

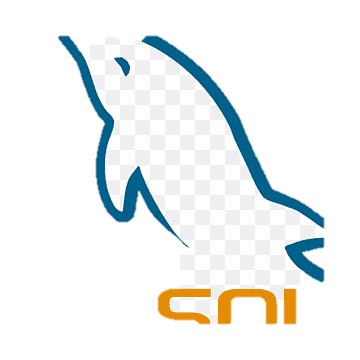
**AIRLINE ERD DIAGRAM**

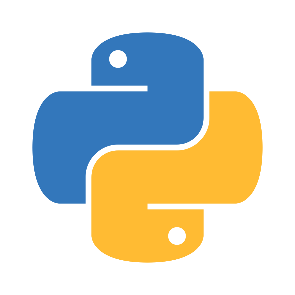
**Tables used for the process of Flight CANCELLATION and RESCHEDULING**

* **Transaction**: This table captures financial transactions pertaining to flight reservations. It includes essential details such as the unique transaction ID, corresponding reservation ID, and method of payment used (Method\_ID), transaction amount, date, and time of the transaction. It serves as a crucial component in tracking the monetary aspects of flight bookings and payments.
* **Route:** The Route table catalogs flight routes, providing vital information for flight planning and scheduling. It encompasses details such as the unique route identifier, departure airport ID, arrival airport ID, and the distance between these locations. This data is fundamental for optimizing flight operations, determining travel times, and facilitating route analysis.
* **Reservation**: Reservation maintains a comprehensive record of flight reservations made by passengers. It includes essential data like the reservation ID, corresponding flight ID, passenger ID, reservation date, seat number, and reservation status. This table plays a pivotal role in managing passenger bookings, seat allocations, and reservation status tracking throughout the flight booking lifecycle.
* **Rescheduled Flights**: Rescheduled Flights table tracks flights that have undergone rescheduling due to various factors. It encompasses details such as the rescheduled flight ID, associated cancellation ID, reimbursement fee, departure and arrival dates and times, departure and arrival locations, and the operating airline. This table facilitates the monitoring and management of flight rescheduling processes, ensuring smooth operations amidst disruptions.
* **Payment Method**: Payment Method table catalogs the various methods utilized for processing flight transactions. It includes entries for each payment method, comprising a unique method ID, method name, and a descriptive summary. This table aids in understanding the diversity of payment options available to passengers and facilitates financial transaction management within the airline's system.
* **Passenger**: Passenger table contains detailed information about passengers utilizing the airline's services. It includes data such as passenger ID, first name, and last name, date of birth, gender, email address, and phone number. This table serves as a central repository for passenger information, enabling personalized services, communication, and passenger profiling for operational purposes.
* **Flight**: Flight table provides comprehensive details about individual flights operated by the airline. It encompasses essential flight information such as flight ID, departure and arrival airport IDs, aircraft ID, departure and arrival dates and times, flight duration, and ticket prices. This table forms the backbone of flight management systems, facilitating flight scheduling, tracking, and passenger itinerary management.
* **Cancelled Flights Compensation**: The Cancelled Flights Compensation table records compensation details provided to passengers for cancelled flights. It includes entries for compensation ID, passenger ID, flight ID, compensation amount, compensation date, time, and the type of compensation provided. This table ensures transparent handling of flight cancellations and facilitates compensation management in compliance with regulations and airline policies.
* **Cancelled Flights**: Cancelled Flights table maintains records of flights that have been cancelled, along with relevant details such as cancellation ID, flight ID, reason for cancellation, cancellation date, and time. This table aids in tracking flight disruptions, analyzing cancellation patterns, and implementing measures to minimize future cancellations.
* **Airports**: The Airports table offers detailed information about airports served by the airline. It includes data such as airport ID, name, city, country, longitude, and latitude. This table serves as a valuable resource for flight planning, route optimization, and operational decision-making, ensuring efficient navigation and service provision across different airport locations.
* **Airport Status**: Airport Status table tracks the operational status of aircraft at various airports. It contains entries for status ID, aircraft ID, current status, last maintenance date, and next maintenance date. This table facilitates real-time monitoring of aircraft availability, maintenance schedules, and operational readiness, contributing to safe and efficient flight operations.
* **Airport Maintenance History**: Airport Maintenance History table maintains a historical record of maintenance activities conducted on aircraft at airports. It includes information such as maintenance ID, aircraft ID, and type of maintenance performed, maintenance date, and a descriptive summary. This table supports maintenance planning, compliance monitoring, and performance analysis for aircraft maintenance operations.
* **Aircraft**: The Aircraft table provides comprehensive details about the aircraft utilized by the airline fleet. It includes data such as aircraft ID, aircraft type, manufacturer, model, year manufactured, and capacity. This table serves as a central repository for aircraft information, facilitating fleet management, maintenance planning, and operational decision-making processes.

