

# Towards Development of Spark Based Agricultural Information System including Geo-Spatial Data

Purnima Shah

School of Engineering and Applied  
Science,  
Ahmedabad, India  
purnima.shah@iet.ahduni.edu.in

Deepak Hiremath

School of Engineering and Applied  
Science,  
Ahmedabad, India  
deepak.hiremath@iet.ahduni.edu.in

Sanjay Chaudhary

School of Engineering and Applied  
Science,  
Ahmedabad, India  
sanjay.chaudhary@ahduni.edu.in

**Abstract**— Voluminous and variety of disparate information is generated and consumed in agriculture domain at a higher velocity. In agriculture, information is available in the form of weather and soil conditions reports, GPS mapping, water resources, fertilizer/ pesticide use, field characteristics, and commodity market conditions. Big data technology has a huge potential to refer these information and produce comprehensive insight via Geo-spatial processing, remote sensing, advance analytics algorithms, cloud resources, and advance storage systems. The paper, proposes a spark based information management system for agriculture and intend to reduce the technological gap between agro users and information. The system is proposed to collect, query, analyze, and visualize heterogeneous and distributed data including Geo-spatial data at scale using open source. The implementation is done on big data open source architectures by developing various web based analytical and visualization services for cotton crop in Gujarat state, India. The analytical results are explored through interactive maps and Restful ad-hoc APIs.

**Keywords**- Big Data; Big Data Analytics; NoSQL Database; Geo Spatial Data; Spatial Analytics.

## I. INTRODUCTION

By 2050, a global population of 9.7 billion will demand 70% more food than is consumed today<sup>1</sup>. Feeding safe, nutritious and affordable food to billion plus population by sustainable food production is the biggest challenge for agriculture. In India, 68% of population is rural and living in small villages, and out of 58% of population is sustained by agriculture<sup>2</sup>. Big data analytics has a huge potential to support the development and delivery of agricultural information and services to make farming economical and sustainable (socially, economically and ecologically).

### A. Spatial Analytics in Agriculture

Spatial data are playing an important role in many application areas such as, agriculture, finance and banking, disease surveillance, retail and E-commerce, etc. For example, the technologies like, Global Positioning System (GPS), remote sensing, mobile devices enable agro-users to collect high resolution spatial data in agriculture applications. The GPS, which can very accurately locate a

geographical position, has made development of spatial data in agriculture. Big data technology plays a key role in spatial analytics and helps to enhance decision making in real time. GIS tools for Big Data processing facilitate deep insights and predictive modeling for policy making in agriculture domain. It supports spatial analysis of unstructured data (meteorological data, sensor data, e-mails, blogs, social media data, etc.) in real time. It also enables efficient query processing and data mining on huge spatial-temporal data.

Geo-spatial data is very important to develop flexible and versatile functions and applications, for developing multidisciplinary applications, helping in planning, managing and utilizing natural resources efficiently using spatial analysis. Spatial data in agriculture can help farmers to enhance knowledge of farmland, increase value of farmland, and market potentials. Location-based and Geo data applications are major contributors to Big Data.

### B. Need for Agricultural Information Systems – Indian Context

In the world of digitalization, Indian farmers are still considering fellow farmers, agro retailers, various digital media channels, and agriculture experts as preferred source of information; and make decisions based on them. The information provided by extension services are perceived to be either biased (e.g. agro marketing companies) or less actionable due to lack of consistency, accuracy and personalization. Hence, farmers never reach near 100% production potential. Farmers have a dearth of information on weather, soil nutrients, optimal market price, availability of agro products, and timely advice on pest/disease management.

There is a large technological gap between farmers and information due to: 1) Lack of adoption of modern technologies due to some unresolved issues still persists in the existing systems; for example, lack of recommendations in local languages, 2) Farmers reluctance to share their personal/farm-level data, 3) High cost that farmers have to pay to get sophisticated and customizable solutions provided by proprietary and monolithic tools, 4) Farmers are getting piecemeal of information rather than comprehensive insight and personalized solutions on time using existing information systems.

