



Analysis Precipitation to Seek a Good Place for Farm



Yawei Yang, Sereyvut Mao, Jie Zhou,
 Advisor Prof. Jeongkyu Lee
 Department of Computer Science and Engineering
 University of Bridgeport, Bridgeport, CT

Abstract

In this poster, we present how to employ Hadoop System including HDFS and MapReduce to analyze the precipitation data to find good places for farming. The precipitation data are collected from National Oceanic and Atmospheric Administration (NOAA) and some formulas from Food and Agriculture Organization of the United Nations (FAO) to help to find places which have good precipitation for specific plant.

In order to address this, we employ Hadoop Distributed File System (HDFS) and MapReduce programming with banana as an example. Combining the weather data with the precipitation data, we can figure out the places which are good for banana grows. The implemented system uses two MR programs and Google Earth to implement the visualization.

For cluster, we choose from beginning. First, we create new Virtual Machine with Ubuntu 12.04 Server operating system. Download the Hadoop 2.2.0 source code from hadoop.apache.org, then compile it. Install the Java and SSH for run the Hadoop. Generate authority keys. Setup Hadoop configuration file Duplicate the VM image to different 5 computer. Reset the hosts and hostname. Finally, start HDFS.

After launch Hadoop Map-Reduce Job, we get result part-r-00000, then we use another Java program to convert Hadoop result to KML file. Then import KML file into Google Earth, and graphical check the results.

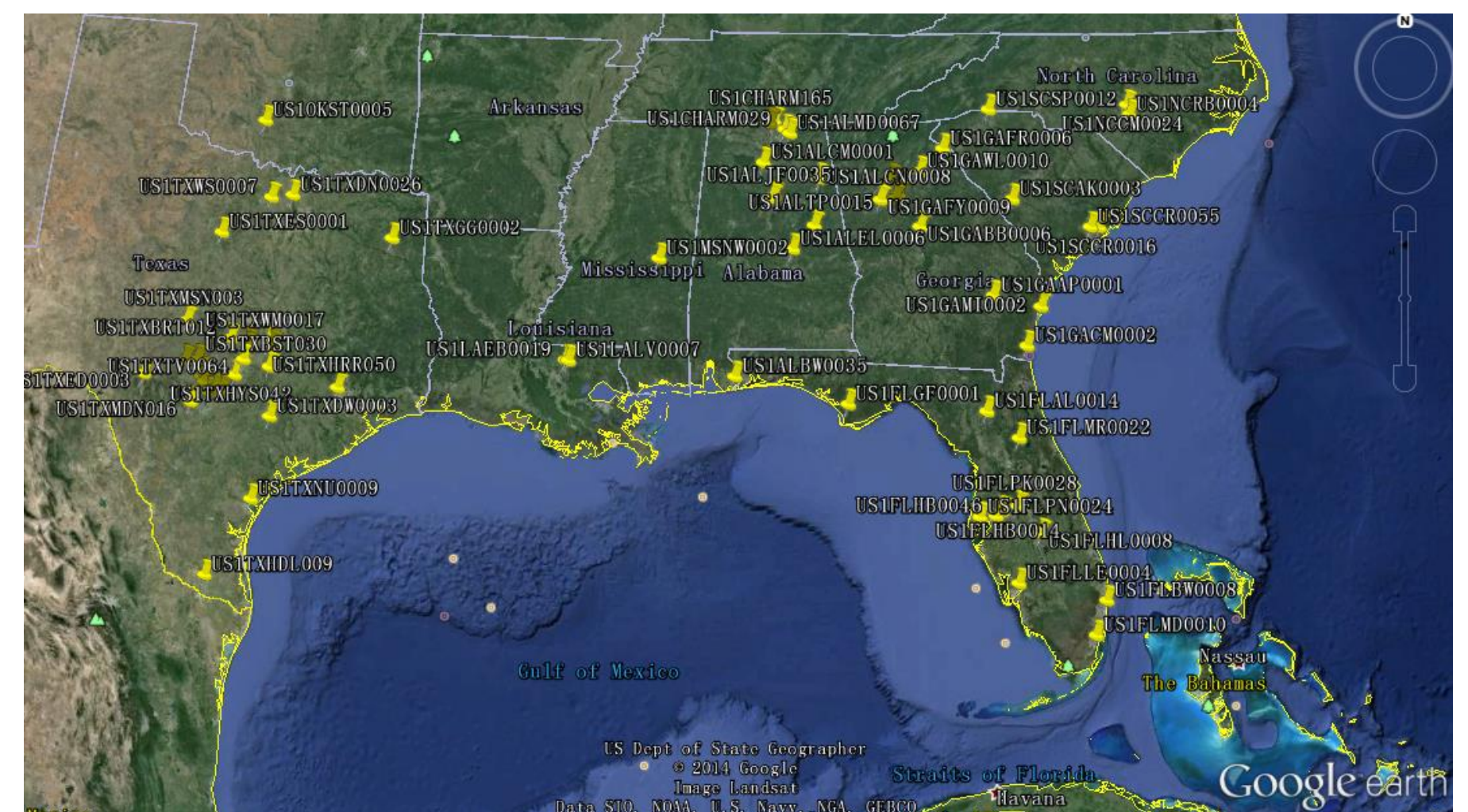
Introduction

Nowadays, to find good place for farm is not easy job. Even right now we have a high technology, but if we just obey the nature, build a farm place is not easy to achieve, and the population is still growing, earth resources is too limit to assigned for everything. For factory, working, education, living, we can build skyscraper, but for farm, just only on earth. So get more farming place is very important.