**Assumptions:**

1. **Compensation Policy for Flight Cancellations**: If a flight is cancelled less than 12 hours before its scheduled departure time, passengers will be entitled to compensation since they were informed at short notice. However, if the cancellation occurs with more than 12 hours’ notice, no compensation will be provided to passengers.
2. **Rescheduling and Collaboration Protocol**: In the event of a flight cancellation, the airline will attempt to reschedule the affected flight either using its own available aircraft or by collaborating with other airlines. The airline may have to pay a Reimbursement fee to collaborating airlines to cover the costs incurred for accommodating passengers and managing the disruption.
3. **Reasons for Flight Cancellations**: Flight cancellations may occur due to several predefined reasons:
   * + Aircraft Maintenance: When aircraft require maintenance or repairs that cannot be completed within the scheduled timeframe.
     + Crew Unavailability: Insufficient crew members available to operate the flight due to scheduling conflicts or unforeseen circumstances.
     + Insufficient Bookings: If the number of bookings for a particular flight falls below a predefined threshold, making the operation economically unfeasible.
     + Law and Order Disruptions or Emergencies: Situations such as civil unrest, security threats, or unforeseen emergencies that pose a risk to passenger safety or disrupt normal flight operations.
     + Weather Conditions (Arrival Location): Adverse weather conditions at the arrival airport that prevent safe landing or operation of the flight.
     + Weather Conditions (Departure Location): Adverse weather conditions at the departure airport that hinder aircraft departure or pose risks to flight safety.

