Big Data analytics and Cloud Computing in Agricultural Research

A. K. Mishra[3]
Principal Scientist,
Div. of Ag. Engineering,
ICAR- Indian Agricultural Research Institute,
New Delhi-110012, India

Abstract: Big data analytics deals with large and complex heterogeneous datasets and the challenge is how to collect, analyze, secure, share, store, transfer, update and visualize the huge datasets. Cloud Computing technique is used to handle enormous agricultural data by analyzing and organizing it in meaningful manner for knowledge discovery. It will provide great facility to store heterogeneous data remotely and further use that data based on the demand of the user. Day to day data grows rapidly from different fields which create the problem to store and analyze it. Analysis of data sets will give new correlations and opportunities in business, medical and agricultural research. In the paper the characteristics, issues and challenges of big Data is discussed in detailed. In addition to that the benefits of integration between big data and cloud computing is also discussed.

***Keywords*:** Big data; Cloud computing; Data management; Analytics, Machine learning:

I. Introduction

After computer technology and internet technology, the new emerging technology comes from vast data generated by the combination of these two, termed as Big Data Analytics. Big data is the huge amount of data sets that is vast, varied, structured and unstructured and arrives much faster than you think. The data comes in various forms such as documents, emails, images, graphs, videos, personal information, data transactions and much more data obtained from various new technologies [1]. Cloud computing, a delivery model for Big Data Analytics, is a type of modern technology that uses remote servers and internet for data and application maintenance [2]. The integration of big data and cloud is beneficial to better insights and performance. Due to huge amounts of data available in today world, they need to analysed for knowledge discovery hence the cloud computing can handle these data with fast response times and real time processing of the available data. Data storage using cloud computing techniques is a viable option for small to large sized Industry considering the use of Big Data analytic. Cloud computing is on-demand access to computing resources (network) provided by an outside entity and require little management effort by the user. A number of architectures and deployment mechanism are available for cloud computing, and these are able to be used with other technologies and approaches. Users who are unable to afford adoption of large infrastructure can consider a number of cloud computing models to meet their big data needs. Users need to consider the optimal cloud computing architecture in order to remain both competitive and profitable. The real-time application of Cloud computing service is ranging from Medical to Agricultural research including manufacturing, pharmacy, medical imaging, environmental services, climate change study and many other relevant scientific fields because of data production at large [3]. India being an agriculture based as more than 70% of the country's population depends on farming and allied sectors. We need to increase our agricultural productivity to meet the ever growing supply-demand issues to feed the nation and export the surplus. Cloud computing is an effective measure in

the ICT (Information and Communication Technology) era, which is changing the Indian agricultural sector scenario at large [4]. With the Next generation sequencing techniques, huge amount of data is being generated from research labs varying from macromolecular sequencing, metabolic pathways, molecular interactions and mass spectroscopy. These biological data resources are very important from agricultural aspect [4]. The main purpose of this article is to provide in depth insight into the Cloud computing and its application in Big Data analytics and use and benefits of the integrated frame work in context of Indian agriculture. The paper focuses on implication of big data analytics and cloud computing services in present agricultural scenario with the outline of future prospects and challenges ahead.

II. Big Data

Big Data refers to data volumes in the range of exabytes .Big data is defined as an accumulation of huge and complex datasets and it is difficult to process the datasets using traditional database management and data processing Tools. Major challenges in big data processing is its collection, storage, integration, sharing, transfer, analysis, and visualization.

Every day we create 2.5 billion or quintillion bytes of data; 90% of the data in the world today has been created in the last two years alone. The data comes in various forms such as documents, emails, images, graphs, videos, personal information, data transactions and much more data obtained from various new technologies.

Big data analytics is a modern technique for collecting, organizing, maintaining and analyzing large datasets to discover new patterns and to uncover information which hidden inside the data by using remote servers and internet.

Big data that is at the exabyte level and that comes from various sources and heterogeneous in nature. Data growth is very fast and exponential in nature. Tools and techniques of small and medium size database management will not applicable for big data analytics. Big data here is discussed mainly in two contexts: Public and private data.

- ***Public-level big data:*** Public-level data consists of records that have been generated, maintained and analyzed through publicly available funding agencies, specifically some federal agencies. Such records are of Soil Survey data, Weather reports, etc. collected and maintained publicly.

- ***Private-level big data:*** Private-level data consists of records generated by direct involvement of farmers at production level. Crop yield, soil analysis report, etc, reports consists of such data in them.

