SCRIPTS DE CRIAÇÃO DAS TABELAS E INSERÇÃO DOS DADOS

Criando tabelas de dimensão e fato

```
-- Tabela de Dimensão: Data
CREATE TABLE DimDate (
    date_id INT PRIMARY KEY AUTO_INCREMENT,
    month INT,
    day INT,
    day_of_week INT
);

-- Tabela de Dimensão: Aeroporto
CREATE TABLE DimAirport (
    airport_id INT PRIMARY KEY AUTO_INCREMENT,
    airport_code VARCHAR(10),
    airport_name VARCHAR(255),
    airport_city VARCHAR(50),
    airport_country VARCHAR(10)
);

-- Tabela de Dimensão: Companhia Aérea
CREATE TABLE DimAirline (
    airline_id INT PRIMARY KEY AUTO_INCREMENT,
    airline_code VARCHAR(10),
    airline_name VARCHAR(255)
);

-- Tabela de Dimensão: Voo
CREATE TABLE DimFlight (
    flight_id INT PRIMARY KEY AUTO_INCREMENT,
    origin_airport_id INT,
    destination_airport_id INT,
    airline_id INT,
    scheduled_departure_time TIME, -- Horário programado de partida
    scheduled_arrival_time TIME,   -- Horário programado de chegada
    departure_time TIME,           -- Horário real de partida
    arrival_time TIME,             -- Horário real de chegada
    FOREIGN KEY (origin_airport_id) REFERENCES dimairport(airport_id),
    FOREIGN KEY (destination_airport_id) REFERENCES dimairport(airport_id),
    FOREIGN KEY (airline_id) REFERENCES dimairline(airline_id)
);

-- Tabela de Fatos: Atrasos de Voos
CREATE TABLE FactFlightDelays (
    fact_id INT PRIMARY KEY AUTO_INCREMENT,
    date_id INT,
    flight_id INT,
    origin_airport_id INT,
```

```
    destination_airport_id INT,
    airline_id INT,
    arrival_delay INT,
    air_time INT,
    scheduled_time INT,
    distance INT,
    departure_delay INT,
    FOREIGN KEY (date_id) REFERENCES dimdate(date_id),
    FOREIGN KEY (flight_id) REFERENCES dimflight(flight_id),
    FOREIGN KEY (origin_airport_id) REFERENCES dimairport(airport_id),
    FOREIGN KEY (destination_airport_id) REFERENCES dimairport(airport_id),
    FOREIGN KEY (airline_id) REFERENCES dimairline(airline_id)
);
```

Populando tabelas de dimensão e fato

```
-- Populando a tabela de dimensão DimAirline com dados da tabela airlines
INSERT INTO dimairline (airline_code, airline_name)
SELECT IATA_CODE, AIRLINE
FROM airlines_data;

-- Populando a tabela de dimensão DimAirport com dados da tabela airports
INSERT INTO dimairport (airport_code, airport_name, airport_city, airport_country)
SELECT IATA_CODE, AIRPORT, CITY, STATE
FROM airports_data;

-- Populando a tabela de dimensão DimDate com dados da tabela flights
INSERT INTO dimdate (month, day, day_of_week)
SELECT DISTINCT MONTH, DAY, DAY_OF_WEEK
FROM flights_data;

-- Populando a tabela de dimensão DimFlight com dados da tabela flights utilizando JOINs
INSERT INTO dimflight (origin_airport_id, destination_airport_id, airline_id,
scheduled_departure_time, scheduled_arrival_time, departure_time, arrival_time)
SELECT
    da_origin.airport_id AS origin_airport_id,
    da_destination.airport_id AS destination_airport_id,
    dl.airline_id AS airline_id,
    TIME_FORMAT(STR_TO_DATE(REPLACE(LPAD(f.SCHEDULED_DEPARTURE, 4, '0'),
'2400', '0000'), '%H%i'), '%H:%i:%s') AS scheduled_departure_time,
    TIME_FORMAT(STR_TO_DATE(REPLACE(LPAD(f.SCHEDULED_ARRIVAL, 4, '0'),
'2400', '0000'), '%H%i'), '%H:%i:%s') AS scheduled_arrival_time,
    TIME_FORMAT(STR_TO_DATE(REPLACE(LPAD(f.DEPARTURE_TIME, 4, '0'), '2400',
'0000'), '%H%i'), '%H:%i:%s') AS departure_time,
    TIME_FORMAT(STR_TO_DATE(REPLACE(LPAD(f.ARRIVAL_TIME, 4, '0'), '2400',
'0000'), '%H%i'), '%H:%i:%s') AS arrival_time
FROM
    flights_data f
```

```
JOIN dimairport da_origin ON f.ORIGIN_AIRPORT = da_origin.airport_code
JOIN dimairport da_destination ON f.DESTINATION_AIRPORT =
da_destination.airport_code
JOIN dimairline dl ON f.AIRLINE = dl.airline_code;

-- Populando a tabela de fatos FactFlightDelays com dados da tabela flights_data
INSERT INTO factflightdelays (
    date_id,
    flight_id,
    origin_airport_id,
    destination_airport_id,
    airline_id,
    arrival_delay,
    air_time,
    scheduled_time,
    distance,
    departure_delay
)
SELECT
    dd.date_id,  -- Obtém o ID da data
    df.flight_id,  -- Obtém o ID do voo
    da_origin.airport_id,  -- Obtém o ID do aeroporto de origem
    da_destination.airport_id,  -- Obtém o ID do aeroporto de destino
    dl.airline_id,  -- Obtém o ID da companhia aérea
    f.ARRIVAL_DELAY,  -- Atraso na chegada
    f.AIR_TIME,  -- Tempo de voo
    f.SCHEDULED_TIME,  -- Tempo programado de voo
    f.DISTANCE,  -- Distância percorrida
    f.DEPARTURE_DELAY  -- Atraso na partida
FROM
    flights_data f
JOIN dimdate dd
    ON f.MONTH = dd.month
    AND f.DAY = dd.day
    AND f.DAY_OF_WEEK = dd.day_of_week
JOIN dimairport da_origin
    ON f.ORIGIN_AIRPORT = da_origin.airport_code
JOIN dimairport da_destination
    ON f.DESTINATION_AIRPORT = da_destination.airport_code
JOIN dimairline dl
    ON f.AIRLINE = dl.airline_code
JOIN dimflight df
    ON df.origin_airport_id = da_origin.airport_id
    AND df.destination_airport_id = da_destination.airport_id
    AND df.airline_id = dl.airline_id
    AND df.scheduled_departure_time =
TIME_FORMAT(STR_TO_DATE(REPLACE(LPAD(f.SCHEDULED_DEPARTURE, 4, '0'),
'2400', '0000'), '%H%i'), '%H:%i:%s')
```

AND df.scheduled_arrival_time =
TIME_FORMAT(STR_TO_DATE(REPLACE(LPAD(f.SCHEDULED_ARRIVAL, 4, '0'), '2400', '0000'), '%H%i'), '%H:%i:%s');

Diagrama ER