The main contributions of the paper are:

1. To propose spark based agricultural information system including Geo-spatial data.
2. To realize the proposed system on top of open source architectures and perform relevant analytics on real life datasets in agriculture including Geo spatial data.

<sup>1</sup> <http://www.fao.org>

<sup>2</sup> <http://www.mospi.gov.in>

3. To develop web based analytical and visualization services for cotton crop in Gujarat and explore analytical results through Restful ad-hoc APIs and interactive maps.

The paper is organized as follows. Section II lists the existing agricultural systems and applications using GIS and big data, Section III describes the proposed system, and Section IV shows the experimental results

## II. LITERATURE SURVEY

Agricultural systems that have been built and realized through the concept of big data analytics are quite limited, compared to the existing systems. Big data analytics application in agriculture has a huge potential which has been untapped so far, especially in the Indian context.

[7, 8, 9, 10, 11] have demonstrated and implemented web-GIS based information systems for agriculture using traditional GIS tools and technologies. These technologies are often insufficient to provide a complete picture of analytics in a geographic context.

P. Shah et al. [12] have demonstrated big data analytics architecture for agro advisory system and implemented an analytical engine on top of open source architectures Spark and Cassandra. Garg, R. et al. [13] developed a framework for generating recommendation solutions for the paddy leaf blast in Punjab state, India using Hadoop and HIVE. S. Lamrhari et al. [14] suggested a profile based architecture for precision agriculture to improve decision making in real time based on profile by introducing dynamic big data service selection and composition method. Claudia Vitolo et al. [15] presented an overview of the current state-of-art of web-based tools and technologies for processing large and heterogeneous data sets and investigated most relevant web-based environmental data processing tools in the big data era. Peisker, A. et al. [16] proposed a conceptual framework for an integrated analytics approach towards rural development by supporting many sectors like agriculture, health, governance, banking, etc. in India. Xie, N. F. et al. [17] presented a conceptual system for an agricultural information system hierarchy based on big data technology. Chalh, R. et al. [18] has presented a conceptual architecture of Big Data Open Platform used for supporting water resource management. The Platform has been designed to provide effective tools that allow water system managers to solve complex water resources systems, water modeling issues and help in decision making. Y. O. Sayad et al. [19] has presented an efficient strategy for crops and water resources management through big data and remote sensing to process satellite images over large areas and extract insights from the collected data to help farmers in managing their crops and taking irrigation decisions.

Compared with the existing agricultural systems that are demonstrated after a human-in-loop decision process, the proposed system is characterized by following technical features:

1. Provides a big data integration and analytical platform built on open source architectures for enrichment of large data sets including geo-spatial data.

2. To collect and process complex data coming from heterogeneous and distributed data sources as inputs to corresponding services to perform relevant analytics.
3. Capable to store heterogeneous and distributed data with diverse data model support.
4. Generate high quality data reports through value added data processing and management. The data can be accessed via web based analytical services.
5. Develop web based analytics and visualization services for agriculture and provide analytical solutions to end users through interactive maps and Restful ad-hoc APIs.

## III. PROPOSED SYSTEM

Big data solutions may be implemented on a range of technological platforms like, stream processing engines, relational DBMS, analytical DBMS or on non-relational data management platforms, e.g. a column database. The challenge is to build and implement a robust and efficient system for massive scale data management including Geo spatial data using state-of-the-art big data tools and technologies. Open source big data solutions are coming with very limited functionality compared to the commercial and proprietary solutions. Hence, the selection and integration of big data open source architectures to manage data at scale is the main issue for data architects/engineers. The paper presents a spark based agricultural information system to reduce the technology gap between agro users and information. The proposed system is shown in figure 1.

