IJCTA, 9(22), 2016, pp. 257-264

A Novel Cloud based Data Analytics Framework for effective Crop Management

Balaji Prabhu B.V* Dr. M Dakshayini**

Abstract : Today in developing countries like India, in the field of agriculture, farmers are facing many problems. Some of them are, imbalance in demand for and production of crops (Demand Supply problem) due to which, either farmer fails to get good market prices for their crops or consumer suffers with high prices. In addition to that, Indian agriculture has been defined as "gambling with monsoon" and also there is no reliable and easy access to required information like predicted demand for crops, weather forecasts, pest control and management, fertilizer usage and agriculture expert/ scientist consultation. As there is no provision for farmers to have expert's consultation in crop cycle and management, farmers fails to take a right decision in growing appropriate crop for which there could be a demand. Farmers are not well connected to the agricultural scientists/experts through common forums where they can discuss/exchange their good practices and issues. Hence there is a need for a consolidated farmer's crop calendar system through which farmer community can be provided with an educated decision support mechanism with relevant information while choosing crops. This paper proposes a novel Cloud based data analytics framework and an effective crop management system providing an opportunity for farmers to get connected themselves to the expert's consultation forum. So that, demand of the consumers and supply from the farmers could be greatly balanced.

Keywords : Agriculture, Farmer, Supply and Demand, Weather Forecast, Pest Management, Fertilizer, Crop, Cloud.

1. INTRODUCTION

Big Data plays a key role in integrating the technologies being developed across industries in the recent years. Big Data is a collection of appropriate data from the huge number of sources and also translating the same into actionable information to improve various professional methods and insightfully solve problems at scale and speed.

By considering agriculture as a major living in India, where farmers and agribusinesses have to make numerous decisions every year to make farmers, market people and the public happy. In this regard, agriculture also has been an obvious target for big data. In the current scenario, perhaps unpredictable climate, unorganized and un-matching supply-demand for required crops and conditional prices of the crops are making it all the more relevant for agriculturalists to use any beneficial information in making critical farming decisions. This paper proposes a novel Cloud enabled data analytics framework for integrating IoT, Cloud technology with Big Data providing data analytics as a key service to achieve smart Agriculture tackling the problem being faced by farmers.

The rest of the paper is organized as follows. Section 2 gives a brief introduction to related work and also states the need for proposed Cloud based Data Analytics Framework for effective Crop Management. Section 3 discusses a proposed Data Analytics Framework for effective Crop Management. Section 4 describes the Data Analytics model to match the demand for and supply of crops. Section 5 concludes the paper.

* Research Scholar, Dept. of ISE BMSC, Bangalore, Karnataka, India balajitiptur@gmail.com

** Professor, Dept. of ISE BMSCE, Bangalore, Karnataka, India dakshayini.ise@bmsce.ac.in

2. RELATED WORK

2.1. Big-Data

Big - data can be referred as a massive volume of data which is not easy to store, manage and process with the traditional data management tools [1-2]. According to Gartner big data can be defined in terms of 3Vs (Figure 1) as high velocity, high volume and high variety of data [3]. In three Vs volume refers to size of the data which is very large in terms of petabytes and zetabytes, velocity refers to the speed of the data which is arriving at a higher speed, and variety refers to the different forms of data which is beyond the structured data includes unstructured and semi structured data.

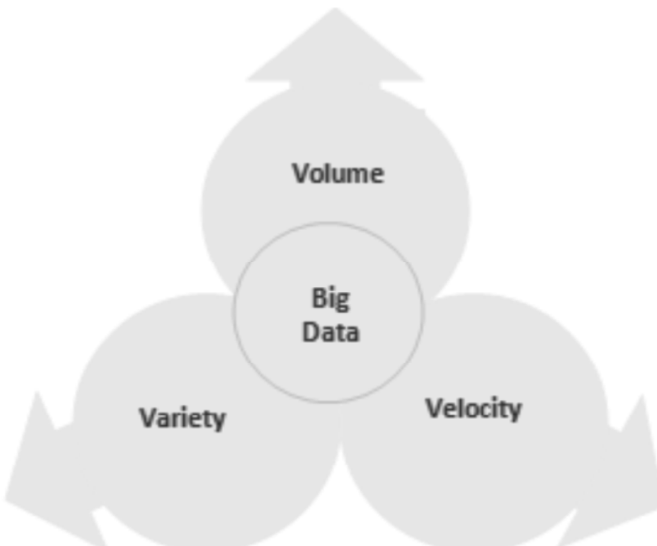

Fig. 1. The Three Vs of big data.