Both, private and public data play the major role in making use of technology and analytical services for the benefit of the user. But they have their own challenges with respect to privacy and security is concerned.

III. Cloud Computing

"Cloud Computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management efforts or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models. The emerging cloud abstracts infrastructure complexities of servers, applications, data, and heterogeneous platforms [12]".

Cloud computing have five main universal values as the pillars of service:

- **Cost reduction:** User just need to pay per usage for time, service and storage needed. This way they can reduce financial load. Just like we can reduce electricity bill by proper and optimal use of

electrical devices. Anyone can monitor, control, and report the usage of the resource so the system is transparent.

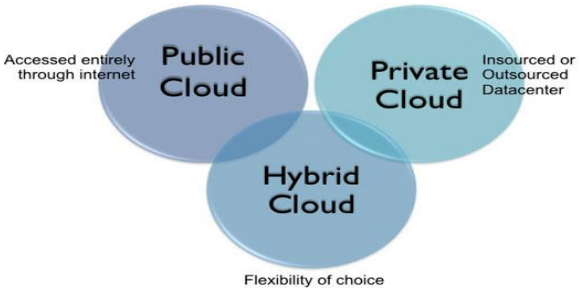
- **On demand resource allocation**: A user can itself cope up with the computing capabilities and resources, such as additional systems and network resources. The system is elastic so that the user can increase or decrease the resource based on the requirement.

- **Easy maintenance:** Upgrades and maintenance being performed as the back-end work. Cloud computing can be scaled up as and when the need arises. The user can also reduce the resources accordingly. The maintenance of resource is easy as it is the pain of the service provider.

- **Inter-Cloud collaboration:** The resources are shared at various levels such as, application, host, and network level. The pooled resources are used by multiple users where various physical and virtual resources are assigned and reassigned on demand of the user.

- The services are delivered as URL, HTTP and IP which provides benefits from the huge global network [5, 13]. With such an advanced technology, Cloud computing have become the major tool for delivering IT-related services with all of its complexity in a much simpler way, where all the users can remotely access all the information and data from the Cloud at their workplace without losing much of their time and cost at the same time[5]. Therefore, this technology is of significant importance in managing the agricultural information eco system, a great combination of agricultural information and modernization [6].

The cloud delivers its services on four cloud environments:



Public Cloud: The cloud services are made available on pay-per-use basis (on demand) or free, owned by an organization selling cloud services and very popular model.

Community Cloud: It shares services for commonly concerned community, managed either internally or externally.

Hybrid Cloud: It is a composition of more than one cloud, grouped together to enable data and application portability for user benefit.

Private cloud: The cloud operated particularly for a single organization. It can be managed by the organization itself, or may also be shared for specific purposes.

- ***Service Models***

Cloud computing comprises three major services:

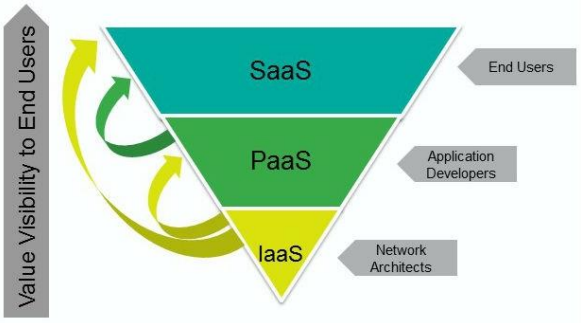


Fig 1. Service Models of Cloud Computing.

- ***Infrastructure as a service (IaaS):*** It provides delivery of hardware as a service i.e. operating systems, servers, storage network etc. where the clients can buy them as an outsourcing service directly.

- ***Platform as a Service (PaaS):*** It provides computer platform combining hardware and software as a service, where web based applications can be developed and tested on any software for approval and fulfils all the hardware requirements.

.

- ***Software as a Service (SaaS):*** It provides applications as a service for customers and provides software on demand. [8, 9].

IV. Agro-Cloud System

It is a Cloud service system in agricultural domain, and is based on MAD-Cloud Architecture. It is used to monitor the system functionalities and rendering the required services. For render its services Agro System have following services:

- Provides space for a comparative analysis of complete demand and supply chain.
- The system will provide information in the local Indian languages spoken across the country.
- The system incorporates mobile services and helps the farmers to acquire information from the data bank at real time.
- The system also shares information to the farmers, useful for them if any information is not available locally from the centralized database.
- It will help researchers to extract data from e-bank and analyses them to contribute in agriculture domain [10].
- Data analyzer service for farmers and agro-experts and officers in centers like Kisan call centres and KVKs.
- Data processing services like data sharing, computing, conversion, reporting and security, etc.

