Big Data in Weather Forecasting

COURSE: BIG DATA TECHNOLOGIES

PROFESSOR: DAVID BELANGER

NAME: JAYKUMAR PATEL

(CWID:20008512)

# Abstract

The interpretation of a vast amount of data, just like the temperature and weather, is crucial. The data must be obtained, examined, and organized. Numerous human activities, including agriculture, pollution, production, and the creation of electricity, are weather-dependent. The growing body of evidence supporting the existence of global climate change is driving the need for greater information on the weather, from what will happen tomorrow to what will exist here next year. Using big data in weather forecasting offers various advantages, including the potential to save lives, increase profitability, and improve quality of life. This is because we must analyze a tremendous quantity of data to predict the weather. Precision agriculture, large-scale crop production projections for global food security, solar power forecasting for utility operations, and space weather are some of these industries. In order to comprehend how various weather forecasting applications may impact standard operations, this article defines a number of related technical challenges.

# Introduction

The term "big data" describes the exponential rise of both structured and unstructured data since data is coming in from all directions. As an illustration, social media, movies, digital pictures, sensors, etc. Features that make it difficult to acquire data, analyze it, organize it, and process it quickly utilizing software tools. Three very important characteristics of big data exist. volume, rapidity, and abundance of variety.

As mentioned, Bryson asserted that "Weather is the first large data challenge." Although it has already been mentioned, any approach is used; the first value problem is weather forecasting. Forecasting accuracy rises with larger initial data sets. Forecasting based on temperature is essential for agricultural and commodity markets. As a result, utility companies project demand for the following days using temperature predictions. Every single day People frequently look to weather forecasts to determine what to wear on a particular day. As a result of outdoor activities The snow, rain, and wind all created significant obstacles. Weather forecasts can also be used to plan ahead, coordinate activities with these events, and endure them. We have suggested A Prediction Framework of Weather with Big Data to help solve all of these issues and predict weather very effectively. Big data has an edge over other weather prediction methods when utilizing the MapReduce algorithm since it uses a variety of techniques to reduce error and provide us with a forecast value that is close to the actual value.

The six Vs—volume, variety, velocity, value, veracity, and variability—are the finest ways to define big data. The amount of big data is an evident aspect and mostly relates to the relationship between size and processing speed. The V of variety refers to the enormous variety of data that is being saved but still needs to be processed and examined. Velocity is used to gauge the transitory value of data. This V (value) diagram explains the potential value of each type of data as well as how big data enhances the results of previously gathered data.

Veracity provides information on problems you're unsure how to handle as well as the quality and source of the data, allowing it to be viewed as suspect, contradictory, or impure. Last but not least, variability: How much and how quickly is the structure of your data changing? And how frequently does the nature or format of your data alter?

Graphical user interface, application

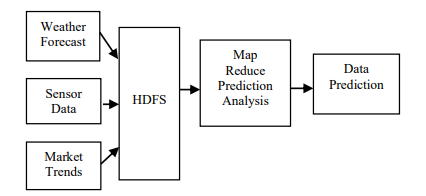
Description automatically generated

**Implementation of Weather Forecasting**

The MapReduce algorithm is used to implement big data weather forecasting. Map and Reduce are two crucial operations that the algorithm carries out. A map is implemented by the Mapper class, and a reduce is implemented by the Reducer class. Data processing can be scaled across multiple data nodes more easily and reliably thanks to MapReduce. It examines the supplied data and forecasts the day's weather. Consequently, in this design, it is able to forecast the weather, thunderstorms, snowfall, a lovely day for a stroll, and other significant details that most people consider at the beginning of the day.

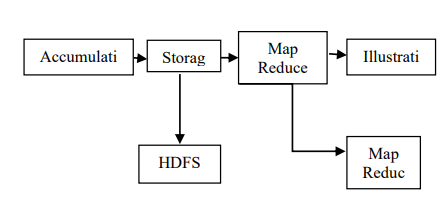
It’s done in 4 parts:

1. Accumulation 2) Storage
2. Map Reduce 4) Illustration



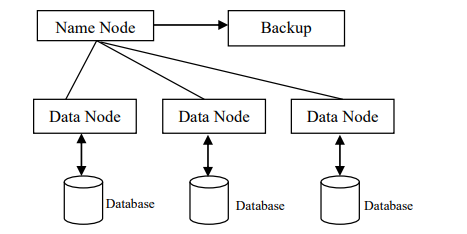
1. Accumulation

The design begins by compiling data from a range of sources, such as social media, sensor data, market trends, and weather predictions. This design's information was compiled from a variety of online resources. The database management system stores the acquired data. The data is still kept in the database system until it is moved to a storehouse. During transfer, the data is unconsciously deleted. The flow diagram for the entire process is shown in the image below.



