Using Big Data and GIS to Model Aviation Fuel Burn

Gary M. Baker
USDOT Volpe Center

2015 Transportation DataPalooza
June 17, 2015

The National Transportation Systems Center
Advancing transportation innovation for the public good

U.S. Department of Transportation
Office of the Secretary of Transportation
John A. Volpe National Transportation Systems Center

Presentation Outline

- Big Data Processing
- Input Data and Fuel Burn Model
- Analysis

Presentation Outline

- Big Data Processing
- Input Data and Fuel Burn Model
- Analysis

Background: Map Reduce

- 2004 publication from Google
  "MapReduce: Simplified Data Processing on Large Clusters"

- "MapReduce is a programming model and an associated implementation for processing … large data sets."

- **Map** function and **Reduce** function

- Many real world tasks are expressible in this model

- The MapReduce system orchestrates distributed servers, runs various tasks in parallel, manages all communications and data transfers between the various parts of the system, and provides for redundancy and fault tolerance."

http://en.wikipedia.org/wiki/MapReduce

Background: Hadoop

- Apache Hadoop is open-source software used for reliable, scalable, distributed computing.
- Framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. (i.e. Map Reduce)
- Designed to scale up from single servers to thousands of machines, each offering local computation and storage.
- Designed to detect and handle failures at the application layer, so delivering a highly-available service on top of a cluster of computers, each of which may be prone to failures.

http://hadoop.apache.org

Traditional Databases vs. Hadoop

RDBMS

Sports Car:

- refined
- has a lot of features
- accelerates very fast
- pricey
- expensive to maintain

Hadoop

Freight Train:

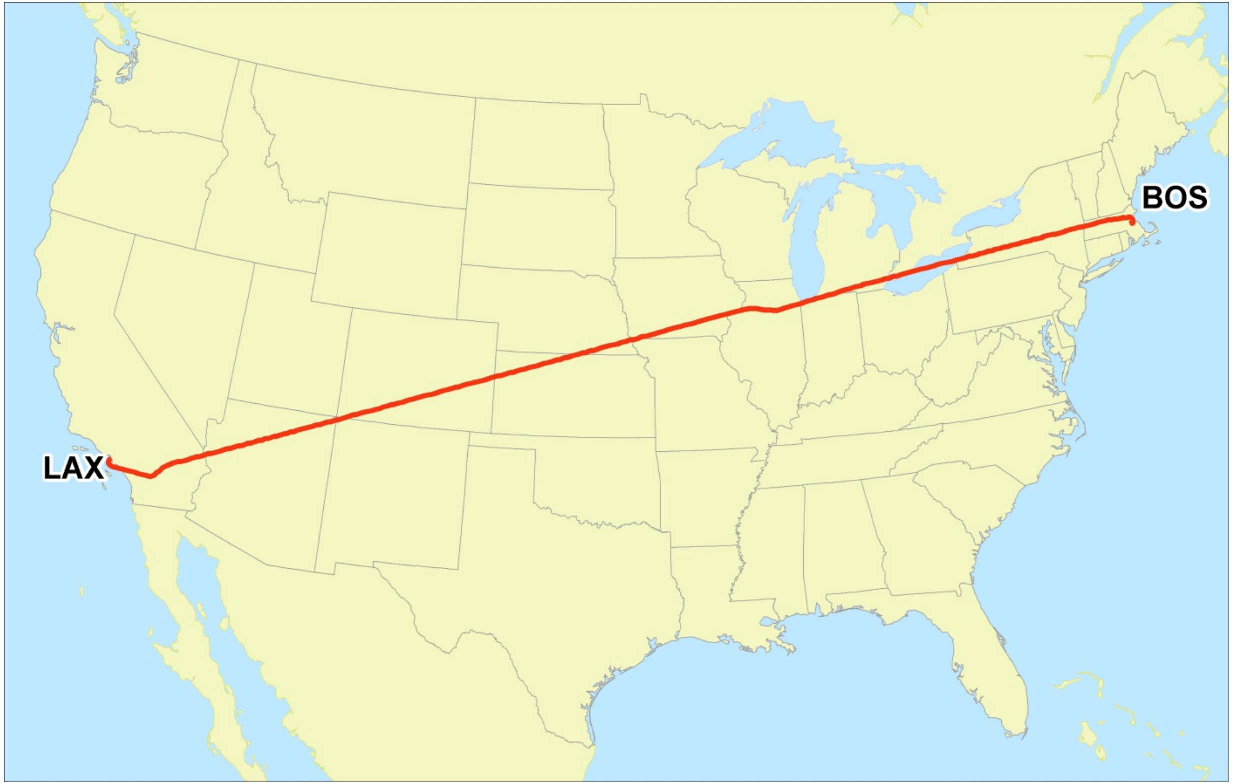
- rough
- missing a lot of "luxury"
- slow to accelerate
- carries almost anything
- moves a lot of stuff very efficiently

