

# **Machine Learning Techniques**

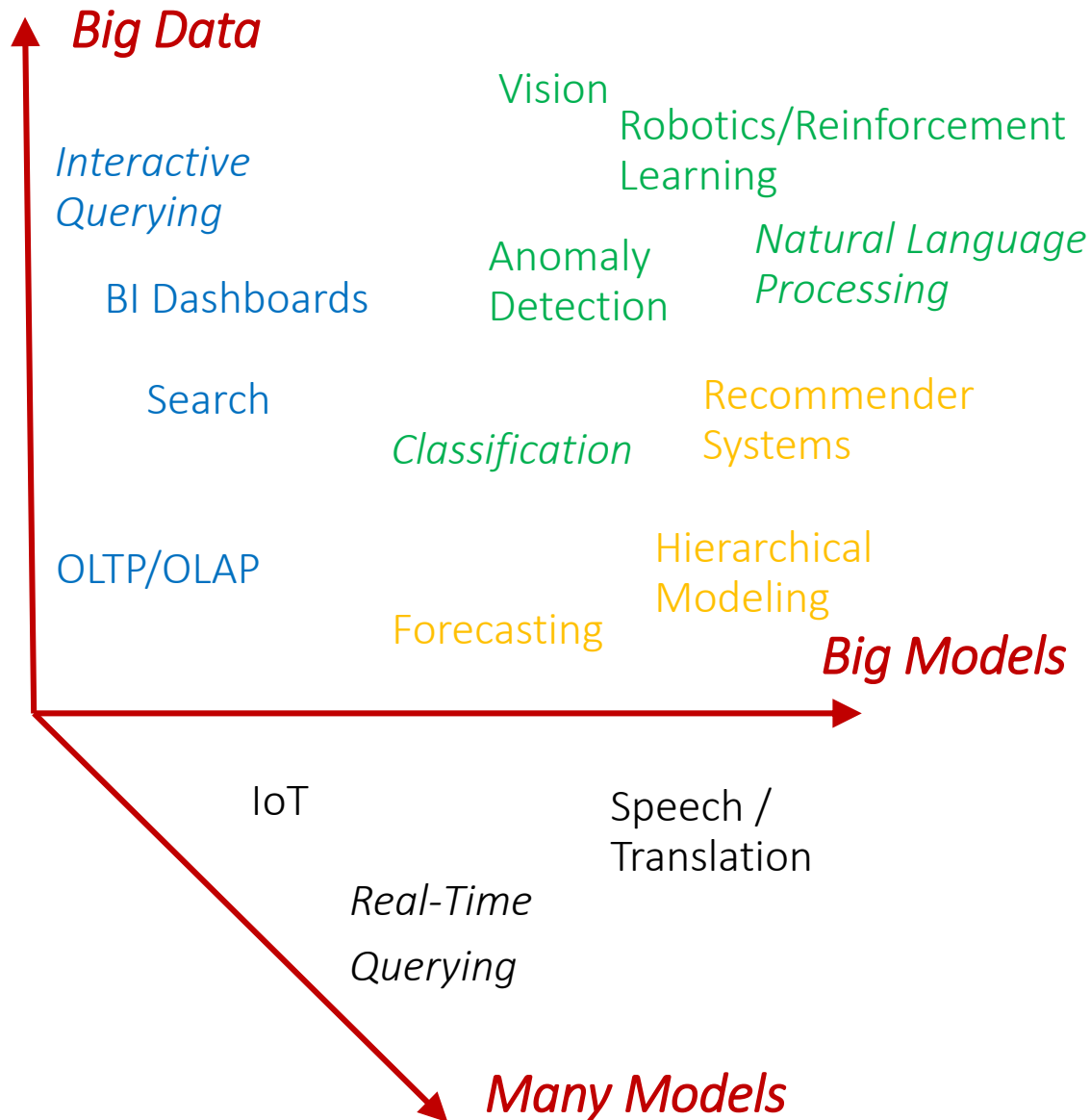
## **DATASCI 420**

Lesson 01-3: Machine Learning Use Case

# 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

# 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

# Case Study: Open hydrological Data Interoperability



# National