6406 VISUALIZATION PROJECT PROPOSAL
VISUALIZING AIRPORT DATASET

Abstract

The project is about implementing an interactive D3 visualization tool for accessing the airport dataset collected from *"http://openflights.org/data.html"*. This dataset consists over 10,000 airports, train stations and ferry terminals operating over the globe. The dataset has Airport ID, Name, City, Country, IATA code, ICAO code, Latitude, Longitude, Altitude, Time zone in hours offset from UTC, DST (Daylight saving Time), Time zone dataset in tz (oslon) format, Type of the airport, Source of data. This data is updated recently on January 2017 and has an online portal for updating new flights. The project will be implemented by fetching data directly from online link which stores the updated data. This will make the visualization more live and updated. This project is based on the International Journal *"Analysis of Airport Data using Hadoop-Hive: A Case Study"*. This journal mainly states the uses of Big data and how Airport data is processed to provide various information using Big data analytics. But the concentration of this project is only on efficiently visualizing data on a very cool and interactive system.

Introduction

Implementing a visualization system for the airport dataset collected live which is updated as of January, 2017. This dataset has various details and each column is described in detail below:

Column	Description
Airport ID	This is a unique identifier of OpenFlights, for a particular airport.
Name	It is the name of the airport.
City	It is the main city name where the airport is located
Country	Country where airport is located
IATA	IATA code which is in 3 letter format, can be Null when unknown
ICAO	ICAO code which is 4 letters, can be Null when unknown
Latitude	Latitude in degrees, which is usually up to 6 decimal digits. Negative values are for South, positive are North.
Longitude	Longitude in degrees, which is usually up to 6 decimal digits. Negative values are for West, positive are East.
Altitude	Height in feet at which the flight travels.
Time zone	Time zone in Hours.
DST	Daylight saving time. Single alphabet assigned for each country
Tz dataset time zone	Time zone in "tz" format.
Type	Type of entry. Eg. "airport" for air terminals and "station" for train stations. But for airports dataset all values will be "airport" by default.
Source	Source of the data entry and for airports dataset its "OurAirports" by default.

This dataset is implemented based on the research journal *"Analysis of Airport Data using Hadoop-Hive: A Case Study".* The concentration on this project is only on the visualization part and the Big data discussed on the paper are not considered as part of this project. This paper also considers other datasets in the website such as Routes and Airlines and how it can be processed in Big data to get more analytics on the same.

Timeline

Part 1

Week 1 – Analysing airport.csv dataset and pre-processing if required.

Week 2 – Figuring the best visualization technique for the proposed dataset.

Week 3 – Implementing the idea.

March Update - The visualization part should be 80% completed.

Part 2

Week 1 – Testing the implemented code and making required corrections.

Week 2 – Implementing the interactive part to the visualization

Week 3 – Testing the entire code and preparing report.

Final Submissions – The complete visualization and Interaction of the entire dataset will be completed.

Future work

The visualization tool can be made more useful when all the connected datasets are coupled with the airport.csv dataset implemented for this project. Also, this project can be connected to dataset directly for making updates on the flights.

References

Research Journal

[1] International Journal of Computer Applications, "*Analysis of Airport Data using Hadoop-Hive: A Case Study* ", 2016, http://research.ijcaonline.org/ncrtit2016/number2/ncrtit201634.pdf

Airport dataset

[2] Github, https://raw.githubusercontent.com/jpatokal/openflights/master/data/airports.dat

[3] Contentshare – "Airport, airline and route data", 2017, http://openflights.org/data.html

[4] Contentshare – "Airport, airline and route data - search and update", http://openflights.org/html/apsearch