From: http://www.basas.com/newsletters/20110705/Presentation/Kerem%20Tomak.ppt

Presentation Outline

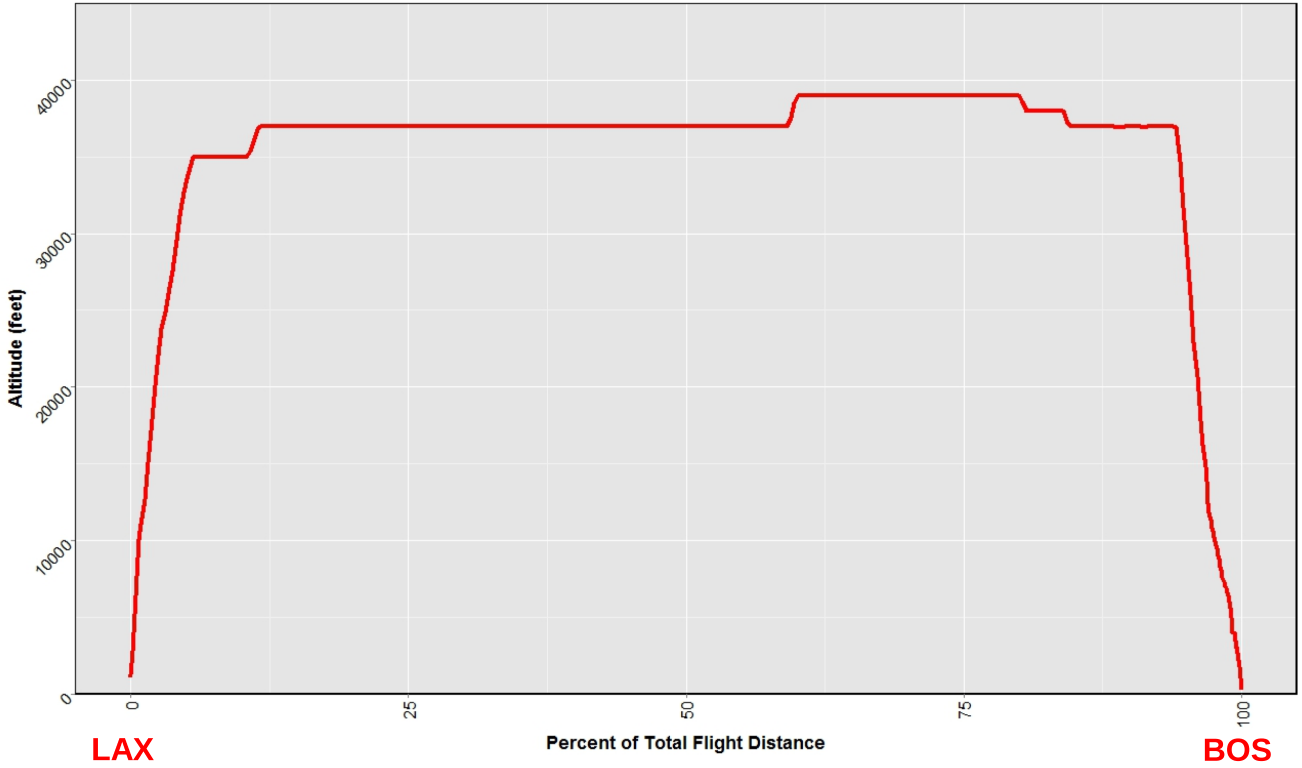
- ❑ Big Data Processing
- ❑ Input Data and Fuel Burn Model
- ❑ Analysis

