

Productivity Improvement in Agriculture sector using Big Data Tools

Ch. Chandra Sekhar
Asst. Prof, Dept. of IT,
AITAM,
Srikakulam, AP, India.
dlaxmics@gmail.com

Ch. Sekhar
Asst.Prof, Dept. of CSE,
Vignan's IIT,
Visakhapatnam, AP, India.
sekhar1203@gmail.com

Abstract—The traditional methods used by the farmers, particularly in India, are very slow, undependable and large amount of crops are damaged in fields due to bacterial attacks and lack of information resources. Annually, such loss exceeds 40% in total. The main goal is to provide feasible solution for the finding problems and to enhance the productivity of the agriculture sector. The main awareness of this work is focused on Indian farmers as it addresses the key problems of getting the market status of different products, weather alerts and also provides multiple language support. This will effectively help farmers to sell their products in global market and earn remarkable profit. At the same time it's a big challenge of getting information about weather situations, kind of fertilization need to use based type of crop. Due to lack of communication to farmers production was not good, also government sanctioning various schemes, benefits towards agriculture sectors and farmers towards their wealth. There is scope to maintain the information of all these and analyze properly and communicate with farmers. This kind of analysis can be done with latest technologies such as big data analytical tools. A real monitor system needed to communicate with farmers, time to time with support of mobile based application.

Keywords—Agriculture; Big Data; Mobile application;

I. INTRODUCTION

Agriculture is the primary source of livelihood for about 58 per cent of India's population. [3]Agriculture, with its allied sectors, is unquestionably the largest livelihood provider in India, more so in the vast rural areas. It also contributes a significant figure to the Gross Domestic Product (GDP) Sustainable agriculture, in terms of food security, rural employment, and environmentally sustainable technologies such as soil conservation, sustainable natural resource management and biodiversity protection, are essential for holistic rural development.

Rural Farmers do not have knowledge over the latest methods for harvesting and machinery to operate. Still most of the farmers follow the tradition methods and outdated machinery. Hence the productivity is low. Need agriculture institutional arrangement to farmers to educate them with short term course or workshops to educate and update the latest methods and machinery with the hope that it will deliver results is quite common in the context of Indian agriculture. Another issue is simple things like a warehouse receipt system is a complicated issue in India. There is no proper monitoring system available to know the availability of cold storage and warehouse to store the final items about to sale in the market. It leads the most of the goods to be damaged.

Another major reason is for poor productivity in India in most crops can be largely attributed to lack of soil fertility. Soil fertility in India is further going down due to erroneous application of fertilizers (more Nitrogen (N) and less of Phosphorus (P) and Potash (K) than the recommended mix). NPK mapping at each field is necessary for the right prescription including type of seed, seed rate, irrigation, plant growth regulators, fertilizers etc.

II. BIG DATA

“A collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. Big data means really a big data with huge are in size in terms tera bytes, peta bytes. The birth of big data, as a concept if not as a term, is usually associated with a META Group report by Doug Laney entitled “3-D Data Management: Controlling Data Volume, Velocity, and Variety” published in 2001 [1]. Further developments now suggest big data problems are identified by the so-called “5V”: volume (quantity of data), variety (data from different categories), and velocity (fast generation of new data), veracity (quality of the data), and value (in the data)

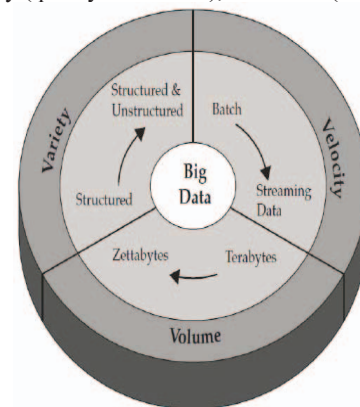


Fig. 1. Big Data Characterization

A. Why Big Data Now?

Big data can able to handle any kind of data as: Structured Data: Data that has a pre- defined format, data will be managed as per the format only. e.g. Employee information, banking transactions, etc.. Unstructured Data: Data that has no pre defined format. Data can receive any format that need to be stored. e.g. Audio, text files, social media. Semi-Structured Data: Unstructured data that can be put into a structure by available format descriptions. e.g. xml document.