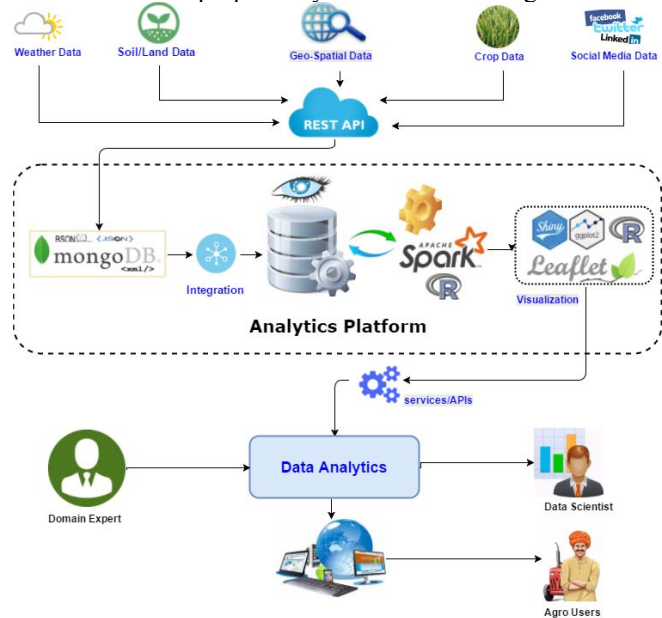


Figure 1. A spark based agricultural information system

### A. Data collection

Compared to open datasets and digital data available over the web and Internet, majority of the agriculture based data like crop data, soil data, farmer/farm level data, etc. are archived by agriculture universities, research organizations, and meteorological departments as dark data in India. These noisy, inconsistent and poor quality data are available in legacy formats (spread sheets, PDFs and hand written).

Collecting and cleaning such data and prepare them for analytics is the prime issue for big data application developer in agriculture domain.

The proposed system is able to collect and process complex data coming from heterogeneous and distributed data sources as inputs to corresponding services to perform relevant analytics. Data collection and extraction is done using available services like OpenWeatherMap APIs, GIS web services etc. and third party Restful APIs implemented by system developer.

### B. Analytics platform

The proposed system aims to realize big data integration and analytical platform built on open source architecture for enrichment of large data sets including geo-spatial data. The analytics platform is built to connect structured and unstructured data in real time or batch processing for data management and analytics of systems to make decision in operational environment, e.g. weather analytics applications should be able to process weather streams in a real time, extract relevant information and identify values that do not follow the general trends. The analytics platform is built on following open source architectures for data management in big data pipeline.

#### 1) MongoDB – data/metadata store

MongoDB<sup>3</sup>, is an open source document database written in C++. It provides high availability, high performance with automatic scaling. It scales well horizontally using sharding. It stores the data in documents and documents are similar to JSON format. Data collected and extracted from various sources using Restful APIs are mainly coming in semi-structured format (XML, JSON/BSON). MongoDB is the best option to store such data and Meta data. These data can be imported into MongoDB using GridFS specifications, Nodejs and R. The extracted data needs to be integrated into a unified format such that it can be searched, queried, and analyzed for advance analytics.

#### 2) Cassandra – persistent store

Cassandra [20] is a massive scale, high available and high performance columnar data store. It is mainly used to store and process structured data at scale. It has a peer-to-peer architecture with no single point failure. It is the only database which supports write durable operations compared to CouchDB, Hbase, and MongoDB. It is used as persistent storage to store attribute data sets on weather, soil and crop production.

#### 3) Spark – big data processing framework

Apache Spark [21] is a general-purpose cluster computing framework with advanced in-memory programming model support. It offers language-integrated APIs in Scala, Python, Java, and R. Main components of Apache Spark are Spark core, and upper level libraries. These libraries are used for graph processing (Graphx), stream processing (Spark Streaming), machine learning (MLib) and structured data processing (SparkSQL). Its core data abstraction, the Resilient Distributed Dataset (RDD), has made developers to design scalable algorithms and

pipelines conveniently with better performance. It can be easily integrated with different data sources like HDFS, Cassandra, MongoDB, and other Hadoop based data source. It outperforms Hadoop by 10x in disk and 100x in memory.

