

Emerging Trends

Lesson 10

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON



Lecture Overview

Distributed Processing

- Technologies for Distributed Processing
- Scaling with Spark Ecosystem

Quantum Computing

- Implications for Machine Learning

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Distributed Processing

Lesson 10 – Section 1

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

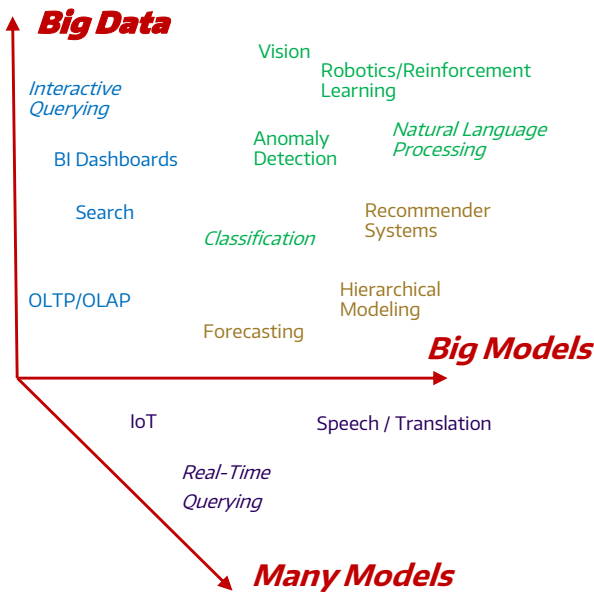


Why do we need distributed processing?

- Large data can come from many sources, be stored in many formats and be stored in geographically disparate locations
- Transferring it to a central location for processing can be expensive, time consuming and potentially impossible

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Dimensions of Scalability – Use Cases



- Scalability means and requires different things depending on the use case
- Enterprise Reporting with Big Data
 - Fault tolerance
 - Replicability
 - Data Compliance/Security
- Natural Language Translation
 - Low Latency
 - Large memory
 - Efficient linear algebra libraries
- Self-driving cars / mobile assistants
 - Compressed algorithms
 - Little/no latency

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON



Case Study: Open Hydrologic Data Interoperability

National Drought Interoperability Project



UNIVERSITY of WASHINGTON

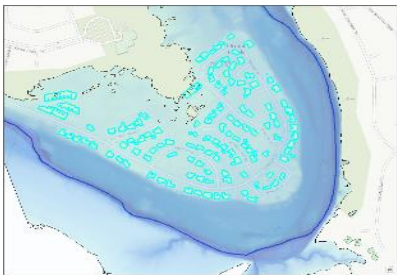
Halloween Flood, Onion Creek, Austin, Texas, October 2013



A Massive Flash Flood Occurred at 5:00a



Upstream watershed



A stream gauge indicated a problem, but then it stopped working...

