

Real time GPU accelerated analysis of Big Data

Jeff Adie
Principal Systems Engineer

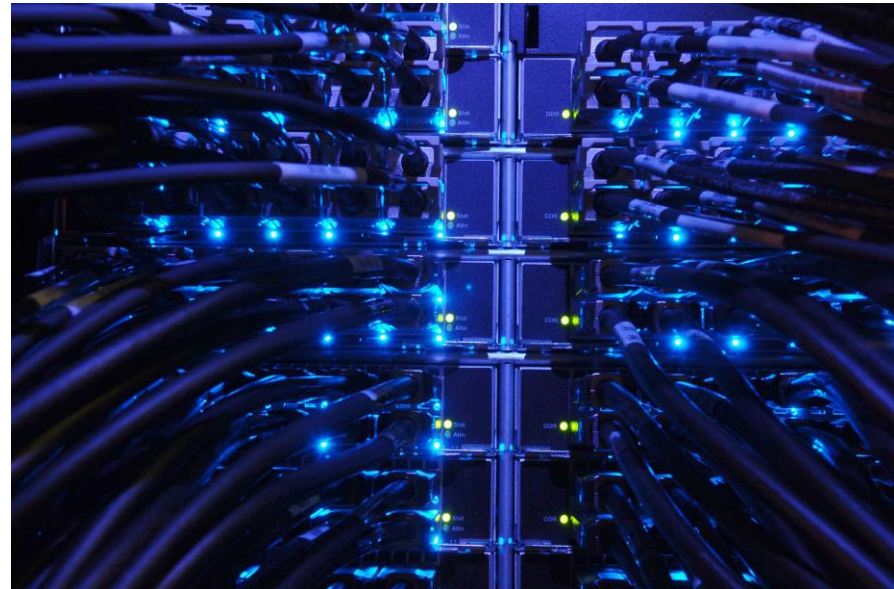
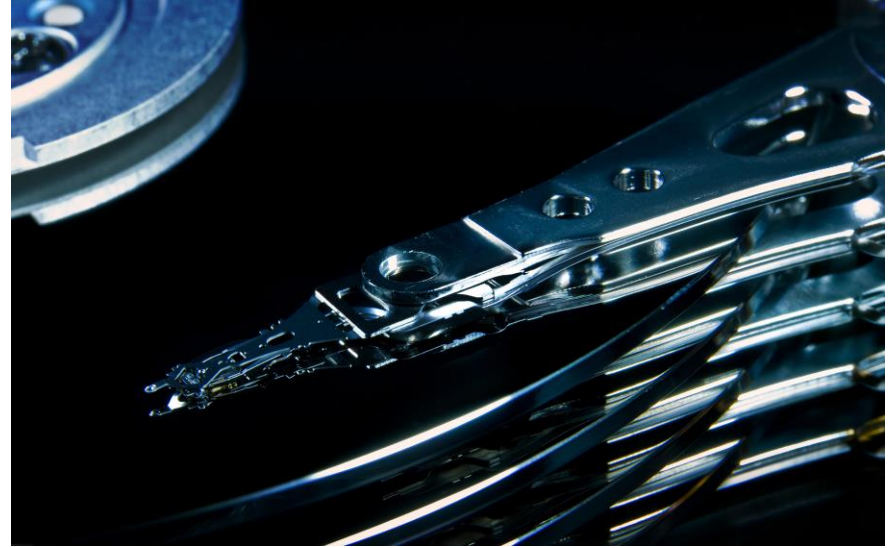
jeffadie@sgi.com

The background of the slide is a photograph of a server rack. The server panels are dark with a hexagonal perforated pattern. A white sticker with the SGI logo is affixed to one of the panels. The logo consists of the lowercase letters 'sgi' in a blue, sans-serif font, with a small green dot above the 'i'.

sgi

Agenda

- Real Time analytics
 - Analysis classes
 - Analysis methods
 - GPUs and Analytics
 - GPUdb and GAIA
- Example cases:
 - USPS
 - Ebay/Paypal
 - INSCOM
- Summary