Drought Interoperability Project

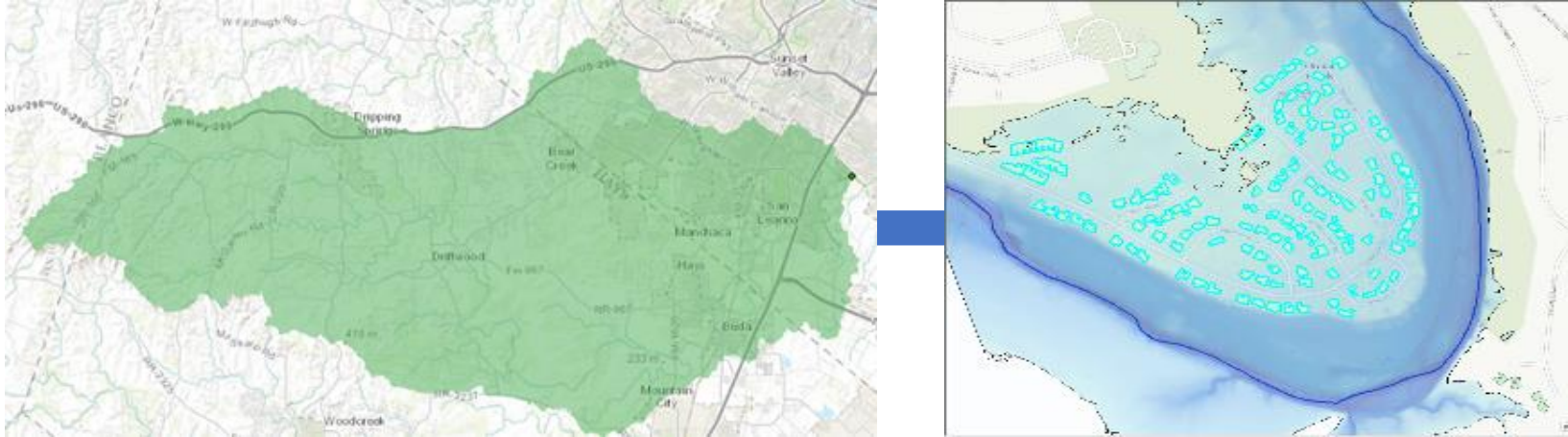




# Halloween Flood, Onion Creek, Austin, Texas, October 2013



# 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

May 27-29<sup>th</sup>, 2015 48 people died in the Oklahoma and Texas Floods Including two first responders



