Weather Prediction Based on Big Data Using Hadoop Map Reduce Technique

[1]Raswitha.Bandi, [2] M.Sai Sudeepthi, [3]B.Sneha, [4]N.Sai Krishna Goud
[1]Asst. Prof, MLRIT, Dundigal, [2,3,4] Student of IT, MLRIT, Dundigal

Abstract- The helps of enormous information gathers extensive volume of information, it is extraordinary computational test for the huge information Hadoop to keep up and process this information and furthermore separates valuable data in a productive way. Remembering these things there is requirement for planning framework engineering that predicts climate estimate for future. It causes individuals to take choice ahead of time for their any open air occasions. Subsequently in our proposed design we are concentrating on the disconnected information that is put away in the NCDC to anticipate the climate investigation by Hadoop delineate structure the yield of result comprises of, least temperature, most extreme temperature, number of sweltering days and frosty days and furthermore foresee future climate gauge, which brings the considerable centrality of our work. Climate expectation is the use of innovation to anticipate the activity of the air for a given area. There are a few confinements in better execution of climate anticipating for instance in information mining procedures; it can't foresee climate here and now productively. They utilized little constrained zones for climate estimating. Since environmental change has been looking for a considerable measure of consideration since long time. It is troublesome errand to anticipate climate because of dynamic changes in the environment.

I. INTRODUCTION

Big Data is the procedure of inspect vast informational collections most extreme temperature and least temperature of year, containing assortment of information composes. The enormous information keeps up and ready to anticipate the future climate estimate. At long last, we the enormous measure of information and process them. It is customary plot the chart for the acquired MAX and MIN temperature information examination; it can process the organized information, yet for every month of the specific year to imagine the not unstructured information. In huge information it can process the two temperatures. In light of the earlier year information climate information organized and unstructured information. Enormous information more often than excludes of coming year is anticipated informational collections with sizes past the capacity of normally utilized programming apparatuses to catch, clergyman, oversee and process the information. Enormous data size ranges from terabytes to numerous petabytes of information.

Climate expectation is the use of innovation to foresee the activity of the air for a given area. It is imperative fundamentally for business agriculturist, ranchers, fiascos administration and so forth climate expectation is a standout amongst the most intriguing and interesting area and assumes critical part in meteorology. There are a few restrictions in better execution of climate estimating for instance in information mining strategies; it can't foresee climate here and now effectively. They utilized little restricted territories for climate anticipating. It is troublesome assignment to anticipate climate because of dynamic changes in the air. Environmental change has been looking for a considerable measure of consideration since long time. The hostile impact of this atmosphere is being felt in all aspects of the earth. There are numerous cases for these, for example, ocean levels are rising, less precipitation, increment in mugginess. The propose framework defeats the a few issues that happened by utilizing different systems. In this venture we utilize the idea of Bigdata Hadoop. In the proposed design we can process disconnected information, which is put away in the National Climatic Data Center (NCDC). Through this we can discover the most extreme temperature and least temperature of year, and ready to foresee the future climate figure, in view of the earlier year information climate information of coming year is anticipated.

The forecast of the environmental change dependably has demonstrated vital and valuable. In the United States of America there are frequently numerous occasions composed in various urban communities. These occasions may incorporate the auto hustling, celebrations, shows, and so forth. As these are the open air shows, they experience the ill effects of the incessant climate changes, which is expanding because of a dangerous atmospheric devation. To maintain a strategic distance from these issues, they have to pre-design and pick the information for their occasion ahead of time. This can work out just in the event that they have any forecasts of the atmosphere information utilizing the Hadoop and appropriated framework and MapReduce. By utilizing MapReduce, we can likewise ascertain the most extreme and the base temperature for the hot days and cool days. So as the outcome we can find valuable data about occasion arranging, for example, area, time and factual information.