Here it is used to import/export agriculture datasets to and from Cassandra using Spark-Cassandra-Connector API and perform analytics. Spark-Cassandra-Connector<sup>4</sup> is an open source project developed by DataStax and implemented in Scala. It allows to import data into and export from Cassandra using spark by exposing Cassandra table into Spark RDD. Spark core libraries are used to perform attribute join and aggregation operations. Spark ML libraries are used to implement various analytics algorithms in Agriculture.

#### 4) GeoSpark– Spark Extension for Spatial analytics

GeoSpark<sup>5</sup> is an official third party project of Apache Spark. It is an open source in-memory cluster computing system for large-scale spatial data processing. It extends RDDs to form Spatial RDDs (SRDDs) and efficiently partitions SRDD data elements across machines. It can load, process and analyze large scale spatial data by parallelized spatial transformations and actions using SRDD abstraction. GeoSpark supports spatial objects (Point, Rectangle, Polygon and LineString), geometrical operations (Minimum Bounding Rectangle, PolygonUnion and Self-join) and spatial query operations (spatial range query, spatial join query and spatial KNN query).

However GeoSpark can only perform spatial analytics on spatial coordinates with no user-friendly programming interface like SQL or the dataframe API, the spatial operations are performed separately using Scala libraries. The resultant data are integrated into R environment for further analytics and visualization. GeoSpark spatial operations can be used to implement agriculture applications like, a) find aggregated weather per agro climatic zone, b) find number of agro markets in particular region, c) find number of regions having rainfall below the threshold value, and d) find the nearby warehouse. Some of the applications are implemented and shown in results.

#### 5) Shiny– web application framework for R

Shiny<sup>6</sup> is a web application framework for R. It facilitates R developers to write interactive web applications with no web development skills (HTML, CSS, and JavaScript) are required. Shiny package of Rstudio comes with attractive user interfaces, highly customizable slider widget with built-in support for animation, and pre-built output widgets for displaying tables, plots and printed output of R objects.

It also supports fast bidirectional communication between the web browser and R using httpuv package. The user interfaces can be built entirely in R. Visualization libraries like leaflet and ggplot integration with shiny applications enables to display Geo spatial data in the web browser. Web based analytical and visualization services are built and implemented on top of shiny framework.

<sup>4</sup> <https://github.com/datastax/spark-cassandra-connector>.

<sup>5</sup> <http://www.geospark.datasyslab.org>

<sup>6</sup> <http://www.shiny.rstudio.com>

<sup>3</sup> <http://www.mongodb.org>

6) *Sparklyr<sup>7</sup> – R interface for Apache Spark*

Sparklyr is an R interface for Apache Spark which offers much more functionality compared to the existing SparkR interface developed by Databricks. It allows dplyr-based data transformations on spark datasets, as well as access to the machine learning libraries of both Spark and H2O Sparkling Water from R. It is used to deploy Spark jobs to the Spark cluster locally or remotely.

C. *Web based analytical and visualization services*

The system provides a web based user interface by developing analytical and visualization services through Restful ad-hoc APIs, and interactive maps. Web based interactive maps are implemented using R libraries (leaflet, ggplot, and shiny) in the form of choropleth map with user interface, pulse markers and plots. Restful web services are implemented to explore data and analytical results using SparkSQL interface on top of analytics platform. Logical flow diagram of Restful API implementation is shown in figure 2.

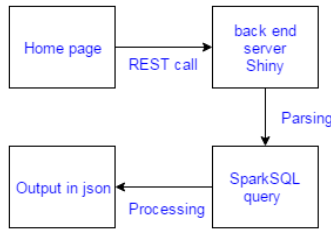


Figure 2. Restful API implementation – logical flow diagram

IV. EXPERIMENTAL RESULTS

The proposed system is implemented by developing web based analytics and visualization services for cotton crop in Gujarat, India. Cotton is an important non-food crop which provides lint to textile industry, high protein feed to livestock, oil for human consumption, byproducts used as fertilizer, produce paper and cardboard, etc. [5]. India is the second largest cotton crop producer and consumer [6] and Gujarat stands second in cotton production after Maharashtra. Cotton is a major cash crop in Gujarat. Effective advisory for cotton crop may lead productivity growth in cotton, so as the economic growth in Gujarat. The implementation is carried out as follows:

A. *Data collection, preprocessing and integration*

Spatial and non-spatial data on weather, crop and market are collected from many sources like meteorological departments, agriculture universities and web portals. The summary of data collection is given in Table 1.