Performing operations like storage, access, analysis and processing of big complex data with traditional algorithms is becoming much more intricate and inefficient. Hence, new tools and methodologies need to be developed in order to store, analyze, manage and process this big data in a secure and efficient manner. Arriving rate of such varied data is very high and hence many organizations may not be capable of (*i*) handling this huge data flow to make the information useful and (*ii*) investing for huge infrastructure required. Resulting big data environments require new tools and methodologies to be developed to process the huge volumes, high velocity, and diverse formats of big data.

2.2. Cloud Computing

Cloud computing is the ideal technology for storage and computing services over the Internet [4]. Cloud services permit individuals and businesses to use software and hardware that are managed by third parties at remote locations. The cloud computing model allows access to information and computer resources from anywhere that an internet connection is available. Cloud computing provides a shared pool of resources, including data storage space, networks, computer processing power, and specialized corporate and user applications.

2.3. Cloud Is A Better Enabler For Big Data Analytics

Cloud computing technology provides an ideal platform for big data storage, dissemination and interpreting with its great computation power. The flexibility and elasticity of the cloud technology makes it essential for big data analytics, the practice of chomping huge volumes of unstructured data to identify patterns and develop better corporate strategies makes it suitable for Big Data analytics. Cloud has already deployed on pools of servers and networking resource, processing and analysis can be supported with these valuable resources through provisioned cloud solutions [5-6].

In the cloud computing environment, the analytics can be offered as a service [7- 9] which provides a cost effective alternative from the point of both developers and consumers. The cloud platform transforms analytics into a ‘utility’ that can be accessed by end users on a use-and-pay scenario. All these beneficiary features of a cloud make it as an efficient platform in addressing the solutions for big data problems.

Most Cloud based business applications have used data analytics to help direct their strategy to maximize profits. Data analytics helps eliminate much of the guesswork involved in trying to understand clients, instead systemically tracking data patterns to best construct business tactics and operations to minimize uncertainty. This is also true in the field of agriculture where weather, market prices, crop yield, diseases and soil management etc. varies in a fast phase in which farmer needs an expert’s valuable suggestion and support mechanism to take an educated decision. Currently there is no such forum to provide the formers with the timely relevant information and hence there is no synchronization in production and demand due to which either farmer fail to get good market prices or consumer suffers high prices due to less production (Demand Supply problem).

Based on the survey made, today one of the main reasons for Indian farmers to incur loses is due to the variation in demand supply of crop production in the market. In spite of reliable agriculture relevant data available from government sources [10-12], farmers in India currently facing the following problems

- **Demand Supply management Problem :** No synchronization in production and demand due to which either farmer fail to get good market prices to their produce or consumer suffers high prices due to less production.
- **Inefficient of Crop management :** There is no reliable and easy access to weather forecast services due to which farmers incur productivity loses while managing crops (pest management, fertilizer usage etc.)
- **Lack of Expert/Agricultural Scientist Consultation :** There is no expert’s consultation in crop management due to which farmers fail to control diseases, take appropriate actions based on weather changes and soil tests etc. (mostly standard reminder services based on individual crop habits)
- **Lack of Knowledge about new techniques :** By providing relevant and easily accessible training resources farmers can gain/share knowledge in new techniques and methods of crop management.
- **Lack of access to reliable market :** There is a need for an easily accessible and reliable platform through which users can purchase or barter farmer equipment or accessories.
- **Nonexistance of common forum for information exchange :** Farmers are not well connected through common forums where they can discuss/exchange their good practices and issues.

