Term Project Proposal  
**Database of US Airports and Local Airline Routes**

Frederico Santos (santfd01@pfw.edu)  
Michael-Anthony Noman (normmj02@pfw.edu)  
  
Computer Science Department, Purdue University  
Fort Wayne, IN, 46805

**Project Description**

This project combines various datasets to compile a database of US airports and the routes between them. With this data, we can then design and implement a dashboard application which can be used to easily view headline statistics about the data, as well as perform other analytic operations.   
  
The key tasks in this process will be divided into two main parts: the design and implementation of the database and the design and implementation of the dashboard application around said database. For the database, the key datasets will have to be compiled and then based on the data available, we can design the database; building the models and relational schema and then implement the schema in our chosen database technology. Once the database is implemented, we can then automate the process of populating the data from the datasets compiled.   
  
Once the database has been designed, we can then also design the dashboard application. In this process we will have to identify key statistical data which we will need to extract from the data and the key operations which will be provided to the users. This will then guide the design process as the application will be built around these key operations.

**Related Works**

**Graphic Visualization of Specific Dashboards in Transport Technologies***Pokorný & Stokláska (2016)*

This paper details the development of a web application for the generation of interactive graphs for the purpose of city traffic monitoring. After an investigation of the existing libraries, namely Charts, D3 and Flot, the authors determined that they were too complex and were not suited for this application, leading to their decision to develop their own solution.   
  
In light of this they developed their own JavaScript library, *inCharts*, which followed object-oriented principles. It is composed of individual component classes and static classes which are common to all the desired graphs. The component instances are all associated with a Canvas which manages their presentation in the DOM.

This paper, and the inCharts library it describes, provides an outlook which we may take in the development of our own system. It discusses some other resources which we may consider using in our own project, and if we come to the same conclusion as the authors, may provide an ideal data visualization tool for our needs.

**A database perspective on geospatial data modeling***Voisard & David (2002)*

This paper discusses the use of database management systems (DBMSs) for the representation and manipulation of geospatial information. This model allows for the efficient definition of collections of geographic objects and the relationships between them, and for the manipulation of spatial and other non-basic data types.

The authors discuss the encapsulation of the geospatial data into two layers: the higher Map level which is represented by geographic objects and a lower Geometric level, consisting of the spatial components of the geographic objects. They also discuss the ubiquity of aggregation and disaggregation operations in GIS applications and how the standard relational model is ill-suited for these operations. To remedy this, they propose a mapping a general geographic model unto the models of existing DBMS, namely the relational and object-oriented models.

This paper provided an interesting outlook as to the design of a database with geospatial data. While all of the considerations discussed are not necessarily applicable to a project of our scope, it does give us an understanding of the potential issues we could face with a naïve database implementation.

**Data Requirement**

A model of US local air routes, based on the following requirements:

* Each route is represented by a RouteID and consists of an Originating airport, a Destination airport and is offered by a airline/air carrier.
* An Airport has a name, IATA code, City, State, Altitude and Geo-positional coordinates which consist of a longitude and latitude.
* An Airline has an AirlineID, Name, Alias, Callsign and is registered in a Country.
* For each Route there is a set of monthly records, which give the number of passengers, available seats, airtime flown, weight of mail and freight carried, type of aircraft flown, its configuration and the service class of the airline route during that month period.
* An Aircraft has an Aircraft ID, a name and belongs to an Aircraft Group.
* Service Class defines the type of service offered by that route, and has an SCID and a description.
* Aircraft Configuration defines how the aircraft was modified for that route, and has an ACID and a description.
* Aircraft Group describes a range of aircraft with similar uses and features, and has an AGID and a description.

