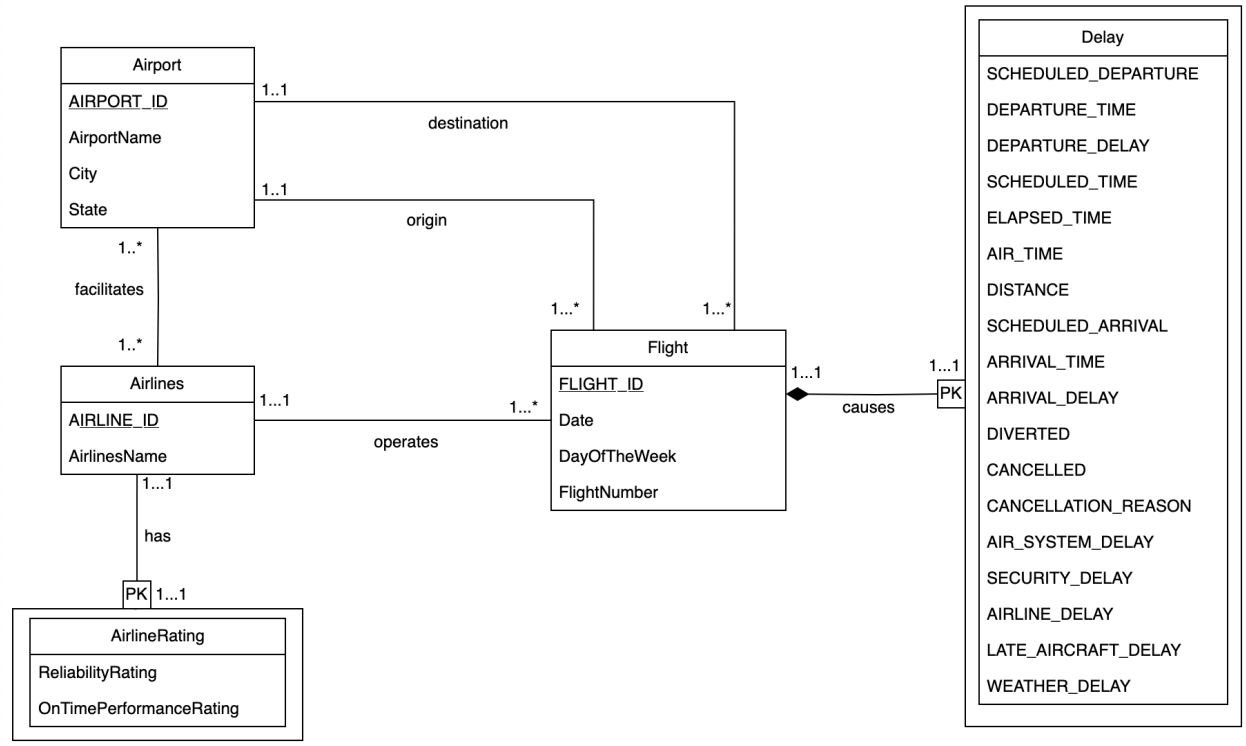
FlightShield Data Schema

ER/UML Diagram:



Normalization:

We have chosen to normalize our data using 3NF since BCNF does not guarantee dependency preservation, which 3NF guarantees. Also BCNF does not allow any sort of trivial dependencies, which reduces flexibility.

Airport.AIRPORT_ID->Airport.AirportName
Airport.AIRPORT_ID->Airport.City
Airport.AIRPORT_ID->Airport.State

Airlines.AIRLINE_ID->Airlines.AirlinesName

Airport.AIRPORT_ID->Airlines.AIRLINE_ID (facilitates relationship)

Airlines.AIRLINE_ID->AirlineRating.AIRLINE_ID (weak entity relation)

AirlineRating.AIRLINE_ID->AirlineRating.ReliabilityRating
AirlineRating.AIRLINE_ID->AirlineRating.OnTimePerformanceRating

Airlines.AIRLINE_ID<-Flight.AIRLINE_ID (operates relation) - FOREIGN KEY

Flight.FLIGHT_ID->Flight.Date
Flight.FLIGHT_ID->Flight.DayOfTheWeek
Flight.FLIGHT_ID->Flight.FlightNumber
Flight.FLIGHT_ID->Flight.OriginAirport - Foreign key
Flight.FLIGHT_ID->Flight.DestinationAirport - Foreign Key
Flight.FLIGHT_ID->Airlines.AIRLINE_ID

