

# **WEATHERPIPE**

## **A HISTORICAL RADAR ANALYZER**

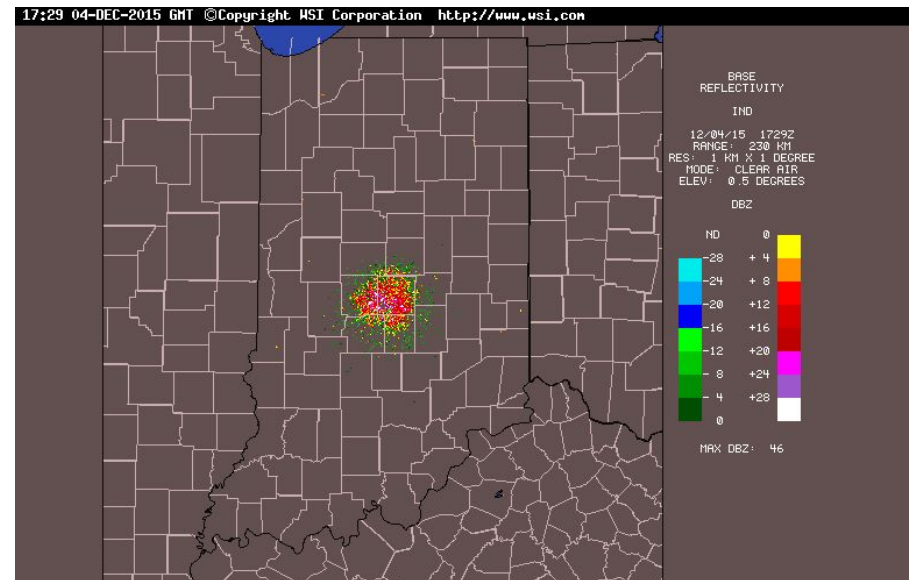
**RADAR DATA FROM THE AMAZON WEB SERVICES CLOUD  
COMPUTING WITH ELASTIC MAP REDUCE**

**Team: Stephen Harrell, Lala Vaishno, Hanqi Du, Xiaoyang Lin**

Advising: Dr. Baldwin, Associate Professor, Earth, Planetary and Atmospheric Sciences

# THE PROBLEM

- Radar data Requires large amounts of storage and computation to process
- Must have access to a specialized cluster to do historical analysis on Radar data



# NEXRAD DATA

- NEXRAD data is data collected from radar stations across the US.
- NOAA now has an agreement in place with Amazon for Amazon to host the entire radar dataset from 1990 to today.
- This historical data can be used to improve forecasts today.
- This tool is designed to find patterns and training data to better understand some of our biggest weather problems today. (tornadoes, severe storms, extreme winter weather)



# TECHNOLOGIES USED

- JAVA!
- Hadoop/MapReduce
- NetCDF file formats
- Amazon services
  - S3
  - EMR



# DESIGN

## Part 1

- Command line tool (built so API can be used elsewhere)
- Few required inputs
  - Radar station
  - Start date and time
  - End date and time
  - AWS credentials
- Using MapReduce as an easy to use compute scheduler and embarrassingly parallel API



# DESIGN

## Part 2

- Using Amazon EMR to reduce the barrier for entry (no hardware, special accounts, system administration skills)
  - Just need a credit card
- Analysis easily editable in one java file
  - Abstract away the messiness of MapReduce
- Point of the tool is a frame work to do science
  - First analysis is a simple average over time which is not scientifically relevant

# MAIN USER STORIES

- Select different time periods from the data such as a range of dates or as a specific scheme of dates
- Retrieve required radar data from s3 for every analysis
- Perform simple operations (such as calculating the average) on the data selected to perform the analysis
- Use EMR for the analysis.
- Save results of the analysis in a file from a certain run
- Handle and report failures

# SPRINT 3 USER STORIES

- Output NetCDF files for visualization
- Run the tool on a typical MapReduce cluster (in addition to EMR)
- Add a config file
- Add Control-C ending of EMR job



# NEXT STEPS

## AFTER CS307

- Amazon wants to write about and demo this tool at American Meteorological Society meeting in January
  - Implement a common radar algorithm (storm edge detection)
- Write paper with Dr. Baldwin for the International Workshop on Advances in High-Performance Computational Earth Sciences: Applications and Frameworks (IHPCES)