Day 9: Airline Operations & Flight Disruption Management System

Domain: Aviation / Operations Analytics

Difficulty Level: 🔥 Expert+ (Complex Multi-Table Relations, Nested Queries, Window Functions, and Edge Cases)

Goal: Build and analyze a real-world **airline management and disruption system** that tracks flights, crew assignments, airport operations, delays, and maintenance issues to derive insights and optimize airline performance.

✈️ Scenario Description

An international airline operates hundreds of daily flights across multiple countries.

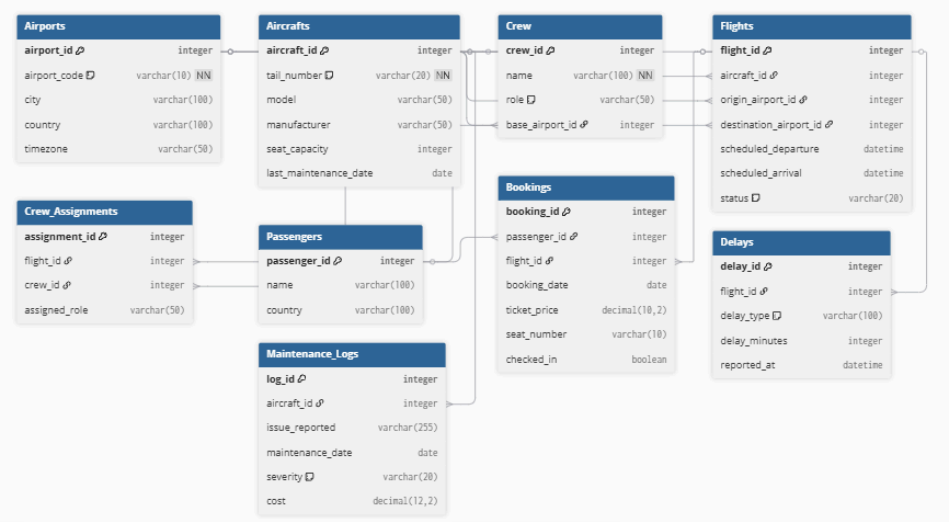
They need a system to track **flight schedules**, **crew assignments**, **passenger loads**, **aircraft maintenance**, and **disruption causes (weather, technical faults, air traffic delays, etc.)**.

The analytics team must answer critical questions such as:

- Which aircrafts or routes are most delay-prone?
- Are delays related to crew or maintenance issues?
- Which flights consistently exceed capacity utilization?

This dataset simulates a **complete airline operations analytics system** with deep inter-table relationships and realistic edge cases.

🧩 Database Schema Overview



Tables

1️⃣ Airports

Column	Type	Constraints	Description
airport_id	INT	PRIMARY KEY	Unique airport ID
airport_code	VARCHAR(10)	UNIQUE NOT NULL	IATA code (e.g., DEL, JFK)
city	VARCHAR(100)		City where the airport is located

Column	Type	Constraints	Description
country	VARCHAR(100)		Country name
timezone	VARCHAR(50)		Local timezone

2 Aircrafts

Column	Type	Constraints	Description
aircraft_id	INT	PRIMARY KEY	Unique aircraft ID
tail_number	VARCHAR(20)	UNIQUE NOT NULL	Aircraft registration number
model	VARCHAR(50)		Aircraft model (e.g., A320, B737)
manufacturer	VARCHAR(50)		Manufacturer name
seat_capacity	INT		Total seats
last_maintenance_date	DATE		Last maintenance performed

3 Crew

Column	Type	Constraints	Description
crew_id	INT	PRIMARY KEY	Unique crew member
name	VARCHAR(100)	NOT NULL	Crew member name
role	VARCHAR(50)		Role (Pilot, Co-Pilot, Cabin Crew)
base_airport_id	INT	FOREIGN KEY → Airports(airport_id)	Home base

4 Flights

Column	Type	Constraints	Description
flight_id	INT	PRIMARY KEY	Unique flight number
aircraft_id	INT	FOREIGN KEY → Aircrafts(aircraft_id)	Assigned aircraft
origin_airport_id	INT	FOREIGN KEY → Airports(airport_id)	Departure airport
destination_airport_id	INT	FOREIGN KEY → Airports(airport_id)	Arrival airport
scheduled_departure	DATETIME		Planned departure time
scheduled_arrival	DATETIME		Planned arrival time
status	VARCHAR(20)		Scheduled, Completed, Cancelled