Input Flight Data: Flight Track



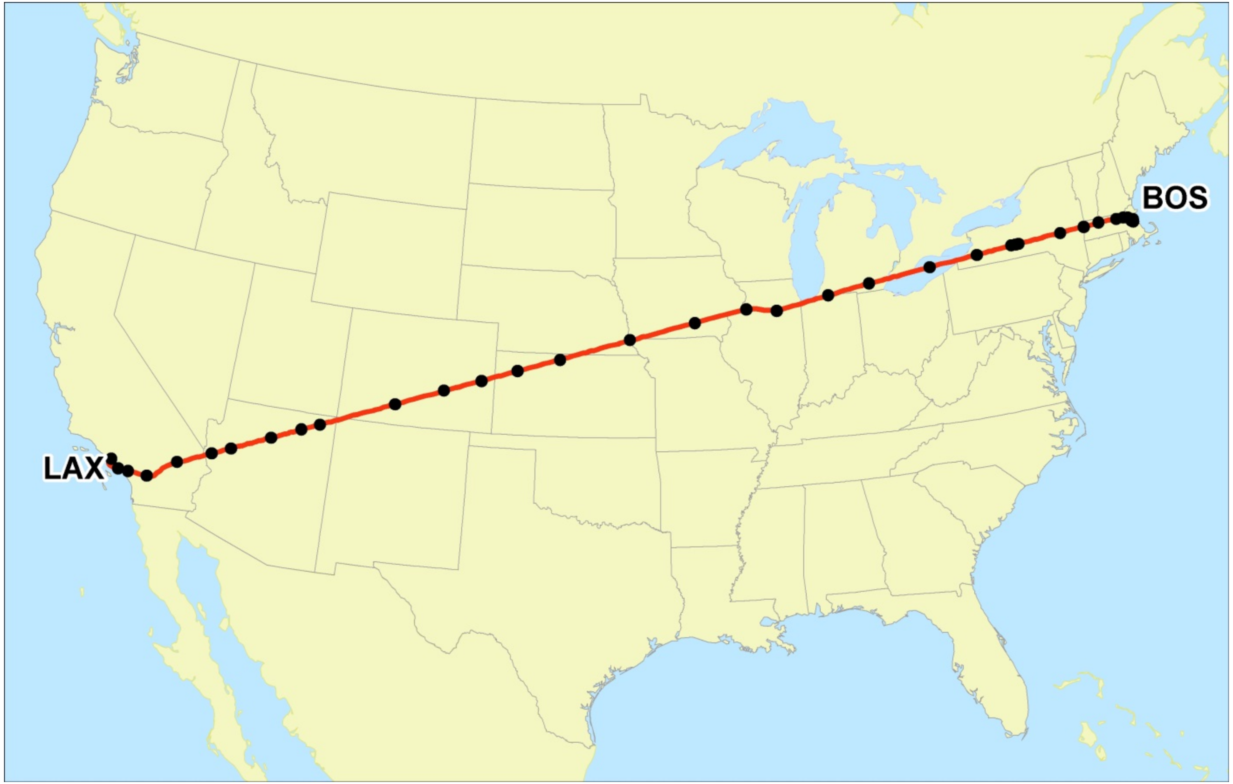
- National Airspace System (NAS)
  - radar based
  - latitude, longitude, altitude, time
- Other parts of the world
  - airways, waypoints
  - great circle

