WALAILAK JOURNAL
http://wjst.wu.ac.th MiniReviews

A Big Data Virtualization Role in Agriculture: A Comprehensive Review

Sandeepkumar MATHIVANAN and Prabhu JAYAGOPAL*

Department of Software System and Engineering, School of Information Technology and Engineering, VIT University, Vellore-632014, Tamil Nadu, India

(*Corresponding author; e-mail: jprabhuit@gmail.com)

Received: 3 March 2017, Revised: 16 June 2018, Accepted: 13 July 2018

Abstract

Big data is a collection of large volumes of data sets which are more complicated to analyze using standard data processing methods. It also emphasizes parameters like data variety and velocity data. Big data will play a most significant role in our daily life regarding applications like healthcare electronic commerce, agriculture, telecommunication, government, and financial trading. In the agriculture domain, big data is an optimal method to increase the productivity of farming by gathering and processing information like plant growth, farmland monitoring, greenhouse gases monitoring, climate change, soil monitoring and so forth. Virtualization is an emerging technique that can be combined with big data in agriculture. Virtualization has been used extensively in research for a long time, the term "virtual" entities affecting a real-life form. In agriculture, it has many more physical objects, sensors, and devices. This physical object is virtualized and has digital representation to store, communicate and process via the internet. The information from the virtual object has a large volume of data which helps meaningful data analysis or aspects to make application services like decision making, problem notification, and information handling. This paper provides a comprehensive review of big data virtualization in the agriculture domain. The virtualization methodology, and tools used by many researchers is surveyed.

Keywords: Virtualization, big data, agriculture, virtual object, decision-making methods

Introduction

Big data primarily consists of 3 V'S: volume, velocity, and variety. For the specific issue of big data, duration is not essential in processing using the right technology and technique in order to expresses the value [1]. Some other characteristics of V's are variability, veracity, value, venue, verbosity, verification, visualization, validity, vocabulary, etc. The fundamental components of big data are processed within a local computing structure using standard approaches and techniques to handle the datasets capably. Data virtualization is a comprehensive method to deal with a large volume of data from various sources like storage, database, system file, and primary memories. The main advantage of data virtualization includes reducing the risk of data error, workload, development, data storage, policies and also an increase in the speed of data access. Data virtualization supports SQL virtualization for unified access, query, reporting, predictive analytics, back-end data repository like Hadoop, and NoSQL. Virtual data is present in the abstract layer and also works in infrastructure layers. The storage and server platforms are also in multiple cloud applications. Due to the increasing complexity of big data, we require data virtualization, and its infrastructure.

In the agriculture domain, the big data virtualization concept plays a significant role in a variety of data such as weather data, GPS data, soil specifics, seed, fertilizer, and sensor data. Nowadays the growth of the internet as well as the development of modern society, digital agriculture development is also necessary. Big data virtualization helps the farmer to make better decisions through decision-making support tools or methods to attain needed resources. Agriculture plays an important role in the growth of the economy in India. It mainly depends on 58 % of rural household in agriculture [2]. Challenges lie in delivering sustainable strong outcomes for farmers and require new ideas in agriculture. Big data analytics and investigation deserve temporal changes in the cropping pattern in the agriculture field. Some of the technology used in the agriculture field includes satellite navigation, sensor networks, and grid computing used for improving monitoring and decision making capabilities [3]. The decision making process in the agriculture domain needs to grow by combining current local environment and agro monitoring in GIS (Geographic Information System) and WSNS (Wireless Sensor Networks) [4]. Variation sensors used in the agriculture domain like precision agriculture, precision farming, site specific crop management, and variable rate technology [3]. Some of the issues in a wireless sensor network are energy consumption, data acquisition, sampling, transmission, fault tolerance, the size of nodes in the sensor, and sensor placement.

Related work

Goya *et al*. [5] described the decision making for supporting agriculture with a cloud computing platform, distributed processing, and technology in the framework of big data. Weather data do the metric calculation, and the solution to test in the different virtual machine scenario, and configuration like a cluster in helping the farmer to get a better prediction. In the future, big weather solution replication will be applied to the physical machine in the form of clustering.

In MCC (Mobile Cloud Computing) for better cultivation and marketing, Ghosh *et al*. [6] describes an application called Agro mobile cloud computing; it mainly focuses on MCC on crop image analysis which consumes more memory; causing high power consumption making mobile devices fail. The idea of MCC is a crop image analysis that helps to limit the data storage, processing power and server called Agro mobile server is established. The application service providers demand software called Software as a Service (SaaS) which acts as a bridge between the cloud developer and customer where information is transferred. The advantage of this proposed work is the reduction of the workload and complexity of image processing on the cloud server. Future work on the Agra mobile architecture could apply algorithms for crop analysis, cloudlet, virtualization concepts on Android OS and the Internet of Things (IoT) application platform.

Hauhui *et al*. [7] described the VIDB (Virtualized Information Database) used for solving issues for large agriculture data, data replication, concurrency, transmission. It gives service in support of organizations, sharing, and broadcast in agricultural research. The XML metadata service, resource information service, resource server monitoring service is the 3 virtualized database access interface. It smartly deals with failed storage nodes due to more scalability and fault tolerance in the database.

Xie *et al*. [8] presented the collection, storage, analysis and visualization of agriculture big data. Data collection used for getting resource from various places like the web, sensor, and network. For storing and maintaining large data, NoSQL was efficiently utilized. MapReduce and Hadoop were used for analyzing big data and mining. Presentation of virtualization was determined based on attributes, variables and required information. It used extreme data technology like Spoop, HDFS, Hive, Mahout, and Karmasphere.

Verdouw *et al*. [9] presented a virtualization concept applied to the IoT domain for food supply chain. Virtualization is a well-versed approach to maintaining complexity, wherein the food chain scenario makes the decision easy to handle. Virtual is assigned objects, network, control, and process. Architecturally designed with an F1 space platform, this platform helps enable genetic technology to build virtualization in IoT and cloud computing.

