ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

THE ROLE OF BIG DATA ANALYTICS IN AGRICULTURE SECTOR : A SURVEY

P.Surya
Ph.D Research Scholar
Dept. of Computer Science
Periyar University,
Salem-636011.
suriyaa14@gmail.com

Dr.I.Laurence Aroquiaraj
Assistant Professor,
Dept. of Computer Science
Periyar University
Salem-636011.
laurence.raj@gmail.com

M.Ashok Kumar,
Ph.D Research Scholar,
Dept. of ComputerScience,
Periyar University,
Salem-636011,
williamashok@gmail.com

ABSTRACT

We are 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 agriculture data are 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 agriculture factors. Here in this paper, we are done a survey about the factors of agriculture under using big data analytics. And also gave a suggestion for implement a big data analytics to give a better performance.

Key words: Big data, Agriculture, Hadoop.

1. INTRODUCTION

Agriculture is the cultivation of animals, plants, fungi, and other life forms for food, fiber, medicinal and other products used to sustain and enhance human life. The study of Agriculture is known as Agricultural science [2].

1.1. AGRICULTURE IN INDIA

India is an indomitable country with more than billion plus people, and also one of the world's rapidly flourishing economies. Out of the huge population, 58.4% are innocent agricultural assemblage. India's recent accomplishments in crop yields while being inspiring, are still just 30% to 60% of the best crop yields reachable in the farms of developed as well as other developing countries[1][7].

India ranks second worldwide in farm productivity. Agriculture and allied sectors like forestry and fisheries accounted for 13.7% of the GDP (gross domestic product) in 2013, about 50% of the workforce. The economic contribution of agriculture to India's GDP is steadily declining with the country's broad-based economic growth. Still, agriculture is demographically the broadest economic sector and plays a significant role in the overall socio-economic fabric of India. As per the 2010 FAO world agriculture statistics, India is the world's largest manufacturer of many fresh fruits and vegetables, milk, major spices, select fibrous crops such

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

as jute, staples such as millets and castor oil seed. India is the second largest producer of wheat and rice, the world's major food staples.

A **Crop** is any cultivated plant, fungus, or alga that is harvested for food, clothing, livestock fodder, medicine, or other uses. In contrast, animals that are raised by humans are called livestock, except those that are kept as pets [3].

Crop Categories

* Food crops
* Plantation crops
* Commercial crops
* Horticulture crops
* Forage crops
* Manure crops

Major crops areas in India

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

1.2. SEASONS IN INDIA

Seasons		Duration	
		Sowing	Harvesting
Kharif		June-July	September- October
Rabi		October - December	March-April
Zaid	Kharif	August-September	December-January
	Zaid		
	Rabi	February-March	April-May
	Zaid		

Seasons in India

Kharif crops or **monsoon crops** are domesticated plants cultivated and harvested during the rainy (monsoon) season in the South Asia, which lasts between April and October depending on the area. Kharif means "autumn" in Arabic. Main kharif crops are millet and rice. In India, The kharif season varies by crop and state, with kharif starting at the earliest in May and ending at the latest in January, but is popularly considered to start in June and to end in October. Kharif crops are dependent on the quantity of rain water as well its timing. Too much, too little or at wrong time may waste the whole year's efforts [4].

Rabi crops or **Rabi harvest** is agricultural crops sown in winter and harvested in the spring in the South Asia. The term is derived from the Arabic word for "spring", which is used in the Indian subcontinent, where it is the spring harvest (also known as the "winter crop").The Rabi crops are sown around mid-November, after the monsoon rains are over, and harvesting begins in April/May. Their main source of water is rainwater that has percolated into the ground; they require irrigation. A good rain in winter spoils the Rabi crops but is good for kharif crops. The agriculture crops produced in India are seasonal in nature and highly dependent on these two monsoons [5].

Zaid crops (also written as **Zayad crops)** is on the Indian sub-continent, the crops grown on irrigated lands which do not have to wait for monsoons, in the short duration between Rabi and Kharif crop season, mainly from March to June. These crops are grown mainly in the summer season during a period called the "Zayad crop season."