Input Flight Data: Altitude Profile



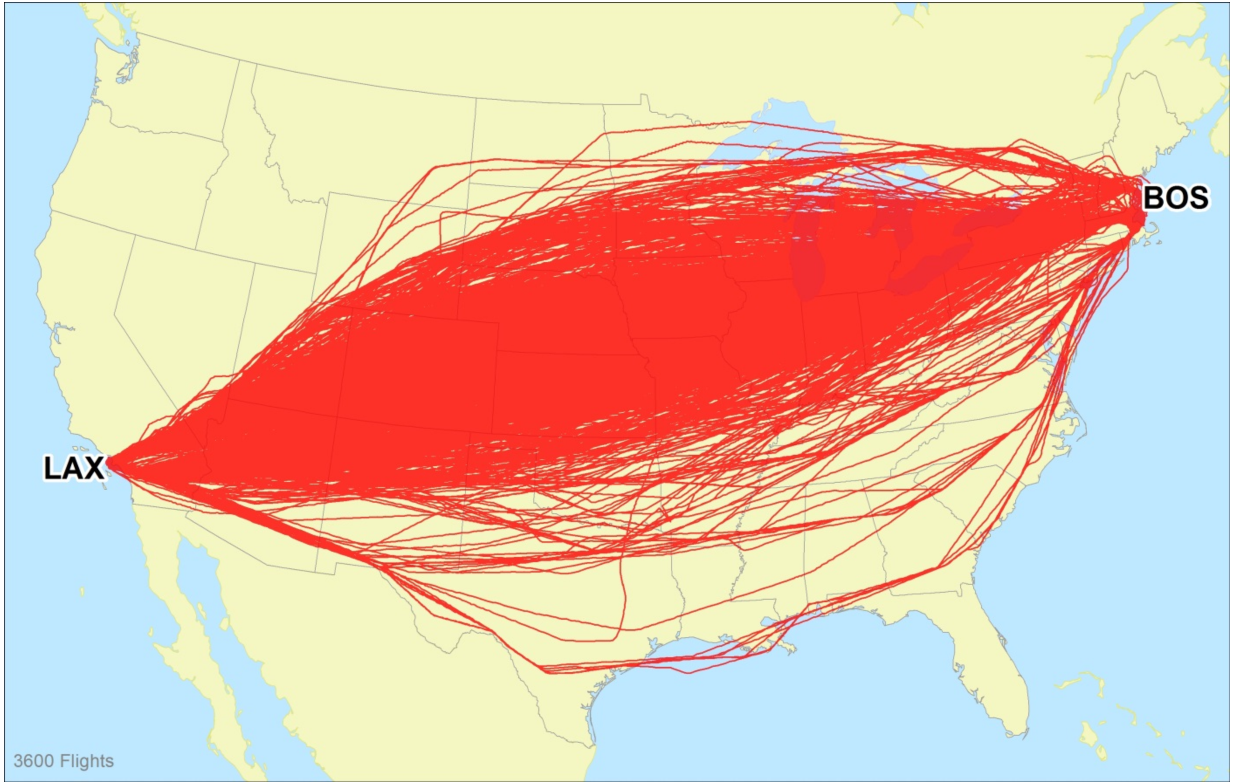
- Altitude varies throughout the flight affecting fuel burn

Fuel Burn Model



- Model outputs many segments for each flight
- Fuel burn is uniform over segment
- Average of over 100 segments per flight

