

# Spatiotemporal Applications of Big Data

Khalid Bin Muhammad

College of Computer Science & Information  
Systems, IoBM  
Karachi, Sindh, Pakistan

Tariq Rahim Soomro

College of Computer Science & Information  
Systems, IoBM  
Karachi, Sindh, Pakistan

## ABSTRACT

Big data has emerged as a field of study and gained huge importance these days for both industry and researcher's point of view. Initially database management systems (DBMS) developed to solve data management and relevant queries. Relational DBMS (RDBM) gave another innovation to the database field. Through the passage of time, it observed that some issues remained unsolved and required some more dimensions be added to the data. One of those was time and the other was location. Spatiotemporal aspects of data gained importance and scientists thought of incorporating these in the upcoming databases. This paper covers the inclusion of these dimensions in a database and its applications in today's world. It also compares some of the tools used these days and suggests a combination for better results in an efficient and cost-effective way.

## Keywords

Spatiotemporal, Big Data, RDBMS, Hadoop.

## 1. INTRODUCTION

The term "geographic data", "geo-data", "geographic information" and/or "geo-spatial data" is usually known as, spatial data, which stores the data in the form of coordinate and topology and can be easily mapped. Features and boundaries on our planet earth identified by this data. Some features exist in nature and others are synthetic [1]. The term "temporal" data refer to the data, which varies over time [2]. It may identify the change in characteristic of an object with the passage of time. This characterizes data elements related to time. The term spatiotemporal databases connects both space and time aspects of data together [3]. Geometries may change continuously in relation to time for an object. If it requires to study movement of an object than it is considered as a point. If it requires dimensions of challenges in data management [4]. Big Data is visualized by some researchers as data possessing characteristics of having Volume, Velocity, Variety, Variability and Complexity [5]. Big Data is large sets of structured and unstructured data which needs be processed by advanced analytics and visualization techniques to uncover hidden patterns and find unknown correlations to improve the decision making process [6].

Volume is the size of the data. Usually size is in multiple of terabytes or petabytes. Variety refers to the variation of data types in a dataset. Structured data, semi-structured and unstructured data all may be used. Structured data is usually the data in form of spreadsheets and relational databases while text, images, audio, and video are unstructured data, which is difficult to handle by machines. Velocity is the speed of data generation. If the rate of data generation is high than it should be analyzed in, similar speed and acted upon to take timely decisions. Data created immensely these days due to mobile devices and electronic sensors. It is now a need to plan and take decisions online to be successful. Mobile devices are

generating huge data of spatiotemporal nature. On the other hand GIS systems also contributing largely in this area. Raw data can used to deduce spatial relations i.e. distance as well as direction and shape data and temporal data, which includes occurrence time, duration, occurrence before or after events etc. Relational Databases incorporate some aspects related to time and space, but are not able to deal with complex data structure and behaviors. This paper written keeping in view, analyzing the uses of spatiotemporal data in selected industries; such as transportation; medical/health; agriculture; and manufacturing industries. It gives a comparison of industry wise application of spatiotemporal tools and their year wise development both horizontally and vertically. This paper is organized as follows; section 2 covers the literature review; section 3 covers material and methods; section 4 discusses results and findings; and section 5 discussion and future work.

## 2. LITERATURE REVIEW

Spatial data may include objects that may be complex in nature. They may be points, lines, polygons and other shapes having different sizes and elevations. Satellite data may be raster in nature, while vector data requires different methods to process. The properties of spatiotemporal data are to be interpreted to make it easy to understand and intuitive. It may converted to visual form for human understanding. Spatiotemporal big data is use in variety of applications; such as transportation; medical/health; agriculture; manufacturing etc. A brief of the work done in above-mentioned industries is as follows:

### 2.1. Transportation

Spatiotemporal data is use by transport industry in many ways. New types of geospatial applications are being developed to be used by position-aware devices installed in vehicles. Large movement data is produced and is available for analysis. New Spatial databases are required to handle such data with new techniques [7]. Application domains for moving object databases are expanding and new challenges are being face. It is required to propose powerful query predicates [8]. GIS applications involving time may also include changes in regions with the passage of time. Vehicle position tracking may include cars, trains, aircrafts or vessels that may require monitoring. Similar evolution is also occurring in the all other fields. It is a difficult to model such applications having spatial aspects as well as incorporate newer challenges. One issue with this approach is feasibility of implementation becomes low as it provided with big data. Fleet management requires memory space and query from spatiotemporal data. Cabs needs to be tracked, in real-time to improve their services. It may require to answer closest taxi to a calling customer; how many taxi's in proximity to a client in next 10 minutes [9] etc. these questions are not easily answered in available applications handling spatiotemporal

data. Limitations of current applications is that they consider moving objects as point objects changing their position continuously over road networks hence called moving objects. The road networks represented with the help of maps, which usually updated on regular basis. Moving objects connected with road networks to make it meaningful. A new set of powerful predicates are required to act on these systems. The language proposed in this research handles many reasonable questions at a time related to such objects [10].

A research in [11] carried out on the traffic volume predictions by incorporating the spatial and temporal characteristics of data. Neural networks used to perform this analysis. The reason was to optimize traffic volume from more than one control point to provide accurate forecasts. A system that can evolve self-optimizing in time and space can be helpful in such predictions. Traffic conditions information is collected using advanced sensor systems. Intelligent Transportation Systems (ITS) is the name given to such systems. Videos often recorded for traffic analysis for traffic planners. Vehicles can be classified in this way as well. Traffic flow analysis can also be done in addition to incident detection and analysis. Vehicle tracking is also an added advantage in such system. Multimedia augmented transition network model is developed with the help of unsupervised image segmentation method and vehicle tracking algorithm. The learning process running in the background is simple and very effective. Majority of objects identified successfully through this framework [12]. The data related to transportation has unique characteristics containing features of both spatial and temporal nature. This data is useful in the transportation planning area. Object oriented method used in this case to manipulate spatiotemporal information. Time stamps can be added to the objects. This works for simple vector data structure [13].

In example of transportation, urban transportation system was focus by researchers in the literature and ITS approach mainly applied. The issues related to trip planning and navigation also addressed in literature. Concepts can be abstracted from spatiotemporal sensors data; mobile data; social networks data, as well as can also be used via crowdsourcing. IntelliDrive is an effort by Transports department in USA. Social networks used for information crowdsourcing. Travelers as well as traffic administrators kept informed about the spatiotemporal information changes. The issues of traffic congestion, availability of parking, ride sharing options and joint driving opportunities are resolved [14]. Mobile devices are everywhere and are generating major portion of Web traffic these days. Smart phones well used and liked during travelling by people in majority of countries. Network coverage issues or low bandwidth make people suffer during commuting. To study the characteristics of web traffic is a challenge for researchers. Studies in literature show that trendy content include social networking and news during travel. Content similarity also found high between specific routes. Caching of content reduced the load of bandwidth usage in these cases [15].

## **2.2. Health / Medical**

Another industry utilizing spatiotemporal data in health. Hospitalization related risks are mainly measure with the spatiotemporal data. In [16] three conditions were selected and evaluated; in terms of risks in hospitalization. The ages of patients considered above 20 years of age data taken for years 2002 to 2013. Bayesian method applied to the data to

enable the estimation of the function of different types of judgments on events like hospital admissions. Another study conducted to distinguish spatiotemporal trends of AIDS death discrepancy in 67 Florida jurisdictions. Reason was to find out if trends vary as per race, age or gender. In this study data of AIDS related deaths were analyzed from 1987 to 2004. Results geographically referenced in maps using software used by the “Centers for Disease Control”. Results revealed many interesting facts hidden in the data. Time trends also showed that AIDS death was spiky in 1995 and then sharply dropped till 1998. These findings were very useful as medical resources distribution and improved spotlight on community wellbeing [17]. Another study used space-time scan statistics, which can examine spatiotemporal effects unlike other Spatial-only clustering methods. The weakness in the current study was that it didn’t discriminate deaths directly related to AIDS from non-HIV related deaths in the dataset. Another restraint was increasing number of illegal immigrants to Florida. It was good effort to visualize the spatiotemporal characteristics of AIDS and enabled community to provide better medicinal care delivery [17].