B. Benefits of Big Data in Agriculture [4]:

Big data has been a key driver of the progress made in cutting edge technological interventions in farming like precise agriculture and automated farm machinery operations whereby farmers and agribusinesses are keen to use these big data resources at their disposal in the most efficient way and attempts to get maximum yield. With support of Big data Tools and frame work, can get optimum decisions in farming, Crop recommendations, Intercropping recommendations, Selection of suitable Hybrids, Farming practices, Pests prediction and Management, Forecast the Agri commodity prices ahead of the season, Profitability Analysis, Policy recommendations

III. BIG DATA FRAMEWORK[4]

HDFS has two nodes, Name node and Data node. Name node acts as Master and maintain the Meta information of all the data nodes data. Data node maintains the actual data of various application inputs. Data node communicates with the name node with support of heartbeat message system. To process the data by using big data performance and tools such as Map Reduce, Pig script, Jaql etc.. The analyzed output will generate in various output formats such as Reports, files, graphs.

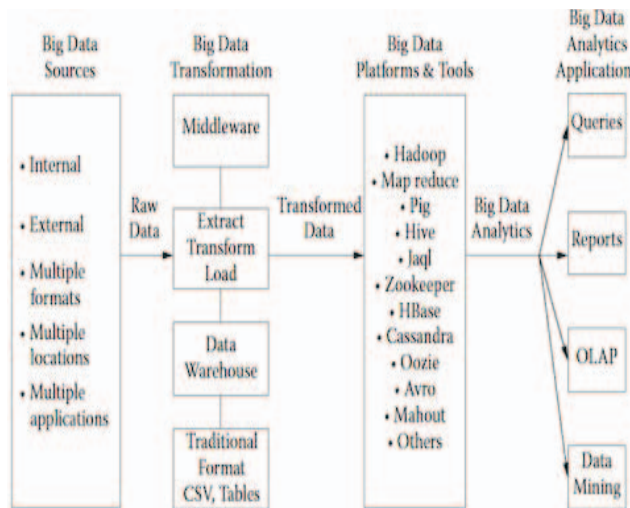


Fig. 2. An applied conceptual architecture of big data analytics.

The conceptual framework for a big data analytics is similar to that for a traditional business intelligence or analytics project. The Conceptual architecture show that work flow mechanism. Data received in as input from various sources in different formats (unstructured data also) and receiving from distributed network. Once the data received as raw format that will be pre- processed and made it standard format with support of ETL tools. The data will be stored in HDFS (hadoop distributed file system).

A. Functionality need to do[7]

- **Information Loading:** A software must be created to load information from numerous, different information sources. This framework needs to manage the distributed way of Hadoop on the one side and the non-distributed nature of

the information source. The framework needs to manage defiled records and need to give observing administrations.

- **Information Parsing:** Most information sources give information in a specific arrangement that should be pursued into the Hadoop framework. For instance, we should consider parsing a log document into records. A few configurations are convoluted to parse like JSON where a record can be on many lines of content and not only one line for each record.
- **Data Analytics:** All together for information to be appropriately broke down, a major information investigation arrangement needs to Data Analytics bolster fast emphases.
- **Visualization:** All together for an expert to see the bits of knowledge, information should be envisioned. Coordinating perception is troublesome in light of the fact that middleware should be worked to convey the information out of Hadoop and into the representation layer.
- **Scheduling:** All the items discussed above need to be orchestrated and scheduled. Scheduling needs to be easily to configure. In addition, the scheduling needs have monitoring services to notify administrators of jobs that fail.

B. Big Data Tools

Enormous data is a term utilized for a gathering of informational indexes so vast and complex that it is hard to process utilizing customary applications/devices. It is the information surpassing Terabytes in size. Due to the assortment of information that it envelops, enormous information dependably conveys various difficulties identifying with its volume and intricacy. A current study says that 80% of the information made on the planet is unstructured. One test is the manner by which this unstructured information can be organized, before we endeavor to comprehend and catch the most imperative information. Another test is the manner by which we can store it. Here are the top tools used to store and analyze Big Data. We can arrange them into two (stockpiling and Querying/Analysis).

- **Apache Hadoop[6]:** Apache Hadoop is a java based free programming system that can viably store expansive measure of information in a group. This structure keeps running in parallel on a group and has a capacity to permit us to process information over all hubs. Hadoop Distributed File System (HDFS) is the capacity arrangement of Hadoop which parts huge information and disseminate crosswise over numerous hubs in a bunch. This likewise repeats information in a bunch along these lines giving high accessibility
- **Microsoft HDInsight:** It is a Big Data arrangement from Microsoft controlled by Apache Hadoop which is accessible as an administration in the cloud. HDInsight utilizes Windows Azure Blob stockpiling as the default document framework. This additionally gives high accessibility minimal effort.
- **NoSQL:** While the conventional SQL can be successfully used to deal with substantial measure of organized

information, we require NoSQL (Not Only SQL) to deal with unstructured information. NoSQL databases store unstructured information with no specific mapping. Each line can have its own particular arrangement of section qualities. NoSQL gives better execution in putting away huge measure of information. There are many open-source NoSQL DBs accessible to investigate enormous Data