Huge information surpasses the scope of normally utilized equipment situations and programming devices to catch, oversee, and process it with in a bearable slipped by time for its client populace. Huge information alludes to informational indexes whose size is past the capacity of run of the mill database programming devices to catch, store, oversee and investigate. Enormous information is a gathering of informational collections so huge and complex that it ends up plainly hard to process utilizing close by database administration apparatuses. Big Data includes everything from click stream information from the web to genomic and proteomic information from organic research and

prescriptions. Huge Data is a heterogeneous blend of information both organized (customary datasets – in lines and sections like DBMS tables, CSV's and XLS's) and unstructured information like email connections, manuals, pictures, PDF archives, therapeutic records, for example, x-beams, ECG and MRI pictures, frames, rich media like designs, video and sound, contacts, structures and reports. Organizations are basically worried about overseeing unstructured information, on the grounds that more than 80 percent of big business information is unstructured and require noteworthy storage room and push to manage.—Big data alludes to datasets whose size is past the capacity of run of the mill database programming devices to catch, store, oversee, and break down.

- Volume
- Variety
- Velocity
- Variability
- Complexity

Proposed System

The Proposed System focus on the application of the weather report using previous studies with the concept of Big data Hadoop. It will give us the deep insight towards the weather forecast. It analyzes each day's climate record and predicts the same day's climate using data sets. In this project, we use Big Data tools to collect large number of datasets like past 50-100 years of weather reports so that based on the previous year data weather data of coming year is predicted.Aim and Objective of the Project

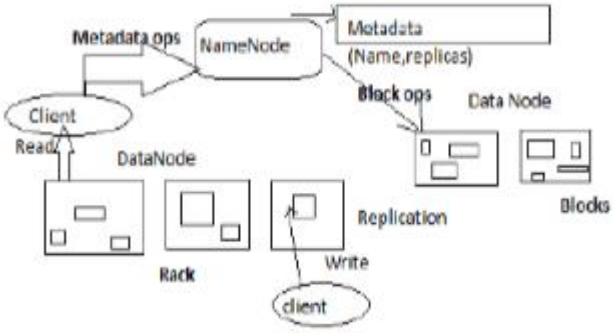
- To provide the maximum and minimum temperature of city for year.
- To predict the climate changes obtained from the map reduce.
- To be able to provide schedule the events based on this climate data.
- To be prepared for the different natural calamities like humidity and cold.
- To provide visualization of the obtained data and compare the increase and decrease in global warming

II. METHODOLOGY

Technologies used: HDFS

As the name shows i.e., Hadoop circulated document framework, the extensive measure of information is conveyed, stores and gives less demanding access. The records put away here are done in the excess mold with the goal that they can reuse the framework from the conceivable information misfortune and henceforth maintaining a strategic distance from disappointment. As the information is appropriated among many machines, the HDFS gives the parallel processing.The Hadoop Distributed File System (HDFS) is a circulated document framework intended to keep running on ware equipment. It has numerous similitudes with existing disseminated record frameworks. Be that as it may, the distinctions from other conveyed record frameworks are noteworthy. HDFS is exceedingly blame tolerant and is intended to be conveyed on minimal effort equipment. HDFS gives high throughput access to application information and is reasonable for applications that have huge informational indexes. HDFS unwinds a couple of POSIX prerequisites to empower gushing access to document framework information. HDFS was initially worked as framework for the Apache Nutch web index venture. HDFS is presently an Apache Hadoop sub project.