According to the survey made, most of the researcher has dealt with various data analytics domains like commerce, business, society administration, scientific research and health care issues of big data. Most of the research work is moving towards developing frameworks, algorithms and basic technologies for providing solutions to the different domains of data analytics. No work has been found out to provide a solution to Agricultural issues with data analytics. This paper proposes an idea of using cloud as better enabler in providing efficient framework for developing a novel cloud based data analytics framework for effective crop management. Demand supply problem is the basic challenge being faced by farmer’s community in India due to which, they do not get good market prices for the their crops. So the idea of integrating the cloud as a better framework for developing and offering a data analytics as a service could address this problem. Hence there is a need for a consolidated farmer’s crop management system through which farmer community can be provided with an educated decision support mechanism with relevant information while choosing crops.

3. PROPOSED DATAANALYTICS FRAMEWORK

Table 1 shows the sample demand and supply data [in Million tons] for some selected crops of state of Karnataka for the financial year 2015-16[10]. Table 1 also shows the gap between the demand for and supply of some sample crops, due to which, either farmers are failing to get good market prices for their crops or consumer suffers with high prices due to less production. The graphical representation for the table is shown in the Figure 2.

Table 1. Demand and Supply of crops for the year 2015-16.

Crops	*Demand in 2015-16 Million tons*	*Supply in 2015-16 Million tons*	*Gap in Demand SupplyMillion tons*
Rice	110	102	-8
Wheat	89	98	+9
Tomato	55	30	25
Coarse Cereals	36	45	+9
Cereals	235	245	+10
Pulses	22	20	-2
Food grains	257	265	+8
Oilseeds	59	36	-23
Sugarcane	279	380	+99
Onion	62	35	-27
Potato	25	12	-13

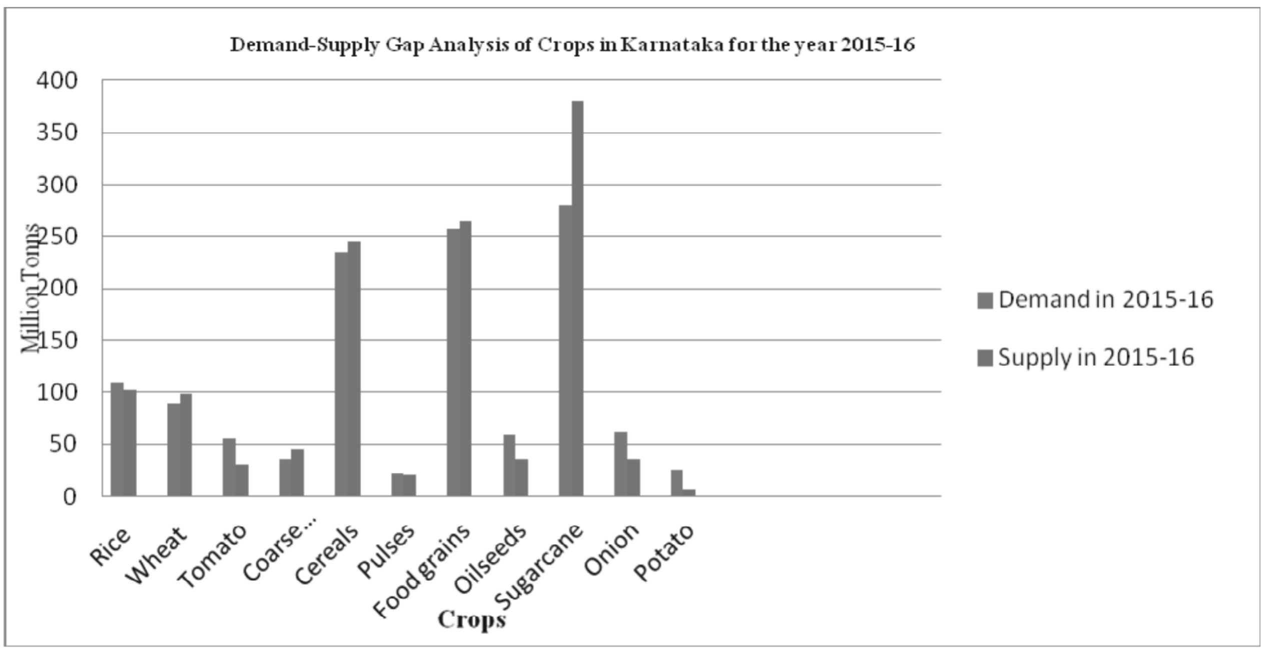

Fig. 2. Bar graph shows the demand and supply of the crops for the table 1.