Data preprocessing and integration techniques like handling missing values, data duplication removal, data validation, similarity search, joins are applied to get consistent data sets on weather, crop and market.

TABLE 1. DATA COLLECTION

Datasets	Data source	Description
Weather data (1992 – 2007)	<ul style="list-style-type: none"> <li>Archived data</li> <li>www.Indiastat.com</li> </ul>	Data collected for seven districts of Gujarat in Spread sheets/document format
Crop data for cotton crop (1960 – 2007)	<ul style="list-style-type: none"> <li>www.Indiastat.com</li> <li>http://apy.dacnet.nic.in/crop_fryr_toyr.aspx</li> <li>Archived data</li> </ul>	Data collected for eighteen districts of Gujarat in Spread sheets/document format
Market data	http://agmarknet.gov.in/	Data collected for 429 agro markets in Gujarat in PDF document
Current weather data	Openweathermap API	Json format
Spatial data for Gujarat	www.diva-gis.org	Shape files

B. *Analytics using spark/Geospark*

The system is implemented on an Ubuntu 14.04 machine with Intel® core™ i3-4130 CPU @ 3.40GHz processor and 4GB RAM. Spark applications with Cassandra data store built on the proposed system and deployed using Sparklyr. The architecture for local deployment of spark applications in R is shown in figure 3.

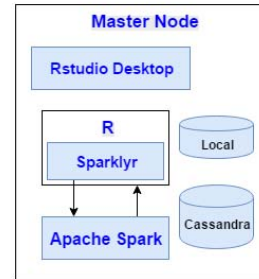


Figure 3. Local deployment using Sparklyr

The system facilitates end users to explore analytical results and data using web based interface through interactive map and ad-hoc APIs. The analytical applications built on spark Cassandra cluster are as follows:

1) *Crop yield prediction for cotton crop*

A Spark based application is developed to predict cotton crop yield based on weather parameters (Average temperature and Rainfall) using Multiple Linear Regression algorithm described in [22][23]. The implementation is done by performing various transformations and aggregation on weather and crop data (1992-2007) using spark core and spark ML libraries. Experiments are performed on seven districts of Gujarat. The yield prediction results are shown in Table 2.

The resultant crop prediction map for cotton crop is shown in figure 4. Crop production, land usage and yield trend from 1960-2007 as well as actual vs. predicted crop yield trend will be displayed on user interface on polygon click event.

<sup>7</sup> <http://spark.rstudio.com/>



TABLE 2. EXPERIMENTAL RESULTS

District	R-squared	Actual yield in 2007 Bales/'00 Hectare	Predicted yield in 2007 Bales/'00 Hectare	Difference Bales/'00 Hectare
Vadodara	0.44	447	359	88
Bharuch	0.53	285	176	109
Jamnagar	0.83	695	682	13
Amreli	0.68	643	476	167
Surat	0.78	545	489	56
Panch Mahals	0.56	469	301	168