## **2.3. Agriculture**

Agriculture is one of the industry in which Spatiotemporal data is being used in many ways. China cropland services developed spatiotemporal database system for its cropland resources. The system receives request from clients and processes the data. The data is stored in a large data warehouse. SQL Server 2000 was used which is highly scalable as it also accommodates small megabytes of personal databases and adapts hundreds of millions of users enabling them to use huge terabyte database. Land use data and covered land data is stored and hence many queries related to land coverage, graphical queries and spatial analysis were resolved. Tabulation of data was made easier now and updating of data can be done on real-time basis [18].

A study was conducted to develop a statistical model for agricultural yield data. A joint modeling was done for spatiotemporal data using hierarchical Bayesian model. Prior to this there were many uncertainties in the predictions being made for the insurance of crops as well as their premium that were resolved. Posterior predictive criterion was used to assess different model stipulations. This method can be applied on limited data as well, but gives better results for huge data. A new legislation was passed to give incentives or subsidies to agricultural producers for purchasing insurance contracts. Also it was used for the forecast of corn crop for some selected counties for the given data [19]. Another study was done to demonstrate that spatial panel data models can be used well for an econometric analysis of data related to agricultural farms. Control of drainage water equipment’s use on corn production was assessed. A yield equation was formed with the help of data coming from fields and spatial methods. Some more factors were added to it to improve the results. The results proved that restricted drainage had a noteworthy impact on corn production [20].

Internet of Things has enabled various applications help in crop growth monitoring and selection and decision making of irrigation related issues. In [21] authors discussed a newly developed OpenIoT platform for a Digital Agriculture use case named Phenonet. They include proposal for data coming from sensors to be collected smartly, annotated, validated, processed and stored. In this study authors presented the

semantically enhanced digital agriculture use case Phenonet built with the OpenIoT platform (Phenonet-OpenIoT).

## **2.4. Manufacturing:**

Manufacturing industry grows when industrialization is in the boom in a country and hence economic development is a by-product. Industrialization results in economic growth of the society and positively contributes to the commerce, trade, communication, human resources, social and cultural environment, availability of technical assistance and the indigenous research as well. A study was conducted to study the district wise spatiotemporal development (1984 to 2012) of industrial hubs in India (Himachal Pradesh). It considered the role of Large and Medium scale industries, keeping in mind that both the central and state government had taken active steps to expand, encourage and enhance their expansion. District wise spatiotemporal analysis of registered industrial units and the industrial workers employed was done in 2017 to find out the developmental status of the Medium and Large Scale Industries in Himachal Pradesh from the duration of 1984 to 2012. This study suggests that in order to support efficient and methodical industrial growth the industry department of the state government should devise effective industrial development policies and strategies on a regular basis as done between years 2002 to 2012. The packages of incentives like interest subsidy, power subsidy, land subsidy, grants in aid for artisans, State capital investment subsidy, incentives for export oriented units, State transport subsidy on plant & machinery etc can attract perspective entrepreneurs [22].

RFID is being used during inventory management in production processes. Production facilities are now smart with tags placed everywhere to identify objects during assembly. It's the latest development that has enabled smart production with reduced errors. In such facilities huge data is generated and to be stored for inventory planning and arrangement for production. Authors in [23] suggests RFID-Cuboids to be created to develop a data warehouse so that the inventory data could be highly integrated. A Map Table can be developed to link those cubes for better information granularity. Spatiotemporal inventory path is defined so that the inventory operators and machines could be evaluated. Another study was done to forecast the tactical sales predictions in tire industry. Human forecasts are known to be biased. A new method was to identify major important indicators for sales among many macroeconomic indicators spread in various regions to make better forecasts. The forecast developed was based on a model. By the help of this method managers are able to understand market dynamics and result in better sales. Prediction intervals can be generated and uncertainty of prediction is covered as well. The solution can be applied to other industries also [24].

## **3. MATERIAL AND METHODS**

This study is conducted to compare the application of various big data tools on selected industries. It is a qualitative study. Qualitative Research is mainly done for exploration and deduce some results. It is used to gain an understanding of underlying reasons, opinions, and motivations. It provides insights into the problem or helps to develop ideas or hypotheses for potential quantitative research. In this study four industries were analyzed including transportation, medical/health, agriculture and manufacturing. The tools being discussed are Hadoop, MapReduce, Apache Pig,

NoSQL, Hive, Cassandra, HBase, Spark, Kafka and Apache Tez; and shown in Figure 1 below.