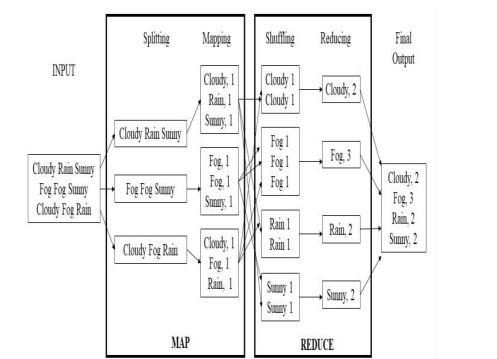
1. Storage

The gathered data is delivered to the main data storage system. This technology is known as the Hadoop Distributed File System (HDFS). The HDFS uses the MapReduce programming language to carry out the processing. The HDFS is a Master-Slave system. The fundamental building component of the system is a single name node that is synchronized with a backup node. The backup node is responsible for compressing and altering operating system-stored files into new files. However, there might be a lot of data nodes. The data nodes break down the transferred data and store it in data blocks. The system can multitask and still perform the read/write tasks. Map Reduce is the second most important Hadoop component after the HDFS. HDFS is scalable and uses parallel processing to avoid failures. The architecture of HDFS is depicted in the figure below.



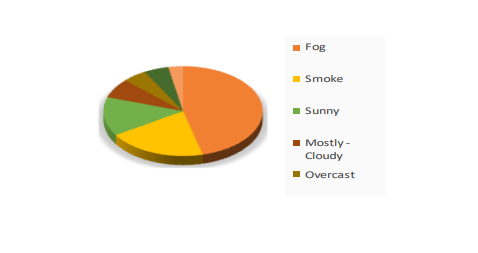
1. Map Reduce

At this point, the stored data is determined using the map-reduce architecture. The compilation of numerous weather forecasts is the first step of the algorithm. The next step is to separate the input into different data blocks. Data is divided, sorted, and then compared to data from the other set. This method is known as mapping. Following that, the information is moved around during the shifting process. The data is mixed and then divided into distinct blocks. The final forecast is made using the most frequent data. The final forecast is FOG, which is shown to be the most common occurrence in the graph below. The entire process is an effective way to divide a group of data into different sets. The mentioned operations are carried out by the MapReduce using a variety of mathematical techniques.



1. Illustration

In the final stage, the filtered data are displayed. Focusing on the most accurate and relevant weather forecasts is made easier with the help of filters. Visual representations of the data include pie charts, line graphs, bar graphs, histograms, and bar graphs. as demonstrated by the Figure below.



## Algorithm

Each intermediate key-value pair of the weather report is produced by the mapper. The reducer totals each count of the object of temperature.

Input: // Weather dataset

Begin

1: class Mapper

2: method Map(LongWritable, Text, Text, IntWritable)

3: for all Text, IntWritable do

4: Emit(string temp; line 10)

5: if stringUtils is Numeric then

6: output (datepart , Temperture)

End.

Begin

7: class Reducer

8: method Reduce(Text, IntWritable, Text, IntWritable)

