa.Airport_Name AS Arrival_Airport,
       AVG(ff.Departure_Time_Delay) AS Avg_Departure_Delay,
       AVG(ff.Arrival_Time_Delay) AS Avg_Arrival_Delay
FROM Flight_Activity_FactT ff
LEFT JOIN Date_DimensionT dd1 ON ff.Scheduled_Departure_Date_Key = dd1.Date_Key
LEFT JOIN Date_DimensionT dd2 ON ff.Scheduled_Arrival_Date_Key = dd2.Date_Key
LEFT JOIN Airport_DimensionT da ON ff.Start_Airport_Key = da.Airport_Code
LEFT JOIN Airport_DimensionT aa ON ff.End_Airport_Key = aa.Airport_Code
GROUP BY ff.Flight_Number, dd1.Datee, dd2.Datee, da.Airport_Name, aa.Airport_Name;
```

FLIGHT_NUMBER	SCHEDULED_DEPARTURE_DATE	SCHEDULED_ARRIVAL_DATE	DEPARTURE_AIRPORT	ARRIVAL_AIRPORT	AVG_DEPARTURE_DELAY	AVG_ARRIVAL_DELAY
TICKET7193	07/04/2020	02/04/2015	Miami International Airport	Incheon International Airport	11	8
TICKET8257	03/05/2015	14/07/2015	Amsterdam Airport Schiphol	Dallas/Fort Worth International Airport	66	3
TICKET7313	15/07/2017	30/07/2015	Guangzhou Baiyun International Airport	Sydney Kingsford Smith Airport	110	95
TICKET5621	06/09/2018	30/09/2015	San Francisco International Airport	Miami International Airport	2	83
TICKET7554	18/09/2015	14/02/2016	Miami International Airport	Charles de Gaulle Airport	95	117
TICKET5845	22/06/2018	27/02/2016	Incheon International Airport	Miami International Airport	111	91
TICKET9991	22/10/2023	13/05/2016	O'Hare International Airport	Tokyo Haneda Airport	0	31
TICKET4691	10/08/2026	04/07/2016	Incheon International Airport	Hong Kong International Airport	105	94
TICKET2796	08/03/2017	04/07/2016	Heathrow Airport	Frankfurt Airport	115	83
TICKET8379	09/09/2025	12/07/2016	Sydney Kingsford Smith Airport	Singapore Changi Airport	41	47
TICKET7926	01/10/2021	10/09/2016	Frankfurt Airport	John F. Kennedy International Airport	12	102

```
-- Query 3: Retrieve passenger interactions along with their sentiment and ratings, and the total number of interactions per passenger
SELECT p.First_Name, p.Last_Name, i.Interaction_Type, i.Sentiment, i.Rating,
       COUNT(*) AS Total_Interactions
FROM Passenger_DimensionT p
LEFT JOIN Customer_Care_FactT cc ON p.Passenger_Key = cc.Passenger_Key
LEFT JOIN Interaction_DimensionT i ON cc.Interaction_ID = i.Interaction_ID
GROUP BY p.First_Name, p.Last_Name, i.Interaction_Type, i.Sentiment, i.Rating;
```

FIRST_NAME	LAST_NAME	INTERACTION_TYPE	SENTIMENT	RATING	TOTAL_INTERACTIONS
Courtney	Moran	inquiries	positive	1	1
Jennifer	Delgado	inquiries	neutral	3	1
Angel	Thompson	inquiries	neutral	5	1
Timothy	Phillips	inquiries	negative	1	1
Brett	Werner	inquiries	positive	2	1
Clinton	Hebert	inquiries	positive	3	1
Brian	Rogers	inquiries	positive	4	1

```sql
-- Query 4: Retrieve the top 5 airports with the highest number of departures
SELECT Airport_Name, City, Country, Region, Departure_Count
FROM (
    SELECT ad.Airport_Name, ad.City, ad.Country, ad.Region, COUNT(*) AS Departure_Count,
        RANK() OVER (ORDER BY COUNT(*) DESC) AS Rank_Departures
    FROM Flight_Activity_FactT ff
    LEFT JOIN Airport_DimensionT ad ON ff.Start_Airport_Key = ad.Airport_Code
    GROUP BY ad.Airport_Name, ad.City, ad.Country, ad.Region
)
WHERE Rank_Departures <= 5;
```

AIRPORT_NAME	CITY	COUNTRY	REGION	DEPARTURE_COUNT
Frankfurt Airport	Frankfurt	Germany	Europe	52
Dubai International Airport	Dubai	UAE	Middle East	48
Tokyo Haneda Airport	Tokyo	Japan	Asia	47
Dallas/Fort Worth International Airport	Dallas	USA	North America	46
Denver International Airport	Denver	USA	North America	44
O'Hare International Airport	Chicago	USA	North America	44
San Francisco International Airport	San Francisco	USA	North America	44