International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 7, Issue 4, April 2017)

Big Data Analytics for Climate- Smart Agriculture

Ruchita Thombare[1], Shreya Bhosale[2], Prasanna Dhemey[3], Anagha Chaudhari[4]
Department of Information Technology, Pimpri Chinchwad College of Engineering, Pune, India.

***Abstract*- With the digital advancements in the field of agriculture, a large amount of data is being produced constantly as a result, agricultural data has entered the world of big data. We have entered in to the era of big data. Big data is the term used for data sets so huge and complicated that it becomes hard to process using traditional data management tools or processing applications. As with many other sectors the amount of agricultural data is increasing on a daily source. Big data is an increasingly important concern in modern agriculture. The use of electronic and smart technologies, now make it possible to collect vast amount of digital information about agricultural factors. In our project the analysis of this data would be carried out to analyze the best suited way of agriculture methods in that specific region and this useful data would be again given to farmers for the better results of crop yields and green agriculture.**

***Keywords*-- Big data, agriculture, MapReduce, clustering, massively parallel processing**

I. INTRODUCTION

Agriculture is the backbone of Indian Economy. In India, majority of the farmers are not getting the expected crop yield due to several reasons. The agricultural yield is primarily depends on weather conditions. Rainfall conditions also influences the rice cultivation. In this context, the farmers necessarily requires a timely advice to predict the future crop productivity and an analysis is to be made in order to help the farmers to maximize the crop production in their crops. Yield prediction is an important agricultural problem. Every farmer is interested in knowing, how much yield he is about expect. In the past, yield prediction was performed by considering farmer's previous experience on a particular crop. The volume of data is large in Indian agriculture. The data is highly useful for many purposes. Adoption of big data in agriculture considerably decreases the possibility of crop failure and farmer's crucial concerns and recommends the soil sensing and crop yield information to be stored in data centres[3].

It is analysed from the literature review that with the digital advancements in the field of agriculture, a large amount of data is being produced constantly as a result agriculture data has entered the world of big data [4]. An initiative in Columbia have found that data-driven climate adaption could revive rice yields in Columbia. From the research article, [3] the researcher express that large amount of data which is collected and stored for analysis. Making appropriate use of these data often leads to considerable gains in efficiency and therefore economic advantages. There are several applications of Data Analytics techniques in the field of agriculture. The researchers implemented [8] K-Means algorithm to forecast the pollution in the atmosphere, the K Nearest Neighbour is applied [9] for simulating daily precipitations and other weather variables and different possible changes of the weather scenarios are analysed [10] using Support Vector Machines. Soil profile descriptions were proposed [11] by the researcher for classifying soils in combination with GPS based technologies. They were applied K-Means approach for the soil classification. One of the researcher used [12] an intensified fuzzy cluster analysis for classifying plants, soil and residue regions of interest from GPS based color images. Weeds were detected [13] on precision agriculture. The researchers worked [14] on rainfall variability analysis and its impact on crop productivity. The effect of observed seasonal climatic conditions such as rainfall and temperature variability on crop yield prediction was considered [15] through an empirical crop model. Furthermore, there are approaches to investigate the impact of climate change on crop production which include the crop suitability approach and the production function approach [16]. The data is given as input to MapReduce functions [5] and then this big data is stored among different clusters using clustering algorithms [6].The speed of the processing can be increased using massively parallel processing [7].

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 7, Issue 4, April 2017)

II. System Architecture

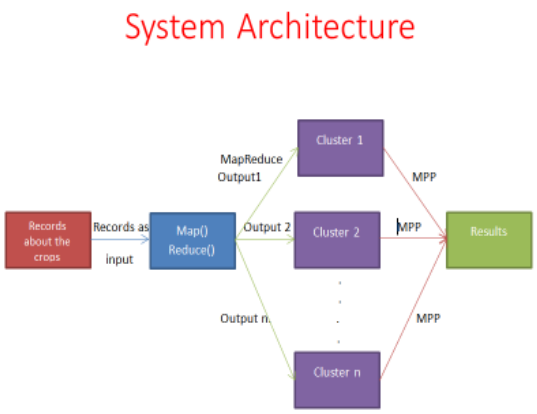

Fig. 1: System Architecture

1. Module 1 : Crops records
2. Module 2 : MapReduce
3. Module 3 : Clustering
4. Control line 4 : Massively Parallel Processing(MPP)
5. Module 5 : Results

1. Module 1 : Crops Records

The given data contains all the information regarding particular region's past 10 year's attributes such as weather, rainfall, soil type, soil fertility, average crop yield, pesticides used, crop type, water availability, electricity availability. Which is further send to MapReduce model for further operations on data.

2. Module 2 : MapReduce

Map Reduce is a processing technique and a program model for distributed computing based on java. The MapReduce algorithm contains two important tasks, namely Map and Reduce. Map takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key/value pairs). Secondly, reduce task, which takes the output from a map as an input and combines those data tuples into a smaller set of tuples. As the sequence of the name MapReduce implies, the reduce task is always performed after the map job.