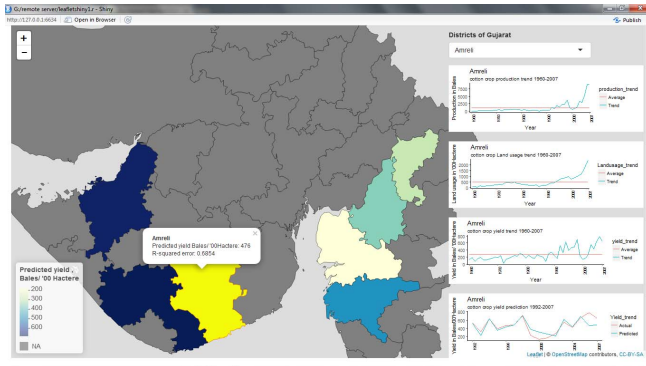


Figure 4. Crop yield prediction map for cotton crop, Gujarat

2) *Current weather with historical weather trend*

Current weather map for average temperature is shown in figure 5. Daily and monthly data for average temperature from 1992 to 2007 will be displayed on polygon click event. It also depicts the hottest location with pulse marker. The hottest location is identified using SparkSQL window functions. Openweathermap API is used to collect current weather at particular locations.

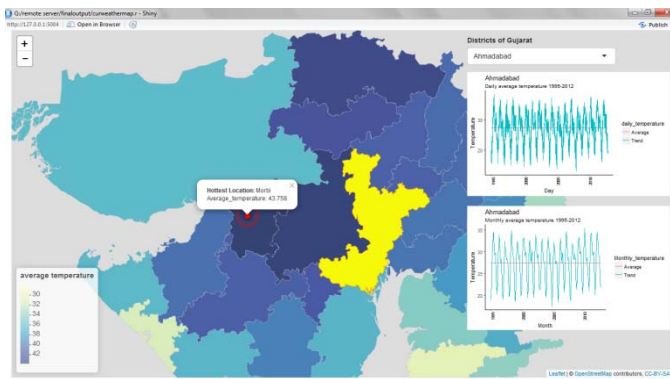


Figure 5. Current weather map for average temperature

3) *Analytics on agro market data*

Choropleth map of Gujarat depicts number of agro markets per district which are trading cotton crop. The filtered agro markets for cotton crop are depicted on map using markers. Geo referencing of market location is done using Google Map Service. Some of the market locations have to be geo referenced manually due to lack of data. 'Points in region' function is implemented using GeoSpark

spatial join operation to integrate point and polygon data. The results are shown in figure 6.

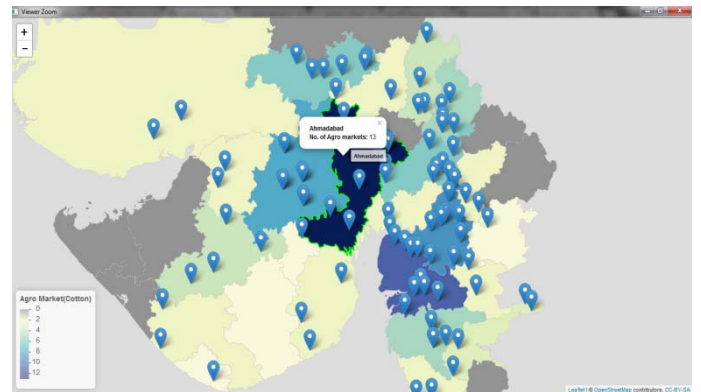


Figure 6. agro markets per district information at polygon click event

4) *Crop data aggregation over the period of time*

Choropleth map of Gujarat for cotton crop production that is averaged over the years 1960 to 2007 shown in figure 7. It depicts the trend in popup on polygon click event.

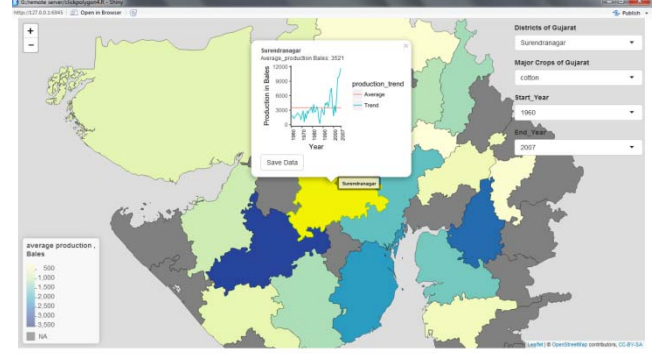


Figure 7. Cotton crop production map with popup

5) *Restful web services to explore analytical results*

Various Restful web services are built and implemented to explore analytical results. A sample RESTful web service is shown in figure 8.