**D Design Diagram of Pipeline**



**OLTP SQL Database (AIRLINE)**

**Dashboard Monitoring and Automatic Update**

**Power BI Dashboard**

**STAGING AREA**

**Data** **Transformation** **and Cleaning**

**Fact Tables Creation**

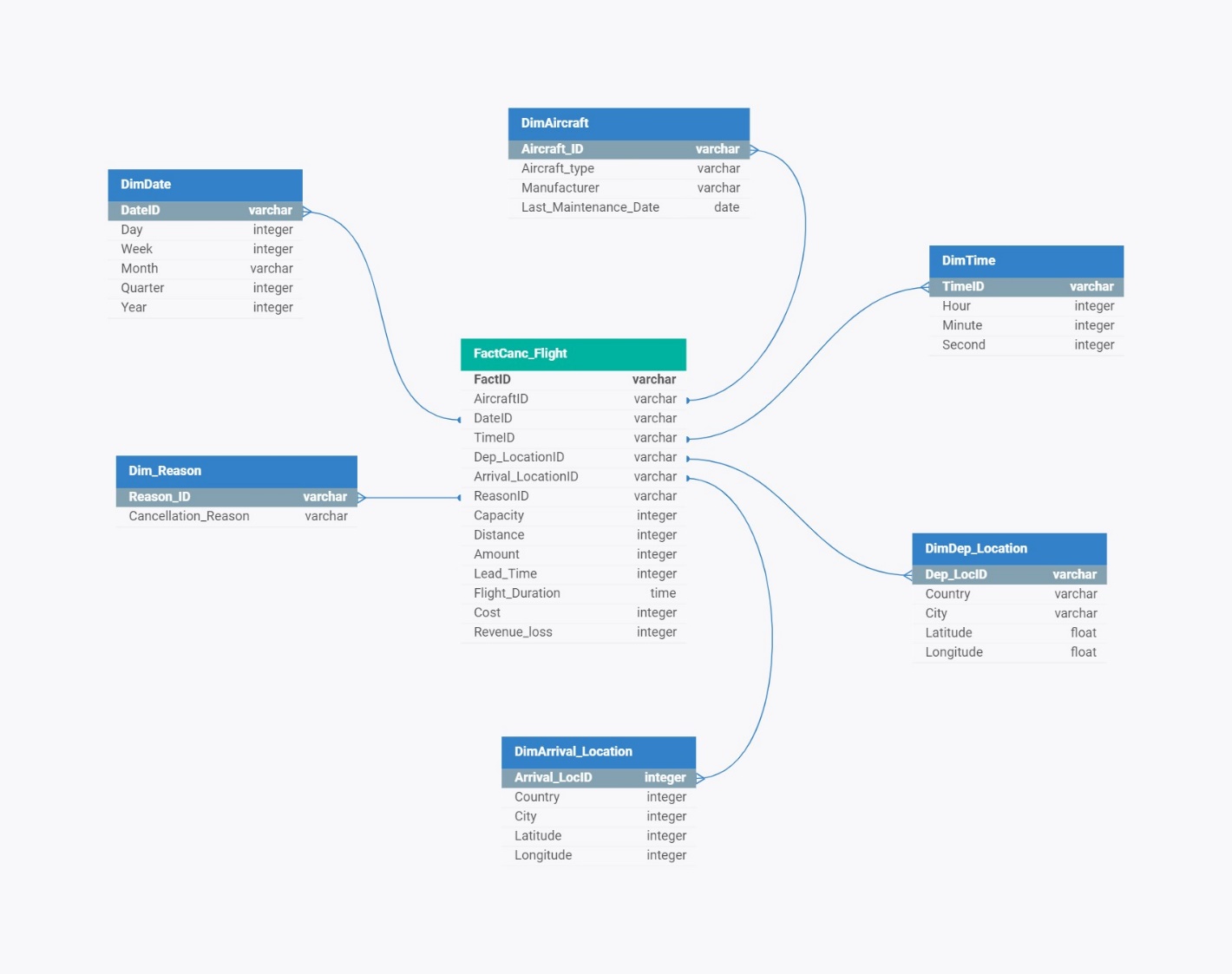
**Dimension Tables Creation**

**Star Schema Creation and Data Mapping to Star Schema**

**Data Extraction to Python**



**1. Extract, Transform, Load (ETL) Process:**

* **Extraction from MySQL**: Data is extracted from the OLTP (Online Transaction Processing) relevant tables stored in MySQL, including transactional, reservation, route, flight, passenger, and other pertinent tables and then stored in a staging area (dataframe).
* **Transformation in Python**: Extracted data undergoes Exploratory Data Analysis (EDA) and cleaning processes in Python. This involves handling missing values, data type conversions, data quality checks, creation of new attributes and other transformations to prepare the data for analysis.
* **Mapping to Star Schema**: Star Schema tables is created, and the transformed data is mapped into the Star Schema database called ‘data warehouse’ in MySQL. This schema is designed to optimize query performance and facilitate data analysis in a data warehousing environment.

**2. Data Warehousing:**

* **Star Schema**: A dimensional model composed of a central fact table surrounded by denormalized dimensional tables. It simplifies queries and improves performance for analytical purposes.
* **Fact Table**: The central table in a Star Schema that contains quantitative data (facts) about a business process or event. In this project, the fact table contains metrics including Lead time, cancelled flights, and compensation amounts.
* **Dimensional Tables**: Tables that store descriptive attributes (dimensions) related to the entities in the fact table. Examples include passenger details, flight information, airport data, and payment methods.

**3. Staging Area:**

* **Intermediate Data Storage**: During the ETL process, a staging area in python in the form of a data frame was used to temporarily store the extracted data before transformation. This staging area ensures data integrity and facilitates data manipulation without affecting the source system.

**4. Export to SQL Database:**

* **Loading Transformed Data**: The transformed data, including the fact table and dimensional tables, is exported back to a SQL database. This database serves as the repository for the denormalized Star Schema.
* **Normalization**: OLTP in normalized database structures to minimize redundancy and maintain data integrity, the denormalized Star Schema is designed for analytical querying and reporting purposes.

**5. Business Intelligence (BI) Visualization:**

* **Power BI Development**: The fact table snapshot stored in the SQL database is connected to Power BI for visualization and reporting purposes. Power BI offers intuitive tools for creating interactive dashboards, reports, and visualizations.
* **Data Modeling in Power BI**: The fact table and related dimensional tables are modeled within Power BI to establish relationships and hierarchies. This enables users to explore and analyze data across multiple dimensions.
* **Dashboard Creation**: Using Power BI, dashboards and reports were designed to provide insights into various aspects of the airline's operations, such as cancellation reasons, revenue, and flight cancellation over time and aircraft analyses.

**6. Publishing to Power BI Services and web:**

* **Cloud-Based Deployment**: The Power BI report, along with its underlying dataset, is published to Power BI Services, a cloud-based platform for hosting and sharing BI content, using gateway that connects local database to power BI services and update the data periodically as specified.