Route Variations

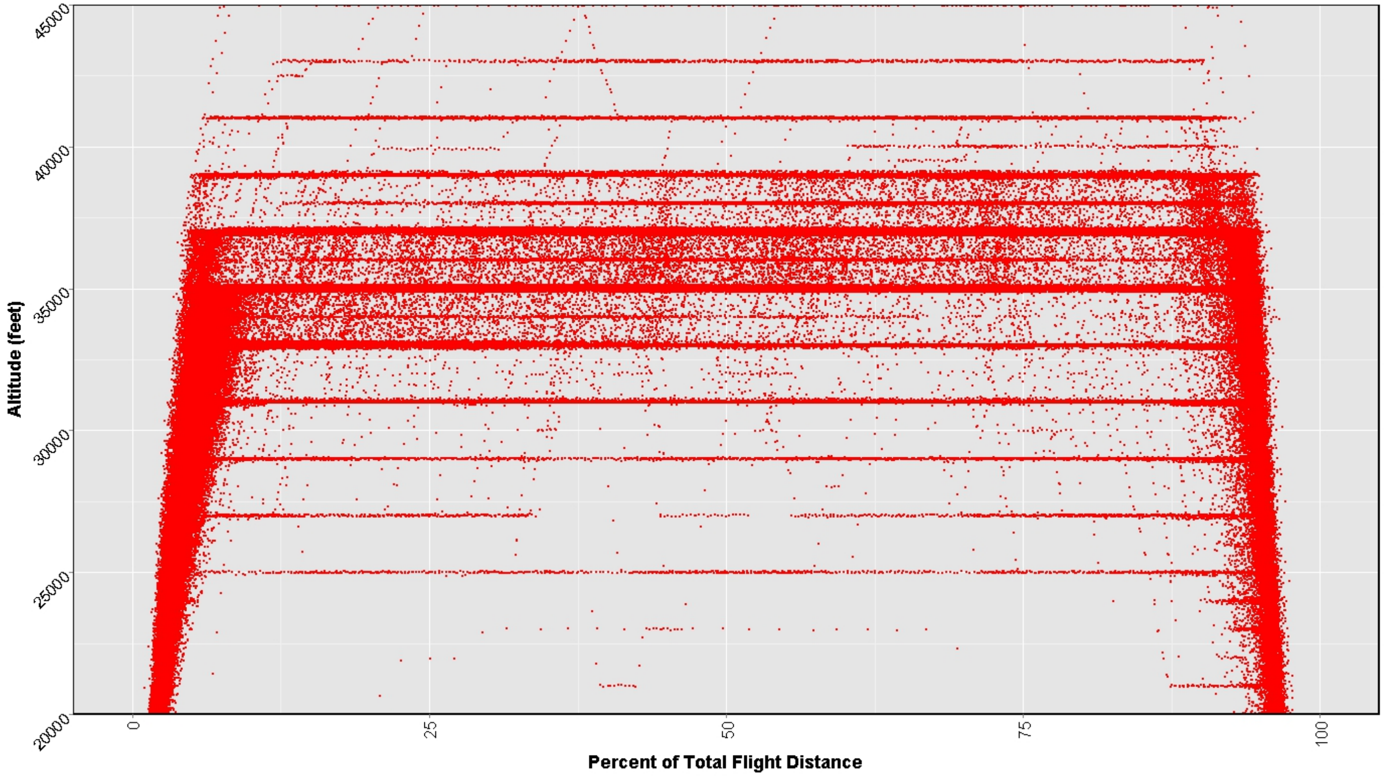


One year of flights from LAX to BOS

- ❑ Winds
- ❑ Weather
- ❑ Traffic
- ❑ Restrictions
- ❑ Procedures

This is only one of over 850,000 origin destination pairs!