Kruize *et al*. [10] described a farm software ecosystem reference architecture which provides access structure, control, map in contribution with FMIs. The features of this approach in connecting with ICT (Information communication technology) components like hardware, software, and service modules. Mapping is with the essential component similarity and difference. The main idea behind this system is to upgrade in configuration and make a difference in the ICT component. The primary usage of a reference architecture for better interaction and collaboration with various vendors in real time software ecosystem. We need to know system performance with join, form, and upgrade in the software ecosystem. Primary research in the development of farm software ecosystems like technology, vision, farm information model, collaborative tools, and programming interface. FMIs (Farm Management Information system) with ICT component with one or more application in the gathering of data, processing, and storage on farms. In the future the author decided to enhance the configuration of the farm software system.

Ojha *et al*. [11] described wireless sensor network issues, and challenges related to improving the performance of farming. It mainly focuses on some requirements like device, sensor and communication techniques related to the wireless sensor network in agriculture. A TWSN (Terrestrial Wireless Sensor Network) was deployed above the surface using modern MEMS (Micro electro mechanical system) technology, it can perform with small size sensor at low cost. The dominant sensor node will get more accurate in gathering data from the surroundings. For example, in precision agriculture, platforms where WSN are performing agriculture field for self-regulation in the irrigation system. An entire sensor performs to get moisture from the soil, then makes a decision, time in irrigation schedule in the agriculture domain. The decision is sent to the sensor node that is combined with a water pump. Gutierrez *et al*. 2014, described a self-regulation irrigation system using a WSNs and GPRS system. WUSNs (Wireless underground sensor network) are placed in the soil, and with a limited communication radius, extra nodes are needed in large fields. The wireless sensor is high when the network coverage requires a low number of sensors.

Risk and water management, irrigation, and improper infrastructure are some of the issues faced by farmers while cropping; it leads to reduced crop yield and a loss for the farmer. To overcome this issue Patil *et al*. [12] introduced smart agriculture in IoT. They have specific features namely temperature, humidity detection, soil moisture detection, wind speed, etc that are monitored allowing the farmer to handle the crop in a better way. Shen *et al*. [13] describes multi-net interconnection, virtualization [14], integration, and engineering technology as part of agriculture information and technology development.

This paper aims to comprehensively review big data virtualization's role in agriculture. The remaining of this paper is organized as follows: In the next section 2: related work. In section 3: Importance of virtualization in big data and requirement of virtualization is discussed. Section 4: Represents big data in agriculture. In section 5: New trends in big data agriculture is explained. In section 6: the Decision-making techniques used in big data are distinguished. In section 7: Research gaps of virtualization and finally section 8: describes the conclusion.

Importance of virtualization in big data

For solving challenges of big data we need a broad amount of data distribution in storage, use of a computer, and a data-intensive application. Virtualization affords an additional level in analyzing big data in reality. Virtualization is not technically required for analysis of big data. A software system is more productive on a virtual platform. Virtualization mainly consists of 3 characteristics for an affordable, scalable operating system for big data processing [15].

- **Partition:** It affords a single physical system, it partitions only based on available resources.
- **Isolation:** Each VM is isolated with a physical system and other virtualization. When a virtual instance crashes, other VM will not be affected. Data will not be shared between one virtual instance and another.
- **Encapsulation:** VM mainly depends on a single file, it is easy to find based on the service provider.

Requirement for using virtualization

Figure 1 shows the wireless sensor network distributed on the field for agriculture applications. Mainly the field consists of sensor nodes that supply information with the application, particularly on the sensor board. The nodes in the field sensor network transmit using Radio Frequency (RF), links of Industrial, Scientific and Medical (ISM) and radio bands. Then the gateway nodes provide both RF and GSM (global system for mobile communications). A remote user will monitor the agricultural field and power on the field sensor and actuator devices. For example, a user can switch on/off or pump/value the water level employed in the field to reach some threshold value. Users will move forward to a mobile phone to monitor and control the on-field sensors. It connects through GPRS and SMS (Short message service). Simultaneously information gets an update from the sensor and system to power both types of user. For saving time, resource optimization, reduce power system, installation of software quickly in the environment, and maintenance, utilization of the CPU increases from 15 to 80 % [16].