5 people drowned and 700 homes were flooded

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

May 27-29th, 2015

48 people died in the Oklahoma and Texas Floods
Including two first responders

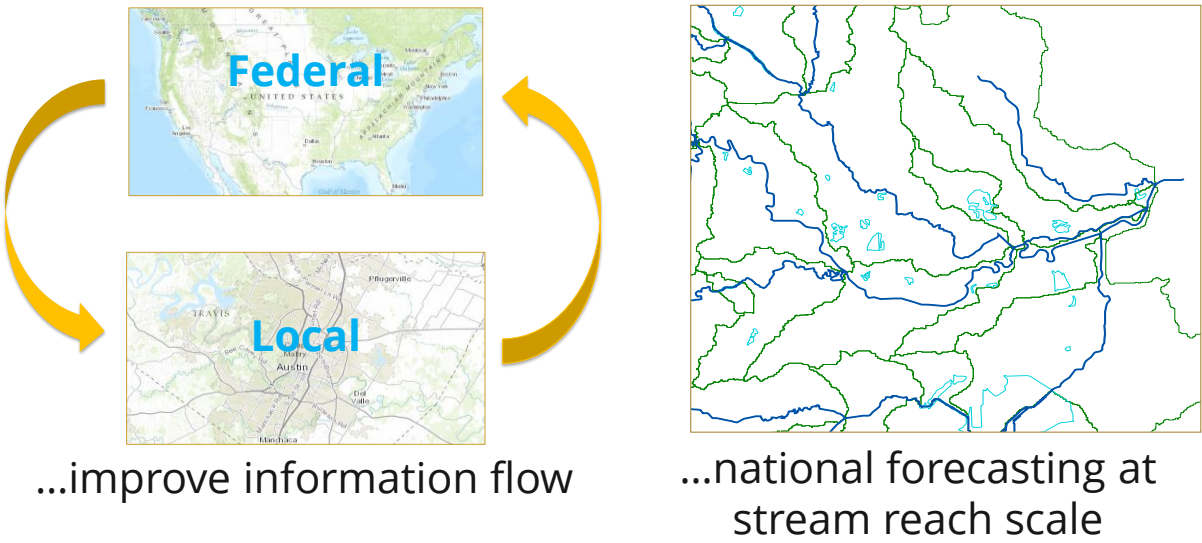


Deputy Jessica Hollis

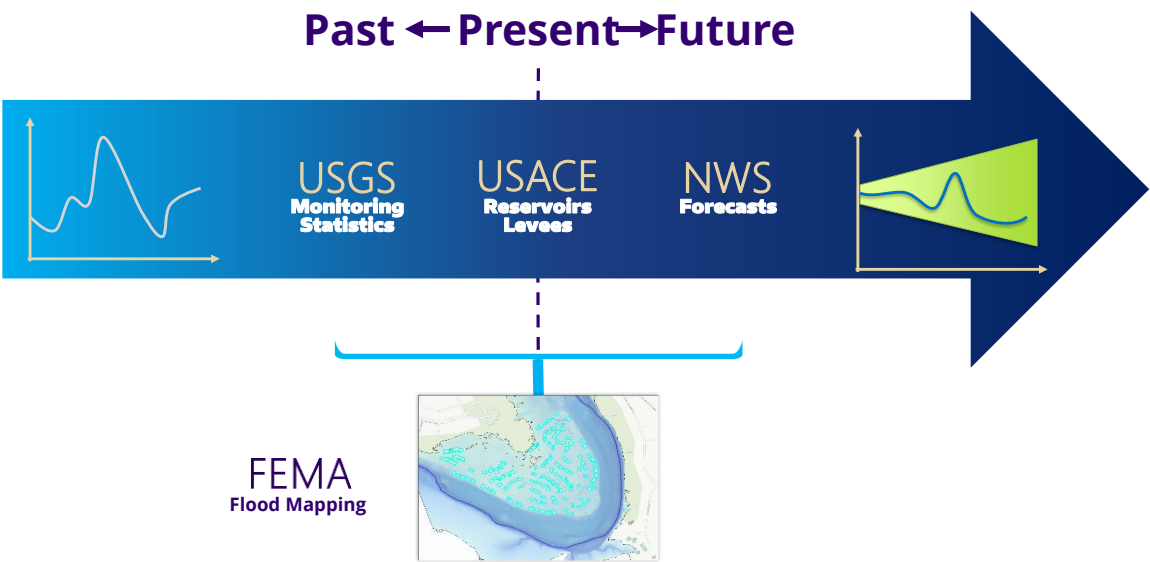


Capt. Jason Farley

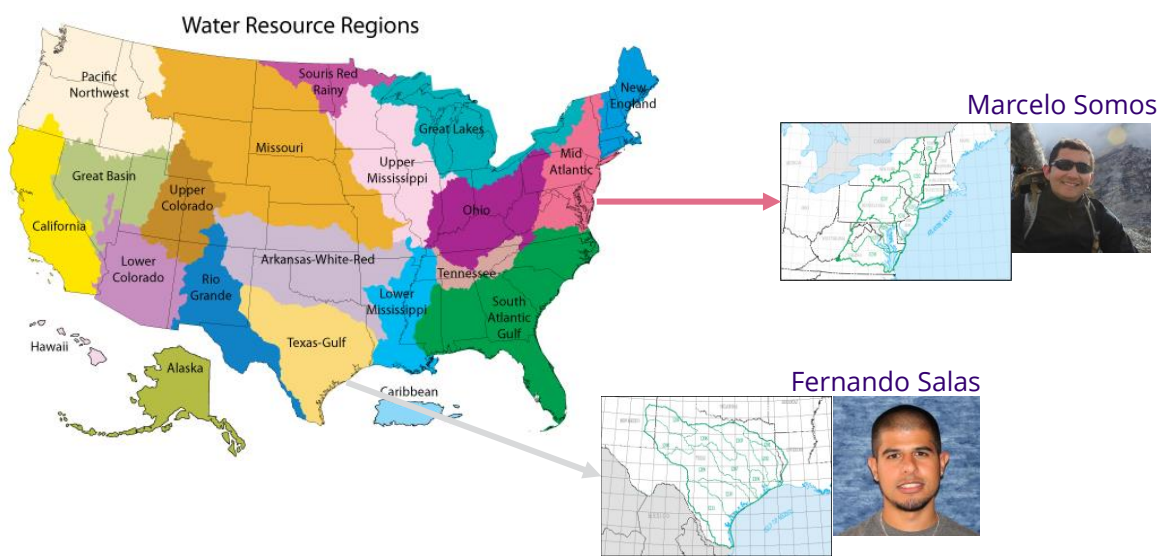
Close the gap between national flood forecasting and emergency response