SGI® and Big Data

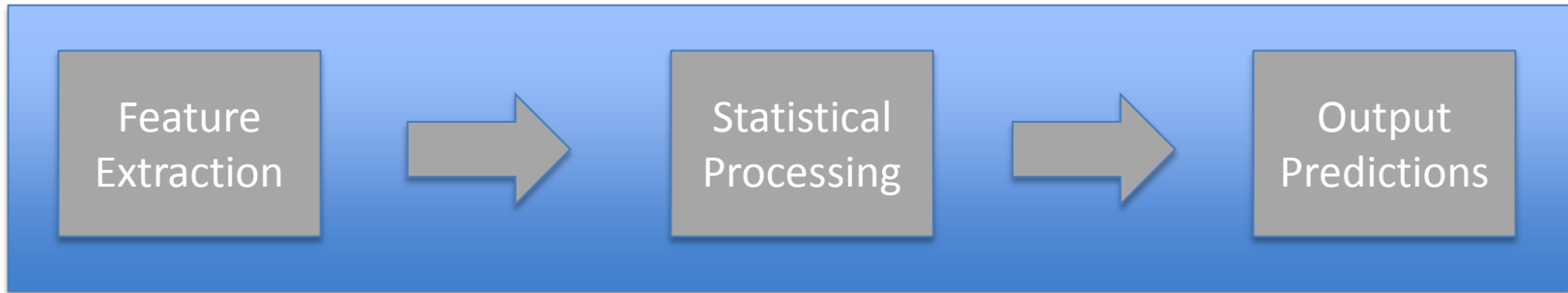


Supercomputing '96

Big Data Analysis classes

- Two fundamental methods:
 - Object analysis
 - Relationship analysis
- Object Analysis – “needle in a haystack”
 - Problem can be easily decomposed
 - Ideal for map-reduce or hadoop type analysis
- Relationship analysis – “relationships between entities”
 - Not easily decomposed
 - Interested in the links rather than the objects themselves
 - Ideal for large-shared memory/graph analysis

Big Data Analytics



- Feature Extraction
 - Conform data to input requirements
- Statistical Processing
 - Analysis
- Output Predictions
 - Derive insight from results

Big Data Analytics

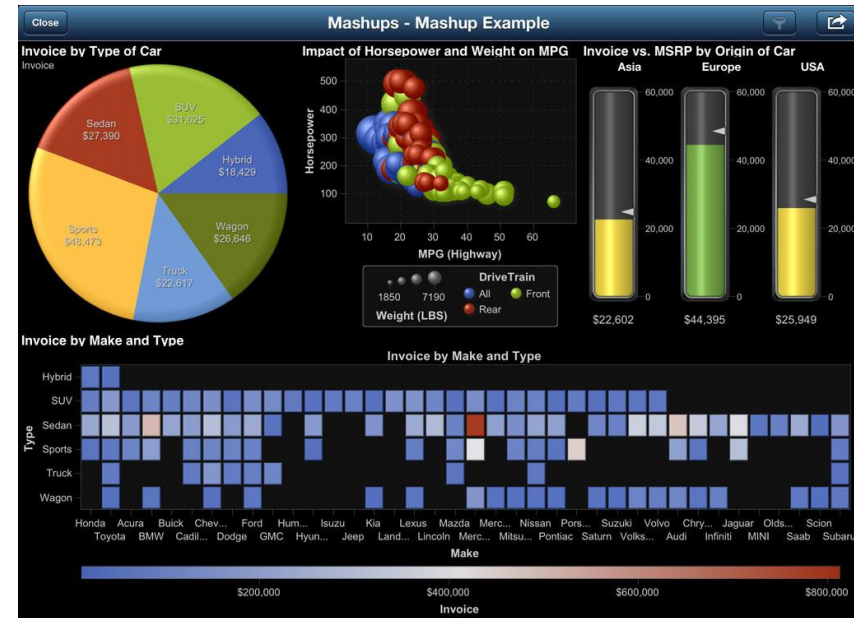
- Many analytics techniques can be framed as convex
 - An optimisation problem where the objective function is convex
 - Exceptions include a-priori and some graph mining algorithms
- In data analysis, convex problems are attractive
 - Local solutions are always globally optimal
 - Problem definition can be decoupled from the solver
 - Many well studied algorithms exist
 - Generally linearly separable

Big Data Analytics

- Most popular algorithms for convex problems are gradient methods
 - Conjugate Gradient
 - Newton Method
 - Incremental Gradient Descent
- These are all ideally suited to GPU processing:
 - Fast Gradient Gradients with Multiple GPUs – Cevahir, et al, ICCS 2009
 - Accelerating the CG method with CUDA – Matt Pennybaker, U of Arizona
 - A GPU framework for solving systems of linear equations, Kruger et al, GPU Gems 2