From the survey and the analysis made it has been found that there exist a gap between supply and demand for and supply of crops. For example the demand for the vegetable tomato for the financial year 2015-16 is 55.70 lakh tons, but the supply for the same year is 30.49 lakh tons, it shows 25.21 lakh tons of shortage in the supply of tomato for the year 2015-16.

Due to this incompatibility between demand and supply, either formers may not get expected rate/profit or consumers could suffer with high price. This problem could be tackled by developing appropriate data analytics

The aim is to design and develop a dedicated, reliable, cost effective, cloud based decision support system with efficient data analytic frame work through which farmer community can obtain all their day to day information at one place without too much technological hassles and preferably in their language. The figure 3 shows the proposed framework for data analytics to help the farmer community to get the better market price for theirs crops.

This proposed data analytic frame work consists of 4 modules, they are Data sources module, Unstructured-structured Data-base module, and Distributed cluster based data storage module and Data processing and analytics module.

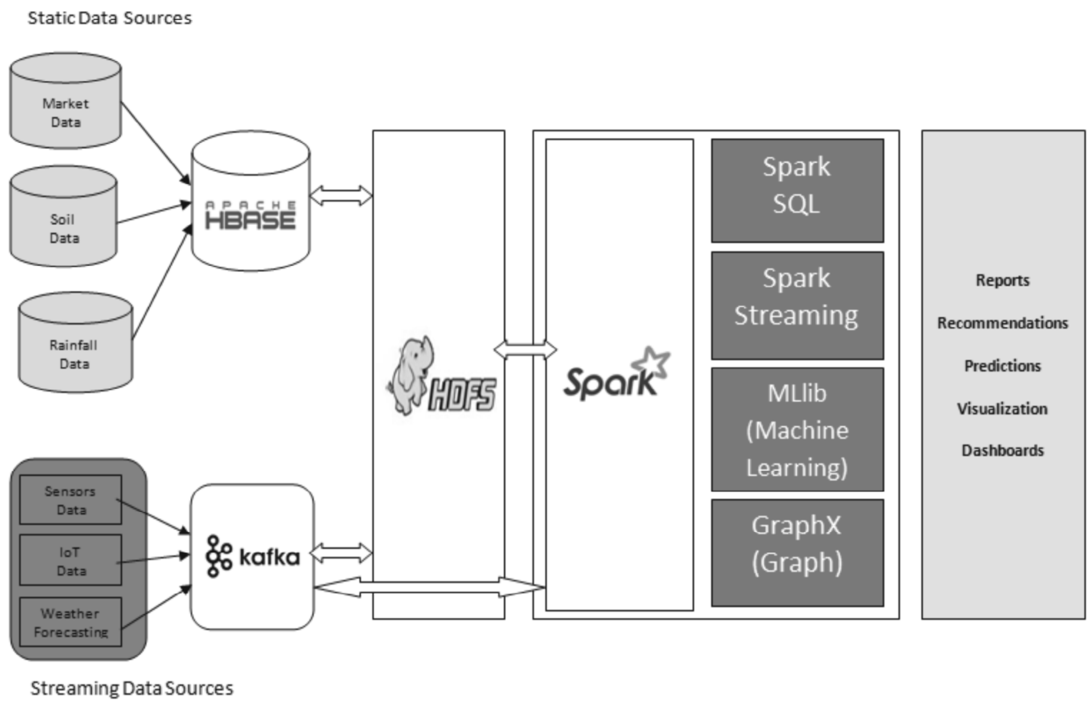

Fig. 3. Proposed Framework for data Analytics.