In earlier method to find the place is based on experience. Even with the other technology, but it still limit, it cannot figure with huge atmosphere data, just small part.

In this project, it combine precipitation information, geography information, analysis 30 years precipitation information to define whether this area is good for farm grows.

Final Result



Design and Implementation of Hadoop cluster

For programing, we use Map-Reduce program, read each row, just take station in US, and count days about effect rainfall, based on Banana's water needs it. Count how many days is raining and calculate rating. Value are showed n FAO websites (<http://www.fao.org/docrep/s2022e/s2022e02.htm>). For reduce part, we read first result into cache, then we pick another file which include station geography information like latitude and longitude. If match, we write it into content. After join, we get final result. This progress is showed in figure 1.

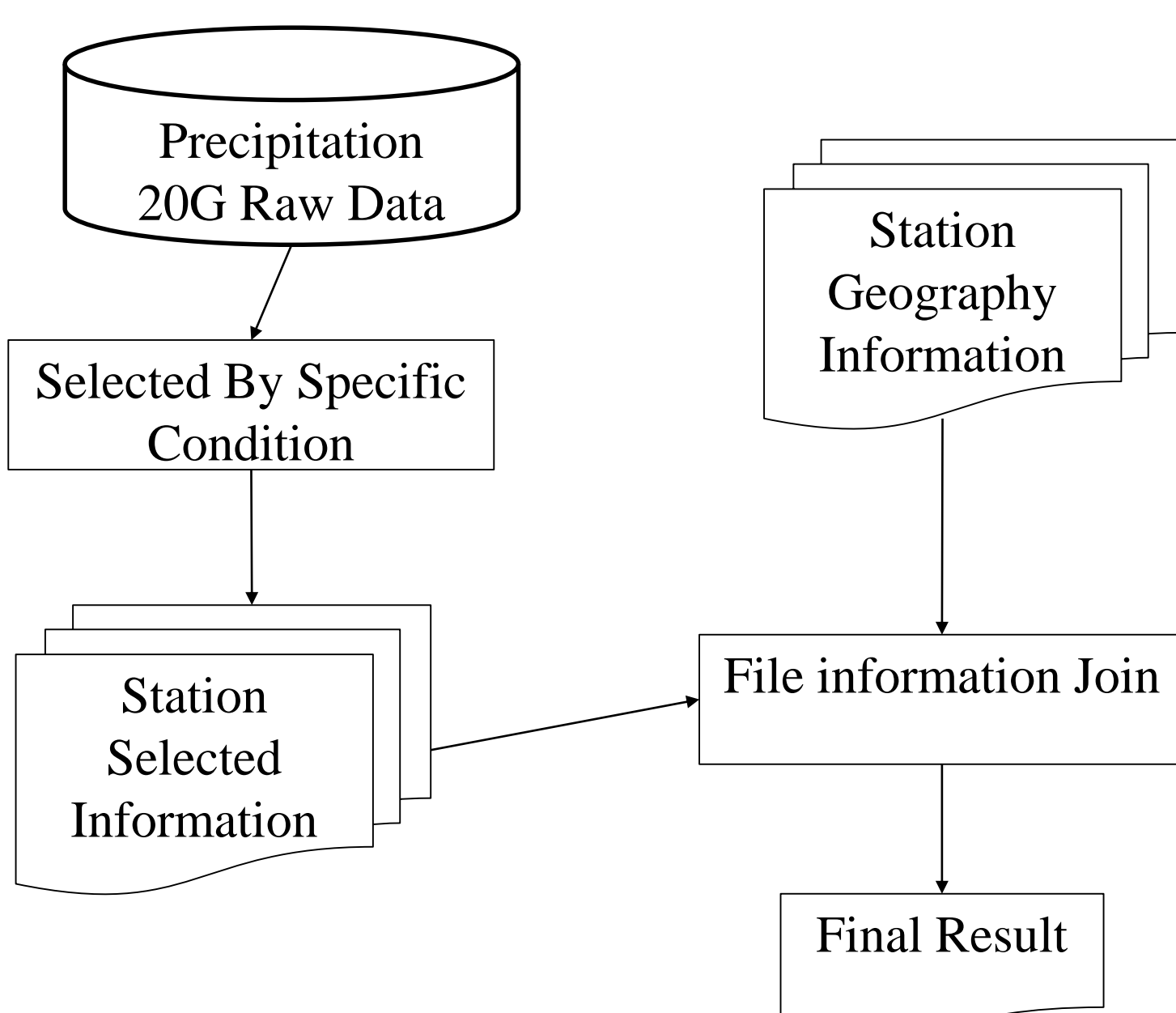


Fig. 1 Project progress

Conclusion

In this project, the most problem is set Hadoop cluster.

First problem is connection between them. We try to use SSH authority keys to connect, but when duplicate the image the authority key is to public. So when run the Hadoop MR programming, it cannot connect with other image. We have to "chmod 700 ~/.ssh".

Second problem is switch hostname, if not it always use Ubuntu to Ubuntu, it cannot define which computer is who.

Third problem is running setup, because 20G data is huge data, so when run 20G data together, the container is too small to hold the inter-media data.

Fourth problem is speed. We combine most of records file into one file to avoid too much IO operation, then increase the map threads to 20. Final the to process 20G data almost 5mins.

Future Job

This project is not perfect to find a good place for farm, because this method didn't include climate, temperature, wind humidity, wind, sunshine, landscape, soil type, technique and so on.

For other information, we need to learn which plant need which resource, which is harmful we need avoided. So besides data, we also need agricultural and geography knowledge.

Then we can make a finish a good project for agriculture.