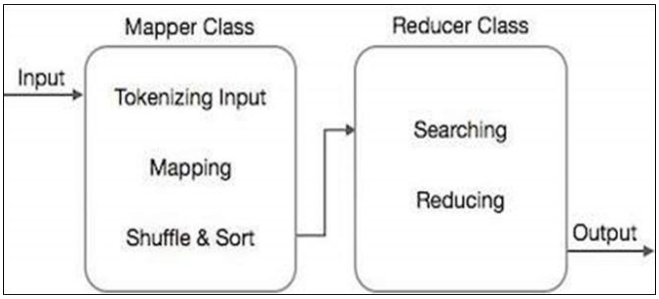

Fig 2. Mapper Class And Reducer Class

The algorithms used in MapReduce are as follows.

A. Sorting
B. Searching
C. Indexing
D. TF-IDF

The raw data is given as input to the map and reduce functions which segregates data and sends the data to appropriate cluster which is based on the attributes of agriculture.

3. Module 3: Clustering

The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering. Cluster is a collection of data objects. They are similar to one another within the same cluster and dissimilar to the objects in other clusters.

Cluster analysis is grouping a set of data objects into clusters [6].Our project works best with Hierarchical type of clustering. Hierarchical clustering method produces a set of nested clusters organized as a hierarchical tree and finds successive clusters using previously established clusters. They can be visualized as a dendrogram; a tree-like diagram that records the sequences of merges or splits [6].

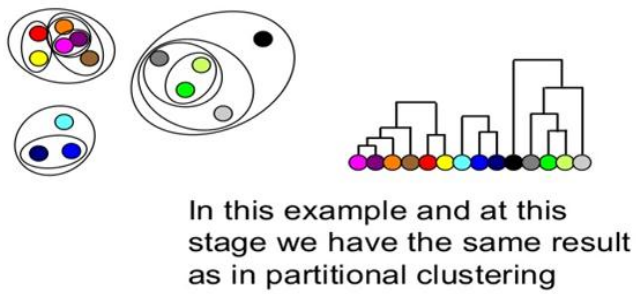

Fig 3. Hierarchical Clustering.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 7, Issue 4, April 2017)

Two types of Hierarchical approaches are as follows.

Type 1: Agglomerative Approach

In Agglomerative approach, we start with each object forming a separate group. We start with each object forming a separate group.

Type 2: Divisive Approach

In Divisive approach, we start with all the objects in the same cluster. In continuous iteration, a cluster is split up into smaller clusters.

Clustering will perform operations to get optimum results. So each cluster will be showing best record out of all records depending upon their attributes.

4. Control line : Massively Parallel Processing

MPP (massively parallel processing) is the coordinated processing of a program by multiple processor s that work on different parts of the program, with each processor using its own operating system and memory . Typically, MPP processors communicate using some messaging interface. In some implementations, up to 200 or more processors can work on the same application[17].An "interconnect" arrangement of data paths allows messages to be sent between processors. Typically, the setup for MPP is more complicated, requiring thought about how to partition a common database among processors and how to assign work among the processors. An MPP system is also known as a "loosely coupled" or "shared nothing" system[17].

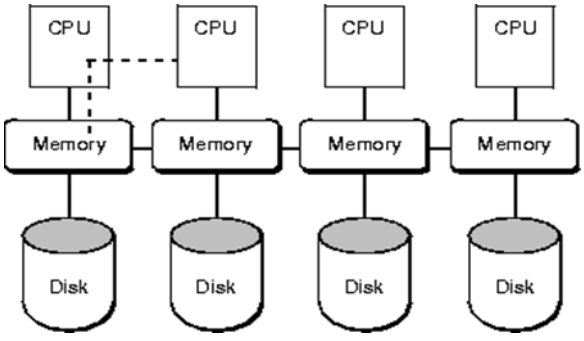

Fig 3. Memory Distribution Mpp

Techniques used for MPP are:

A. Divide and conquer.
B. Greedy Method.
C. Dynamic Programming.
D. Backtracking.
E. Branch & Bound.
F. Linear Programming.

The results obtained in clustering are processed using MPP

5. Module 4: Results.

The result sent by MPP nodes is stored in this module which is final analyzed result obtained. This result after given to a farmer will help him to develop better yield.

Precision is the proportion of retrieved documents that are actually relevant and Recall is the proportion of relevant documents that are actually retrieved.

III. Existing Crop Production System

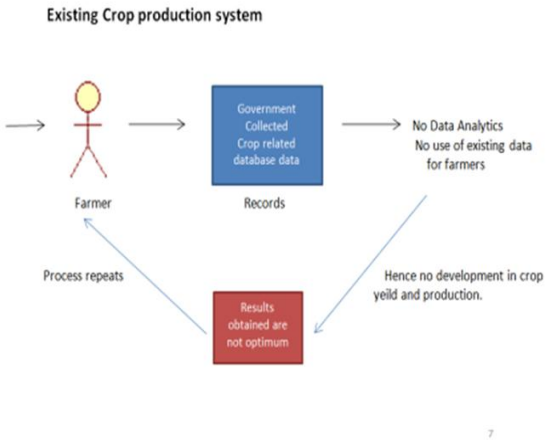

Fig 4. Existing Crop Production System.