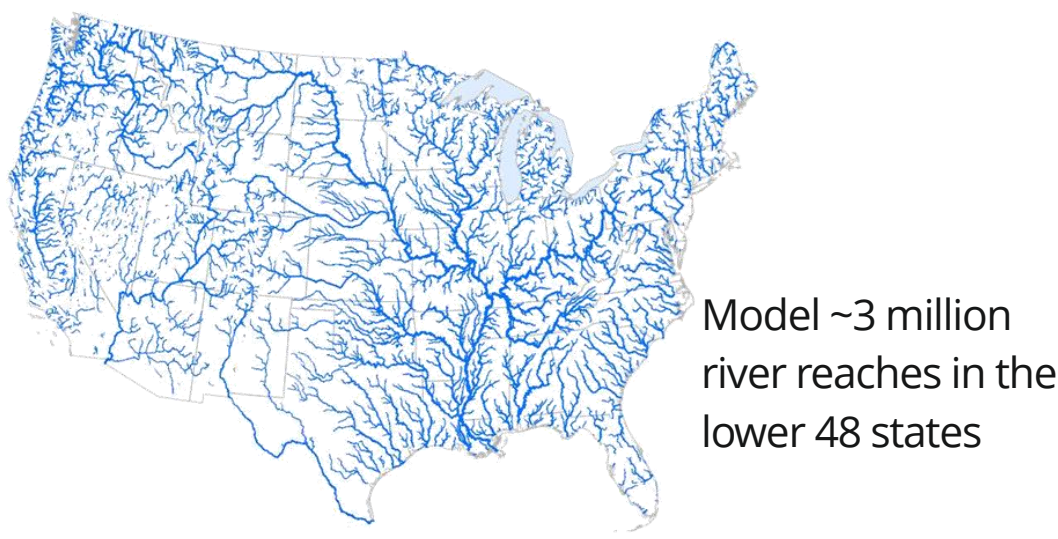
Open Government Data Sources



Where to start: Pilot with 2 Water Regions

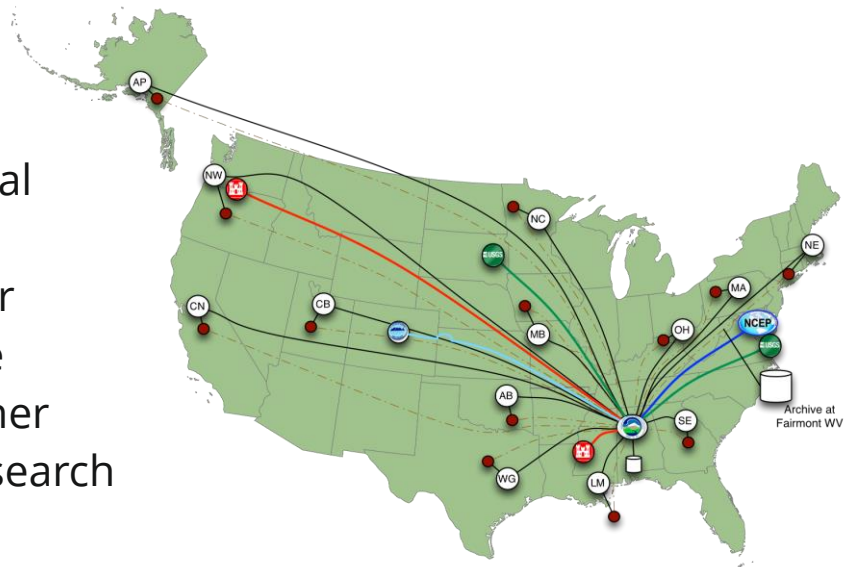


Next step... Scale to the US



National Water Center, Tuscaloosa, AL

A hub for national open water data infrastructure for reprocessing the “products” of other agencies and research institutes



Case for Distributed Processing

Data held by different organizations

National Water Center can share information:

- >National Oceanic and Atmospheric Administration (NOAA)
- >Army Corps of Engineers
- >United States Geological Survey (USGS)