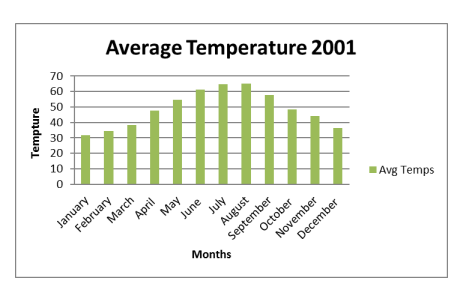
9: sum 0

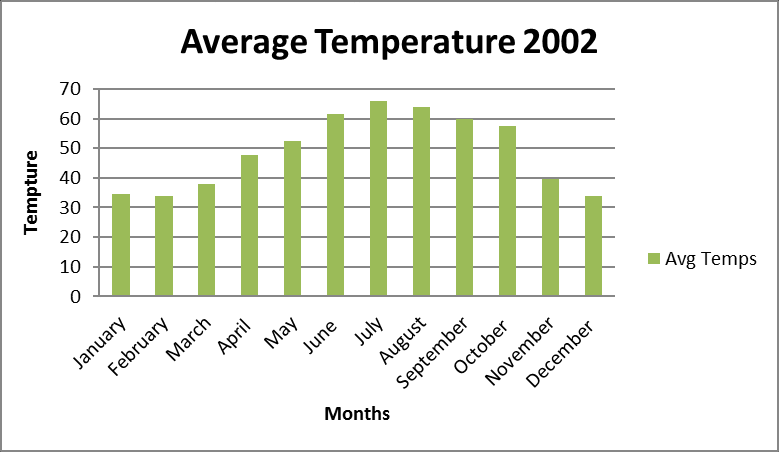
10: while all count temp counts [temp1 ; temp2; and tempn] do

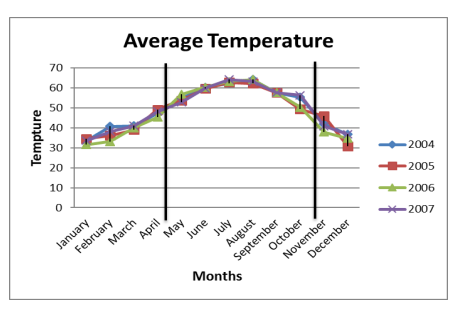
11: sum = sum + temp

12: Emit(temp ; count sum)

End.

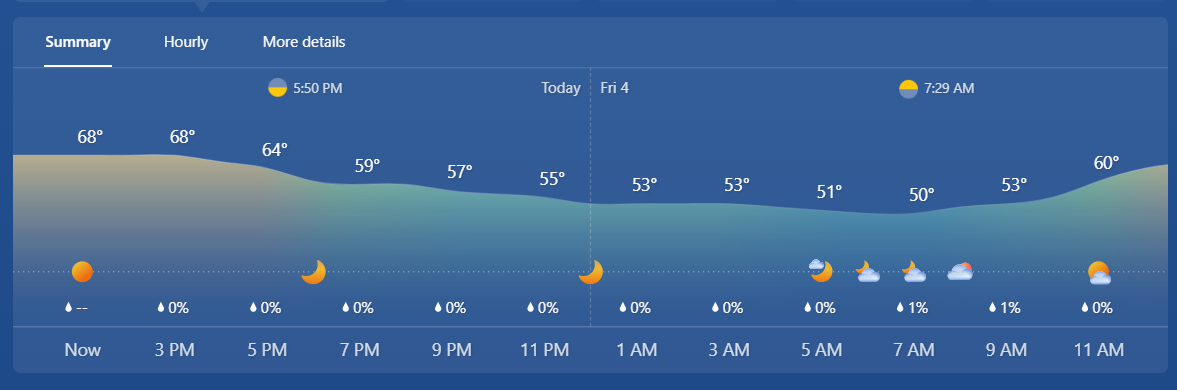






The average temperature for each month over a period of seven years is shown in the aforementioned numbers. The figure clearly depicts all of the ups and downs. As shown in the figure below, we can also predict the future with the aid of big data technology.

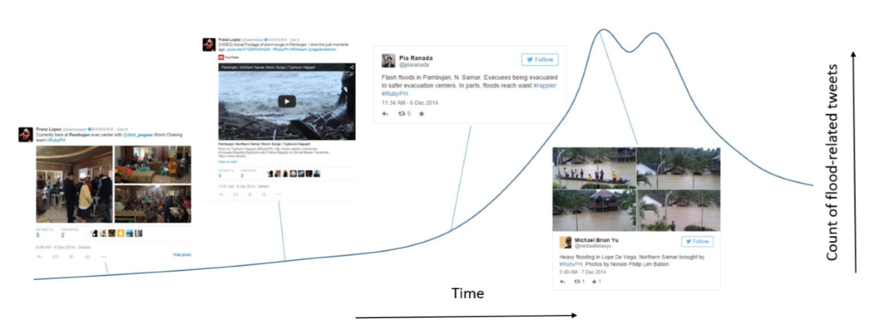




**Examples of Big Data usage in Weather Forecasting**

1. Flood

Big data also helps in the early detection of floods. The position, timing, and effects of floods were mapped in real-time using information from Twitter and satellite observations by a group of scientists led by De Groeve, Kugler, and Brakenridge. The constantly updated map is accessible online.