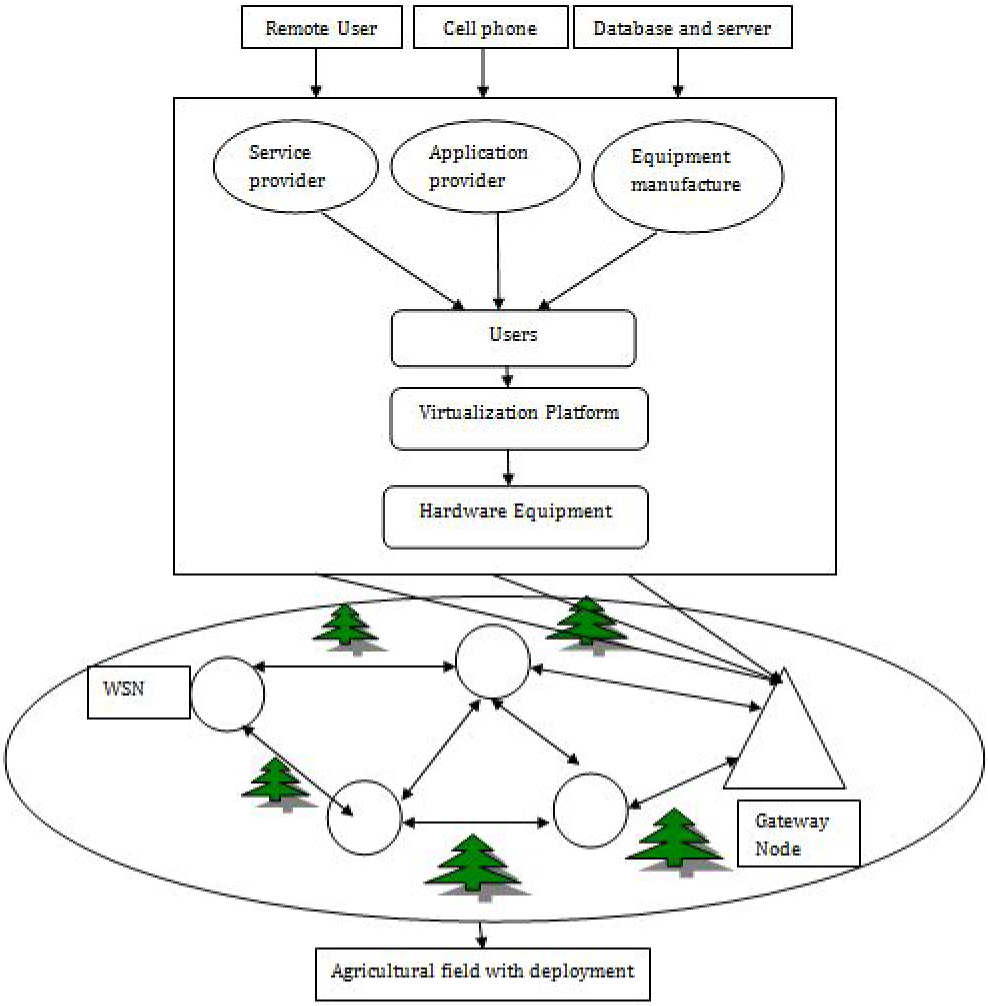

Figure 1 Big data virtualization in agriculture.

Role of big data in agriculture

Most of the data from sensors depends on the moisture, nutrients, atmospheric, and pressure. Sensors help the farmer to locate density seeding for optimizing sowing. Local conditions help to find the regions where more water and fertilizer are needed. Self-driving vehicles give data regarding the location, fuel, consumption, and fertilizer, provided by IoT, or telematics. Farm machinery can be implemented without any help. Most of the issues in agriculture are due to instability; change in climate, soil erosion, and loss of biodiversity. The current agriculture system quickly becomes complex when the economic problems, nutrition factors and ecological conditions are taken into consideration [17].

Comparison of big data virtualization in agriculture

Table 1 Comparison of big data virtualization in agriculture.

S.No	Author	Description	Architecture/Methodology	Benefits
1	Goya *et al*. [5]	A decision-making system for supporting agriculture with cloud, processing technology in big data	Big weather architecture introduced based on 3 essential elements of a web portal, data server, and a Hadoop cluster. It used to get new matrices like average humidity, and temperature.	It helps the farmer to improve decision making and the number of virtual machines in the cluster.
2	Ghosh *et al*. [6]	Mobile Cloud computing used for better cultivation and production. Analysis limitation in data storage, power, and server.	The Agro mobile system is mainly focused on analyzing image due to power, memory; it makes mobile development fails.	The main benefit of this work is a decrease in workload, and the complexity of image processing in the cloud.
3	Hauhui *et al.* [9]	Used for solving issues with last data sets, replication of data, concurrency, transmission, etc.	VIDB architecture consists of 3 layers like a database, access layer, and logical storage. Some other layers are physical storage and agriculture resources.	It provides multi-resource and organization etc.
4	Verdouw *et al.* [10]	The virtualization concept applies to the food supply chain. In this scenario, the decision can be taken quickly, but virtualization used for maintaining complexity.	Virtualization supplying food chain architecture has process control, virtual and the real object. It also provides an information system; it consists of an application service, information integration, data exchange, and sensing, actuating.	The system provided self-control, operation, decision and learned without the necessity for humans.
5	Kruize *et al.* [15]	FMLs (Farm management information) with ICT components with one or more application in collecting data processing, storage on the farm.	Farm software ecosystem is based on actor, platform, ICT component and business service. The technology used is like vision, farm information model, collaborative tools, a programming interface.	It identifies the difference and similarity in farm software ecosystem. Better growth in the configuration of the ICT component.
6	Wang *et al.* [18]	Built on a cloud computing server, storage with various OS stages like environment, application in differentiable with digital agriculture information.	Xen tool is used for server virtualization and storage virtualization in digital agriculture. It provides scalability and flexibility.	Xen software has low-cost, virtual structure, storage in digital agriculture.
7	Ke [19]	IoT and RFID mixed with cloud computing and smart agriculture data. Resources needed such as a controller, load balance, distributed resource, etc.	Agriculture information cloud based on planting, productivity, control, security, and growth monitoring. Other architectures like cloud hardware based on data center using presentation layer, user resource, and management system.	Useful in solving agriculture issues. Quick improvement in materials, agriculture, and smart agriculture.
8	Saguy [20]	Represents food engineering challenges and opportunities. Enginomies introduces in the field based on consumption, SR, human operation and food products.	Food engineering consists of 4 components: modeling, open innovation, social responsibility, and virtualization.	Sustainability needs of customer, expectation, and SR.
9	Xing *et al.* [21]	By using virtualization in cloud computing, we acquire more advantages and challenges in data protection and security.	Cloud computing architecture provides layers like user, service, virtual, resource, physical layers	It offers improving security for cloud computing.

Crop rotation methodology in big data

A crop assists the farmer to change the soil for an alternate approach for cultivating shallow root crops. A crop of plant food is tapped in various depths at a different time. The advantage of control of weeds, pests, and disease, is that soil fertility is maintained. Resource-constrained agriculture crop scheduling has the advantage in being time-consuming, complicated process in determining accuracy in the irrigation process. Irrigation is based on the duration of water for demand, the various types of soil, and the geographic climate [2].

Big data in digital agriculture using satellite data

- **Disruptive technologies**: Potential in farming productivity in agriculture is suitable for the remote satellite. Target design innovation in service and strategies in management. By combining these estimates information can be provided on the scale the cropping region, the potential yield and the gap. The yield is based on climate, nutrient stress, etc. Annual crop estimates are based on the production of food proportional to the area cropped. The measuring conditions affect the estimated crop yield in a specific time and space. Remote sensor data, soil data, climate, weather conditions and stage model in the growth of crop development in performance outcomes depends on the scenario range in management and environment, and the measure yield can perform in the remaining season [22].