Real Time Analytics

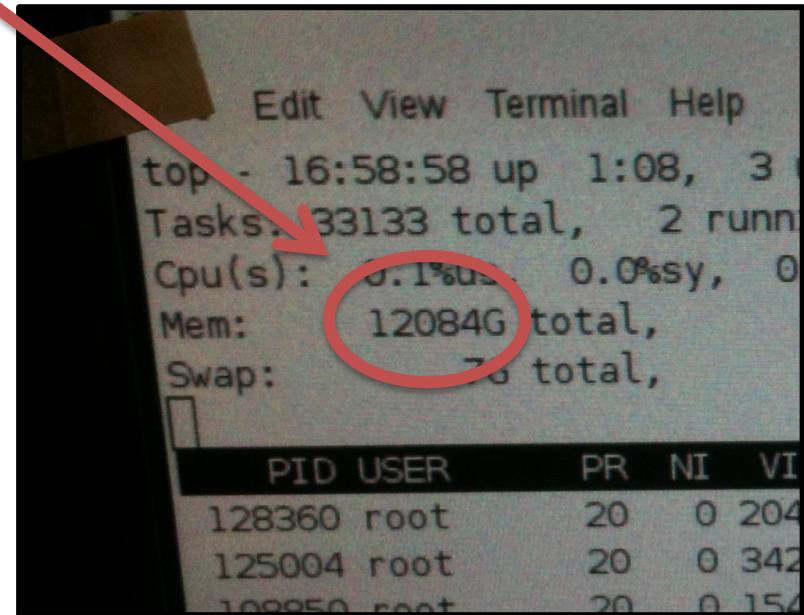
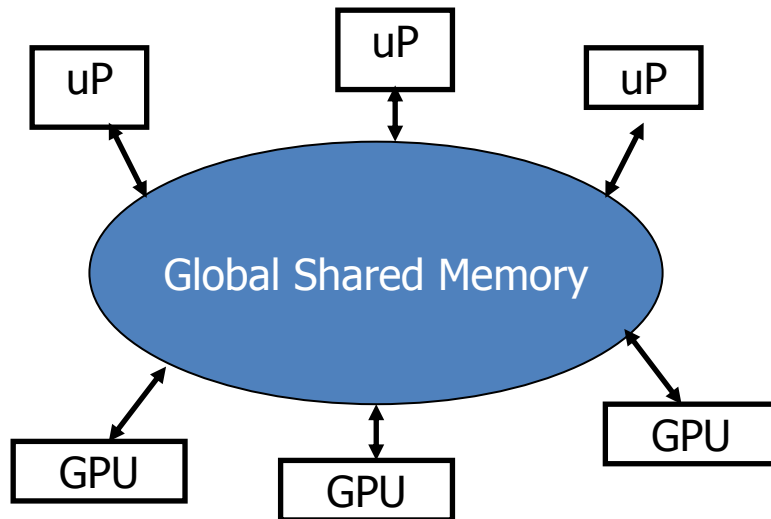
- Real time analysis provides:
 - Fast insight
 - Interactive what-if?
 - Interactive steering
 - Up to date information
 - Support instant yes/no go/no go
 - Online fraud detection



- Real Time has special requirements
 - Data must be immediately available & accessible
 - Processing time must be very fast
- Need a data-centric approach

Data Centric Analysis

- Move the compute, not the data!
 - Surround the data with compute elements (CPUs and GPUs)
- Disks are too slow => In-memory databases
 - Need lots of RAM

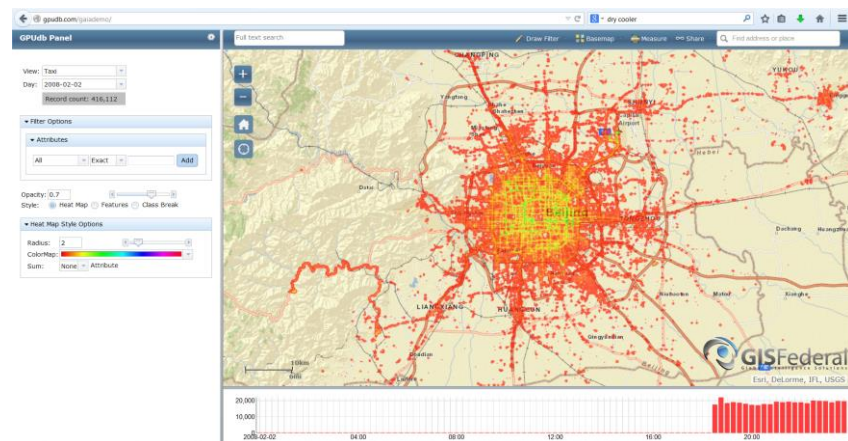


In memory data

- Accessing data on disk (milliseconds)
- Accessing data in flash (microseconds)
- Accessing data in memory (nanoseconds)
- Many methods:
 - In memory databases (memSQL, TimesTen, HANA)
 - In memory tables (most DBMS support)
 - Directly into shared memory (/dev/shm)
 - GPUdb – GPU Accelerated Database

GPUdb

- A scalable distributed database for many core devices
- SQL like query language
- Designed for big data 10-100 million+ rows
- Not to be confused with gpudb by Yuan Yuan @ code.google.com/gpudb