Figure 8. Restful service to explore monthly weather of Ahmedabad district from 2000 to 2003

The experimental results show that Saurashtra region of Gujarat is a main contributor in cotton crop production and Surendranagar district is producing the highest cotton in Gujarat. It is quite intuitive that weather affects the crop production. We noticed that there is a sudden drop in production in 2000 due to drought in Gujarat and sudden rise in production after 2003 due to adoption of BT cotton

variety in Gujarat. That concludes that favorable weather is not only the reason to get the higher crop production; but effective recommendation may also be the leading factor in production growth. Effective advisory on ‘what to grow’, ‘how to grow’ and ‘where to grow’ may drive the farmers to near 100% production potential.

## V. CONCLUSION

Huge amount of diverse and complex data is generated from various sources like, weather stations, meteorological departments, department of agriculture, state government and research organizations, and consumed by stockholders in agricultural practices. There is an urgent need to manage agricultural data at scale with specialized systems, techniques and algorithms. The challenge is how we can exploit the full potential of open source technologies and resources and able to create a sophisticated and customizable working environment for end users to improve the productivity in agricultural practices.

A spark based agricultural information system is proposed and realized on big data open source architectures. The proposed system is implemented by developing various web based analytical and visualization services for cotton crop in Gujarat.

## ACKNOWLEDGMENT

This work is a part of a research project on ‘Developing Data Analytics Architecture, Applications in Agriculture’, funded by NRDMS and NSDI, Department of Science and Technology, Govt. of India, year 2017-2019.