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

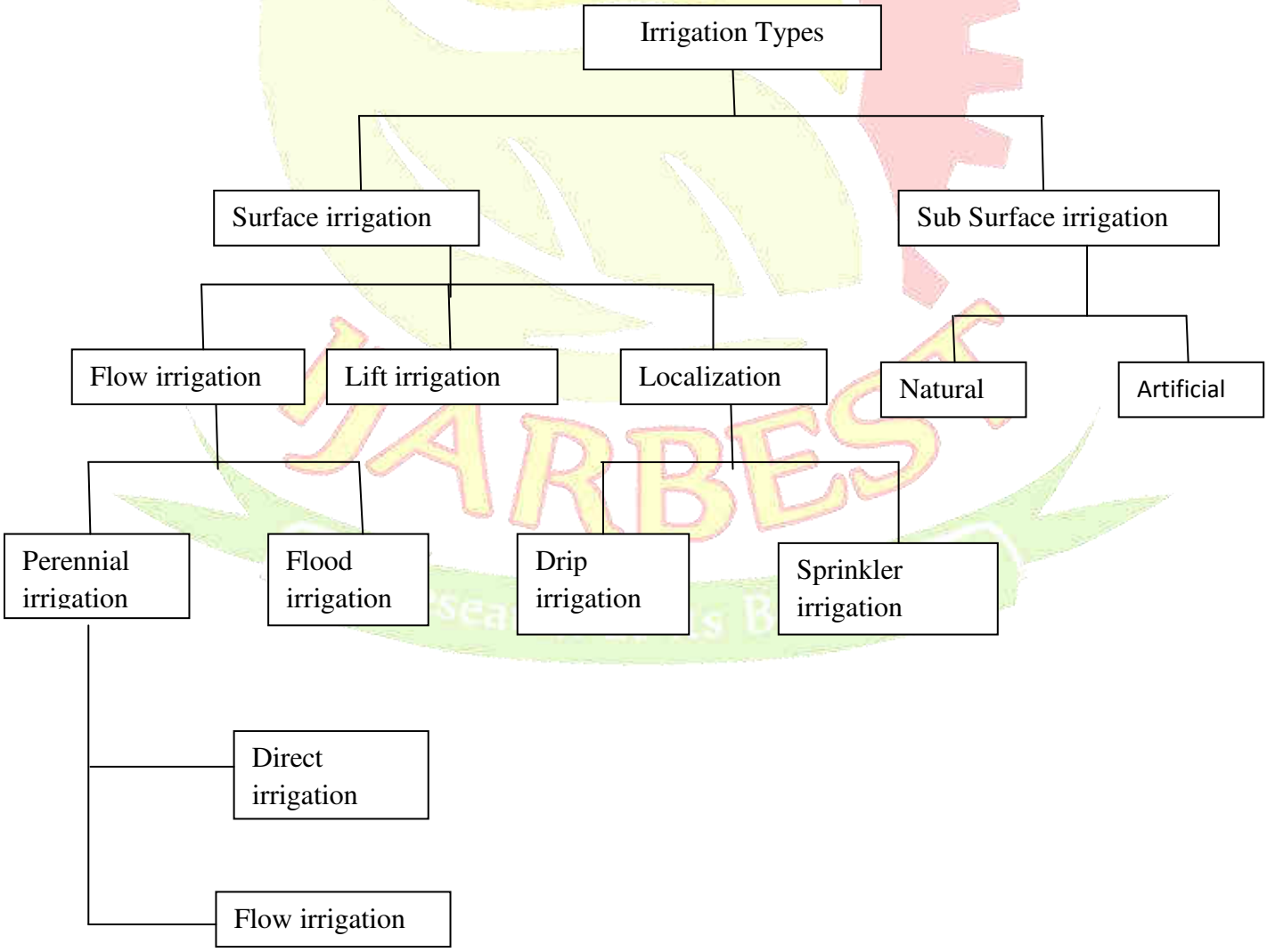
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

They require warm dry weather for major growth period and longer day length for flowering. The main produce is seasonal fruits and vegetables [6].

1.3. IRRIGATION IN INDIA

Indian irrigation infrastructure includes a network of major and minor canals from rivers, groundwater well-based systems, tanks, and other rainwater harvesting projects for agricultural activities. Of these, the groundwater system is the largest [1].



ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

1.4. FACTORS IN AGRICULTURE

* Quantifiable Factors in Agriculture: Seed, land, labour, machinery, fertilizer and manure and capital investment.
* Unpredictable Factors in Agriculture: Climate, rainfall, weather, flood and cyclone.

1.5. BIG DATA IN GENERAL

In general, big data shall mean the datasets that could not be perceived, acquired, managed, and processed by traditional IT and software/hardware tools within a tolerable time. Because of different concerns, scientific and technological enterprises, research scholars, data analysts, and technical practitioners have different definitions of big data [8].

Big data is an abstract concept. Apart from masses of data, it also has some other features, which determine the difference between itself and "massive data" or "very big data".

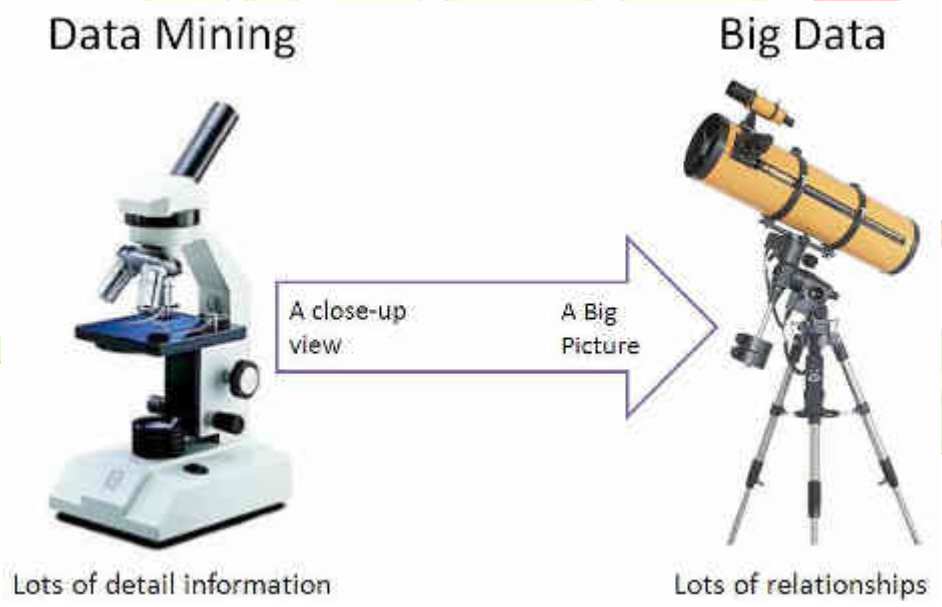


Fig: Difference between Data mining and Big data

Big data is an evolving term that describes any voluminous amount of structured, semi-structured and unstructured data that has the potential to be mined for information. Big data is a set of techniques and technologies that require

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

new forms of integration to uncover large hidden values from large datasets that are diverse, complex, and of a massive scale.

Big Data Analytics refers to the process of Collecting, Organizing, Analyzing, Inspecting, Cleaning, Transforming and Modelling, large sets of data (Big data) to discover patterns and other useful information. To identify the data that is most important to the business and future business decisions. It basically wants the knowledge that comes from analyzing the data.

Big data changes the way of world for management and use big amount of data. Some of the applications are in areas such as medical issues, healthcare, traffic issues, banking management, retail management, and education and so on. Organizations are becoming more reliable & flexible and more open [10].

Characteristics of Big Data Analytics:

Generally 4V's are followed in Big data analytics: they are Volume, Velocity, Variety, Veracity.

Volume: The quantity of data generated as Big Data ranges from Terabytes to Exabyte and Zettabyte of data. The volume has been increasing exponentially: up to 2.5 Exabyte of data is already generated and stored every day. This is expected to double by the end of 2015.

Velocity: Big data is growing rapidly, generating a bizarre of quantities needed to be stored, transmitted, and processed quickly. It refers to the speed of generation of data or how fast the data is generated and processed to meet the strain and the challenges which lie ahead in the path of growth and development.

Variety: This refers to the variation which can be shown by the data at times. In Big data, the variety and heterogeneity of data sources and storage has increased, fuelled by the use of cloud, web & online computing.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

Veracity: Big Data Veracity refers to the biases, noise and abnormality in data. Accuracy of analysis depends on the veracity of the source data. In comparison to Big Data's volume and velocity, veracity is the most challenging feature in data analysis.

