

Clint Valentine

DA5020

Relational Database Design

Objectives

1. Design a relational data storage structure
2. Normalize data to at least 3NF
3. Create a visual model

Introduction

A normalized third normal form relational schema is used to ensure referential integrity. We have many candidate keys to begin evaluating before deciding on the number of tables needed to conform to 2NF and 3NF normalization.

First, the primary key for all bird strike records can logically be chosen to be the `Record ID` as this value is unique in the dataset and is likely to never collide to bird strike records. This can be validated with the following Python code:

```
import csv
from collections import Counter

infile = 'c:/dev/coursework/data_collecting_storing_retrieving/module09/Bird Strikes.csv'
ids = Counter()

with open(infile) as handle:
    reader = csv.reader(handle, delimiter=',')
    header = next(handle)
    for line in reader:
        ids[line[10]] += 1

print(max(ids.values()) == 1)
```

To begin breaking this single spreadsheet into tables that conform to 2NF we must examine all fields and remove any dependency candidate keys have on a subset of other fields. For example, information exists that is specific to the airplane in the bird crash event and those details are specific to that of the airplane and not to the crash itself.

An abstraction that was obvious at first is that one flight should consist of one airplane model but could have multiple bird strike events. To conform to 3NF a table for aircrafts and a table for flights should be made. Finally, there are many fields that characterize the event itself that are not determined by any non-prime attributes.

Summary of Attributes

Field Name	Attribute Of
Aircraft Type	Aircraft
Airport Name	Flight
Altitude	Event
Aircraft: Make/Model	Aircraft
Wildlife: Number struck	Event
Effect: Impact to flight	Event
Effect: Other	Event
Location: Nearby if <i>en route</i>	Event
Flight Number	Flight
Flight Date	Flight
Record ID	Event
Effect: Indicate damage	Event
Aircraft: Number of engines	Aircraft
Airline/Operator	Flight
Origin state	Flight
Conditions: Precipitation	Event
Remains of wildlife collected?	Event
Remarks	Event
Reported: Date	Event
Wildlife: size	Event
Conditions: Sky	Event
Wildlife: species	Event
When	Event
Pilot warned of bird strikes	Event
Cost: Aircraft time out	Flight
Cost: Repair	Flight
Feet above ground	Event
Speed	Event

Aircraft Table

The simplest table to create is that of the aircraft. There only three attributes that belong to this table with the Aircraft: Make/Model appearing as prime and is selected to be the primary key for this table. An assumption is made that no Aircraft: Make/Model has multiple engine configurations.

Flight Table

The flight table joins the bird crash events with the airplane that is used for a flight. An assumption is made that a flight can only consist of one airplane yet could have multiple bird crash events. For this reason, attributes like Airline or Operator, Cost: Aircraft Time Out, Cost: Repair, and Pilot Warned of Bird Strikes are a part of the Flight Table and not the Bird Strike Table. The Flight Number is prime and can be used as the primary key for this table. The foreign key Aircraft Make or Model is used to relate to the Aircraft Table.

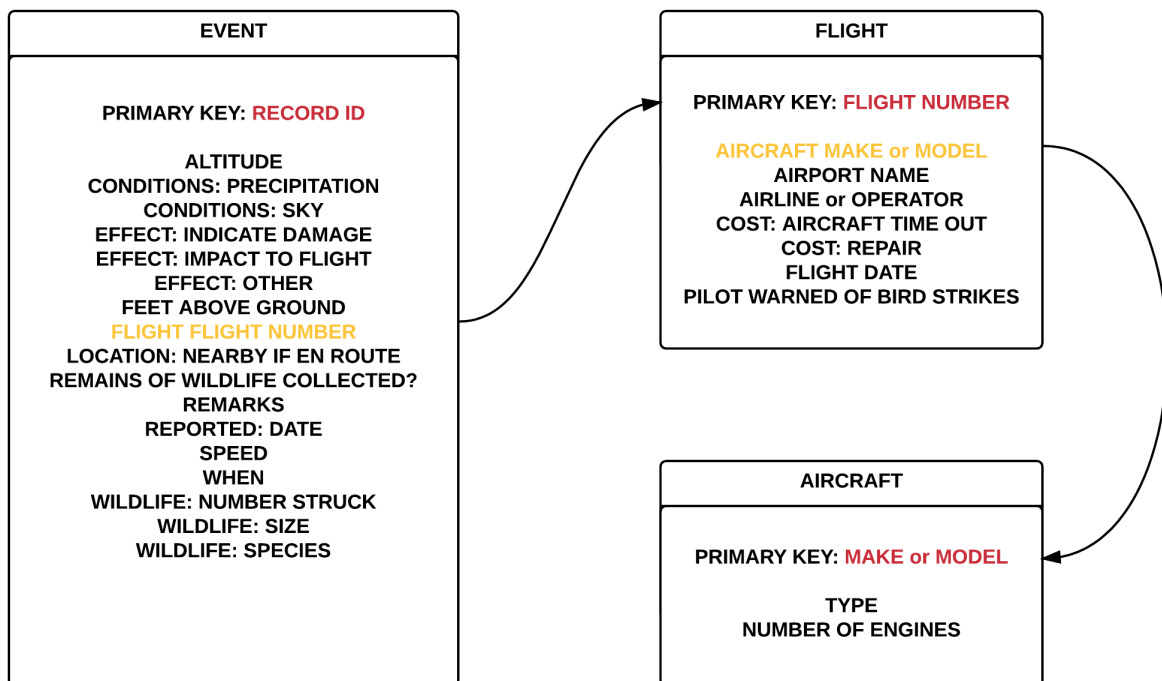
Bird Strike Table

The final table is the largest and holds the records pertaining to each bird strike. A prime attribute, Record ID, is used as the primary key to this table. All attributes assigned to the bird strike table are non-prime, and there are many. The foreign key Flight Flight Number is used to reference a flight record in the Flight Table which then references the aircraft. Since many bird strike events could occur on a flight there are likely multiple bird strike Record IDs referencing the same Flight Flight Number.

Potentially a fourth table could be made for information about the wildlife encountered during a strike which would reduce the footprint of our database albeit by adding a table. After examining the values for the fields Wildlife: Number Struck, Wildlife: Size, and Wildlife: Species it was obvious there was no easy way to abstract a fourth table since multiple species of multiple birds could exist at different sizes and in varying numbers.

Other details that could have been included in the Flight table like Altitude, Conditions..., Effect..., Feet Above Ground, Speed, and When are not a good fit since multiple bird strike events could have different values for these attributes.

The Relational Model



The Relational Model above shows primary keys of tables in red and foreign keys in yellow.