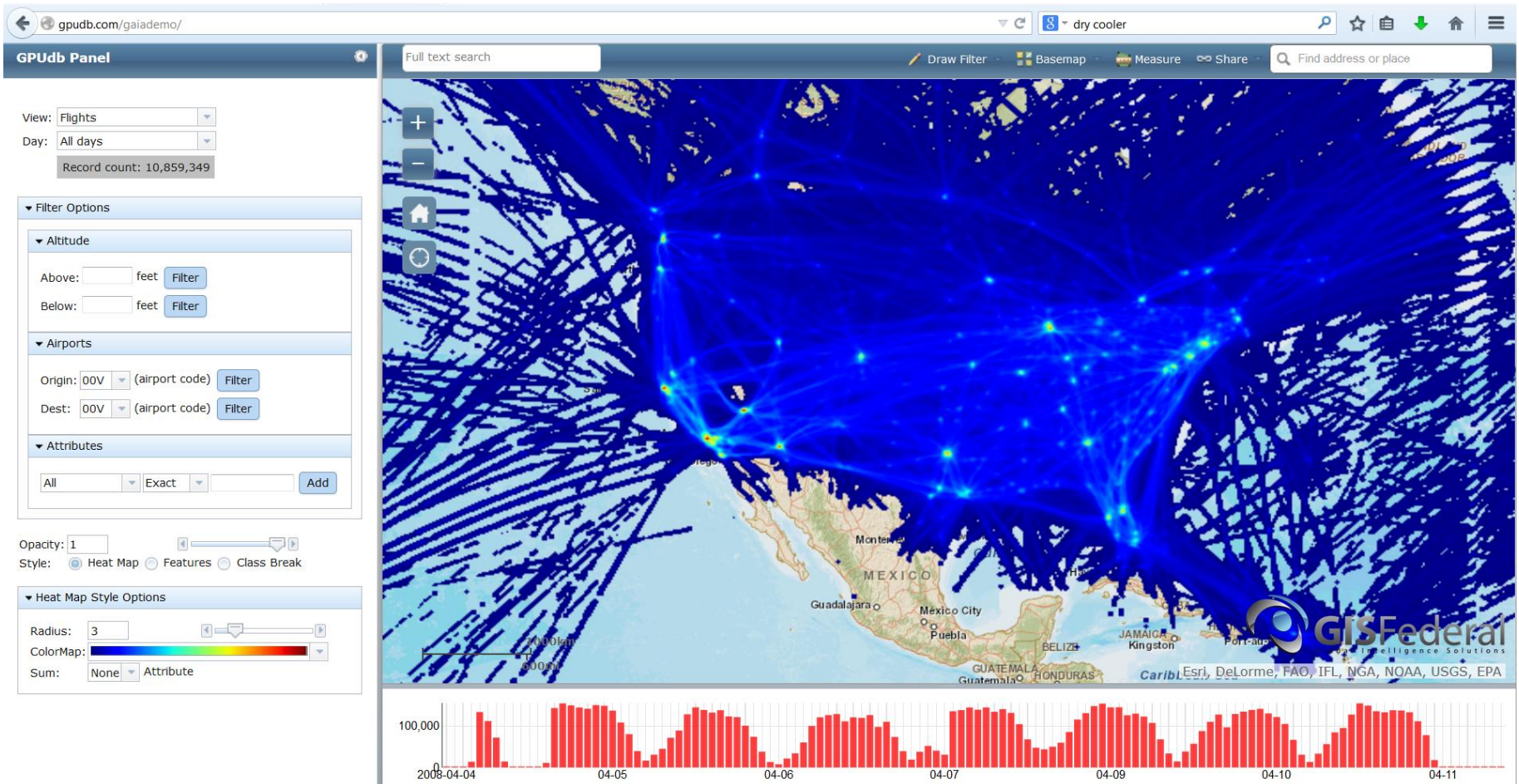
GAIA



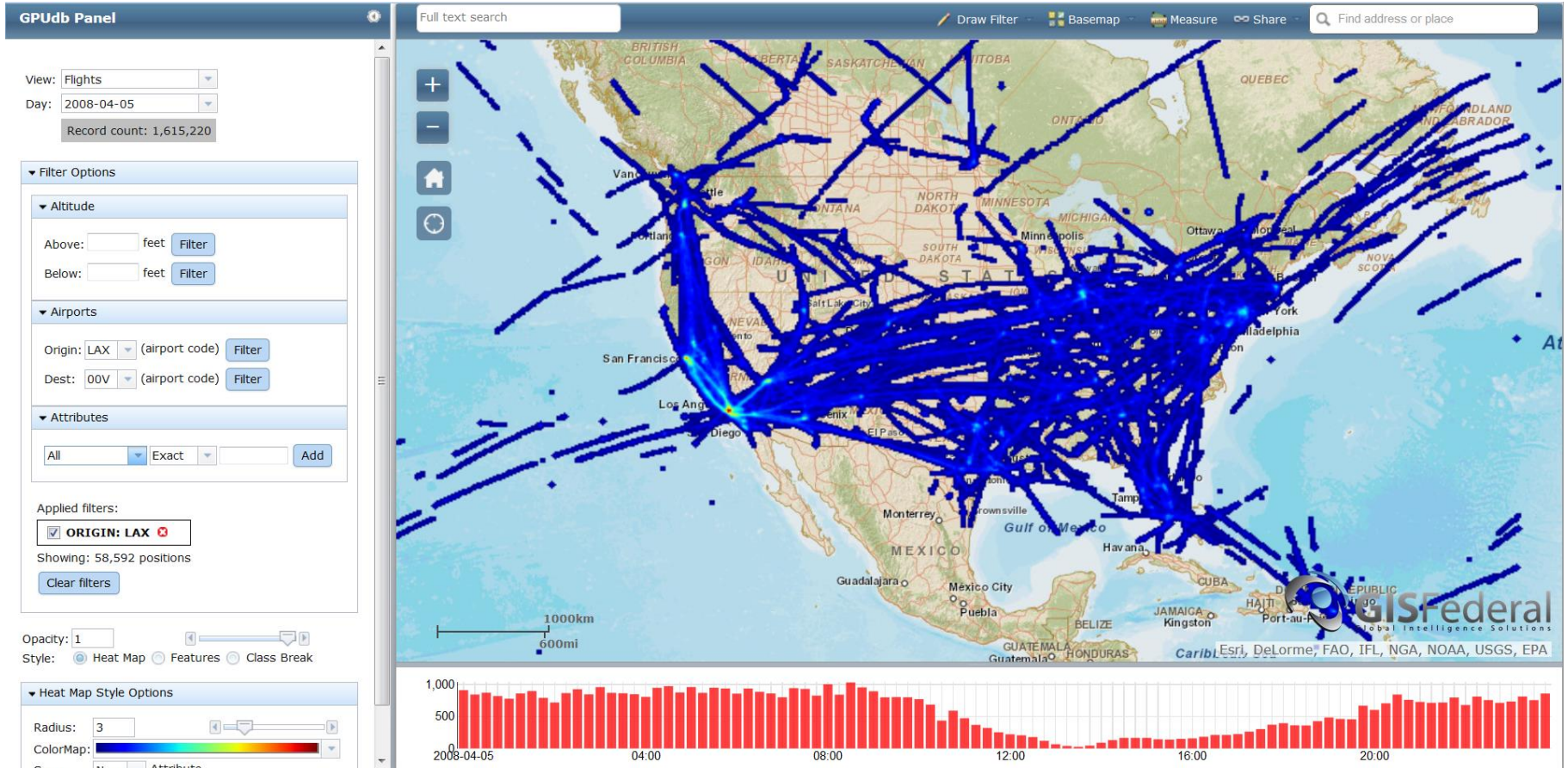
- A comprehensive database that incorporates numerous complex and distinct sources of information, which can be quickly sorted and displayed using easily understood visualization tools.
- Built on top of GPUdb
- Created by Global Intelligence Solutions (GIS) Federal