- **Hive:** This is a conveyed information administration for Hadoop. This backings SQL-like inquiry alternative HiveSQL (HSQL) to get to huge information. This can be basically utilized for Data mining reason. This keeps running on top of Hadoop.
- **SQOOP:** This is an apparatus that interfaces Hadoop with different social databases to exchange information. This can be viably used to exchange organized information to Hadoop or Hive.
- **Big data in EXCEL:** The most of individuals are agreeable in doing investigation in EXCEL, a famous instrument from Microsoft, you can likewise interface information put away in Hadoop utilizing EXCEL 2013. Horton works, which is fundamentally working in giving Enterprise Apache Hadoop, gives a choice to get to enormous information put away in their Hadoop stage utilizing EXCEL 2013. You can utilize Power View highlight of EXCEL 2013 to effortlessly outline the information. (More data)

IV. PROPOSED SOLUTIONS

To overcome the above problems in the existing system, we create a new proposed system that reduces Middle men problem and farmers can learn new methodologies from the experts through this system. They can also plan for planting the best crop with the help of weather alerts that gives information about the climatic conditions and which crop is suitable for that particular climate. This system provides a platform for the farmers where they can know the prices of different markets and can sell their crop for the best price.

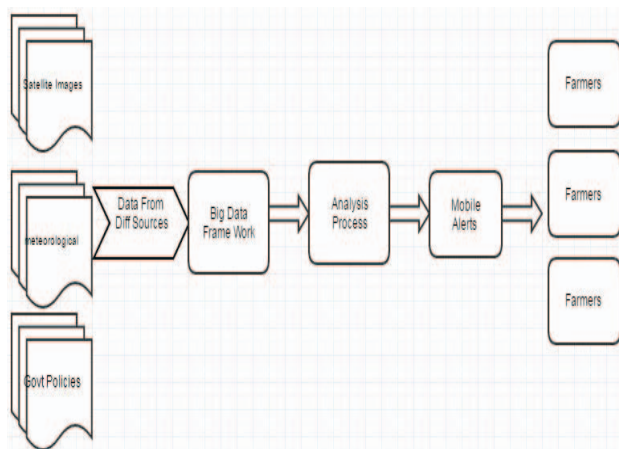


Fig. 3. Proposed Architecture

A. Work Flow

The work flow of proposed system as follows, the data coming from various sources such as, the data on crop imagery can also be captured through farmer's smart phones

and shared with agronomist to find solution of pest attacks or poor growth. Temperature and humidity data from weather stations can be getting it from metrological depts. Data of various benefits and policies provided by government from ministry of agriculture. Data come in different formats that can be taken care by big data frame work.

B. Features Of Proposed System[5]

With support of big data performance tools, analyze the data based on the need of location and type of crop suggestion can be delivered to the farmers through mobile application. The mobile application can contain features as mentioned below.

- News feed: All the updated news can be known through notifications.
 - Online learning: facilitates the farmer to know best practices with the help of tutorials.
 - Weather Alerts: provides the information about the climatic conditions.
 - Buy products: can make order for the products.
 - Sell products: sell the products for enterprises and consumers.
 - Market status: can know the best price dealings in markets.
- Queries: where farmers can clear their doubts.

V. METHODOLOGY

In this edge work, process will be happens in different stages. Information gathering, Data Analysis, Alert System and visualization modules

- **Information gathering:** Here we are getting the information from different sources likes' metrological dept about atmosphere conditions, pesticide data and new seeds in the market. Composers arrive pictures and yield pictures through cell phones. Master and specialists recommendation. Every one of the information can be stacked into appropriated record framework (hdfs). With help of ETL apparatuses we have to do the information cleaning and combination prepare before putting away into to hadoop framework.
- **Data Analysis module:** This module is preparing stage incorporates two sections of information reading/breaking down and foundation of conjecture result. The information perusing is done chiefly by Hive. Hive is a structure for information warehousing on top of Hadoop. It was made to make it workable for experts with solid SQL aptitudes to keep running on the immense volumes of information put away in HDFS. Hive keeps running on workstations and change over SQL queries into arrangement of MapReduce jobs for execution on a Hadoop cluster. MapReduce is an execution motor reasonable for vast information handling and can fundamentally enhance the reaction speed for returning question comes about. . In the second part we implement prediction function for establish forecast data through k-means cluster algorithm. Here we make use of apache mahout. It is an open-source scalable

machine learning library. Mahout provides an efficient way of implementing unsupervised machine learning algorithm. The data of past few years are used to make prediction about future.