Agro-cloud is available for farmers to provide information like crop cultivation, diseases and pest management, fertilizer use etc. Currently, there is no system for plant protection without the use of sensing devices. Relevant information was provided to farmers though SaaS. Using SaaS Sensors were used to gather crop information.

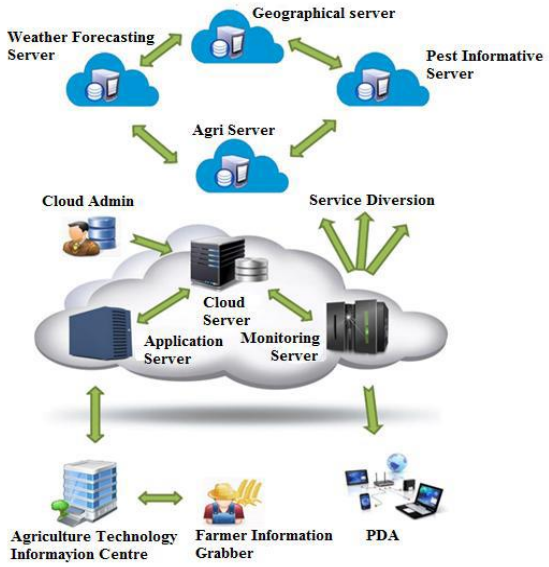


Fig 2. Overall architectural view of Agro-cloud system [7].

Applied Framework

MAD-Cloud framework gives the farmers an consultancy service to control the market pricing, cultivated crops, disease management, seed availability and fertilizers used in the fields, by the use of various modern IoT techniques such as mobile devices, scanners, sensors, etc.

It is a layered architecture containing various layers in it:

- MAD-Data Acquisition Layer (MDAL)
- MAD-Data Processing Layer(MDPL)
- MAD-Data Storage Service Layer(MDSSL)

MAD-Data Acquisition Layer (MDAL): MDAL is deployed as SaaS in Cloud, and its services can be easily used on multiple platforms by the users. It uses internet and IOT to provide user friendly interface to all stake holders like, farmers, experts and researchers to query data on their application devices such as tablets, mobiles, etc. Data is stored in Agri-DB layer.

MAD-Data Processing Layer (MDPL): MDPL is a data format converter. It converts the input data format to standard format and performs several

computation activities on large data sets and reports, in a user friendly manner of MAD-Cloud. It also provides services for Data security, data processing, Expert system generation, Decision making and report generation. It has been deployed as PaaS [5].

MDPL have been further divide into four modules:

- MAD-Secure Data Service module: It contains various libraries and providing authentication, integrity, and maintains secrecy of incoming data from all available sources.
- MAD-Data Processing service module: It contains libraries for data analysis, uniform data conversion format Agri-Cloud Data XML (.ACDXML). It performs computation, processing and data sharing as per requirements.
- MAD-Expert Service Module: This module contains libraries for providing solutions after image processing, error detection, sensor detection, and then generate reports as per the requirements of the user.
- MAD-Solution reporting Service module: It contains libraries to provide reports to the user in desired format as per specific query [5, 11].

MAD-Data Storage Service Layer (MDSSL): MDSSL acts as a data storage layer and this layer supports all the data formats for results to be fetched with higher precision. This layer is deployed as IaaS and allows data usage and information sharing [5].

Some of the popular agricultural databases are:

- Agri-Expert Knowledge Database (AKDB), containing inferences for decision making on real time.
- Image Knowledge Database (IKDB), contains the repository of images from the field to be used for decision making purposes.
- Statistical Knowledge Database (SKDB), allows the decisions to be made according to the available statistics of land.
- Business Knowledge Database (BKDB), is used to make business related decisions regarding market pricing or other commodity comparisons [11].