III. MAP REDUCE

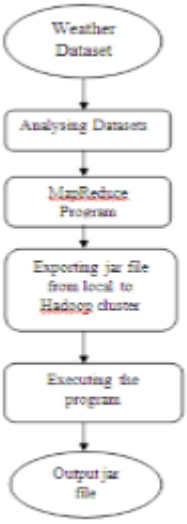


Hadoop MapReduce is a product structure for effortlessly composing applications which process immense measures of information (multi-terabyte informational collections) in-parallel on substantial groups (a huge number of hubs) of item equipment in a dependable, blame tolerant manner. MapReduce is a programming worldview at the core of Apache Hadoop for giving huge adaptability crosswise over hundreds or thousands of Hadoop bunches on ware equipment. The MapReduce display forms huge unstructured informational indexes with a circulated calculation on a Hadoop bunch.The term MapReduce speaks to two particular and unmistakable assignments Hadoop programs perform-Map Job and Reduce Job. Guide work scales takes informational indexes as information and procedures them to create key esteem sets. Decrease work takes the yield of the Map work i.e. the key esteem combines and totals them to deliver wanted outcomes. The information and yield of the guide and decrease employments are put away in HDFS.Hadoop MapReduce is a product structure for effortlessly composing applications which process immense measures of information (multi- terabyte informational collections) in-parallel on substantial groups (a huge number of hubs) of item equipment in a dependable, blame tolerant manner.MapReduce is a programming worldview at the core of Apache Hadoop for giving huge adaptability crosswise over hundreds or thousands of Hadoop bunches on ware equipment. The MapReduce display forms huge unstructured informational indexes with a circulated calculation on a Hadoop bunch.The term MapReduce speaks to two particular and unmistakable assignments Hadoop programs perform-Map Job and Reduce Job. Guide work scales takes informational indexes as information and procedures them to create key esteem sets. Decrease work takes the yield of the Map work i.e. the key esteem combines and totals them to deliver wanted

outcomes. The information and yield of the guide and decrease employments are put away in HDFS.

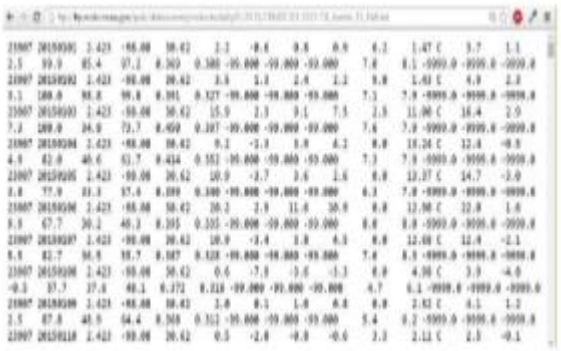
IV. IMPLEMENTATION

Dataflow Design



This project involves the following steps

- Collecting weather datasets
- Analyse the datasets
- Copy the dataset from local to Hadoop Cluster
- Writing a MapReduce program
- Executing MapReduce Program
- Running Hadoop Commands
- Output



V. SOURCE CODE

```
importjava.io.IOExceptionimport java.util.Iterator;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.LongWritable;
 import org.apache.hadoop.io.Text;
importorg.apache.hadoop.mapreduce.lib.input.FileInputFormat;
importorg.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
importorg.apache.hadoop.mapreduce.lib.output.TextOutputFormat;
importorg.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.Job; import org.apache.hadoop.mapreduce.Mapper; import org.apache.hadoop.mapreduce.Reducer; import org.apache.hadoop.conf.Configuration; public class MyMaxMin
{
public static class MaxTemperatureMapper extends Mapper<LongWritable, Text, Text, Text>
{@Override
public void map(LongWritable arg0, Text Value, Context context) throws IOException, InterruptedException
{
String line = Value.toString(); if (!(line.length() == 0))
{
String date = line.substring(6, 14); float temp_Max =Float.parseFloat(line.substring(39, 45).trim());

float temp_Min = Float.parseFloat(line.substring(47, 53).trim());
if (temp_Max > 35.0)
{

context.write(new Text("Hot Day " + date),new Text(String.valueOf(temp_Max)));
}
if (temp_Min < 10)
{
context.write(new Text("ColdDay" + date),new Text(String.valueOf(temp_Min)));
}
}
}
}
public static class MaxTemperatureReducer extends Reducer<Text, Text, Text, Text>
{
public void reduce(Text Key, Iterator<Text> Values, Context context) throws IOException, InterruptedException
{
String temperature = Values.next().toString(); context.write(Key, new Text(temperature));
}
```