5 Crew_Assignments

Column	Type	Constraints	Description
assignment_id	INT	PRIMARY KEY	Unique ID
flight_id	INT	FOREIGN KEY → Flights(flight_id)	Flight assigned
crew_id	INT	FOREIGN KEY → Crew(crew_id)	Crew member
assigned_role	VARCHAR(50)		Role during flight

6 Passengers

Column	Type	Constraints	Description
passenger_id	INT	PRIMARY KEY	Unique passenger
name	VARCHAR(100)		Passenger name
country	VARCHAR(100)		Country of origin

7️⃣ Bookings

Column	Type	Constraints	Description
booking_id	INT	PRIMARY KEY	Unique booking
passenger_id	INT	FOREIGN KEY → Passengers(passenger_id)	Passenger
flight_id	INT	FOREIGN KEY → Flights(flight_id)	Flight
booking_date	DATE		When booked
ticket_price	DECIMAL(10,2)		Price paid
seat_number	VARCHAR(10)		Seat assigned
checked_in	BOOLEAN		Whether passenger checked in

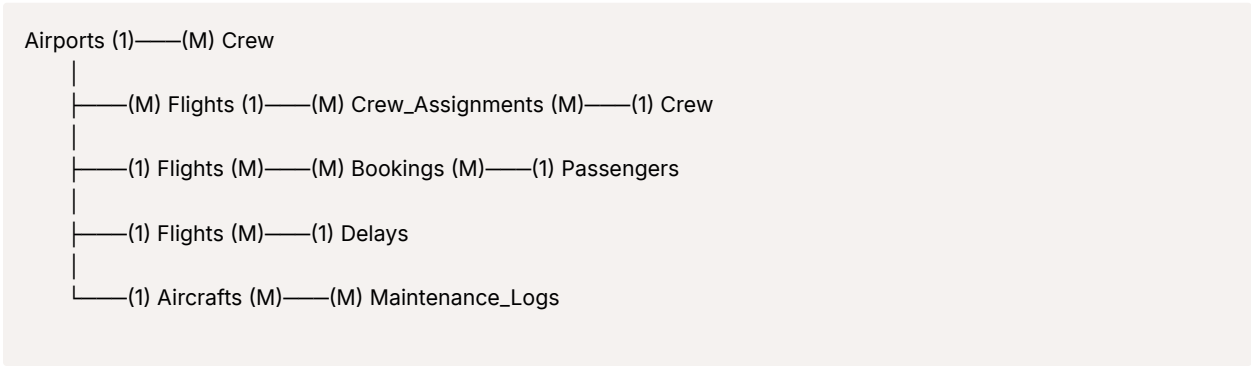
8️⃣ Delays

Column	Type	Constraints	Description
delay_id	INT	PRIMARY KEY	Unique delay ID
flight_id	INT	FOREIGN KEY → Flights(flight_id)	Flight affected
delay_type	VARCHAR(100)		Weather, Technical, Crew, Air Traffic, etc.
delay_minutes	INT		Delay duration
reported_at	DATETIME		When delay was reported

9️⃣ Maintenance_Logs

Column	Type	Constraints	Description
log_id	INT	PRIMARY KEY	
aircraft_id	INT	FOREIGN KEY → Aircrafts(aircraft_id)	Aircraft maintained
issue_reported	VARCHAR(255)		Description of issue
maintenance_date	DATE		Maintenance performed
severity	VARCHAR(20)		Minor / Major
cost	DECIMAL(12,2)		Maintenance cost

🧠 ERD (Textual Representation)



Relationships:

- One flight uses one aircraft but can have many delays.
- One aircraft may have multiple maintenance logs.
- Crew are assigned per flight.
- Passengers have bookings linked to flights.