- **Types of sensor used in big data agriculture**

Agricultural sensors are required in subtle weather conditions and need to be strong. Most of the sensors available in the market are suitable for one type of cultivation. Selection of a sensor is mainly affected by quality, infrastructure, the range of measurement, and response time. Input parameters, material type, technology, transduction are some of the variations in the sensor. The DHT11 sensor will measure both temperature and humidity at the same time. Some of the sensors that measure automatically use electromagnetic, optical, mechanical, electrochemical, airflow, and acoustic sensors [23].

Comparison of the soil-related sensor [24]

Table 2 Comparison of soil-related sensor.

Sensors	Features
Pogo portable soil sensor	Soil moisture, water flow, soil temperature, conductivity, salinity
Hydra probe II soil sensor	Soil moisture, water flow, soil temperature conductivity, salinity, the water level
ECH2OEC-5	Soil moisture
VH-400	Soil moisture, the water level
EC-250	Soil moisture, water flow, water level, soil temperature, conductivity, salinity
THERM200	Soil temperature
Tipping bucket rain gage	Water flow
Aqua Trak500	Water level
WET-2	Soil temperature, conductivity, salinity

Comparison of the plant-related sensor [24]

Table 3 Comparison of plant-related sensor.

Sensor	Features
Leaf Wetness sensor	Moisture
237-L, leaf wetness sensor	Moisture, temperature, wetness
LW100, leaf wetness sensor	Moisture, temperature, wetness
TPS-2 portable photosynthesis	Moisture, temperature, wetness, CO_2, photosynthesis
CI-340 handheld photosynthesis	Moisture, temperature, wetness, CO_2, photosynthesis, hydrogen
PTM-48A photosynthesis monitor	Moisture, temperature, wetness CO_2, photosynthesis

Comparison of the weather-related sensor [25]

Table 4 Comparison of weather-related sensor.

Sensors	Features
CM-100 compact Weather Sensor	Temperature, humidity, atmospheric pressure, wind speed, wind direction
Met Station One (MSO)	Temperature, humidity, atmospheric pressure, wind speed, wind direction
XFAM-115KPASR	Temperature, humidity, atmospheric pressure
HMP45C	Temperature, humidity, atmospheric pressure
SHT71 (Humidity and temperature sensor)	Temperature, humidity, atmospheric pressure
SHT75	Temperature, humidity, atmospheric pressure
CI-340 hand-held photosynthesis	Temperature, humidity
107-L temperature sensor	Temperature

Leveraging in big data

Some of the main factors that contribute towards profitability for the farmer are hybrids, pesticides, air moisture, ground moisture, water availability, temperature, rainfall, price forecasting, government actions and market data and can be leveraged using big data [26].

From above-represented attributes, the big data framework and machine learning algorithm play a critical role. It includes 1) To get optimum decisions for farming, 2) crop and intercropping recommendations, 3) proper selection of suitable hybrids, 4) farming practices, 5) pests predication and management, 6) forecast the argicommodity prices ahead of the season, 7) profitability analysis, and 8) policy recommendations.

By utilizing big data frameworks it could be significant volume, variety and veracity can handle with more computational machine learning algorithm can be introduced. Many advantages are possible utilizing the agriculture big data environment. Optimized farming and commodity pricing are 2 significant benefits.

- **Optimized farming**

Weather, monsoon behavior, groundwater scarcity, soil conditions, labor and machinery cost, intercropping decision, pest's management are some attributes used to prepare in advance and are associated with big data technology and machine learning algorithms. All the characteristics that relate to making an optimized decision in various stages of farming.

The optimized decision can take by decision farming based on weather, soil, crop cutting, plant health, pest management and intercropping. The decision taken at every stage to ensure profitability, low production cost, reducing farmer's risk, high productivity, no pesticide residue is assured and efficient use of land, machinery, labor and time.

- **Commodity pricing**

Farmer get an advantage from forecasted agriculture commodity prices and sharing the present prices of agriculture commodities.

- **Forecasted agriculture commodity process**

The price of commodities move substantially in semi-arid farming and monsoon farming zones. The local government inputs price moves due to the decision like MSP (Minimum Selling Price).

- **Sowing decision by a farmer**

The price forecasting information assists the farmer to know about the right price in advance to make an appropriate decision whether to sow a specific crop or not.

- **A policy decision by the government**

The price forecasting information plays as input to government and other authorities to decide on the minimum selling price.

- **Sharing the current commodity prices**

Information may allow a farmer in many ways namely automation of email or SMS alerts, by utilizing a mobile or internet app, and advertising in media through media analytics.

A new trend in big data agriculture

Drones: Also called an UAVs (Unmanned Aerial Vehicles). For many reasons, it is used in most industries. UAVs are essential and economical in capturing data. Previous data is only available from aircraft satellite images. The orthomorphic model can implement definite drones it can be synced with the program, and it creates the maps. The map can transform with the same applicator it saves more time and the effort for farmers [27]. Cloud computing, wireless sensors, communication, network technology, embedded system, data mining and NANO, IPv6 are some of the essential things to provide to a farmer for innovating in agriculture and in getting innovation technologies.

Sensor: In agriculture, sensors used most frequently in many ways like the availability of water, soil, measuring the temperature of the leaf, detecting disease and insects. The central concept behind the sensor makes farmers cut down on the unwanted application. Maps can show regions of the field where the moisture is okay, hence no need for water at that time. This type of data helps the farmer to save time, resources, and money.

Data analysis: Farmers to check the history of fields, crops, and health issues before they sow. Farmers use the analysis in making a decision, helping them to improve their yield and get more benefit.

Technology Adaptation: Big data technologies need to be adapted from the initial stage to the final stage of crop cultivation which is shown in **Figure 2** [28].