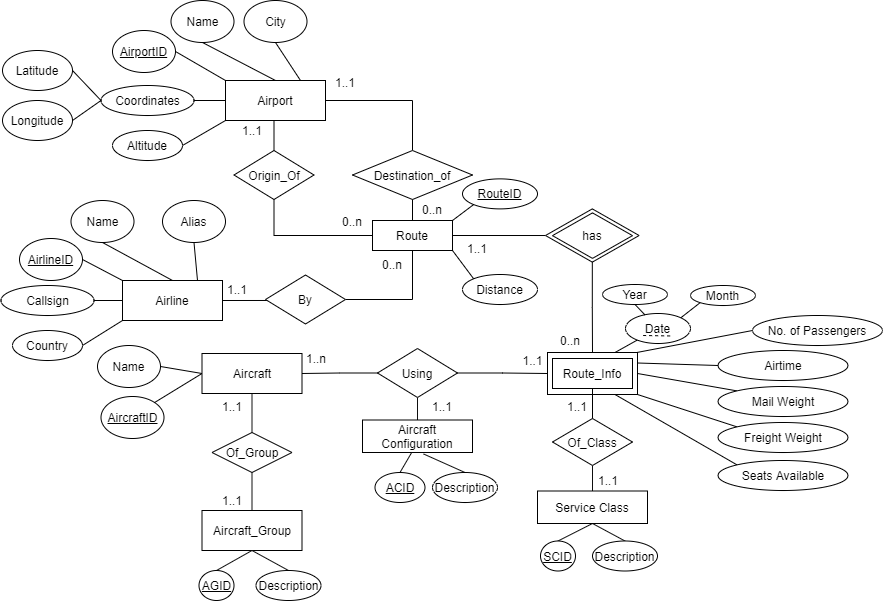
**Conceptual Database Schema**

Figure 1. Showing Entity Relation Diagram for the proposed database

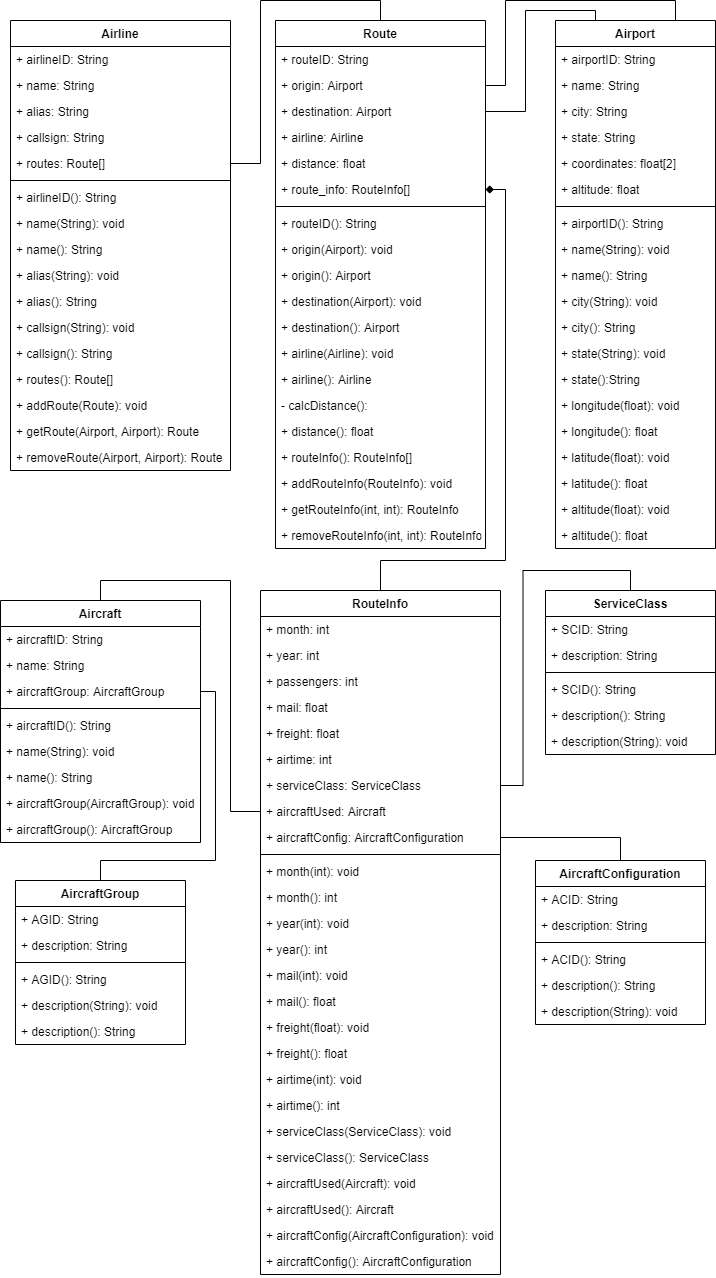
****

Figure 2. Showing Class Diagram modeling the proposed database.