Airlines.AIRLINE_ID->Flight.OriginAirport (origin relation)
Airlines.AIRLINE_ID->Flight.DestinationAirport (destination relation)

Flight.FLIGHT_ID->Delay.FLIGHT_ID (weak entity relation)

Delay.FLIGHT_ID->Delay.SCHEDULED_DEPARTURE
Delay.FLIGHT_ID->Delay.DEPARTURE_TIME
Delay.FLIGHT_ID->Delay.DEPARTURE_DELAY
Delay.FLIGHT_ID->Delay.SCHEDULED_TIME
Delay.FLIGHT_ID->Delay.ELAPSED_TIME
Delay.FLIGHT_ID->Delay.AIR_TIME
Delay.FLIGHT_ID->Delay.DISTANCE
Delay.FLIGHT_ID->Delay.SCHEDULED_ARRIVAL
Delay.FLIGHT_ID->Delay.ARRIVAL_TIME
Delay.FLIGHT_ID->Delay.ARRIVAL_DELAY
Delay.FLIGHT_ID->Delay.DIVERTED
Delay.FLIGHT_ID->Delay.CANCELLED
Delay.FLIGHT_ID->Delay.CANCELLATION_REASON
Delay.FLIGHT_ID->Delay.AIR_SYSTEM_DELAY
Delay.FLIGHT_ID->Delay.SECURITY_DELAY
Delay.FLIGHT_ID->Delay.AIRLINE_DELAY
Delay.FLIGHT_ID->Delay.LATE_AIRCRAFT_DELAY
Delay.FLIGHT_ID->Delay.WEATHER_DELAY

Explanation of normalization:

Our database is normalized to 3NF because of the following reasons:

- Every non-primary-key attribute is fully functionally dependent on the primary key attribute
- None of the non-primary-key attributes are transitively dependent on the primary key
- For every FD X->Y, Y is a part of a candidate key and X is the super key is maintained

Assumptions/Deductions:

1. We think that if the ARRIVAL_DELAY is more than 15 minutes, only then the cause of delay columns (i.e. AIR_SYSTEM_DELAY, SECURITY_DELAY etc) are populated in the table
2. We believe the letters in the columns related to cancellation are cancellation codes used to reflect the scenarios or reasons for cancellation, as detailed in the Federal Register Report. For example, cancellation code A is air carrier issues, B is extreme weather and so on.
3. We assume that all airlines have at least 1 operating flight in the database
4. We have dropped several columns from the original data set (e.g. TAIL_NUMBER, TAXI_IN, TAXI_OUT, WHEELS_OFF, WHEELS_IN from flights table, LATITUDE, LONGITUDE from the airports table) as we do not anticipate using this data either for predicting flight delays or modeling our application
5. In the flights table, we combined the month and day to the date column and we also dropped the year column because the dataset was just from 2015 flight data and it was not giving us any useful information.
6. In the airports table, we dropped the countries column because the flight data was found to be relevant only to the USA, making it redundant.

Cardinality assumptions:

1. We think that there is a many-to-many relationship between Airport and Airlines
2. We think each airline operates one or more flights, so it is a one-to-many relationship between Airlines and Flight
3. We think every airline has exactly a fixed rating (which is a percentage measurement of its timeliness and reliability), which is why we have chosen a one-to-one relationship between Airlines and AirlineRating
4. We think that every flight has exactly one origin airport and one destination airport, which is why there is a one-to-many relationship between Airport and Flight for both origin and destination respectively
5. We think that each flight has a delay schedule associated with it. There can only be one delay entry (since the delay is calculated cumulatively), so there is a one-to-one relationship between Delay and Flight.

Description:

The Airport table is connected to the Airlines table having one to many relationships. All destination and origin airports are connected to the flights from the flight table as a relationship. It has the AIRPORT_ID as the primary key of data type **INT**. It also contains the attributes, AirportName, City in which the airport is located, and the State in which the airport is located of Data type **VARCHAR**.

The Airlines table is connected to the Airport table as 1 to many relationships since the airports facilitate at least one airline. The airlines table is also connected to the flight table since each airline operates the flights and all the airlines are operating at least one flight. It contains the AIRLINE_ID as the primary key with the data type **INT**. The other attribute is AirlinesName with the data type **VARCHAR.**

AirlineRating table is a weak entity that is connected to the Airline table. It contains metrics such as reliability rating and on time performance rating. The reliability rating is calculated using cancellation metrics, and the on time performance is calculated using delay metrics. Both metrics are a percentage value. The primary key is the AIRLINE_ID that uniquely identifies ratings connected to the airlines of data type **INT**.