Big Data: The farmer needs to understand the challenges of big data, tools, and technology from the local level. This approach takes more time, practice, and error. The performance of big data is efficient and is valuable in agriculture. Future precision agriculture needs to perform essential/mandatory tasks; the basis depends on timing, location, data [28].

Biological: Most of the chemical factories rely on elements like weeds, insects, and diseases to perform in the biological organism [28].

Regulation: Due to preventative trends in agriculture, the farmer gets enormous pressure in handling entire commodity and product grown based on demands. Need to have changes and modification in the cultivation of crops in the future. Producers and consumers have more issues than cost only. It requires more than 2 years for constructing the regulation and innovation in DB. Water needs have been appropriately maintained by a farmer only [28].

Demands: When population increases beyond 9 billion by 2050, we get more demand in supplying food, fiber, energy to the world. Agriculture production will play an essential role in improving the world market [28].

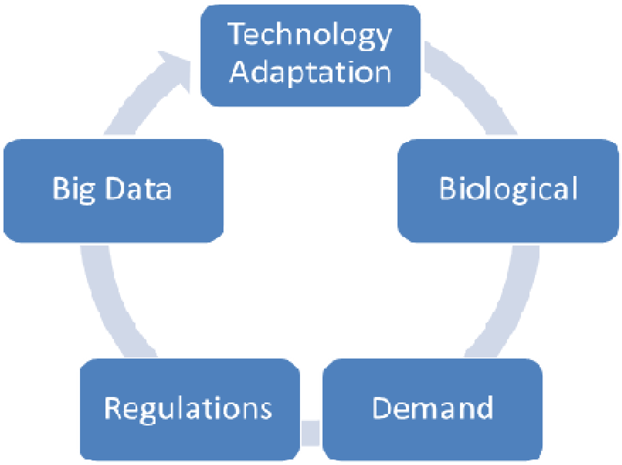

Figure 2 Future trends in big data agriculture.

Table 5 Overview and summary of the development of big data and smart farming in push and pull factors [29].

Push factor	Pull factor
Basic technology developments • IoT and data-driven technologies • Precision agriculture • Push of agritech company	Business driver • Efficiency improved by low-cost price or enhanced market price. • Efficiency in particular management support • Improvised management control and decision-making. • Handle with volatility in weather condition
Sophisticated technology • Satellite imaging • Advance remote sensing • Robots • Worldwide navigation satellite systems • UAVs	Public drivers Food and nutrition security Sustainability

Push factor	Pull factor
Data generation and storage • Advanced data analytics • Process-machine and human-generated • Interpretation of unstructured data	The general requirement is more and improved information
Digital connectivity • Improving availability of agriculture practitioners • Computational power increase	-
Innovation possibilities • Open farm management systems with particular apps • Remote/computer-aided advise and decisions • Regionally grouped data for scientific research advice • Online farmer shops	-

Table 6 Summary of big data application in smart farming and critical issues [29].

Various stages of the data chain	Description	Key issues
Data capture	Sensors, open data, data captured by UAVs [40], biometric sensing, genotype information and reciprocal data [30].	Availability, quality, format
Data storage	A cloud-based platform, Hadoop Distributed File System (HDFS) hybrid storage system and a cloud-based data warehouse.	Fast and safe control of the data cost.
Data transfer	Wireless, cloud-based platform, linked open data	Safety, agreement on responsibility and liabilities
Data transformation	Machine learning algorithms, normalize and visualize anonymize.	Data source, automated data cleaning, and preparation.
Data analytics	Yield models, planting instructions, benchmarking, decision ontologies and cognitive.	Semantic heterogeneity, real-time analytics of data cleaning and preparation.
Data marketing	Data visualization	Ownership, privacy, innovation business models

Table 7 Various stages of smart farming.

Stages of smart farming	Arable	Livestock	Horticulture	Fishery
Smart sensing and monitoring	Robotics and sensors [30,31]	Biometric sensing, GPS tracking [35]	Robotics and sensors (Temperature, humidity CO_2, etc.) greenhouse computer [39]	Automated Identification Systems (AIS) [43]
Smart analysis and planning	Seeding, planting, soil type, crop yield monitoring [32]	Breeding, control [36]	Lighting, energy management [40]	Surveillance, monitoring [44]
Smart control	Precision farming [33]	Milk robots [37]	Climate control, precision control [41]	Surveillance, monitoring [44,45]
Big Data in the cloud	Weather/climate data, yield data, soil types, market information [34]	Livestock movements [38]	Weather/climate, market information, social media [42]	Market data, satellite data [46]

Decision-making techniques in big data

- **Mathematical techniques**: It supports in stages like data curation and data analysis. The central concept of decision making depends on the relationship, correlation, and samples in the statistical approach. Innovation implements for maintaining large data using parallel statistics, statistical computing, and learning [47]. Optimization method used for solving issues at more cost for memory, time consumption data reduction [48] and parallelism respectively. Real-time optimization can define the decision-making problem for the vast scale wireless network [49] and intelligent transportation system [50].
- **Data analysis technique**: It mainly consists of data mining, machine learning, and neural networks. Most of them can process big data. Data mining is categorised into classification, clustering, and regression. It hides knowledge and pattern from the given data [51]. Many classifications and clustering algorithms have substantial data samples as a fuzzy based system, and a clustering colossal application algorithm. Fuzzy reasoning provides uncertainty in both data and output.
- **Machine learning**: It is another artificial intelligence technique that allows both supervised and unsupervised methods. A machine learning algorithm like SVM improves the performance of a large-scale parallel system like map/reduces. It will be used by many applications like biological [52] and sensor data [53].
- **Artificial neural network**: Mostly applied in pattern recognition and adaptive control. The main challenges lie in neural network layers, nodes, performance, and memory and time consumption in the neural network. Sampling techniques used for reducing the size of data. Neural networks are applied in parallel and distributed setting [54].
- **Visualization**: Nowadays the representation of data is too complicated. Its particular focus on significant data and try to identify the proper description of data [55]. Feature extraction provides a reduction in the size of data [56]. Thompson *et al*. [55] show the presentation of data in a compact yet informative approximation for massive data.
- **Cloud computing**: The cloud provides services like platform, infrastructure, decision-making software. Data management, data quality, data currency is the main issues in cloud decision making. Demirk and Delen [57] presented data, information; analytics are described as a service-oriented decision support system.