Altitude Variations

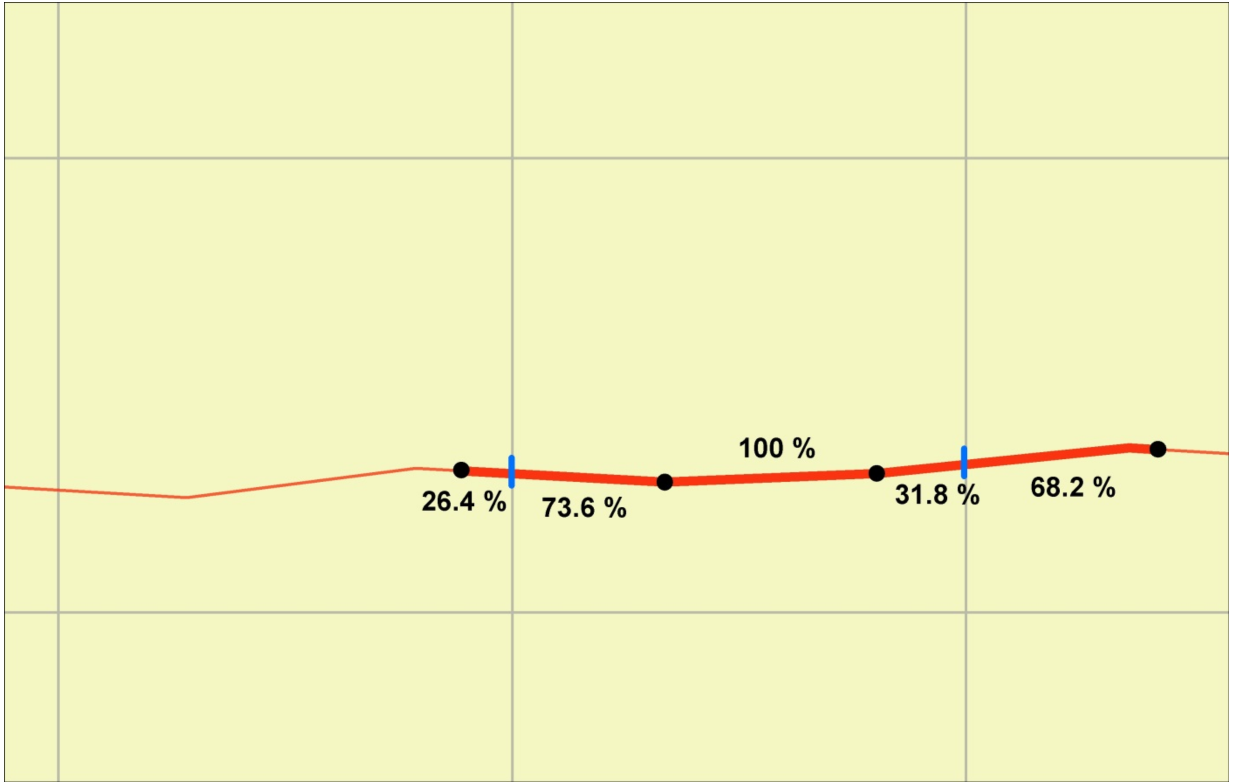


- Winds
- Weather
- Traffic
- Restrictions
- Procedures
- Load
- Equipment

Presentation Outline

- ❑ Big Data Processing
- ❑ Input Data and Fuel Burn Model
- ❑ Analysis

Objective



- Intersect fuel burn segments with 36 km Equal-Area Scalable Earth Grid
- Determine total fuel burn for each grid cell

Sounds Like a Big Data Problem

~ 30 million flights globally per year

*****

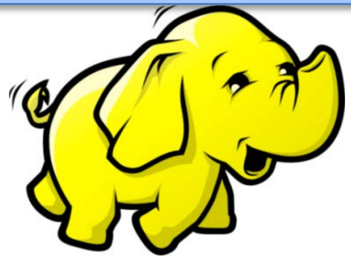
average of over 100 segments per flight

*****

multiple analysis years

*****

different scenarios



We could use traditional statistical approaches to model flight tracks and altitudes but that's not the big data way!