3.1. Data sources Module

1. **Streaming Data Sources could be :**
   - internal records such as sensors data, data from the devices which are deployed in the agricultural fields to measure soil dampness and nutrient density, data from the devices fitted to a tractor to measure crop yields, data from predictive weather stations, and image-capturing satellites and drones mapping out land and measuring crop health. Collectively these data sets can be considered as IoT data sources.
   - Weather data from different weather forecasting websites like AccuWeather.com
2. **Static Data Sources could be :**
- External sources such as government sources, data collected from the markets regarding the consumption of goods.
- Database containing the type of soil which provides information on the soil properties that are present in a given location or planting area.
- Database containing a historical rainfall records.

3.2. Unstructured to structured Data-base module

Data collected from various data sources would be in unstructured format. This unstructured data is converted to structured form and stored in HBase. HBase [13] is a scattered column-based database assembled on top of the Hadoop file system designed to provide rapid random access to huge amounts of structured data. It influences the fault tolerance provided by the Hadoop File System (HDFS). Data could be stored in HDFS either directly or through HBase. user reads or accesses the data from HDFS randomly through HBase.

Apache Kafka [14] is a distributed PUB-SUB [publish-subscribe] message communication system and a strong queue that can handle a huge data and allows users to pass messages from one end-point to another. Kafka is appropriate for both offline and online message communication.

3.3. Distributed cluster based data storage module

HDFS [15] stores and handles huge amount of data and provides flexible access mechanism. In order to store such large volume of data, the files are stored across multiple clusters. These files are stored in redundant manner to rescue the possible data losses in case of any failures in the system. In HDFS applications are also available for parallel processing.

3.4. Data processing and analytics module.

Apache Spark [16] is a very-fast cluster based computing technology, mainly designed for quick computation which is based on the Hadoop Map-Reduce concept. The main feature of Spark is its **in-memory cluster computing** that rises the processing speed of an application.

Spark contains the different tools such as

1. **Apache Spark Core :** Spark Core is the fundamental common execution engine for spark system upon which all other functionality is built. It offers In-Memory computing and referencing datasets in peripheral storage structures.
2. **Spark SQL :** Spark SQL is a module on top of Spark Core that contains a new data abstraction called Schema RDD, which offers a support for structured and semi-structured data.
3. **Spark Streaming :** Spark Streaming controls Spark Core's fast scheduling ability to achieve streaming analytics. It reads the data in mini-sets and performs RDD (Resilient Distributed Datasets) conversions on those mini-sets of data.
4. **MLlib (Machine Learning Library) :** MLlib is a dispersed machine learning module of Spark which offers a machine learning library to the data sets.
5. **GraphX :** GraphX is a dispersed graph-processing module of Spark. It offers an API for conveying graph computation that can model the user defined graphs via Pregel abstraction API.

Data from the different sources like static data and real-time data contains the data in different formats. This data has to be preprocessed before applying the analytics. So the data sets from these data sources are transformed and loaded into the structured format in HBase or Kafka. In the next module the structured data is loaded in to HDFS to the distributed data storage to process the data in distributed way, which enables the parallel processing. The SPARK module is used to apply the analytics to the data sets stored in the HDFS to process the data and extract the useful information to obtain the results. Once the SPARK module produces the analytical results these are sent back to the user in one of the format like report, visualization, prediction, recommendations or dashboards.

Key decisions that Analytics helps farmers to make include when and how much to irrigate a field, based on soil moisture data, weather predictions, and crop health, and planting and harvesting decisions, based on yield data or weather. Fertilizer applications can be much more prescriptive, based on factors such as soil nutrient density, enabling farmers to save money on areas that don't need as much, but also optimize yield across a property.

Big data analytics can also alert farmers to problems on a certain field, such as a pest infestation, or drought conditions, reducing the need for manual checks of every piece of land regularly. With existing and increasing labor shortages in agriculture, the ability for big data analysis to create efficiencies that reduce the need for physical manpower is a big benefit for the industry particularly for very large scale operations.