- **Fuzzy sets and system**: Related to the GrC (Granular Computing) technique, it is more useful for a solving a big data. They are some techniques in fuzzy sets like neutral fuzzy classifiers [58], big data classification represents the linguistic fuzzy rule-based classifier. For clustering of big data the most useful algorithm use a Fuzzy C-system. The fuzzy interface system, the Bayesian process [59], a fuzzy query system is used for the pattern recognition algorithm. For reducing dimensionality a fuzzy neural classifier is used.

Research gaps

1. Change management is an open research problem where the change is not applicable to all kinds of application. The difference does not exist in user sharing of the same virtualization service.
2. Nowadays, more complex analytic models, rules, information converge exists in big data.
3. Research should focus on the development of the infrastructure framework because of some undefined virtual architectures.
4. Data volumes have increased to a terabyte, petabyte and beyond. In information technology, the volume is a big challenge to store the data in the form of logical tables, relation, data modules and to extract the datasets, types, and values.
5. Many applications like Map reduce, Hbase, Simple DB, and Cassandra not as able to solve issues in the repository and to work in big data.
6. Challenge of big data, IoT, virtualization and cloud computing.

Table 8 Challenge of big data, IoT, virtualization and cloud computing.

Big data	**Cloud computing**	**Virtualization**	**IoT**
Decision support tools	Security and privacy	VM sprawl	Sensing a complex environment
Cost	Service quality	Resource and capacity	Connectivity
Quality	Downtime and accessibility	Backup, recovery, and continuity	Power
Curation	Access to data	Security and monitory	Security
Storage	Transition to cloud	Adopting virtualization-VM stall	Complexity
Security	Interoperability	License cost	Cloud is necessary
Disambiguation	Service usage and control	Stuck in storage	Adaptability and scalability

Future research

Table 9 Future research of big data virtualization in agriculture.

Vehicle	Application	Description
BoniRob [60]	Crop, weed finding, plant breeding, weed control	• They are required to handle redundant actuated systems. • Its intervention regions are restricted to space like a robot, limits to flexibility and versatility. • Four independent steerable wheels capable of modifying its track distance to the crop.
Rippa [61] Ladybird [62] Vibro crop (Robotti) [63] AgBot II [64]	Fertilization, seeding, weed control and collecting information	• Under the robots, we utilize sensors, tools, and intervention mechanisms. • In the current situation, some difficulty in working in farmland with medium to high slopes, ditches or in the presence of gully erosion.
Casar [65]	Pest, soil management, fertilization, harvesting, and transport	• It may lack flexibility, robustness, and intelligence to deal with the various processes. • Some of the safety features based on basics and unintelligent (unable to reschedule)
Greenbot [66,67]	Fruit, horticulture, arable farming, urban sector, waterfronts and roadsides	• Weed or soil finding will not possess with any detection system. • It may lack flexibility, robustness, and intelligence to deal with the various processes.

Conclusions

In this paper, the critical concept of big data virtualization in agriculture has been surveyed. Agriculture development based on virtualization tools, virtualization platforms, techniques, methods, tools, platform, and infrastructure is studied. A significant data source gathered, and value extracted for making better decisions and helping farmers to increase their production. Researchers are primarily focused on single VM's and sensors. The current trend in agriculture is WSNs performing in agriculture and farming surroundings, patterns work on irrigation management, vineyard production monitoring, and predicting crop disease. Some factors needs to improve including cost, autonomous operation, intelligence, portability, low maintenance, energy efficiency, interoperability and so forth. Also some of the challenges and specific issues in the Indian scenario of using WSNs in agriculture. More cost used in the sensor, variable climate, and soil it is a crucial part in structuring the WSNs based framework for agriculture in India with various climates and types of soil. Some parameters required to tune the function accurately for each different location. The segmented land structure is based on WSNs used in agriculture. It is similar to irrigation management. Also, some challenges include requirement of a farmer, overall plan, and maintenance.

Acknowledgment

A special note of thanks to the Vellore Institute of Technology University for providing necessary infrastructure facilities to carry out the research work and my friends who have helped me to complete this study.

References

[1] B Schmarzo. *Big Data: Understanding How Data Powers Big Business*. John Wiley & Sons, USA, 2013.

[2] S Senthilvadivu, SV Kiran, SP Devi and S Manivannan. Big data analysis on geographical segmentations and resource constrained scheduling of production of agricultural commodities for better yield. *Proc. Comput. Sci.* 2016; **87**, 80-5.

[3] AU Rehman, AZ Abbasi, N Islam and ZA Shaikh. A review of wireless sensors and networks applications in agriculture. *Comput. Stand. Interf.* 2014; **36**, 263-70.

[4] MS Kumar. Analysis of network function virtualization and software defined virtualization. *Int. J. Inform. Visual.* 2017; **1**, 122-6.

[5] WA Goya, MRD Andrade, AC Zucchi, NM Gonzalez, RDF Pereira, K Langona, TCMDB Carvalho, JE Mångs and A Sefidcon. The use of distributed processing and cloud computing in agricultural decision-making support systems. *In*: Proceedings of the 2014 IEEE 7th International Conference on Cloud Computing, USA, 2014, p. 721-8.

[6] S Prasad, SK Peddoju and D Ghosh. AgroMobile: A cloud-based framework for agriculturists on a mobile platform. *Int. J. Adv. Sci. Tech.* 2013; **59**, 41-52.

[7] Z Haihui, Z Chunjiang, W Huarui, Y Feng and S Xiang. Research of cloud computing based agriculture virtualized information database. *In*: Proceedings of the 2nd APSIPA Annual Summit and Conference, Singapore, 2010, p. 835-8.