- **Alert System:** It's the extension between the legislature and the agriculturists. At whatever point data transferred by rancher, for example, pictures of harvest or land those things need be tell t o Information gathering.
- **Visualization:** The outcomes which are come back from the analysis module will be shown in this module in an imagined mode. Showing complex information with outlines and charts is a fundamental piece of the information data analysis process, and we make utilization of apparatus to make delightful and one of a kind information visualizations.

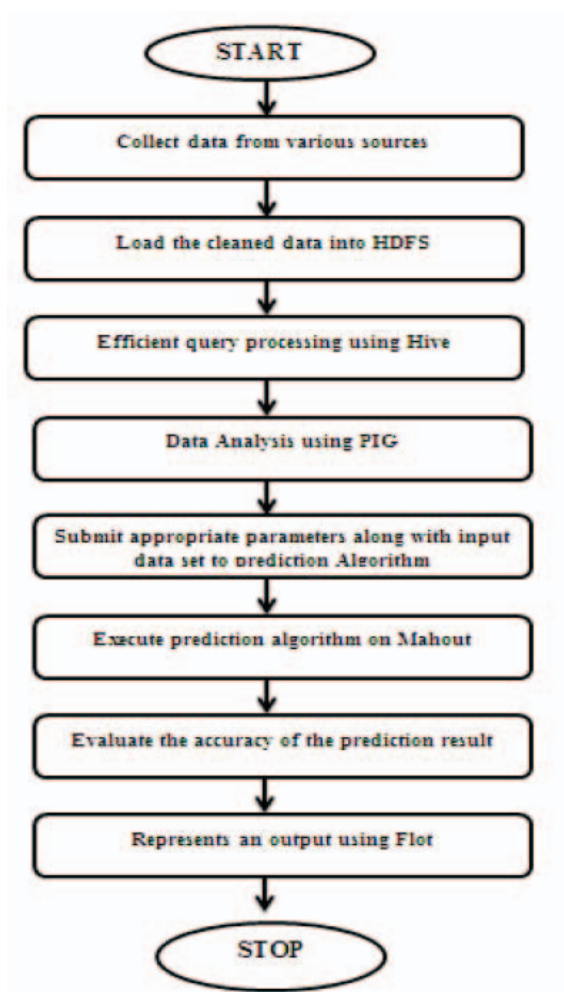


Fig. 4. Proposed System work flow

In the meantime any updates from IMD, agribusiness divisions about different data atmospheres conditions, showcase costs of various harvests and seeds subtle elements should be send time to time alarms to the

agriculturists. It will help to agriculturists towards reaping and enhance generation.

CONCLUSION AND FUTURE WORK

This application can be useful to the farmers who are not much aware about latest technologies in agriculture fields. Can get the latest updates such as whether information. Market for goods, govt polices and benefits. Here we need to design mobile application which is able to understand by the farmers in their local language. And the motivation among the rural people usage of smart phones. Major issue here is availability of internet in villages. Need to develop mobile such a way that at low internet speed also application should work.

REFERENCES

- [1] Big Data, Big Knowledge: Big Data for Personalized Healthcare, IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 19, NO. 4, JULY 2015 1209.
- [2] www.ibef.org/industry/agriculture
- [3] A survey on platforms for big data analytics, Singh and Reddy Journal of Big Data 2014, 1:8 <http://www.journalofbigdata.com/content/1/1/8>
- [4] Big Data Analytics Architectures, Frameworks, and Tools, Wullianallu Raghupathi and Viju Raghupathi
- [5] <https://techwave.net/leveraging-bigdata-in-agriculture/>
- [6] Textbook Hadoop ,The definivite guide by Tom White
- [7] www.datameer.com/pdf/big-data-analytics-ebook

BIBLIOGRAPHY

Mr.Ch. Chandra Sekhar, Asst. Prof in IT Dept, AITAM, Tekkali. His present areas of interests include Cloud computing, Big data, Network Security, Data Mining, Image Processing and mobile computing.

Mr.Ch. Sekhar, Research scholar at JNTUK. He is working as a Sr. Asst Prof in CSE Dept, Vignan's IIT, Vizag Research interests include Data Mining, Big Data Analytics Cloud Computing and Computer Networks