Steps

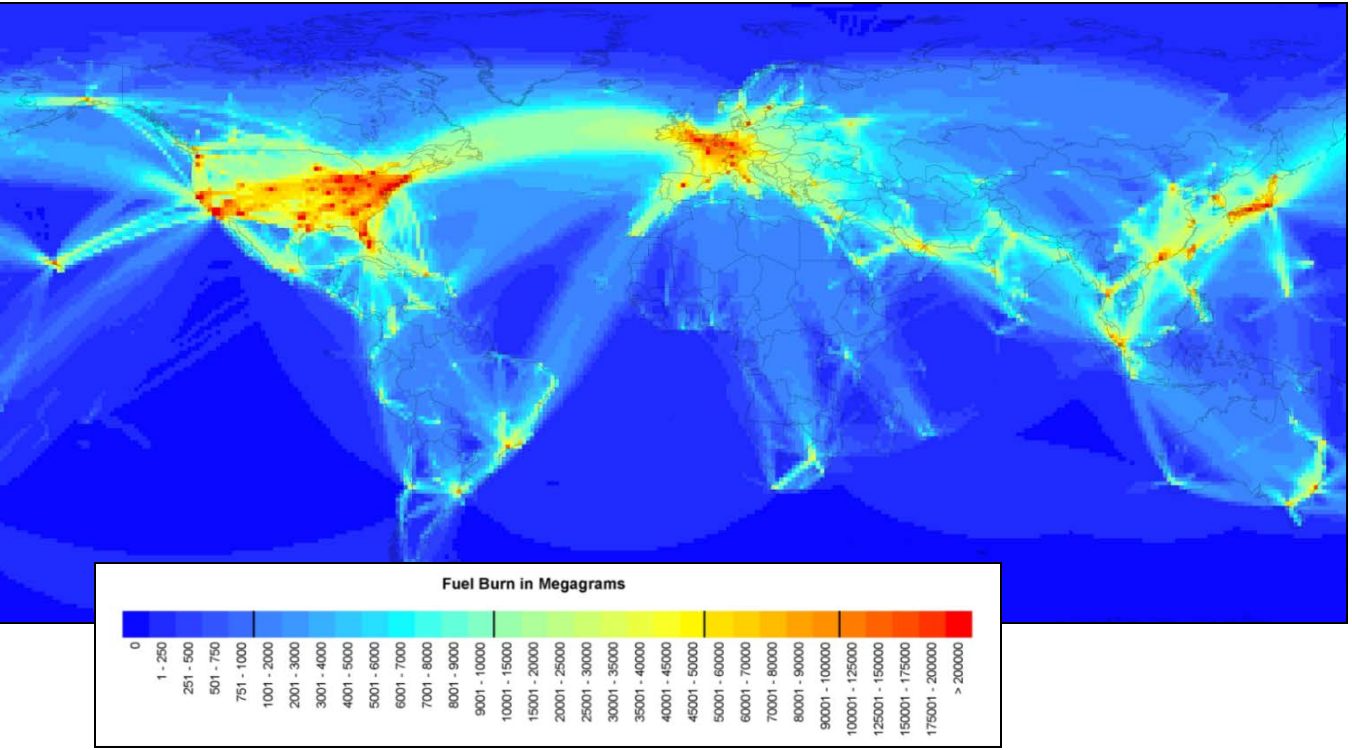
- Preprocess data - reformat, subset, project
- Write and debug program locally (Java)
- Adapt program to MapReduce model and Hadoop environment
- Overcome cloud learning curve
    - connect, execute, debug, automate, etc.
- Upload data to the cloud
- Execute and retrieve results

Results

GRID CELL ID	FUEL BURN
104028	3.0
104029	46489.4
104030	4795.5
104031	985.1
104032	84491.3
104033	622.9
104034	42.9
104035	7736.3
104036	37620
104037	68749.3
104038	12308.8
104039	361.5
104040	9902.7
104041	263496.4
104042	88256.5
104043	14688.3
104044	114664.9
104045	17361.4
104046	1082.5
104047	3.0
…	…

- 36 km EASE-Grid has 391,384 cells. Final result is a text file that contains the total fuel burn for each cell.
- Over 1.5 terabytes/year reduced to a single 5MB result – Data ➡ Information ➡ Knowledge
- 3 billion segments (30 millions flights) can be processed in just over 1 hour on an average 5 computer cluster.
- On a good local server the same was done in about 4 hours.

Results and Further Analysis in GIS



Beyond a Simple Map

- Total fuel burn by region
- annual comparisons
- scenario comparisons
- overlay with other climate datasets
- by aircraft type or class
- altitude slices
- time animations

Thank You