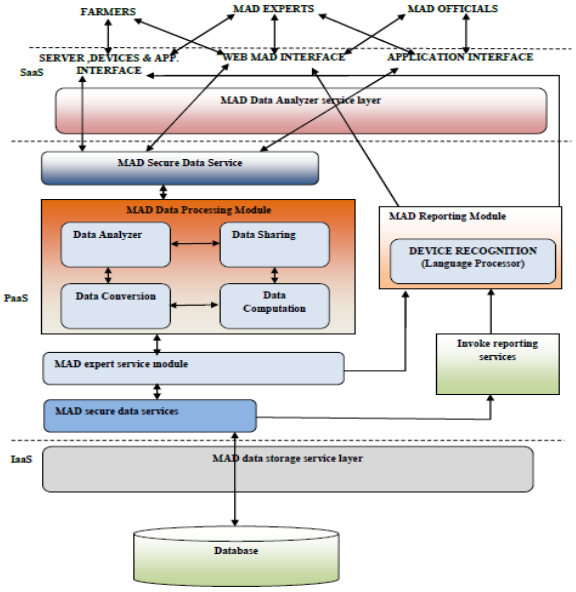


Fig 3. MAD-Cloud Architecture [11].

IT enabled agricultural system using big data analytics has the potential to drastically change the scenario of Agricultural eco system to facilitate the farmers by improving the overall crop productivity. In today world, technology can be used directly as the services and directly accessed by the end user, such as, Satellite technologies, remote sensing and IoT which can improve the crop productivity for enhancing income of farmers.

Advantages of Agriculture Cloud System in Agriculture in India are as follows:

- ***Systematic Data management:*** Data management becomes easier with the IT professionals being specialized in Data management and data warehousing and can be trained with latest data analytics techniques.
- ***Cost Reduction:*** Outsourcing of the agricultural related data analysis activities leads to more use of cloud computing architecture specially software as a service.
- ***Scalability:*** Information can be scaled up to increase the permeability and accessibility of the data.

- ***Security:*** It provides security to the resources as data are stored in the cloud and maintained centrally.
- ***Uniform Communication:*** The use of same cloud among farmers, policy makers and researchers leads to more appropriate and uniform communication leading to optimal outcome.
- ***Information Fusion:*** The sharing of common cloud platform for information sharing by Farmers, Researchers, Traders, Input suppliers, Retailers, research institutions and state agricultural universities, will results in a better way. Combining the resources among all stake holders can give information fusion on cloud environment and all the stake holders will be benefited [14].

Disadvantages of Cloud computing

- ***Security:*** Transferring data to all the stake holders is a big concern due to lack of security layers. It can be resolved by smartly selecting the cloud service providers based on the requirements.
- ***Bandwidth:*** Cloud computing demands very high internet speed. The available bandwidth today is only sufficient for fulfilling the basic and essential needs, but to meet the cloud computing requirements across the whole agricultural chain, more bandwidth is required for deploying the cloud services across the whole agricultural domain.

Future of Big Data analytics

- ***Data explosion:*** The agricultural data will continue to increase exponentially in coming years with the ever increasing internet connections and IT enabled devices.
- ***Data analysis strategy:*** With the emerging new technologies, there will be improvement in the data analysis techniques. As SQL is a standard for relational database queries, but some robust methods like Hadoop Spark will also come up with more advanced data handling capabilities.
- ***AI and Machine learning:*** Artificial intelligence and Machine learning will become the essential element in data representation and prediction modeling.
- ***Actionable data:*** The focus of future will be on actionable data in terms of quality, instead of only the size of data. The better the quality of data, the better will be the knowledge discovery from the data. Making rights decisions with right use of the big data will be the main focus in future. Therefore, Big Data concept will be replaced with actionable and robust data.

V. Conclusion

Big Data is now a value chain product across agriculture and the big data boom has attracted researchers across globe. Various forms of agricultural data are enabling today with analytical insights to help farmers and policy makers to understand the growth trends. In today agricultural world big data analytics is solving problems that were practically impossible to address decades before by enhancing the crop yield, responsiveness to the user and risk management at farm level. Agricultural data mainly relates to weather , soil, seeds, fertilizer, farms machinery , labor, GPS maps, available water resources, market trends, commodity market price, and inputs costs that are being exchanged increasingly from " Lab to land and from land to Market" for enhancing farmers income. The data in agriculture and allied sector are mostly unstructured or very raw, which is being generated at every step of agricultural supply chain management. If the available data is processed rapidly for fruitful information generation and based on the knowledge discovery and insights quick real time decisions will be taken. However, despite of enough resources and availability of data, agriculture sector does not leverage the benefit of modern ICT techniques. Today, pressure is increasing the farm productivity by minimizing input cost and maximizing the commodity price in the market. Big Data analytics can be used with cloud computing architecture to make the accurate, predictive and timely decisions across the Agri value chains. By the year 2050, world population is expected to hit 9 billion, widening the hunger gap. According to scientists this endemic problem can be combated