IV. Proposed Crop Production System

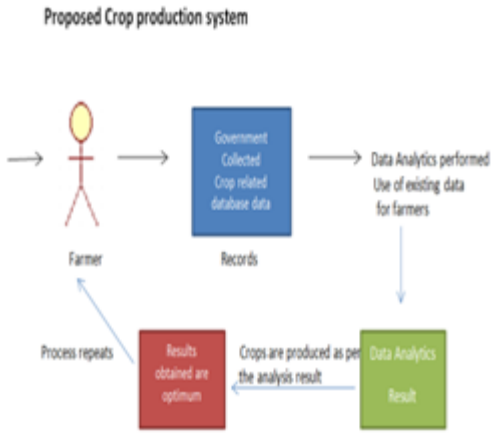

Fig 5. Proposed Crop Production System.

V. CONCLUSION

Agriculture has been transformed into an intelligent world as a result, agricultural data has entered the world of big data. Big data analytics is magical technique that will help us in prediction and decision making. To handle agriculture crop disease problems big data analytics framework has been developed that provides recommendation solutions by using MPP, Clustering and Map Reduce. This is quite helpful for researchers and officers for recommending solutions based on evidence from historical data. We have studied Massively Parallel Processing algorithms, in which data retrieved from Clustering can be processed with MPP and necessary outputs can be driven. The proposed work will help farmers to increase the yield of their crops. The future work involves the following:

A. Storage of big data in clusters by using various clustering algorithms, reduce it to appropriate/valid content using map reduce techniques.
B. Further, MPP technique will be used for parallel processing of data.

REFERENCES

[1] Athmaja S., Hanumanthappa M, "Applications of Mobile Cloud Computing and Big data Analytics in Agriculture Sector: A survey", October 2016.

[2] P. Surya, Dr. I. Laurence and M. Ashok Kumar, "The role of big data analytics in agriculture sector: A survey", March 2016

[3] D Ramesh, B Vishnu Vardhan, "Analysis of crop yield prediction using data mining techniques", January 2015

[4] CCAFS, "Big Data for climate-smart agriculture", 2015.

[5] V. Sajwan, V. Yadav, "MapReduce: Architecture and Internals", October 2015

[6] Odilia Yim , Kylee T. Ramdeen, "Hierarchical Cluster "Analysis: Comparison of Three Linkage Measures and Application to Psychological Data", October 2015

[7] Moufida Adjout, Faouzi Boufares, "A massively parallel processing for the Multiple Linear Regression", September 2015

[8] Jorquera H, Perez R, Cipriano A, Acuna G, "Short Term Forecasting of Air Pollution Episodes", In: Zannetti P (eds) Environmental modeling , WIT Press, UK, 2001.

[9] Rajagopalan B, Lall U, "A K-Nearest Neighbor Simulator for Daily Precipitation and Other Weathe Variables", Wat Res Res 35(10), 1999, pages : 3089-3101.

[10] Tripathi S, Srinivas V V, Nanjundiah R S, "Downscaling of Precipitation for Climate Change Scenarios: A Support Vector Machine Approach", J Hydrol, 2006, pages : 621-640.

[11] Verheyen K, Adriaens D, Hermy M, Deckers S, "High-resolution continuous soil classification using morphological soil profile descriptions", Geoderma Vol.101, 2001, pages : 31-48.

[12] Meyer G E, Neto J C, Jones D D, Hindman T W, "Intensified Fuzzy Clusters for Classifying Plant, Soil and Residue Regions of Interest from Color Images". Computer Electronics Agric Vol. 42, 2004, pages : 161-180.

[13] Tellaeche A, X P Burgos Artizzu, G Pajares, A Ribeiro, "A Vision-Based Classifier for Weeds Detection in Precision Agriculture through the Bayesian and Fuzzy K-Means Paradigms", Adv.Soft. Comp., 2008, pages : 72-79.

[14] Mehta D R, Kalola A D, Saradava D A, Yusufzai A S, "Rainfall Variability Analysis and Its Impact on Crop Productivity - A Case Study", Indian Journal of Agricultural Research, Volume 36, Issue 1, 2002, pages : 29-33.

[15] M Trnka, "Projections of Uncertainties in Climate Change Scenarios into Expected Winter Wheat Yields", Theoretical and Applied Climatology, vol. 77, 2004, pages : 229-249.

[16] M J Foulkes, "Raising Yield Potential of Wheat", Journal of Experimental Botany, vol. 62, 2011, pages : 469-486.

[17] http://whatis.techtarget.com/definition/MPP-massively-parallel-processing.