Social media before, during, and after disasters enable qualitative situational analysis. Information was extracted from Twitter using Floodtags in order to filter, visualize, and map social media content according to geography and keywords. The satellite data were provided by the Global Flood Detection System (GFDS), a system that uses routine passive microwave satellite observations to quickly locate areas that have been flooded. The technique was tested in two case studies, one in Pakistan and the other in the Philippines, and it was found to be extremely helpful for keeping track of significant floods in densely populated areas.

The tweet pattern shown here is related to the floods in the Philippines in 2014. The Dutch government has started experimenting with ways that machine learning can enhance readiness for upcoming floods, where the vast majority of the population lives in flood-prone areas. The New South Wales State Emergency Service in Australia developed an early warning system capable of conducting predictive analysis of flooding in the area based on the integration of external data from the Bureau of Meteorology and additional datasets.

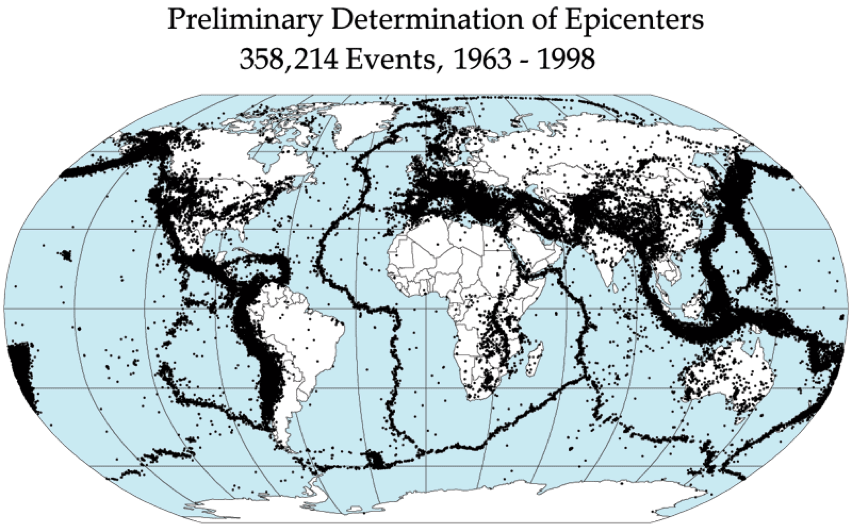
1. Storms

It has been demonstrated that big data is helpful for monitoring and assessing the effects of storms, such as hurricanes, typhoons, and cyclones. In fact, big data technologies provide a never-before-seen capability for processing data from scattered datasets, assisting in the creation of fresh insights into the weather system. For instance, the South Korean government increased the simulation capacity of the meteorological office by 1,000% and provided it with access to the country's most cutting-edge storage system. Combining weather data with social data, mobile telecommunications operator data, and other types of data can be useful for mapping and analyzing meteorological threats.

The Mobile Data, Environmental Extremes and Population (MDEEP) study in Bangladesh examined the possibility that data from the national telecommunications provider Grameenphone may have provided insight into the effectiveness of early warning systems during the occurrence of cyclone Mahasen in 2013. For assisting in real-time weather system sensemaking and raising public awareness of natural risks, big data-driven visualization tools seem to hold the most promise. Earth, a freely available interactive map that includes information from a variety of sources, including OSCAR's ocean currents and NOAA's global forecast system, serves as an obvious example of such possibilities.

1. Earthquakes

Accelerometers in computers and mobile devices that use cloud computing can help identify their occurrence more quickly. A group of scientists (Cochran, Lawrence, Christensen, and Jakka) developed the Quake-Catcher Network (QCN), a seismic network utilizing distributed/volunteer computing to gain crucial insights on an earthquake by linking traditional seismic stations with cutting-edge data sources. In 2009, the group demonstrated that it is possible to detect very small earthquakes using a global network of computers connected via the Internet, demonstrating the ability to develop quick earthquake early warning systems at a low cost due to distributed data gathered via the Internet.