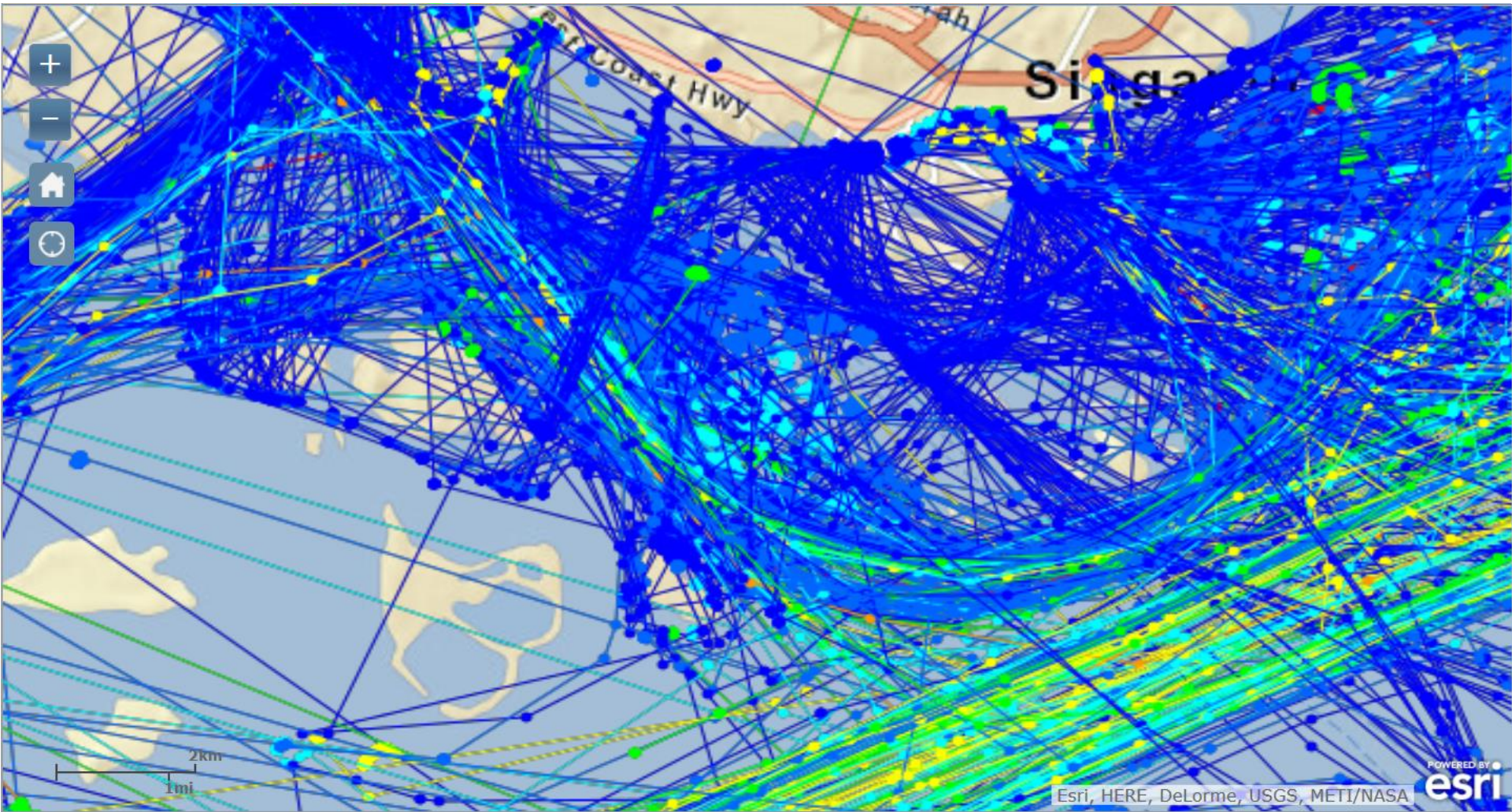
GAIA – 10 Million Flights



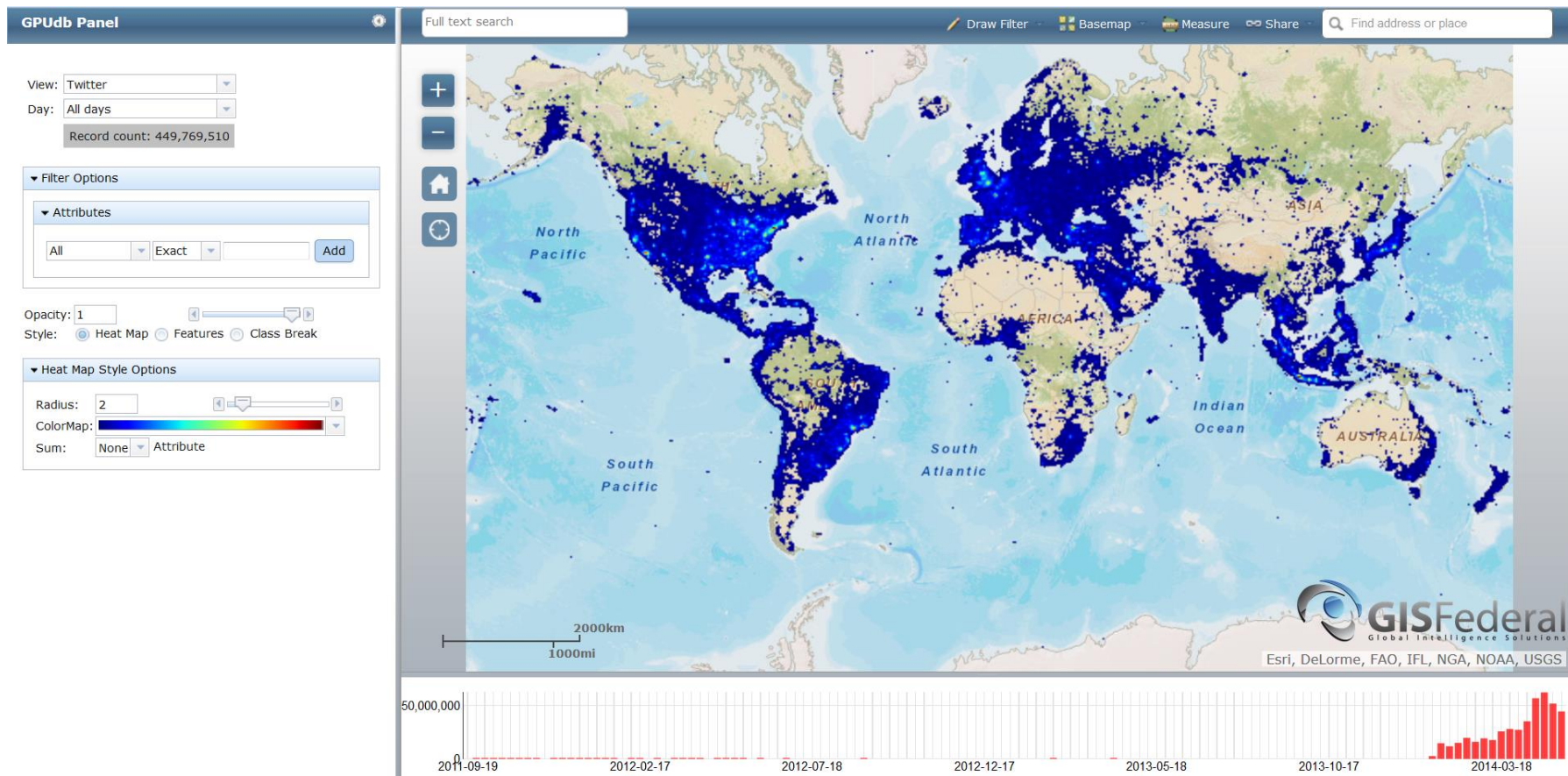
GAIA – From LAX only



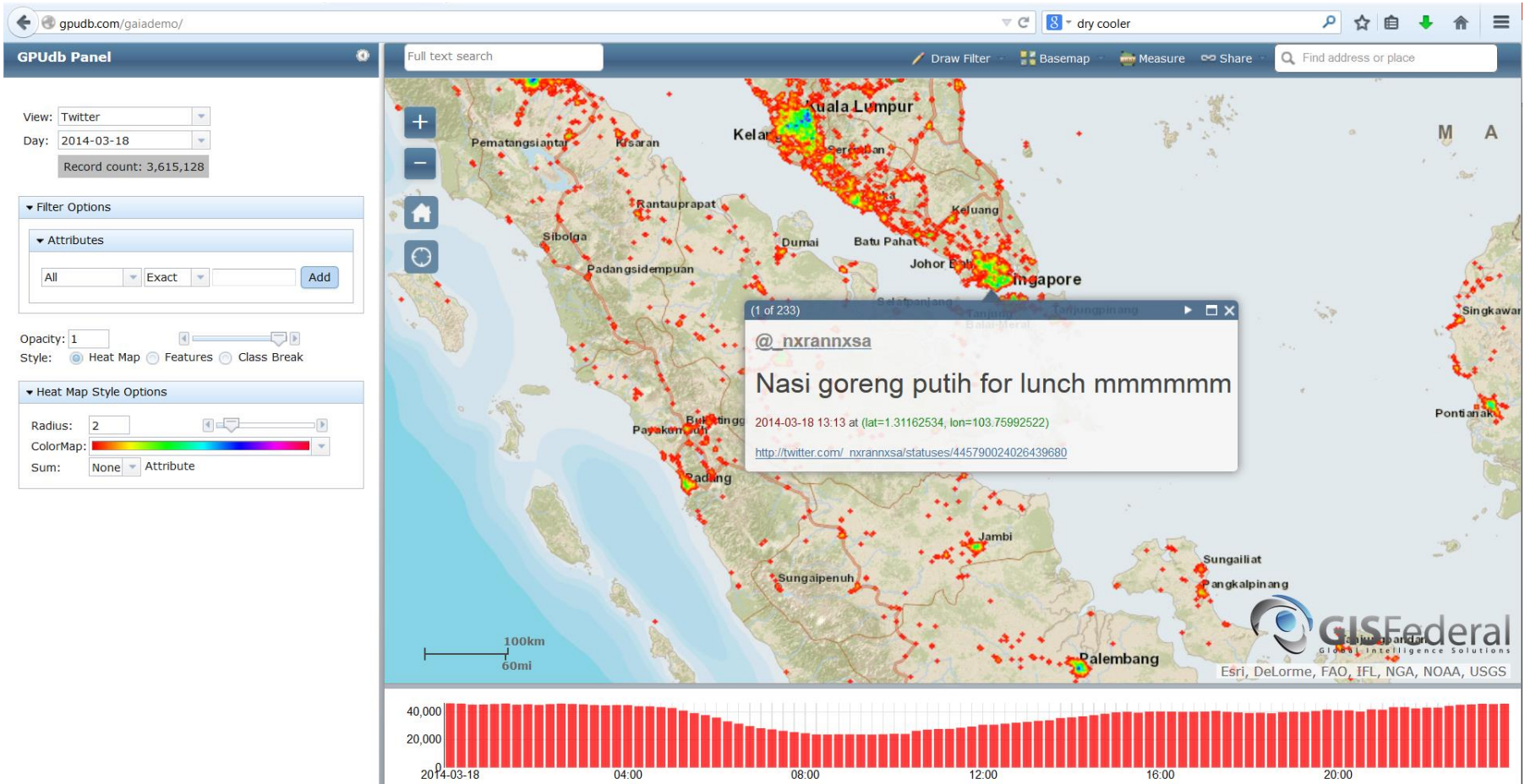
GAIA – Shipping



GAIA – 500 Million Tweets



GAIA – SEA Tweets



Examples



Real Time Fraud Detection @ USPS

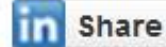
Home » Big Data » HPC » USPS Leverages Big Data To Fight Fraud

USPS Leverages Big Data To Fight Fraud

By: John Rath
May 1st, 2013



47



64



When you think “big data,” you probably don’t think of the **United States Postal Service** (USPS). As it processes more than 528 million pieces of mail each day, the USPS has become an active participant in the big data revolution, and operates one of the most powerful non-classified supercomputing databases in the world.