4. PROPOSED CLOUD BASED DATA ANALYTIC MODEL FOR EFFECTIVE CROP MANAGEMENT SYSTEM

The proposed Cloud based Data Analytics Framework for effective Crop Management consists of following units. Kiosk unit for registration. Data store unit is developed for collecting and storing the data from various internal (sensors), external (Govt. Websites for market data) and Weather forecasting data (AccuWeather.com) sources. Cloud unit is the one where the raw data collected from various sources gets transformed into the required structured form, processed, analyzed and provides the result as an analytics as service for the former as shown in figure 4. This happens by considering the details given by the former, previous and current market supply demand survey details and weather data.

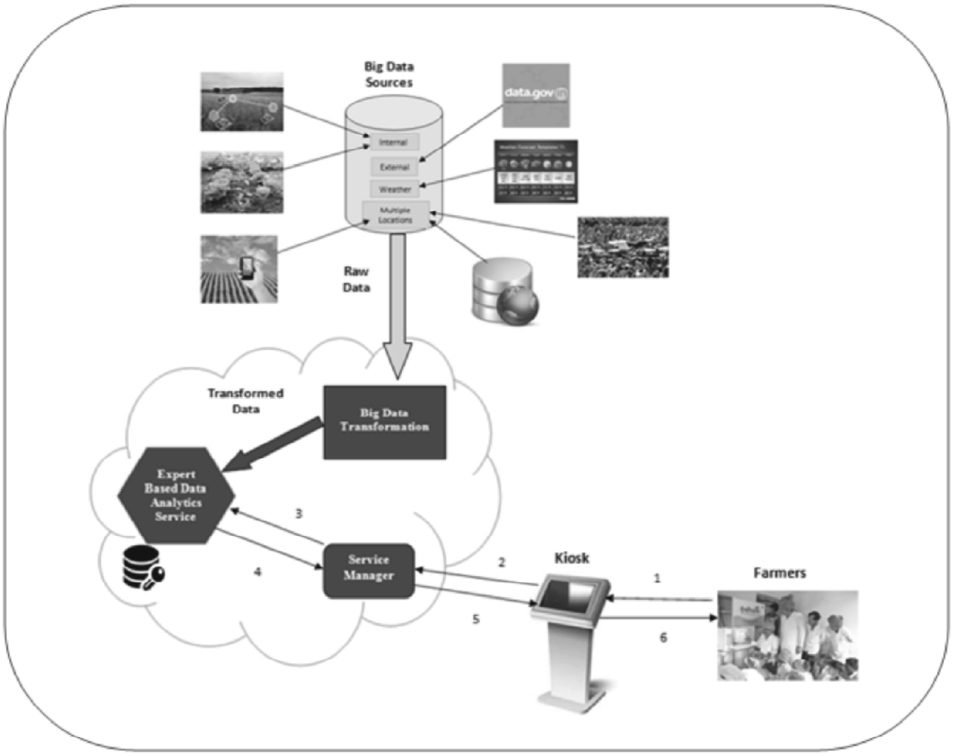

Fig. 4. Cloud based Data Analytic Model for effective crop management system.

4.1. Proposed Algorithm

Phase 1 : Registration

Step 1 : When a farmer think of growing some crop, as a first step a he has to register himself at the kiosk (most common place where people meet very often like dairy in the village) by providing the details like name, location, ID proof number (Ex: Adhar) etc.

Step 2: Once a farmer registered successfully, he will get a unique ID that must be used for the later communication.

Phase 2: Enquiry

Step 3 : Before deciding the crop to grow farmer has to crosscheck with the data base whether there will be a demand [How many have planned to grow the same crop] for that particular crop when he gets the yield.

For example if a farmer X wants to sow tomato crop in January and in the absence of market demand supply information system he has no idea about other farmer's choices and also the market demand when his tomato crop harvesting starts. If more number of farmers grow tomato which results in excess supply there by demand goes down resulting in price drop and farmer incurs loses.

Step 4 : farmer has to submit the crop details (number of acres he is going to sow a particular crop, whether a land is irrigated or non-irrigated and crop variety etc.) which he wants to grow at the kiosk.