🧾 Sample Data

Due to complexity, small subset only:

Airports

airport_id	airport_code	city	country	timezone
1	DEL	Delhi	India	IST
2	BOM	Mumbai	India	IST
3	DXB	Dubai	UAE	GST

Aircrafts

aircraft_id	tail_number	model	manufacturer	seat_capacity	last_maintenance_date
101	VT-A100	A320	Airbus	180	2024-11-01
102	VT-B777	B737	Boeing	200	2024-10-20

Flights

flight_id	aircraft_id	origin_airport_id	destination_airport_id	scheduled_departure	scheduled_arrival	status
201	101	1	2	2024-12-01 08:00:00	2024-12-01 10:00:00	Completed
202	102	2	3	2024-12-01 12:00:00	2024-12-01 16:00:00	Completed
203	101	1	3	2024-12-02 07:30:00	2024-12-02 11:30:00	Cancelled

Delays

delay_id	flight_id	delay_type	delay_minutes	reported_at
1	201	Weather	45	2024-12-01 07:30:00
2	202	Technical	120	2024-12-01 11:30:00

Maintenance_Logs

log_id	aircraft_id	issue_reported	maintenance_date	severity	cost
1	102	Engine Oil Leak	2024-10-15	Major	250000.00
2	101	Landing Gear Noise	2024-11-01	Minor	45000.00

🧩 SQL Questions

🟢 Easy:

List all flights delayed by more than 60 minutes with origin, destination, and aircraft model.

```
SELECT f.flight_id, a1.city AS origin, a2.city AS destination, ac.model, d.delay_minutes
FROM Flights f
JOIN Delays d ON f.flight_id = d.flight_id
JOIN Airports a1 ON f.origin_airport_id = a1.airport_id
JOIN Airports a2 ON f.destination_airport_id = a2.airport_id
JOIN Aircrafts ac ON f.aircraft_id = ac.aircraft_id
```

```
WHERE d.delay_minutes > 60;
```

🟡 Medium:

Find total maintenance cost per aircraft and flag those that exceeded ₹200,000 in the last 3 months.

```
SELECT ac.tail_number, SUM(m.cost) AS total_cost
FROM Aircrafts ac
JOIN Maintenance_Logs m ON ac.aircraft_id = m.aircraft_id
WHERE m.maintenance_date >= DATE_SUB(CURDATE(), INTERVAL 3 MONTH)
GROUP BY ac.tail_number
HAVING SUM(m.cost) > 200000;
```

🔵 Hard:

Identify crew members who were on flights that were delayed due to **technical reasons**.

```
SELECT DISTINCT c.name, c.role, f.flight_id, d.delay_type
FROM Crew c
JOIN Crew_Assignments ca ON c.crew_id = ca.crew_id
JOIN Flights f ON ca.flight_id = f.flight_id
JOIN Delays d ON f.flight_id = d.flight_id
WHERE d.delay_type = 'Technical';
```

🔴 Difficult:

Find top 3 airports with the **highest total delay time** for departing flights.

```
SELECT a.airport_code, a.city, SUM(d.delay_minutes) AS total_delay
FROM Delays d
JOIN Flights f ON d.flight_id = f.flight_id
JOIN Airports a ON f.origin_airport_id = a.airport_id
GROUP BY a.airport_code
ORDER BY total_delay DESC
LIMIT 3;
```

⚫ Expert (Analytical):

Compute the **average flight punctuality score** (100 - delay minutes / 10) per aircraft model,

and determine which **aircraft model performs worst** on average.

```
SELECT ac.model,
    ROUND(AVG(100 - (d.delay_minutes / 10)), 2) AS avg_punctuality_score
FROM Aircrafts ac
JOIN Flights f ON ac.aircraft_id = f.aircraft_id
JOIN Delays d ON f.flight_id = d.flight_id
GROUP BY ac.model
ORDER BY avg_punctuality_score ASC
```

```
LIMIT 1;
```

⚙️ Optimization & Insights

- **Indexing:** Use indexes on `flight_id`, `aircraft_id`, and `delay_type` for query speed.
- **Partitioning:** Large tables like `Flights` and `Delays` should be partitioned by month or route.
- **Data Cleaning:** Handle cancelled flights carefully — exclude them from delay calculations.
- **Analytical Extensions:** Combine this schema with `Weather_Reports` or `Air_Traffic_Control` tables for richer insights.