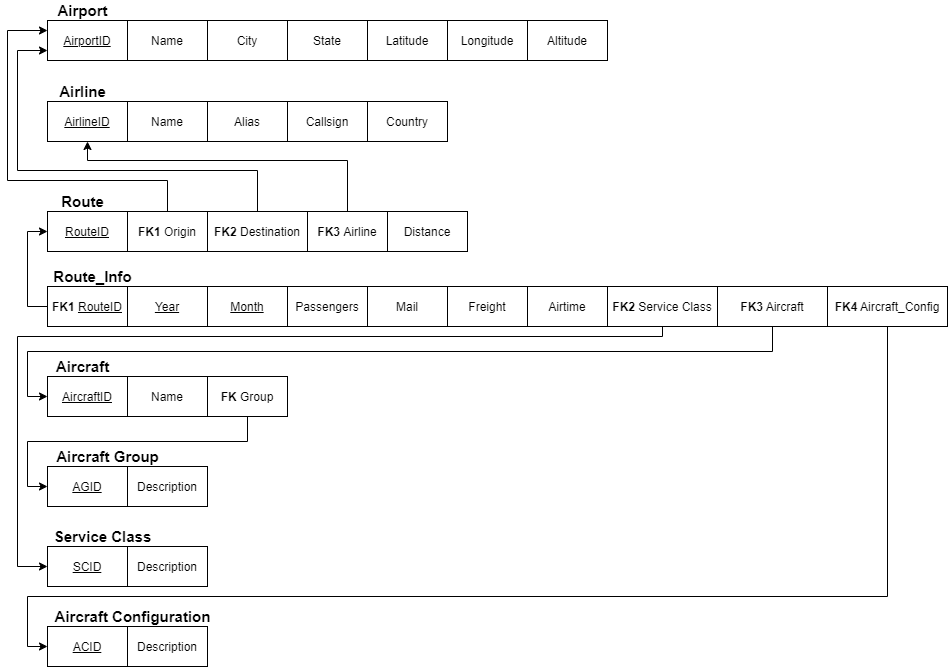
****

Figure 3. Showing Relational Schema based on the normalization of the ER Diagram in Figure 1.

**Database Technology**

A key datatype in this domain will be geographic/spatial data, and as a result we have decided to use PostGIS as the main database technology for this system. PostGIS is a set of extensions to Postgres, a common Object-Relational Database Management System (ORDBMS), which enables it to handle spatial and geographic data.

**Data**

The data for this project is compiled from 2 sources. The first is the Bureau of Transportation Statistics (BTS) which provides a monthly report on all routes flown by all air carriers, including freight, mail, and passenger carriers. The second data source is OpenFlights, an online resource which has compiled an extensive dataset of airlines and airports. This data was compiled in a set of Comma Separated Value (CSV) files and processed using a Python script into the classes represented in the desired database solution. The files before this processing are provided along with this document.

**Brief of Prototype System**

Our prototype system will consist of a fully functioning and fleshed database of the US airports and flights amongst them. The main entities we expect to include are the Airports and Airlines, with routes being represented as relations between at least two airports (a source and destination) and the airline associated with that route. This may be extended to include information on the planes themselves based on the information available.   
With the database, there will be a dashboard that allows the user to navigate and perform important operations with the data, such as query specific data and certain geolocation functionalities. On this dashboard the user may expect to view historical data on flights over a given period, easily determine the most popular routes and airlines and perform more specific statistical operations to identify trends in the data.

**Proposed Work Schedule**

|  |  |  |
| --- | --- | --- |
| **Week** | | **Task** |
| 1 | Feb 7 - 13 | Data Acquisition and Database Design |
| 2 | Feb 14 - 20 | Database Design and Implementation |
| 3 | Feb 21 - 27 | Database Design and Implementation |
| 4 | Feb 28 - Mar 6 | Data Processing and Database Population |
| 5 | Mar 7 - 13 | Preparation of Mid-Way Report, Application Requirements Elicitation |
| **6** | **Mar 14 - 20** | **Mid-Way Report Due** |
| 7 | Mar 21 - 27 | Application Design and Implementation |
| 8 | Mar 28 - Apr 3 | Application Design and Implementation |
| 9 | Apr 4 - 10 | Application Design and Implementation |
| 10 | Apr 11 - 17 | Application Design and Implementation |
| 11 | Apr 18 - 24 | Preparation of Final Report, Presentation and Demonstration |
| **12** | **Apr 25 - May 1** | **Project Presentation and Demonstration** |
| **13** | **May 2 - May 4** | **Final Report Due** |

**References**

1. P. Pokorný and K. Stokláska, "Graphics Visualization of Specific Dashboards in Transport Technologies," 2016 Third International Conference on Mathematics and Computers in Sciences and in Industry (MCSI), Chania, Greece, 2016, pp. 203-206, doi: 10.1109/MCSI.2016.045.
2. A. Voisard and B. David, "A database perspective on geospatial data modeling," in IEEE Transactions on Knowledge and Data Engineering, vol. 14, no. 2, pp. 226-243, March-April 2002, doi: 10.1109/69.991714.