[8] NF Xie, XF Zhang, W Sun and XN Hao. Research on big data technology-based agricultural information system. *In*: Proceedings of the International Conference on Computer Information Systems and Industrial Applications, Bangkok, Thailand, 2015, p. 388-90.

[9] CN Verdouw, J Wolfert, AJM Beulens and A Rialland. Virtualization of food supply chains with the internet of things. *J. Food Eng.* 2016; **176**, 128-36.

[10] JW Kruize, J Wolfert, H Scholten, CN Verdouw, A Kassahun and AJM Beulens. A reference architecture for farm software ecosystems. *Comput. Electron. Agric.* 2016; **125**, 12-28.

[11] T Ojha, S Misra and NS Raghuwanshi. Wireless sensor networks for agriculture: The state-of-the-art in practice and future challenges. *Comput. Electron. Agric.* 2015; **118**, 66-84.

[12] P Patil and V Sachapara. Providing smart agricultural solutions/techniques by using Iot based toolkit. *In*: Proceedings of the 2017 International Conference on Trends in Electronics and Informatics, Tirunelveli, India, 2017, p. 327-31.

[13] Y Shen, Z Zhao and H Wang. Agricultural information technology development and innovation path. *In*: Proceedings of the 2011 International Conference on Electronics, Communications and Control, Ningbo, China, 2011, p. 2512-5.

[14] MS Kumar. Research and development of virtualization in wireless sensor networks. *Int. J. Inform. Visual.* 2018; **2**, 96-103.

[15] JS Hurwitz, A Nugent, F Halper and M Kaufman. *Big Data for Dummies*. John Wiley & Sons, USA, 2013.

[16] M Portnoy. *Virtualization Essentials*. Vol. 19. John Wiley & Sons, USA, 2012.

[17] D Haynes, S Ray and S Manson. Terra populus: Challenges and opportunities with heterogeneous big spatial data. *In*: Proceedings of the 13th International Conference on Advances in Geocomputation, USA, 2015, p. 115-21.

[18] G Chen, X Wang, H Chen, C Li, G Zeng, Y Wang and P Liu. Research on digital agricultural information resources sharing plan based on cloud computing. *In*: Proceedings of the International Conference on Computer and Computing Technologies in Agriculture, Beijing, China, 2011, p. 346-54.

[19] TK Fan. Smart agriculture based on cloud computing and IoT. *J. Converg. Inform. Tech.* 2013; **8**, 1-7.

[20] IS Saguy. Challenges and opportunities in food engineering: Modeling, virtualization, open innovation and social responsibility. *J. Food Eng.* 2016; **176**, 2-8.

[21] Y Xing and Y Zhan. Virtualization and cloud computing. *In*: Proceedings of the Future Wireless Networks and Information Systems, 2012, p. 305-12.
[22] H Wang, Z Xu, H Fujita and S Liu. Towards felicitous decision making: An overview on challenges and trends of Big Data. *Inform. Sci.* 2016; **367**, 747-65.
[23] J Yan, N Liu, S Yan, Q Yang, W Fan, W Wei and Z Chen. Trace-oriented feature analysis for large-scale text data dimension reduction. *IEEE Trans. Knowl. Data Eng.* 2011; **23**, 1103-17.
[24] H Shen, L Zhao and Z Li. A distributed spatial-temporal similarity data storage scheme in wireless sensor networks. *IEEE Trans. Mobile Comput.* 2011; **10**, 982-96.
[25] J Zhang, FY Wang, K Wang, WH Lin, X Xu and C Chen. Data-driven intelligent transportation systems: A survey. *IEEE Trans. Intell. Transport. Syst.* 2011; **12**, 1624-39.
[26] S Mithas, MR Lee, S Earley, S Murugesan and R Djavanshir. Leveraging big data and business analytics. *IT Profess.* 2013; **15**, 18-20.
[27] Z Wen, W Zhang, T Zeng and L Chen. MCentridFS: A tool for identifying module biomarkers for multi-phenotypes from high-throughput data. *Molec. BioSyst.* 2014; **10**, 2870-5.
[28] Y Wang, X Jiang, R Cao and X Wang. Robust indoor human activity recognition using wireless signals. *Sensors* 2015; **15**, 17195-208.
[29] S Wolfert, L Ge, C Verdouw and MJ Bogaardt. Big data in smart farming: A review. *Agric. Syst.* 2017; **153**, 69-80.
[30] JB Cole, S Newman, F Foertter, I Aguilar and M Coffey. Breeding and genetics symposium: Really big data: Processing and analysis of very large data sets. *J. Anim. Sci.* 2012; **90**, 723-33.
[31] A Faulkner and K Cebul. *Agriculture Gets Smart: The Rise of Data and Robotics, Cleantech Agriculture Report*. CleanTech Group, USA, 2014.
[32] M Chen, S Mao and Y Liu. Big data: A survey. *Mobile Netw. Appl.* 2014; **19**, 171-209.
[33] A Faulkner and K Cebul. *Agriculture Gets Smart: The Rise of Data and Robotics, Cleantech Agriculture Report*. CleanTech Group, USA, 2014.
[34] S Wolfert, L Ge, C Verdouw and MJ Bogaardt. Big data in smart farming: A review. *Agric. Syst.* 2017; **153**, 69-80.
[35] Z Sun, FX Zheng and SY Yin. Perspectives of research and application of big data on smart agriculture. *J. Agric. Sci. Tech.* 2013; **15**, 63-71.
[36] M Chen, S Mao and Y Liu. Big data: A survey. *Mobile Netw. Appl.* 2014; **19**, 171-209.
[37] S Sonka and I Ifamr. Big data and the ag sector: More than lots of numbers. *Int. Food Agribus. Manag. Rev.* 2014; **17**, 1-20.
[38] JB Cole, S Newman, F Foertter, I Aguilar and M Coffey. Breeding and genetics symposium: Really big data: Processing and analysis of very large data sets. *J. Anim. Sci.* 2012; **90**, 723-33.
[39] S Neethirajan. Recent advances in wearable sensors for animal health management. *Sens. Bio-Sens. Res.* 2017; **12**, 15-29.
[40] SF Wamba and A Wicks. RFID deployment and use in the dairy value chain: Applications, current issues and future research directions. *In*: Proceedings of the 2010 IEEE International Symposium on Technology and Society, Australia, 2010, p. 1-7.
[41] Z Jianmin, L Jingtao, Z Dongting and H Zhiwen. Spherical fruit automatic recognition method based on grey relational analysis and fuzzy membership degree matching. *Chin. J. Sci. Instrum.* 2012; **33**, 1826-36.
[42] L Pingping and J Wang. Research progress of intelligent management for greenhouse environment information. *Trans. Chin. Soc. Agric. Mach.* 2014; **45**, 236-43.
[43] H Luo, P Yang, Y Li and F Xu. An intelligent controlling system for greenhouse environment based on the architecture of the internet of things. *Sensor Lett.* 2012; **10**, 514-22.
[44] CN Verdouw, AJM Beulens and JGAJVD Vorst. Virtualisation of floricultural supply chains: A review from an Internet of Things perspective. *Comput. Electron. Agric.* 2013; **99**, 160-75.
[45] F Natale, M Gibin, A Alessandrini, M Vespe and A Paulrud. Mapping fishing effort through AIS data. *PloS One* 2015; **10**, e0130746.
[46] B Yan, P Shi and G Huang. Development of traceability system of aquatic foods supply chain based on RFID and EPC internet of things. *Trans. Chin. Soc. Agric. Eng.* 2013; **29**, 172-83.