Flight table is a strong entity to which Delay entity is connected to. It has several attributes, including FLIGHT_ID which is the primary key of data type **INT**, the flight number, AIRLINE_ID, date of flight (which includes month and day), day of the week, code of the origin airport, and code for the destination airport.

Delay table is a weak entity connected to Flight's table having a one to one relationship.The primary Key is the FLIGHT_ID which uniquely identifies the delays and cause of delay related to each flight. The table contains attributes such as SCHEDULED_DEPARTURE, DEPARTURE_TIME, SCHEDULED_TIME, ELAPSED_TIME, AIR_TIME, DISTANCE, SCHEDULED_ARRIVAL, ARRIVAL_TIME which are metrics to identify the scheduled trip plan of the flight , time and distance it has taken to reach the destination. All of these attributes are of **INT** data type. The delay attributes which give the time delays of the flight are given by DEPARTURE_DELAY, ARRIVAL_DELAY, AIR_SYSTEM_DELAY, SECURITY_DELAY, AIRLINE_DELAY, LATE_AIRCRAFT_DELAY, WEATHER_DELAY. These attributes are also of

data type **INT**. Lastly, DIVERTED and CANCELED are Boolean indicators of the data type **INT** if a flight was canceled or diverted from its planned schedule. CANCELLATION_REASON gives a **CHAR** indicator of the reason for cancellation.

The letter codes indicate the following based on:
Bureau of Transportation Statistics
A - Carrier Caused
B - Weather
C - National Aviation System
D - Security

Relational Schema:

Airport table

```
Airport(
      AIRPORT_ID VARCHAR(3) PRIMARY KEY,
      AirportName VARCHAR(255),
      City VARCHAR(255),
      State VARCHAR(2)
);
```

Airlines table

```
Airport(
      AIRLINE_ID VARCHAR(2) PRIMARY KEY,
      AirlinesName VARCHAR(255)
);
```

Facilitates table

```
Facilitates(
      AIRLINE_ID VARCHAR(2),
      AIRPORT_ID VARCHAR(3),
      PRIMARY KEY(AIRLINE_ID, AIRPORT_ID),
      FOREIGN KEY(AIRLINE_ID) REFERENCES Airlines.AIRLINE_ID,
      FOREIGN KEY(AIRPORT_ID) REFERENCES Airport.AIRPORT_ID
);
```

AirlineRating table

```
AirlineRating(
      AIRLINE_ID VARCHAR(2) PRIMARY KEY,
```

```
      ReliabilityRating INT,
      OnTimePerformanceRating INT,
      FOREIGN KEY(AIRLINE_ID) REFERENCES Airlines.AIRLINE_ID
);
```

Flight table

```
Flight(
      FlIGHT_ID INT PRIMARY KEY,
      Date DATE,
      DayOfTheWeek INT,
      AIRLINE_ID VARCHAR(3),
      FlightNumber INT,
      OriginAirport VARCHAR(3),
      DestinationAirport VARCHAR(3),
      FOREIGN KEY(OriginAirport) REFERENCES Airport.AIRPORT_ID,
      FOREIGN KEY(DestinationAirport) REFERENCES Airport.AIRPORT_ID,
      FOREIGN KEY(AIRLINE_ID) REFERENCES Airlines.AIRLINE_ID,
);
```

Delay table

```
Delay(
      FlIGHT_ID INT PRIMARY KEY,
      SCHEDULED_DEPARTURE DATE,
      DEPARTURE_TIME DATE,
      DEPARTURE_DELAY INT,
      SCHEDULED_TIME DATE,
      ELAPSED_TIME INT,
      AIR_TIME INT,
      DISTANCE INT,
      SCHEDULED_ARRIVAL DATE,
      ARRIVAL_TIME DATE,
      ARRIVAL_DELAY INT,
      DIVERTED INT,
      CANCELED INT,
      CANCELATION_REASON VARCHAR(1),
      AIR_SYSTEM_DELAY INT,
      SECURITY_DELAY INT,
      AIRLINE_DELAY INT,
      LATE_AIRCRAFT_DELAY INT,
      WEATHER_DELAY INT,
      FOREIGN KEY(FLIGHT_ID) REFERENCES Flight.FLIGHT_ID
);
```