Phase 3 : Grow-Crop-Analytics

Step 5 : Once a farmer submitted the details of the crop, the service manager analyzes the supply [How many have planned to grow the same crop] demand [predicted from the data available from government and other reliable sources (Ministry of Agriculture) and current scenario] and provide a predicted decision back to the farmer

If there is a demand

Then grow the selected crop.

else (if more farmers are selected the same crop)

He will be suggested with some other crops which is having more demand in the future also suits for his field details.

5. CONCLUSION

The farmers in the developing countries like India, are facing so many problems like un-matched demand for and supply of crops problem, inefficient pest management, and fertilizer usage and crop management. Lack of expert's consultations may be a main reason for most of these problems. Through the integration of cloud, IoT and big data technologies with the data analytics it is possible to provide a solutions to the above problems. Which help farmers to plan crop cycles, manage crop diseases, and get access to weather forecasting and also provide a decision support system. This proposed Cloud based Data Analytics Framework for effective Crop Management could help the farmers in making appropriate decision in selecting a crop to yield. So that farmers will not incur loss in the farming or consumer will not suffer from the high prices.

6. ACKNOWLEDGMENT

The authors would like to acknowledge and thank Technical Education Quality Improvement Program me [TEQIP] Phase 2, BMS College of Engineering and SPFU [State Project Facilitation Unit], Karnataka for supporting the research work

7. REFERENCES

1. Sangeeta Bansal, Dr. Ajay Rana, "Transitioning from Relational Databases to Big Data", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 1, January 2014.
2. Jim Gray, "Data Management: Past, Present, and Future", IEEE Computer 29(10): 38-46 (1996).
3. B. Mark, ''Gartner says solving 'big data' challenge involves more than just managing volumes of data,'' Gartner. Archived from the original on Jul. 10, 2011, Retrieved Jul. 13, 2011.
4. F. Xhafa and N. Bessis, "Cloud Computing: Paradigms and Technologies", DOI: 10.1007/978-3-642-35016-0_2, Springer-Verlag Berlin Heidelberg 2014.
5. Prajakta Rawool, Swapnil Salvi, "The Cloud As An Enabler For Big Data Analytics", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 6, June 2015.
6. Assuncao, M. D., Calheiros, R. N., Bianchi, S. and Netto, M. A. S. (2015). Big Data Computing and Clouds: Trends and Future Directions. J. Parallel Distrib. Computing, 79-80 (2015) 3-15.
7. Zibin Zheng, "Service-Generated Big Data and Big Data-as-a-Service: An Overview", IEEE International Congress on Big Data, pp-403-410,2013.
8. Samiya Khan1 , Kashish Ara Shakil and Mansaf Alam, "Cloud-Based Big Data Analytics – A Survey Of Current Research And Future Directions", Department of Computer Science, Jamia Millia Islamia, New Delhi.
9. Yali Zhao, Rodrigo N. Calheiros, Graeme Gange, Kotagiri Ramamohanarao, Rajkumar Buyya, "SLA-Based Resource Scheduling for Big Data Analytics as a Service in Cloud Computing Environments", Department of Computing and Information Systems The University of Melbourne, Australia.
10. Open Government Data(OGD) Platform India, https://data.gov.in/catalogs/sector/Agriculture-9212. Accessed on 24th May, 2016.
11. Government of India, Department of Agriculture and Corporation, http:// http://agricoop.nic.in/State_Agri_Dept.aspx, Accessed on 24th May 2016.
12. Directorate of Economics and Statistics Karnataka, http://des.kar.nic.in/agstat.asp, Accessed on 8th June 2016.
13. Apache HBase- Reference Guide: Apache HBase Team.
14. Reference Guide for Deploying and Configuring Apache Kafka, White Paper by Cloudera.
15. Konstantin Shvachko, Hairong Kuang, Sanjay Radia, Robert Chansler, "The Hadoop Distributed File System", IEEE-2010.
16. Apache Spark Project, Accessed from http://spark.apache.org.