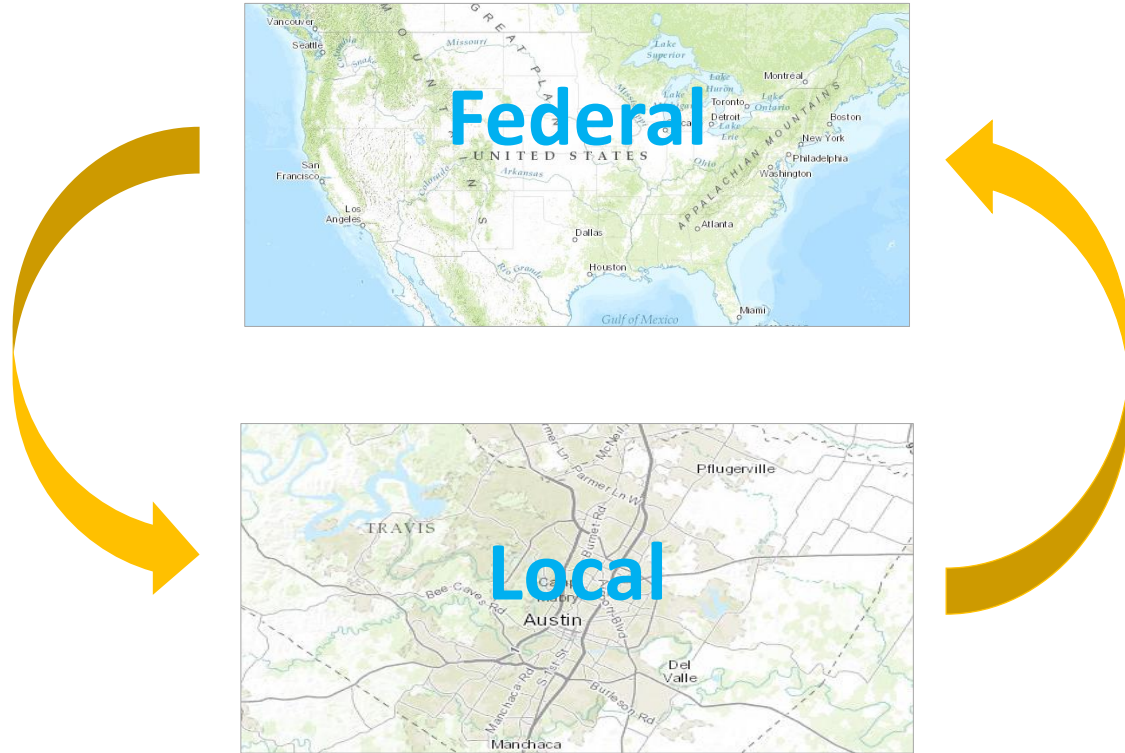
Deputy Jessica Hollis



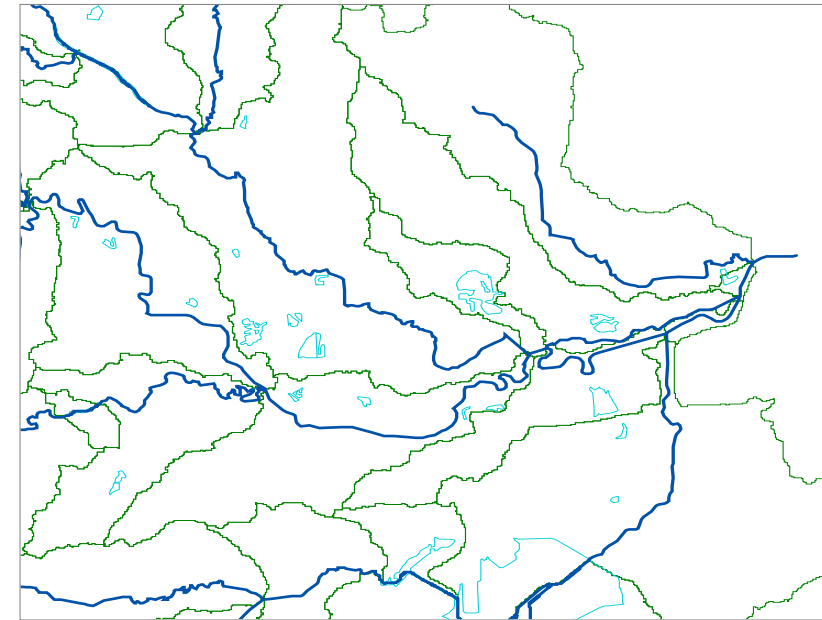
Capt. Jason Farley



# Close the gap between national flood forecasting and emergency response



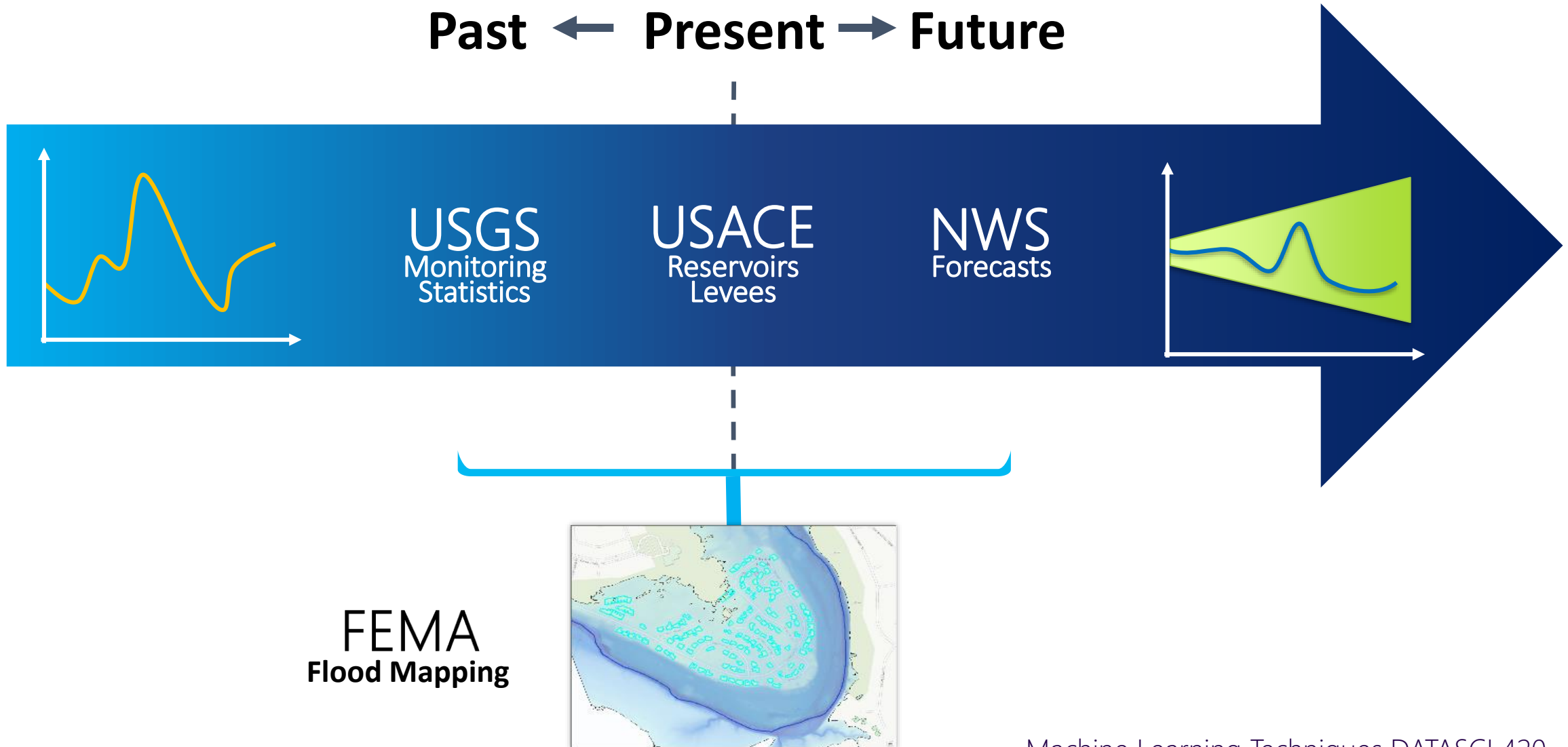
...improve information flow



...national forecasting at stream reach scale

# Open Government Data Sources

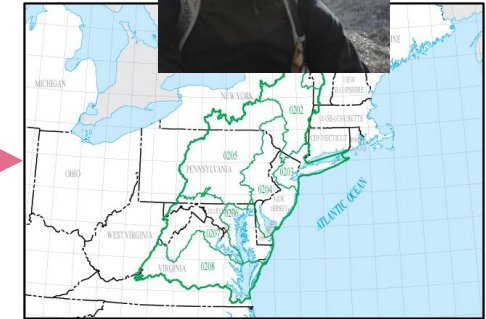
**Past** ← **Present** → **Future**



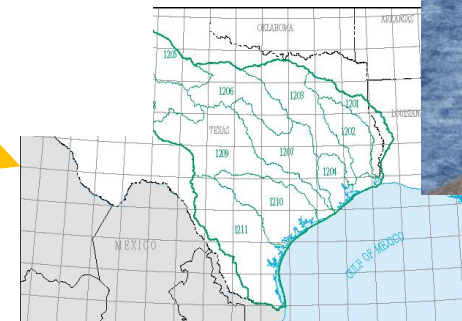
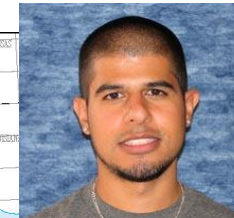
# Where to start: Pilot with 2 Water Regions



Marcelo Somos



Fernando Salas





# Next step... Scale to the US



**Model ~3 million  
river reaches in the  
lower 48 states**

# National Water Center, Tuscaloosa, AL

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

