

Historical Weather Big Data Pipeline

Team 16

Stephen Harrell, Lala Vaishno De, Hanqi Du, Xiaoyang Lin

Problem Statement

In weather science what has happened before can be an analog with what is happening now and what will happen tomorrow. Being able to analyze conditions in the past is a powerful tool in weather prediction. Historically, this has been a hard problem for weather scientists because of the amount of data needed to process. To meet this need we will create a pipeline using MapReduce that analyzes weather data that has been recently published [1] in Amazon S3. [2] This pipeline will be written in Java and allow researchers to write small analyses like they traditionally would for small amounts of data and allow them run that over 30 years or more of radar and/or forecast data in order to pick out outliers, get statistics or a combination of the two.

Project Objectives

- Design a tool that will allow researchers to analyze large amounts of data in a relatively simple manner
- Develop an interface that is intuitive for weather researchers and can work within their normal workflows.
- Work with Dr. Baldwin to create example analyses to test the pipeline with.
- Give to Dr. Baldwin a usable project to do research with.
- Write a poster/paper about this project with Dr. Baldwin.

Stakeholders

Users: Atmospheric Researchers at Purdue

Developers: Stephen Harrell, Lala Vaishno De, Hanqi Du, Xiaoyang Lin

Project Manager: Stephen Harrell

Project Owner: Dr. Michael Baldwin, Professor in Earth, Atmospheric and Planetary Sciences at Purdue

Deliverables

- Pipeline in Amazon EMR [3] for analyzing weather data from s3
- Front-end to submit the analysis code and information about the job
- Delivery method for results (different types of analysis may have different results)
- Example analyses to test with and show different delivery methods

1. NOAA Big Data Project - <https://aws.amazon.com/noaa-big-data/>
2. Amazon Simple Storage Service - <https://aws.amazon.com/s3/>
3. Amazon Elastic Map Reduce - <https://aws.amazon.com/elasticmapreduce/>