In order to detect landslides that follow earthquakes, a group of scientists led by Musaev, Wang, and Pu developed LITMUS in 2014. LITMUS combines data from various sources. The model combined social (Twitter, Instagram, and YouTube) and physical (USGS seismometers and TRMM satellite) data to outperform more traditional USGS real-time hazard mapping techniques. Data from relevant organizations' digital social platforms is being combined to determine when crises happen. By way of illustration, the USGS monitors Tweets mentioning earthquakes with a magnitude of 5.5 or higher in order to recognize them and notify a larger audience via their Twitter Earthquake Dispatch (@USGSted).

**Role of Big Data Analytics in Weather-based application**

1. Agriculture

A forecast is required to schedule when to plant, irrigate, and harvest crops. It is urged to harvest the crop quickly even if only 60% of it is mature since oncoming floods are predicted by weather predictions. In a manner similar to this, a hint that the rainy season has started helps farmers sow their crops on time. Weather forecasts can indicate the likelihood of fungus in the wind, which can be used to treat crops by reducing pest populations and fertilizer use.

1. Sports

Numerous apps provide information on where to play, how many days to wait, the best time to play, the environment in which the game will take place, and other factors depending on the weather at the time. Sports use of weather forecasting is unique.

1. Forestry

Making accurate predictions is essential for problem prevention and management, ensuring the safety of wildlife and wildfires, anticipating the conditions under which dangerous insects will proliferate, etc.

1. Medication

For individuals suffering from conditions such as asthma, allergies, wheezing, colds and coughs, eye-flu, etc., predictions are useful in describing the current state of a specific location. Environmental elements such as temperature, humidity, dust content, air quality, cold or warm climate, etc. are related to patient health; forecasts are useful to provide an up-to-date picture of a specific location.

**IBM’s Deep Thunder: Weather predictions applications**

Big Data is utilized by a well-known weather forecasting tool. In order for local authorities to take action as soon as risk is identified, it gives forecasts for relatively specific places, such a single city or airport.

Deep Thunder can offer a wealth of useful data, such as assessments of regions where floods are more likely to have occurred, forecasts of the size and direction of tropical cyclones, estimates of the amounts

A picture containing text, grass, outdoor, outdoor object

Description automatically generated

# Conclusion

In this work, the temperature is predicted using big data analysis of propagation using the MapReduce algorithm. Through the application of this framework, it is shown how a big data prediction framework and an intelligent system may work together to predict the temperature. The condensed conjugate gradient approach turns out to be what this method does. Due to the fact that there were very few processing errors, MapReduce operated satisfactorily in the huge cluster after being integrated into the framework. The use of big data in weather forecasting to predict temperature can lead to promising outcomes when compared to traditional meteorological techniques. This approach allows for the identification of the nonlinear relationship between the historical data (such as temperature, wind speed, humidity, etc.) supplied to the system during training and the forecasting of future temperatures.

# References

* 1. Cochran, E. S., Lawrence, J. F., Christensen, C., & Jakka, R. S. (2009). The Quake-Catcher Network: Citizen Science Expanding Seismic Horizons. Seismological Research Letters, 80(1), 26–30. <http://doi.org/10.1785/gssrl.80.1.26>
  2. De Groeve, T., Kugler, Z., & Brakenridge, G. R. (2007). Near Real Time Flood Alerting for the Global Disaster Alert and Coordination System. In Proceedings of the 4th International ISCRAM Conference. Delft, the Netherlands.
  3. Musaev, A., Wang, D., & Pu, C. (2015). LITMUS: A Multi-Service Composition System for Landslide Detection. IEEE Transactions on Services Computing, 8(5), 715–726. <http://doi.org/10.1109/TSC.2014.2376558>
  4. Blog on Weather Forecasting, <https://www.analyticssteps.com/blogs/weather-forecasting-how-do-big-data-analytics-magnify-it>
  5. Adam, Khalid & Majid, Mazlina & Fakherldin, Mohammed & Mohamad Zain, Jasni. (2017). A Big Data Prediction Framework for Weather Forecast Using MapReduce Algorithm. Advanced Science Letters. 23. 11138-11143. 10.1166/asl.2017.10237.
  6. Big data and weather forecast : <https://www.zdnet.com/paid-content/article/big-data-and-the-weather-forecast/>
  7. Image of Twitter Pattern Link , <http://www.mdpi.com/2220-9964/4/4/2246>
  8. <http://gcmd.nasa.gov/records/GCMD_NASA_DTAM.html> : Image of Earthquake epicenters.