S. Ryan Quick – Principal Architect, PayPal



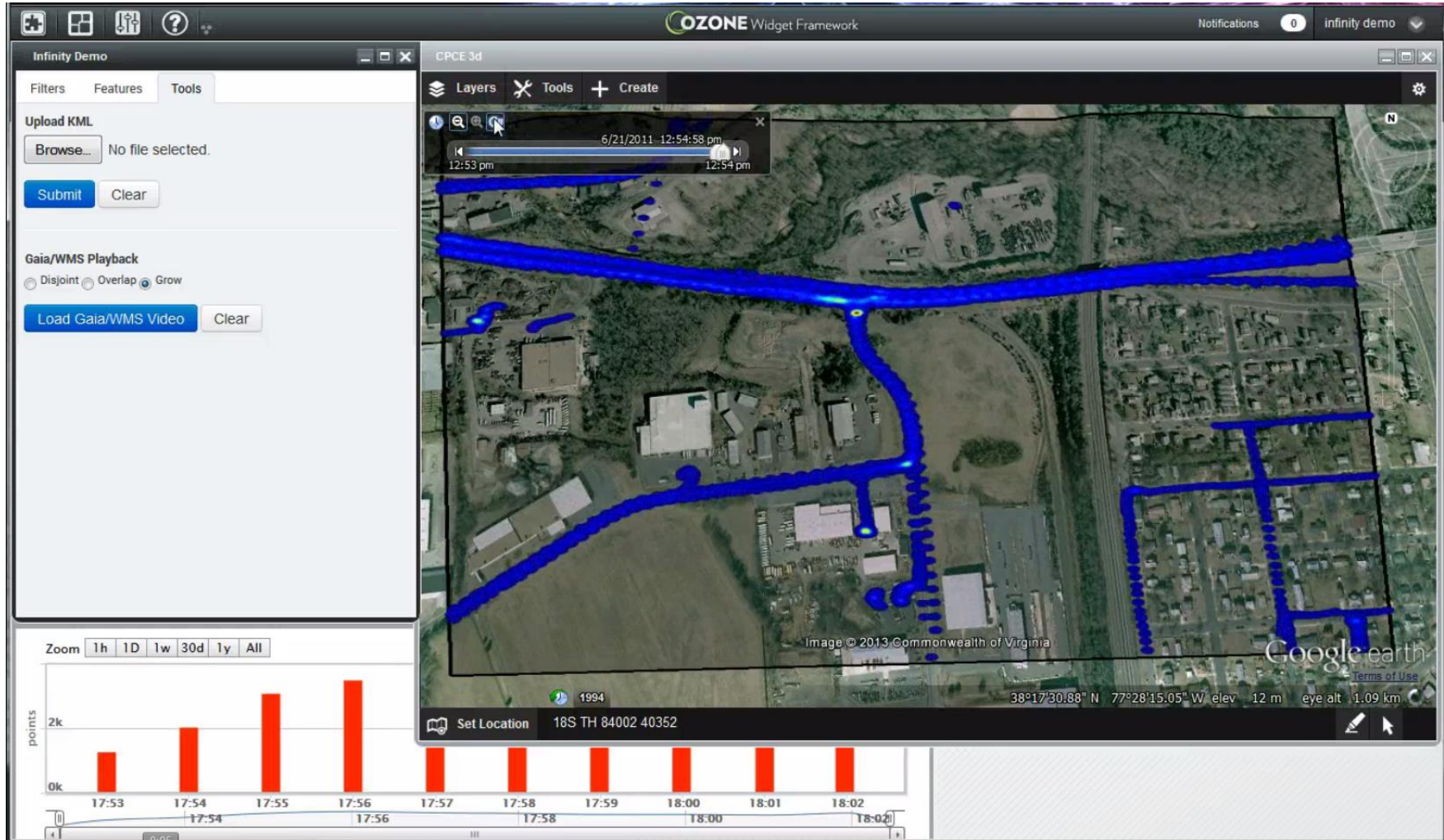
S. Ryan Quick

Principal Architect, PayPal

INSCOM GAIA System

- Partnership between SGI, GIS Federal and NVIDIA
- Customer : U.S. Army INtelligence and Security COMmand (INSCOM)
- Custom system for providing soldiers with real-time actionable insight:
 - Quickly and clearly recognise threats
 - Pinpoint hazards along a route in real time to allow military personnel the opportunity to hange their course of action.

INSCOM GAIA



INSCOM GAIA System

- 10 TB in-core database
- UV2000 SMP system with 16 NVIDIA K20X GPUs
- 2,048 CPU cores
- Live ingest from multiple sensors
- Sub second response for geospatial calculations



Summary

- Big Data provides both substantial challenges and opportunities
- GPU accelerated Data Analytics holds much potential for many typical problems
- Big Data feeds which require immediate filtering/pre-processing can take real advantage of GPU offload
- Data movement, as always, is an important issue

sgi