only by greater agricultural produce. In their quest to provide the sustainable food productivity, experts are constantly looking forward towards big data for food security to provide for future generation needs. Precision Farming contribute in many essential roles such as weather forecasting, cloud-hosted information resources, real-time farm machinery optimization using robotics, fertilizer use optimization, disease and pest management in extreme conditions, automated irrigation with minimum water , input cost optimization price monitoring and market trend analysis using latest IT enabled technologies. Having access to such technologies on a mobile platform enables the farmers to take on spot decisions for the better results. Where, there is no perfect clue as to the use of sophisticated farming tools for solving problem of food scarcity, big data analytics in as a bright streak of light to optimize and track the harvests.

Cloud computing integration with big data will play a critical and vital role in coming years for agricultural growth and overall sustainable development of the nation.

REFERENCES

[1] K. K. Bharathi, "Converging Technologies of Cloud and Big Data", IJCSET, 6(1): 75-77, 2016.

[2] S.B. Jadhav, R. Prasad, S.S. Panhale and C.S. Mohture, "Review of Cloud Computing and Its Application," International Journal of Advanced Research in Computer Engineering & Technology, 2013

[3] M. Hori, E. Kawashima and T. Yamazaki, "Application of Cloud Computing to Agriculture and Prospects in Other Fields," FUJITSU Science Technical Journal, 46(4), pp. 446-454, 2010.

[4] S. Prasad, "Agro Mobile: A Cloud-Based Framework for Agriculturists on Mobile Platform," International Journal of Advanced Science and Technology vol. 59, pp. 41-52, 2013.

[5] Karuna Chandraul and A. Singh, "An Agriculture Application Research on Cloud Computing," International Journal of Current Engineering and Technology, 3(5), 2013.

[6] Y. Zhu, D. Wu and S. Lil, "Cloud Computing and Agricultural Development of China: Theory and Practice," International Journal of Computer Science Issues, 10(1), 2013.

[7] T. Prasath and M. Babu, "Agro-Cloud: Information Providence Based on Monitoring Service Via New Computational Gist," International Journal of Engineering Research & Technology (IJERT) 3(11), 2014.

[8] N. Swarnkar, "Application of cloud computing technology for rural development," International Journal of advanced Computer Technology, 2(5), 2013.

[9] D.T. Satej and S.R. Kumbhar, "The Study of Cloud Computing Challenges in Agriculture with Special Reference to Sangli District (MS)," International Journal of Computer Applications, ICAET, 2015.

[10] R. Patel and M. Patel, "Application of Cloud Computing in Agricultural Development of Rural India," International Journal of Computer Science and Information Technologies, 4(6), 2013.

[11] K. Venkataramana and M. Padmavathamma, "A Design of Framework for AGRI-CLOUD," IOSRJCE, 4(5) pp 01-06, 2012.

[12] R. Panigrahi, M.K. Ghose and M. Pramanik "Cloud Computing: A New Era of Computing in the Field of Information Management," IJCSE, 2(5): 189-195, 2013.

[13] S. Kalghatgi, K.P. Sambrekar, "Review: Using Cloud Computing Technology in Agricultural Development," International Journal of Innovative Science, Engineering & Technology, 2(3), 2015.

[14] B.D. Patel, R.G. Thakkar, B.L. Radadiya. Prospects of Cloud Computing in Agriculture in India. Agricultural Science, 3(11), 2014.

[15] Megan Stubbs, "Big Data in U.S. Agriculture," Specialist in Agricultural Conservation and Natural Resources Policy, 2016.

[16] Shefali Singh, Pooja Sharma, Himansu Awasthi & A.K.Mishra "Big data analytics: A reality in virtual space" Proceedings of Indiacom-2017

[3]Dr Mishra, working as a Principal Scientist (Computer Application & IT) in ICAR-Indian Agricultural Research Institute, Pusa Campus, New Delhi, a premier and autonomous research institute under Ministry of Agriculture and farmer welfare(GOI) and a deemed university .

He is a senior member of CSI, IACSIT and ISAS and presently Vice Chairman, CSI (Delhi), having more than 23 years of experience as a scientist in research and teaching.