

FAA Flight Analysis

The Summary

Big Data systems take in data from many sources, integrate it, correlate it, and then offer a summary of current information using a variety of visualization and reporting mechanisms. Although there are a lot of data sources available, this use-case focuses on publicly available data from the Federal Aviation Administration (FAA). The data will either be live or captured data which shows airport locations and near real-time flight information of various carriers. The real-time aspect provides additional challenges which likely will yield an interesting design.

The Challenge

The data is constantly changing, however summary and answers are needed as soon as possible. While data can be filtered to reduce the number of events, ideally, all data will be stored and processed. Both current and historical analysis is needed. Below are some of the desired goals:

- Find out what is within a geometry with a given time frame. A more advanced query would enable time oriented searches and might generate alerts for incoming new data. Alerting might be out of scope. Given an air corridor, have any flights deviated from it?
- Over time, which airport is the busiest and which airline has the best record for arriving on time. Graph this data as part of a BI tool integration?
- Can you detect airport closures, perhaps as a result of weather activity? Optional, if weather data can be accessed, correlations might be interesting.
- Given all the data, find and merge duplicates based on trajectory and time. Seed the data near duplicate data to test algorithms.
- Predict which plane will land where, based on required runway length, in the event of an emergency at a given time.
- Predict where a given plane might be at a specified time in the future. Indicate where it will land if that's detectable.

Incorporate spatial indexing algorithms, such as Z-Indexing.

The Data

Input Source: FAA KML

Airplane Locations <http://flightwise.com/flighttracking/>

Airport locations

<https://groups.google.com/forum/?fromgroups#!topic/rec.aviation.soaring/EOBD9OO2Qe4> and format thereof: http://www.justsoar.com/public/nasr Apt/nasr_spec/Apt_rf.txt

Airline data: <http://services.faa.gov/docs/tutorial/>

Or <http://stat-computing.org/dataexpo/2009/the-data.html> and

<https://github.com/jseidman/hadoop-R>

Look at Slides from: <http://www.slideshare.net/jeffreybreen/big-data-stepbystep-using-r-hadoop-with-rhadoop-rmr-packag>

The Solution

Leverage: National Day of Hacking: <http://hackforchange.org/>

We need help to create a github site with all the material: Use Case, Design, Data, Implementation for Hadoop. We'll try to get as much done in a day and continue it long term as an open source project.

R2AD will help with documentation, testing, QA, polish, sharing via github or Google project/etc, slides, organization, meetups, etc. We need help with the Hadoop injection (parsing/storage), design, and implementation of analytics to solve some of the interesting challenges.

Skills: Hadoop, NoSQL, Java, KML perhaps