Flights Schema

Dataset

The data in the provided dataset (zipped CSV files) is abridged from the Bureau of Transportation Statistics, and describes a subset of flights that took place in the United States during September 2024.

Flights

The flights table has one row for every marketed domestic commercial flight in the United States during September 2024. There are approximately 577,000 rows.

```
Flights(
    fid INT,                    -- unique id for each flight
    year INT,
    month INT,                  -- 1-indexed (ie, 1-12)
    day_of_month INT,           -- 1-indexed (ie, 1-31)
    day_of_week INT,            -- 1-indexed (ie, 1-7). 1 = Monday,
                               -- 2 = Tuesday, etc
    cid VARCHAR(8),             -- id of the airline, see carriers
    tail_num VARCHAR(6),        -- the "license plate" of the aircraft;
                               -- see also N_Numbers table
    op_carrier_flight_num INT, -- carrier-specific flight number.
NOT
                                -- unique; different carriers can each
                                -- have a flight numbered "123"
    origin VARCHAR(3),          -- airport code where flight started
    origin_city VARCHAR(40),  -- "<city>, <state>", e.g. "Seattle,
WA"
    origin_state VARCHAR(2),
    dest VARCHAR(3),            -- airport code of destination
    dest_city VARCHAR(40),    -- "<city>, <state>", e.g. "Seattle,
WA"
    dest_state VARCHAR(2),
    sched_dep_time INT,         -- scheduled departure time
    dep_time INT,               -- actual departure time
    dep_delay REAL,
    sched_arr_time INT,         -- scheduled arrival time
    arr_time INT,               -- actual arrival time
    arr_delay REAL,
    cancelled INT,              -- 1 if flight was cancelled, else 0
```

```
    cancellation_code VARCHAR(1),-- reason for cancellation, or
empty
                                  -- see cancellation_codes table
    duration_mins INT,
    distance_mi INT,
    price INT                 -- in dollars
)
```

N-Numbers

The N-numbers table includes one row for every publicly registered aircraft in the United States as of January 2025, around 300,000 records. An "N-number" uniquely identifies a physical aircraft in the United States, similar to a car's license plate. On the other hand, a single aircraft may be used for multiple flights; this is similar to how a single physical bus can operate multiple routes.

```
N_Numbers(
  n_number VARCHAR(6), -- the "licence plate" number
  serial_number VARCHAR(30),
  mfr_mdl_code VARCHAR(7),  -- manufacturer and model code, see
                            -- aircraft_types
  year_mfr VARCHAR(4),
  name VARCHAR(50),     -- name of entity that registered the
aircraft
  street VARCHAR(40),   -- the following columns are their address
  street2 VARCHAR(40),
  city VARCHAR(20),     -- all-caps city, e.g. "SEATTLE".  Can be
                        -- concatenated with state to obtain
                        -- Flights.origin_city or Flights.dest_city
  state VARCHAR(2),     -- all-caps state code, e.g. "WA".  Can be
                        -- concatenated with city to obtain
                        -- Flights.origin_city or Flights.dest_city
  zip_code VARCHAR(10),
  region VARCHAR(1),
  county VARCHAR(3),
  country VARCHAR(2)
)
```

Aircraft Types

The aircraft types table contains one row for each model of aircraft ever registered, including its manufacturer; there are around 92,000 records. It is used by the `mfr_mdl_code` column of the N-number table to describe the type of each registered aircraft.

```
Aircraft_Types(
```

```
  atid VARCHAR(7),          -- the id, used by the n_numbers table
  mfr VARCHAR(40),          -- name of the manufacturer
  model VARCHAR(30),        -- name of the model
  num_engines INT,          -- number of engines on this type
  num_seats INT,            -- max number of seats on this type;
                            -- sometimes called the aircraft's
                            -- "capacity".  Never null.
  weight_class VARCHAR(7), -- 1-indexed (1-4)
  avg_speed_mph INT         -- average cruising speed, miles per hour
)
```

Carriers

The carriers table is a table of airlines and their names, roughly 1700 records. Each carrier has a unique carrier code used by `cid` in the Flights table to describe which airline flew the flight.

```
Carriers(
    cid VARCHAR(8),      -- the id of the airline
    cname VARCHAR(100)   -- the full name of the airline
)
```

Cancellation Codes

The cancellation codes table includes four short codes for possible reasons a flight would be cancelled. Used by the `cancellation_code` column of the Flights table. A flight which was not cancelled (*i.e.*, flew as scheduled) will have a null value for its `cancellation_code`.

```
Cancellation_Codes(
    ccid VARCHAR(1),          -- the cancellation id
    description VARCHAR(20)   -- why the flight was cancelled
)
```

Key Constraints

The tables are subject to the following additional constraints, which you should enforce:

- Primary keys:
  - `fid` for the `Flights` table
  - `cid` for `Carriers`
  - `ccid` for `Cancellation_Codes`
  - `n_number` for `N_Numbers`
  - `atid` for `Aircraft_Types`

    - Other than these primary keys, **DO NOT** assume any other attribute is a key and/or unique across tuples.
- Foreign keys:
    - `Flights.cid` references `Carriers.cid`
    - `Flights.tail_num` references `N_Numbers.n_number`
    - `N_Numbers.mfr_mdl_code` references `Aircraft_Types.atid`
    - Note that `Flights.cancellation_code` is NOT declared as a foreign key to `Cancellation_Codes`. This is because it can be blank for non-cancelled flights. However, you may join on it just like a foreign key.