## REFERENCES

- [1] J. Nabrzyski, C. Liu, C. Vardeman, S. Gesing and M. Budhatoki, "Agriculture Data for All - Integrated Tools for Agriculture Data Integration, Analytics, and Sharing," *2014 IEEE International Congress on Big Data*, Anchorage, AK, 2014, pp. 774-775.
- [2] Abhijit Suprem, Nitaigour Mahalik, Kiseon Kim, A review on application of technology systems, standards and interfaces for agriculture and food sector, *Computer Standards & Interfaces*, Volume 35, Issue 4, June 2013, Pages 355-364.
- [3] M. Hirafuji, "A Strategy to Create Agricultural Big Data," 2014 Annual SRII Global Conference, San Jose, CA, 2014, pp. 249-250.
- [4] A. Eldawy and M. F. Mokbel, "The era of Big Spatial Data," 2016 IEEE 32nd International Conference on Data Engineering (ICDE), Helsinki, 2016, pp. 1424-1427. doi: 10.1109/ICDE.2016.7498361.
- [5] Freeland Jr., Thomas B. , Pettigrew, Bill, Thaxton, Peggy, Andrew, Gordon L., "Agrometeorology and Cotton Production" Guide of Agricultural Meteorological Practices (GAMP), 2010, edition (WMO-No.134), Chapter 13.
- [6] Osakwe, Emeka, "Cotton Fact Sheet India", International Cotton Advisory Committee, May 19, 2009.
- [7] Tiago H. Moreira de Oliveira, Marco Painho, Vítor Santos, Otávio Sian, André Barriguinha, Development of an Agricultural Management Information System based on Open-source Solutions, *Procedia Technology*, Volume 16, 2014, Pages 342-354, ISSN 2212-0173, <http://dx.doi.org/10.1016/j.protey.2014.10.100>.
- [8] Kumar SK, Babu SDB (2016) A Web GIS Based Decision Support System for Agriculture Crop Monitoring System-A Case Study from Part of Medak District. *J Remote Sensing & GIS* 5:177. doi: 10.4172/2469-4134.1000177.
- [9] Weiguo Han, Zhengwei Yang, Liping Di, Richard Mueller, CropScape: A Web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support, *Computers and Electronics in Agriculture*, Volume 84, June 2012, Pages 111-123, ISSN 0168-1699, <https://doi.org/10.1016/j.compag.2012.03.005>.
- [10] Z. Zhu, R. Zhang and J. Sun, "Research on GIS-Based Agriculture Expert System," *2009 WRI World Congress on Software Engineering*, Xiamen, 2009, pp. 252-255. doi:10.1109/WCSE.2009.104.
- [11] Zhang, H., Zhang, L., Ren, Y., Zhang, J., Xu, X., Ma, X., & Lu, Z. (2011). Design and implementation of crop recommendation fertilization decision system based on WEBGIS at village scale. *Computer and Computing Technologies in Agriculture IV*, 357-364.
- [12] P. Shah, D. Hiremath and S. Chaudhary, "Big Data Analytics Architecture for Agro Advisory System," 2016 IEEE 23rd International Conference on High Performance Computing Workshops (HiPCW), Hyderabad, 2016, pp. 43-49.
- [13] Garg, R., & Aggarwal, H. (2016). Big Data Analytics Recommendation Solutions for Crop Disease using Hive and Hadoop Platform. *Indian Journal Of Science And Technology*, 9(32). doi:10.17485/ijst/2016/v9i32/100728.
- [14] S. Lamrhari, H. Elghazi, T. Sadiki and A. El Faker, "A profile-based Big data architecture for agricultural context," 2016 International Conference on Electrical and Information Technologies (ICEIT), Tangiers, 2016, pp. 22-27. doi:10.1109/EITech.2016.7519585.
- [15] Claudia Vitolo, Yehia Elkhatib, Dominik Reusser, Christopher J.A. Macleod, Wouter Buytaert, Web technologies for environmental Big Data, *Environmental Modelling & Software*, Volume 63, January 2015, Pages 185-198, ISSN 1364-8152, <https://doi.org/10.1016/j.envsoft.2014.10.007>.
- [16] Peisker, A., & Dalai, S. (2015). Data Analytics for Rural Development. *Indian Journal Of Science And Technology*, 8(S4), 50-60. doi:10.17485/ijst/2015/v8iS4/61494.
- [17] Xie, N. F., Zhang, X. F., Sun, W., & Hao, X. N. (2015, June). Research on Big Data Technology-Based Agricultural Information System. In *International Conference on Computer Information Systems and Industrial Applications*. Atlantis Press.
- [18] Chalh, R., Bakkoury, Z., Ouazar, D., & Hasnaoui, M. D. (2015, June). Big data open platform for water resources management. In *Cloud Technologies and Applications (CloudTech)*, 2015 International Conference on (pp. 1-8). IEEE.
- [19] Y. O. Sayad, H. Mousannif and M. Le Page, "Crop management using Big Data," *2015 International Conference on Cloud Technologies and Applications (CloudTech)*, Marrakech, 2015, pp.1-6. doi:10.1109/CloudTech.2015.7337003.
- [20] Avinash Lakshman and Prashant Malik. 2010. Cassandra: a decentralized structured storage system. *SIGOPS Oper. Syst. Rev.* 44, 2 (April 2010), 35-40. DOI=<http://dx.doi.org/10.1145/1773912.1773922>.
- [21] Matei Zaharia, Mosharaf Chowdhury, Michael J. Franklin, Scott Shenker, and Ion Stoica. 2010. Spark: cluster computing with working sets. In *Proceedings of the 2nd USENIX conference on Hot topics in cloud computing (HotCloud'10)*. USENIX Association, Berkeley, CA, USA, 10-10.
- [22] Fisher, R. A. "Statistical Method for Research Works", Published by Oliver and Boyd (1925). Fisher, R.A. "The inuence of rainfall on the yield of wheat at Rothamsted", *Roy. Soc. London, Phil. Trans.*, B, 213, 89-142 (1924).
- [23] Agrawal, Ranjana, and S. C. Mehta. "Weather based forecasting of crop yields, pests and diseases-IASRI models". *A\_J. Ind. Soc. Agril. Statist.* 61(2), (2007): 255-263.
- [24] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL<https://www.R-project.org/>.