[47] N Nedjah, FPD Silva, AOD Sá, LM Mourelle and DA Bonilla. A massively parallel pipelined reconfigurable design for M-PLN based neural networks for efficient image classification. *Neurocomputing* 2016; **183**, 39-55.
[48] D Thompson, JA Levine, JC Bennett, PT Bremer, A Gyulassy, V Pascucci and PP Pébay. Analysis of large-scale scalar data using hixels. *In*: Proceedings of the 2011 IEEE Symposium on Large Data Analysis and Visualization, 2011, p. 23-30.
[49] J Ahrens, K Brislawn, K Martin, B Geveci, CC Law and M Papka. Large-scale data visualization using parallel data streaming. *IEEE Comput. Graph. Appl.* 2001, **4**, 34-41.
[50] H Demirkan and D Delen. Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decis. Support Syst.* 2013; **55**, 412-21.
[51] V López, SD Río, JM Benítez and F Herrera. Cost-sensitive linguistic fuzzy rule based classification systems under the MapReduce framework for imbalanced big data. *Fuzzy Sets Syst.* 2015; **258**, 5-38.
[52] S Ramachandramurthy, S Subramaniam and C Ramasamy. Distilling big data: Refining quality information in the era of yottabytes. *Sci. World J.* 2015; **2015**, 453597.
[53] A Ruiz-Martinez, F Pereniguez-Garcia, R Marin-Lopez, PM Ruiz-Martínez and AF Skarmeta-Gomez. Teaching advanced concepts in computer networks: Vnuml-um virtualization tool. *IEEE Trans. Learn. Tech.* 2013; **6**, 85-96.
[54] Big Data in Digital Agriculture using Satellite Data, Available at: http://www.slideshare.net/search/slideshow/Big+data+in+Digital+agriculture+using+satellite+data, accessed January 2017.
[55] MK Gayatri, J Jayasakthi and GSA Mala. Providing Smart Agricultural solutions to farmers for better yielding using IoT. *In*: Proceedings of the 2015 IEEE Technological Innovation in ICT for Agriculture and Rural Development, Madras, India, 2015, p. 40-3.
[56] T Ojha, S Misra and NS Raghuwanshi. Wireless sensor networks for agriculture: The state-of-the-art in practice and future challenges. *Comput. Electron. Agric.* 2015; **118**, 66-84.
[57] AZ Abbasi, N Islam and ZA Shaikh. A review of wireless sensors and networks' applications in agriculture. *Comput. Stand. Interfac.* 2014; **36**, 263-70.
[58] Big Data Innovations in Agriculture. Available at: http://new-innovations-in-technology-help-growers-with-big-data, accessed January 2017.
[59] Agriculture Trends worth Watching. Available at http://www.agweb.com/article/5-agriculture-trends- worth-watching-naa-ben-potter, accessed January 2017.
[60] L Emmi and P Gonzalez-de-Santos. Mobile robotics in arable lands: Current state and future trends. *In*: Proceedings of the 2017 European Conference on Mobile Robots, Paris, France, 2017, p. 1-6.
[61] A Ruckelshausen, P Biber, M Dorna, H Gremmes, R Klose, A Linz, R Rahe, R Resch, M Thiel, D Trautz and U Weiss. BoniRob: An autonomous field robot platform for individual plant phenotyping. *Precis. Agric.* 2009; **9**, 841.
[62] R Bogue. Robots poised to revolutionise agriculture. *Indust. Robot Int. J.* 2016; **43**, 450-6.
[63] JP Underwood, M Calleija, Z Taylor, C Hung, J Nieto, R Fitch and S Sukkarieh. Real-time target detection and steerable spray for vegetable crops. *In*: Proceedings of the International Conference on Robotics and Automation: Robotics in Agriculture Workshop, USA, 2015, p. 26-30.
[64] Available at: http://www.kongskilde.com/Agriculture, accessed January 2017.
[65] O Bawden, D Ball, J Kulk, T Perez and R Russell. A lightweight, modular robotic vehicle for the sustainable intensification of agriculture. *In*: Proceedings of the Australian Conference on Robotics and Automation, Australia, 2014.
[66] Fruit Robot, Available at: http://www.raussendorf.de/en/fruit-robot.html, accessed February 2017.
[67] Precision Makers, Available at: http://www.precisionmakers.com/greenbot, accessed February 2017.