INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume 4, Issue 3)

Available online at: www.ijariit.com

Big data: An analytic architecture and prediction using spark for E-agriculture

Patel Jaydeep Pravinbhai
jaydeep.patel@ngivbt.edu.in
Noble Engineering College, Junagadh, Gujarat

Ashutosh Abhangi
ashutosh.abhangi@ngivbt.edu.in
Noble Engineering College, Junagadh, Gujarat

ABSTRACT

Nowadays Big Data has a key role in E-Agriculture. Previous technologies have some limitations so Big Data is very useful for E-Agriculture because agriculture has very large structure and unstructured data. Moreover, big data analytics can be used to increase and improve the productivity of agricultural. The main aim of this paper is to propose an open source, economical, ideal and flexible big data analytics architecture for E-Agriculture. In an implementation, an analytic framework for big data application development is built and implemented. Also, a prototype application prediction base for agriculture in spark frame work. Based on the agriculture prediction model various recommendations can be provided to agro users.

Keywords: *Big data, Prediction system, Spark, E-agriculture.*

1. INTRODUCTION

Expansion of Big Data has reduced the throughput, efficiency and resource utilization. There are various ways to increase the throughput and resource utilization from which one way is the task scheduling. In Big data, many platforms have been introduced by many different researchers. The main goal of any platform is to achieve and organize structure and unstructured data, better utilization of resources and best system throughput. To utilize the system resources efficiently, some factors such as resource usage, network traffic, response time, and other overheads are to be optimized. Spark is one of the best platforms to deal with real-time data. Real-time data is a major concern because it is hard to utilize and function. The agriculture sector has a large amount of data so it is very hard to figure out all data on the traditional system. So big data is very important in this sector and spark deal with real-time data so it is one of the best platform for big data for agriculture system.

The need of research is to develop a more efficient and effective system which can predict and, analytic agriculture unstructured data more effectively and increase the prediction performance and quality prediction so it will helpful of agriculture scientist and farmers. In all types of big data frameworks spark works with live data so it is a very important concept in big data. No one can do work with live data on agriculture field with big data and spark frame work so it is a very new concept in e agriculture.

Today's work big data is very useful to handle the data because with this type of data big data deal very carefully and complete the work very carefully. This all type of agriculture data is complex so it is very important to work with it in a proper format. Big data and its system can do his complex work so it is very important and useful way to conduct the work. Big data technology is under trail so it is very important to show it with proper format and work with IT industry so this is a very large scale complex to show and done deal with data work efficiency. Big data has 5 V's concept volume, variety, velocity, veracity, and value. Data volume is for how big is data. In big data is in the form of tbs or more than that so it is a first important factor in big data. Second is a variety which means different type, size of data which are structure and unstructured data. The third one is a velocity. It is used and referred to the form of speed. How can we deal with data in how much time? The fourth one is a veracity means data is trustworthy or not and if it is trustworthy than in the form of how much ratio. The last one values means it is important than how much it is? It is useful or valuable data or not this is the parameters of it.

Next one is a big data analytic architecture in e-agriculture. Agriculture is one of the very important structures in our system. In this analytic architecture system first, collect all the data from different sources after that work on that and structured it. After that works on this data and collect useful data among all necessary and unnecessary data sets. Data collect from different websites, Agriculture

University, agriculture departments, weather department so on. Put this all data on mongodb. It is the best way to process all the data. After that put this process data into the system which is built in spark framework because it is very fast and unique frame work so it is very important to show that and work with it. Now all output stores at one place and create one web portal so users, farmers, agriculture data scientist all of that access it and use it.

2. LITERATURE SURVEY

A. Big Data Analytics Architecture for Agro Advisory System [1]

Big data is a very new and important trend in agriculture domain. People now realize the importance of Big Data in E-Agriculture. Big data analytics is a very tough thing in the agriculture field. However, how to use big data analytics in agriculture field to improve the productivity in practices. Purpose of this E-Agriculture to reduce technological gape between rural communities and share information via recommendations and decision support system. The main aim of this paper is to propose an open source, cost-effective and scalable big data analytics architecture for the agro advisory system. In this paper authors approach prototype application for crop yield prediction implemented for cotton crop in Ahmedabad district, Gujarat, India

Here, proposed a big data analytics architecture for an Agro advisory system. Also, an analytic frame work built and implemented in spark for crop yield model. Because of this many another model can be implemented for farmers according to them requirements. Because of that farmers take good decisions for their crop and income. In future work, other many architecture and analytical services can be searched and done one the base of this model like crop price prediction, crop disease detection etc… The limitation is that it considers limited parameters. Also, this paper had not specified clearly about the effectiveness of big data on agriculture field. So it can be further improved by considering more parameter.