Big Data in Agriculture: E-Agriculture service data can be considered as a Big Data because of its variety of data with huge volumes flowing with high velocity. Some of the solutions to the e-Agriculture service big data consist of the predominant present technologies like HDFS, Map Reduce, Hadoop, STORM etc [7].

2. RELATED WORK

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 centers [7].

MapReduce is a batch-oriented parallel computing model. There is still a certain gap in performance with relational databases. Improving the performance of MapReduce and enhancing the real-time nature of large-scale data processing have received a significant amount of attention, with MapReduce parallel programming being applied to many machine learning and data mining algorithms[9].

The authors said that, to provide high quality Farming techniques and also aim to ensure increased productivity of the crops to rural people and overcome the problems in the agricultural systems like use of harmful pesticides, excessive use of fertilizers, providing proper irrigation facilities and fraud management in the agricultural system. The proposed concept enables agriculturists, big-data analysts and staff to have role-based access to information on electronic farm records [7].

3. SUGGESTION

Here in this section we have given some of the tools and techniques to implement the big data analytics using agriculture data sets.

- **Hadoop** is an open-source software framework for storing data and running applications on clusters of commodity hardware. It provides massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs.
- **MapReduce** is a programming model and an associated implementation for processing and generating large data sets with a parallel, distributed algorithm on a cluster.
- **HDFS** is a Java-based file system that provides scalable and reliable data storage, and it was designed to span large clusters of commodity servers. **HDFS** has demonstrated production scalability of up to 200 PB of storage and a single cluster of 4500 servers, supporting close to a billion files and blocks.

* **HIVE** supports analysis of **large** datasets stored in Hadoop's HDFS as well as on the Amazon S3 file system. The best part of **HIVE** is that it supports SQL-Like access to structured **data** which is known as HiveQL (or HQL) as well as **big data** analysis with the help of MapReduce.

* **HBase** designed for very large tables with billions of rows and millions of columns, HBase is a distributed database that provides random real-time read/write access to big data. It is somewhat similar to Google's big table, but built on top of Hadoop and HDFS. Operating System: OS Independent.

4. CONCLUSION

Thus we have a tendency to conclude that, there's a growing range of applications and the role of Big data analytics techniques in agriculture and a growing quantity of information that area unit presently obtainable from several resources. This is often comparatively a completely unique analysis field and it's expected to grow within the future.

ACKNOWLEDGMENT

The second Author of this paper acknowledges UGC-Minor Research Project (No.F.41-1361/2012 (SR)) for the financial support.

REFERENCES

[1] https://en.wikipedia.org/wiki/Agriculture_in_India

[2] https://en.wikipedia.org/wiki/Agriculture

[3] https://en.wikipedia.org/wiki/Crop

[4] https://en.wikipedia.org/wiki/Kharif_crop

[5] https://en.wikipedia.org/wiki/Rabi_crop

[6] https://en.wikipedia.org/wiki/Zaid_crops

[7] Rupika Yadav, Jhalak Rathod,Vaishnavi Nair, "**Big Data Meets Small Sensors in Precision Agriculture**", International *Journal of Computer Applications (0975 – 8887) Applications of Computers and Electronics for the Welfare of Rural Masses (ACEWRM) 2015.*

[8] **Min Chen, Shiwen Mao, Yunhao Liu, "Big Data: A Survey",** Springer Science Business Media New York 2014.

[9] Rohit Pitre, Vijay Kolekar, "A Survey Paper on Data Mining With Big Data", **International Journal of Innovative Research in Advanced Engineering (IJIRAE), Volume 1 Issue 1 (April 2014).**

[10] Steve Sonka, "**Big Data and the Ag Sector: More than Lots of Numbers**", *International Food and Agribusiness Management Review Volume 17 Issue 1, 2014*.

[11] PARAG SHUKLA, BANKIM RADADIYA and KISHOR ATKOTIYA, "An **Emerging Trend of Big data for High Volume and Varieties of Data to Search of Agricultural Data**",**ORIENTAL JOURNAL OF COMPUTER SCIENCE & TECHNOLOGY,** ISSN: 0974-6471 August 2015, Vol. 8, No. (2): Pgs. 164-169.