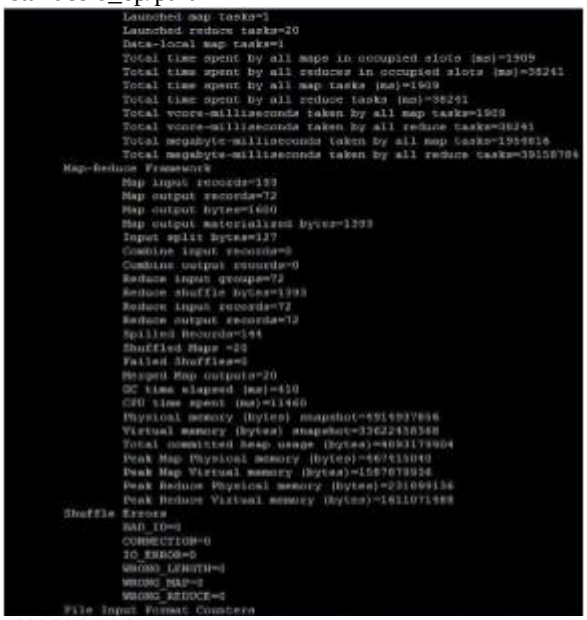
```
}
public static void main(String[] args) throws Exception
{
Configuration conf = new Configuration(); Job job = new Job(conf, "weather example");
job.setJarByClass(MyMaxMin.class);
job.setMapOutputKeyClass(Text.class);
job.setMapOutputValueClass(Text.class);
job.setMapperClass(MaxTemperatureMapper.class);
job.setReducerClass(MaxTemperatureReducer.class);
job.setInputFormatClass(TextInputFormat.class);
job.setOutputFormatClass(TextOutputFormat.class); Path OutputPath = new Path(args[1]);
FileInputFormat.addInputPath(job, new Path(args[0]));
FileOutputFormat.setOutputPath(job, new Path(args[1]));
OutputPath.getFileSystem(conf).delete(OutputPath);
System.exit(job.waitForCompletion(true) ? 0 : 1);
}
}
```