B. Big Data Analytics Framework to Identify Crop Disease and Recommendation a Solution [2]

Due to technology, the term data is replaced by transforming big data in many fields. Big data has structure and unstructured data because of this it is the very useful format in agriculture data. Because of that big data enter in the agriculture field. Traditional technology cannot handle a large amount of unstructured data so big data has a key role to handle, store and analyze this massive amount of data. So big data is a very good solution to all this problem. In this paper, big data analytic framework is developed that identify disease based on high similarity. Hadoop and Hive tools used in this method. Firstly all data collected, cleaned and normalized from using this tools. Data is collected from different websites, laboratories, agriculture universities and private companies. First, collect all the important information from the unstructured data than normalize and clean it. HDFC is useful in normalization to store the agriculture data. After that clean that normalize data. HiveQL is a query language and used to analyze the agriculture data. It finds out disease based on crop disease symptoms and recommends a solution for it. Results are given by graphs.

This is one type of advance technology in agriculture sector it is very useful for farmers. It will handle the crop disease problem on the large scale using Hadoop and Hive tools. It is a location-based prototype for detecting crop disease so I am very helpful for researcher also who research on particular gio location. Very complex and static so it is not helpful on large scale.

C. Survey on Crop and Weather Forecasting based on Agriculture related Statistical Data [3]

India is an agriculture-based nation. It has a key role in nation's GDP. Crop production is mainly based on the season but India has many seasons so it is the effect on crop yield and it directly effects on crop prices. Soil type, temperature, atmospheric pressure, humidity and crop type etc. effect on crop yield and prices. However, us agriculture data and give crop prediction is very helpful for farmers to aids maximum profits. This paper presents a survey on various algorithms used for weather, crop yield and crop cost prediction. Lots of techniques and algorithms comparison in this paper and they all have some pros and cons.

D. SPARK – A Big Data Processing Platform For Machine Learning [4]

Apache Spark is a distributed memory-based computing framework which is naturally suitable for machine learning. Hadoop, the spark has a better way of functionality and ability of computing learning. In this paper, analyze of spark frame work with basic concept means spark's primary framework, core technologies and run a machine learning instance on it. Moreover, analyze the output result and introduce hardware equipment. This paper provides a general introduction of spark's core technology for machine learning with big data. Spark is suitable and fulfilled all requirements at parallel computing and iteration application. It will be very useful in machine learning on large-scale data.

E. A Big Data Architecture for Automotive Application: PSA Group Deployment Experience [5]

Vehicles collect a big amount of data because they are moving and collecting data at all the time when they are active. Vehicles collected data from different sensors like GPS data, real-time traffic data etc. this data is very tremendous for automotive manufacturers and vehicles' owners. It is used to improve many services related automotive. However, traditional technologies cannot handle efficiently this large amount of data so big data is in the picture. In this paper, the PSA group on leveraging big data in automotive context. Big data architecture deployed in depth within the PSA group and the underlying technologies used in each component. This paper shows a basic overview of big data architecture deployed within PSA group for automotive applications. The main concept of this design is to respond specific automotive applications requirement. However, this architecture is unidirectional and future work is the car to car architecture for data send and receive base. In future, next step is used Apache spark process and analytics of data to improve the performance of batch processing.

F. Prediction of crop yield using big data [6]

Food security is a major concern in the third world. This paper mainly providing new way to predict crop yield based on big data. Moreover, this is different from traditional methods in the structure of handling data and in the means of modeling. Big data main advantages are to use existing massive agriculture data set which are growing rapidly. 2nd most important thing big data is it can provide balance results, accuracy and prediction time in advance. This paper takes crop agriculture dataset 1995 to 2014 from China have shown a better performance and improved prediction accuracy of the proposed method compared with traditional ones. This paper proposes an architecture for managing big data in the agriculture area. The main advantage of this method is managed massive data set which is already existing. This technique is faster than any other traditional one.

3. PROPOSED WORK

The biggest challenge in BDA application development is to bring data into action. Selection of state-of-the-art big data technology for big data management is the main issue for data architects/Engineers. The main role of system developer is to establish a big data pipeline. Firstly store data in mongodb and then data inelastic in spark analytic engine. Data collected from different sources and it is implemented in the e-agriculture system. After mongodb work data goes to the spark and create API and after that, all system works properly.

A. Basic Architecture

Here we can present the basic architecture of the proposed system.

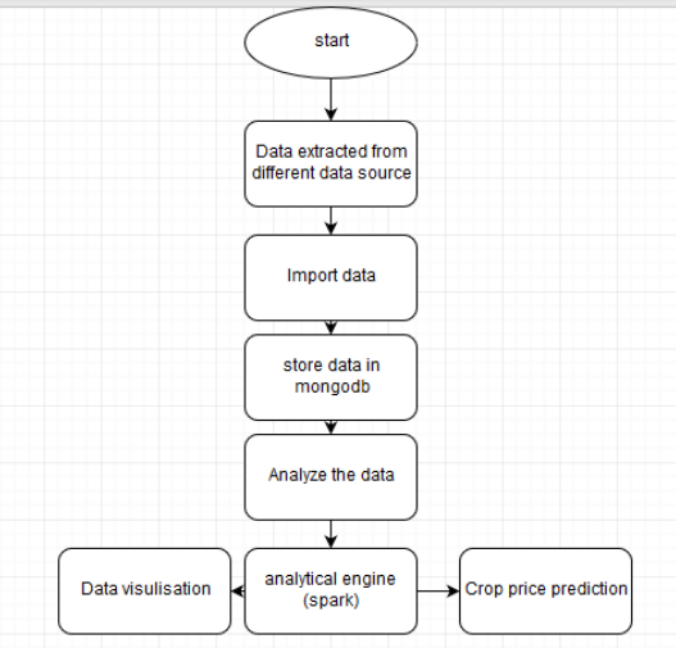
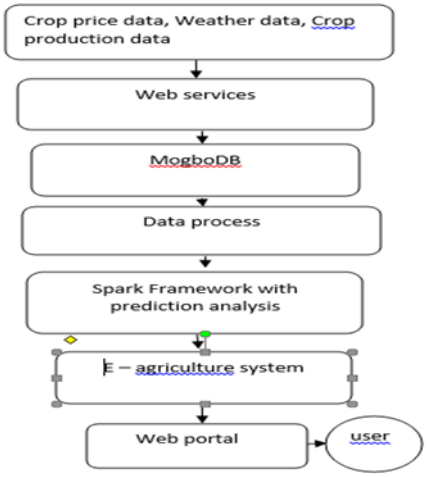
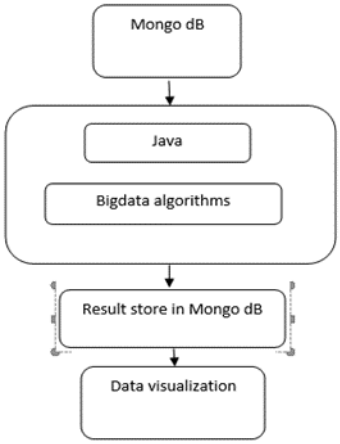


Fig 1. Basic Architecture of the system

B. Proposed system architecture



Here this is a final proposed system architecture which is useful for every data scientist, agriculture scientist and so on, it is an e agriculture system architecture.



This is a spark architecture for the e-agriculture system. Here we can use mongo db database. Java language for coding and prediction algorithm and result store in data base.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots \beta_p x_{ip} + \varepsilon_i \text{ for } i = 1,2, \ldots n.$$

This is a multiple linear equation which ne is used for prediction system and useful in crop price prediction to improve the data set of it.

4. RESULT

R-Squared	Actual in 2014	Prediction in 2014	Difference
1.02	289	211	78

Here, given result is from data which are shows in dashboard results it is a frame work of it and it is the main concept of it. Here given data is taken from various sources and put it into frame work of data base which is a replica of the different data base. It is mainly work on spark framework which is a framework of a big data and accurate for prediction faster than hadoop. Total take 1.02 r- squared land and implement this data on framework after that actual prediction is 289 ton in 2014 according to government report but we got 211 prediction rate it is less than 78 ton.

5. CONCLUSION

There are many techniques and system in big data prediction system but this is a unique system which one work on spark and it is very useful to do with all characteristic to improve the statistic ratio of prediction system so this is a good and nearly accurate system which can be work with agriculture system. In next step, we will work on next prediction system like weather, rainfall, crop dieses, animal location prediction etc…

6. REFERENCES

[1] Purnima Shah, Dipak Hiremath, Sanjay Chudhary, "Big Data Analytics Architecture for Agro Advisory System" 2016 IEEE 23rd International Conference on High-Performance Computing Workshops
[2] Raghu Garg, Dr. Himanshu Aggarwal, "Big Data Analytics Framework to Identify Crop Disease and Recommendation a Solution" 2016 International Conference on Inventive Computation Technologies (IEEE)
[3] Pushpa Mohan, Dr. Kiran Kumari Patil, "Survey on Crop and Weather Forecasting based on Agriculture related Statistical Data", International Journal of Innovative Research in Computer and Communication Engineering (February 2017)
[4] Jian Fu, Junwei Sun, Kaiyuan Wang, "SPARK – A Big Data Processing Platform For Machine Learning" 2016 International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration (IEEE)
[5] Amir Haroun, Ahmed Mostefaoui, Francois Dessables, "A Big Data Architecture for Automotive Application: PSA Group Deployment Experience", 2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing
[6] Wu Fan1, Chen Chong2, Guo Xiaoling2, Yu Hua* Wang Juyun2," Prediction of crop yield using big data", IEEE 2015
[7] Ramya M G, Chetan Balaji, Girish L, "Environment Change Prediction to Adapt Climate- Smart Agriculture Using Big Data Analytics", 2015 International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
[8] P. Chandrashaker Reddy, Dr. A. Suresh Babu, "Survey on Weather Prediction using Big Data Analytics", IEEE-2017
[9] S. Chaudhary, M. Bhise, A. Banerjee, A. Goyal, C. Moradiya, `Agro Advisory System for Cotton Crop', Workshop on Networks and Systems for Agriculture AGRINETS 2015, collocated with 7th International Conference on Communication Systems and Networks COMSNETS, pp1-6, 2015, DOI: 10.1109 /COMSNETS .2015.