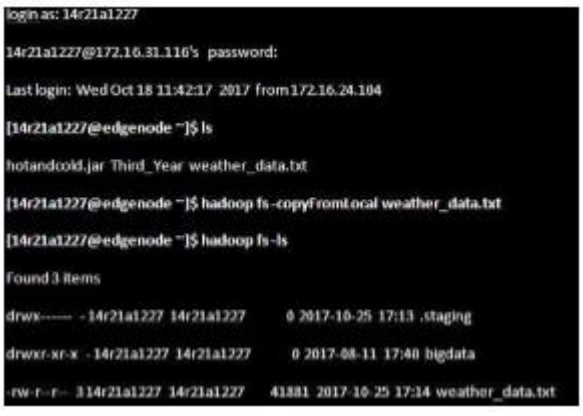
SCREENSHOTS

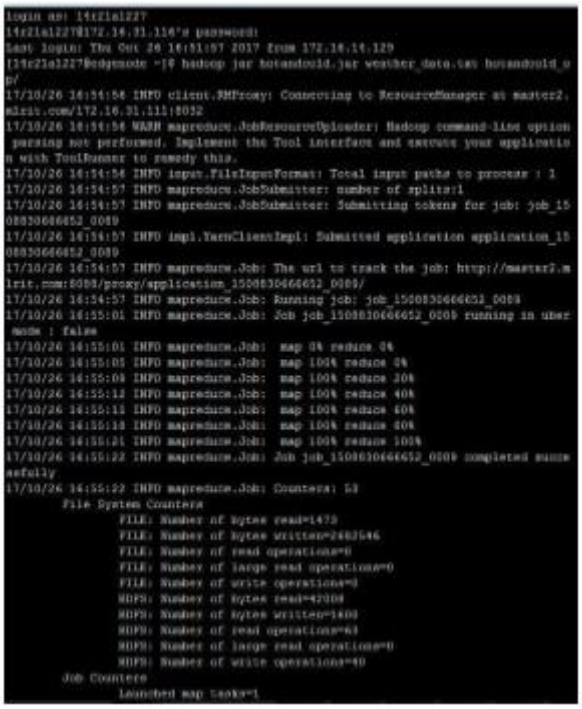
Execution Commands

[14r21a1227@edgenode ~]$ hadoop fs -ls hotandcold_op

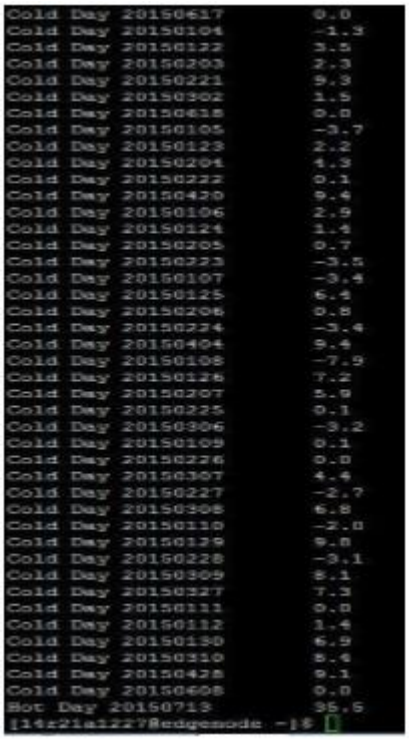
[14r21a1227@edgenode ~]$ hadoop fs -cat hotandcold_op/part-r*







```
[14r21a1227@edgenode ~]$ hadoop fs -cat hotandcold_op/part-r*
Cold Day 20150113       -0.7
Cold Day 20150131       7.4
Cold Day 20150212       5.6
Cold Day 20150429       8.0
Cold Day 20150609       0.0
Hot Day 20150714        36.0
Cold Day 20150114       0.9
Cold Day 20150213       4.7
Cold Day 20150312       9.4
Cold Day 20150115       1.2
Cold Day 20150116       3.5
Cold Day 20150117       5.0
Cold Day 20150216       0.0
Cold Day 20150315       9.5
Cold Day 20150118       7.6
Cold Day 20150217       -0.4
Cold Day 20150613       0.0
Cold Day 20150119       6.7
Cold Day 20150218       -0.4
Cold Day 20150101       -0.6
Cold Day 20150219       5.7
Cold Day 20150615       0.0
Cold Day 20150102       1.3
Cold Day 20150120       9.5
Cold Day 20150201       3.9
Cold Day 20150103       2.3
Cold Day 20150121       6.9
Cold Day 20150202       -1.9
Cold Day 20150301       -0.2
Cold Day 20150617       0.0
Cold Day 20150104       -1.3
Cold Day 20150122       3.5
Cold Day 20150203       2.3
Cold Day 20150221       9.3
Cold Day 20150302       1.5
Cold Day 20150618       0.0
Cold Day 20150105       -3.7
Cold Day 20150123       2.2
Cold Day 20150204       4.3
Cold Day 20150222       0.1
Cold Day 20150420       9.4
Cold Day 20150106       2.9
Cold Day 20150124       1.4
```



VI. RESULTS

The proposed system uses the temperature datasets of 2013, 2014, 2015 . These records are stored in the HDFS and perform map reduce function. Map reduce execution is shown in fig below "the results shows adding more number of systems to the network will speed up the entire data processing". This is one of the major advantage of the map reduce with hadoop frame work.

VII. CONCLUSION

In this project we have proposed weather prediction using big data environment. The method used in our project is Hadoop with map reduces to analyse the sensor data, is an efficient solution.Map reduce is frame work for highly parallel and distributed systems across huge dataset.It is used to analyse for the given data and predict required output to our project. By using map reduce with hadoop helps in removing scalability bottleneck. This type of technology used to analyse large data sets has potential to great enhancement to weather forecast. Hence we predict the future weather forecast, minimum and maximum temperature, hot days and cold days based on the data obtained from datasets. This helps for the people to preplanning for outdoor events based on the weather conditions.

VIII. REFERENCES

[1]. P.Agarwal, S.Das and A.E.Abbadi, "Bigdata and cloud computing: Current state and future opportunities" in Proc Int Conf Extending Database Technol.

[2]. J.Cohen, B.Dolan, M.Dunlap, J.M.Hellenstein and C.Welton, "Mad Skills: New Analysis practices for Bigdata".

[3]. J.Dean and S.Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters" Commun.

[4]. H.Herodotou et al, "Starfish: A Self-tuning System for Bigdata Analytics".

[5]. K.Michael and K.W.Millen, "Bigdata: New Opportunities and New Challenges".

[6]. R.D.Schneider, "Handoop for Dummies", Special Edition. L.Ramaswamy, V.Lawson and S.V.Gogieni, "Towards a quality Centeric Bigdata Architecture for Fractured Sensor services".

.