Hadoop [25] is part of the Apache project sponsored by the Apache Software Foundation supports the processing and storage of extremely large data sets in a distributed computing environment and is Java-based an open source programming framework. MapReduce is also programming model and an associated implementation for processing and generating big data sets with a parallel, distributed algorithm on a cluster [26]. Apache Pig is a high-level procedural language platform developed to simplify querying large data sets in Apache Hadoop and MapReduce. Apache Pig features a "Pig Latin" language layer that enables SQL-like queries to be performed on distributed datasets within Hadoop applications [27]. A NoSQL (originally referring to "non SQL" or "non-relational") database provides a mechanism for storage and retrieval of data that is modeled in means other than the tabular relations used in relational databases [28].

Apache Hive is built on top of Apache Hadoop and considered as data warehouse software project for providing data summarization, query and analysis. Hive gives an SQL-like interface to query data stored in various databases and file systems that integrate with Hadoop [29]. Apache Cassandra is an open-source distributed NoSQL database management system designed to handle large amounts of data across many commodity servers, providing high availability with no single point of failure [30]. HBase is Java based open-source, non-relational, distributed database modeled after Google's Bigtable. It is developed as part of Apache Software Foundation's Apache Hadoop project and runs on top of HDFS (Hadoop Distributed File System), providing Bigtable-like capabilities for Hadoop [31].

Apache Spark is an open-source engine developed specifically for handling large-scale data processing and analytics [32]. Spark offers the ability to access data in a variety of sources, including Hadoop Distributed File System (HDFS), OpenStack Swift, Amazon S3 and Cassandra. Apache Kafka is an open-source stream processing platform developed by the Apache Software Foundation written in Scala and Java [33]. The project aims to provide a unified, high-throughput, low-latency platform for handling real-time data feeds. Apache Tez is an extensible framework for building high performance batch and interactive data processing applications [34], coordinated by YARN in Apache Hadoop. Tez improves the MapReduce paradigm by dramatically improving its speed, while maintaining MapReduce's ability to scale to petabytes of data.

## **4. RESULTS AND FINDINGS**

Transportation industry focuses on fast moving objects and requires powerful query predicates. Traffic flow analysis also requires advanced sensor based systems. Load also needs to be reduced as current systems consumes huge bandwidths on internet. Health industry also faces challenges related to data segregation for better analysis. OpenIoT platform also awaits extensions to diverse services for the development of digital agriculture. Manufacturing industry also requires similar features in their applications. The table-1 below gives a comparison of the various spatiotemporal tools used for data in selected industries:

## 5. DISCUSSION AND FUTURE WORK

Four major industries were considered in this study. Transportation is one of the major industries. The Geo-Spatial position aware vehicles monitoring is being done. Fleet management requires huge memory and powerful query predicates. Moving objects are being considered as point objects. Traffic flow analysis is a major issues these days. Spatiotemporal information change and updating is another issue. Low Internet bandwidth is another hindrance. Health / Medical is another industry. Hospitalization related risks are mainly measure with spatiotemporal data. Studies have been carried out to find the relation between AIDs related deaths in certain regions. Agriculture also depends on spatiotemporal data in various aspects. Land use data and land covered data etc. are stored. Insurance of crops is one the challenges and also yield prediction for crops. Manufacturing industry can also benefit from spatiotemporal data. Studies have been done to find the effect of various incentives given to support manufacturing industry for export oriented products and their result location-wise. Spatiotemporal inventory management is

another challenge. Spatiotemporal data can also be used for sales predictions.

The various Spatiotemporal tools considered in this study Hadoop, Map Reduce, Pig, No SQL, Hive, Cassandra, Hbase, Spark, Kafka and Apache Tez. Apache Tez has proven its significance and can help in the resolution of above mentioned issues and much more. Apache Tez is a purpose built tool 100 times faster than Map Reduce and works well with Hive and Pig. Apache Tez helps convert MapReduce to a more powerful framework as it computes on the basis of dataflow graphs. It can be used to build end-users applications with improved performance and flexibility. Hadoop has been used in batch processing mainly but sometimes on real time query processing as well. Map Reduce is not suitable for machine learning and Apache Tez may help Hadoop address such issues. Apache Tez is highly customizable and requires less effort from users to make it work. It has enabled Hive and Pig significant improvement in their response times.

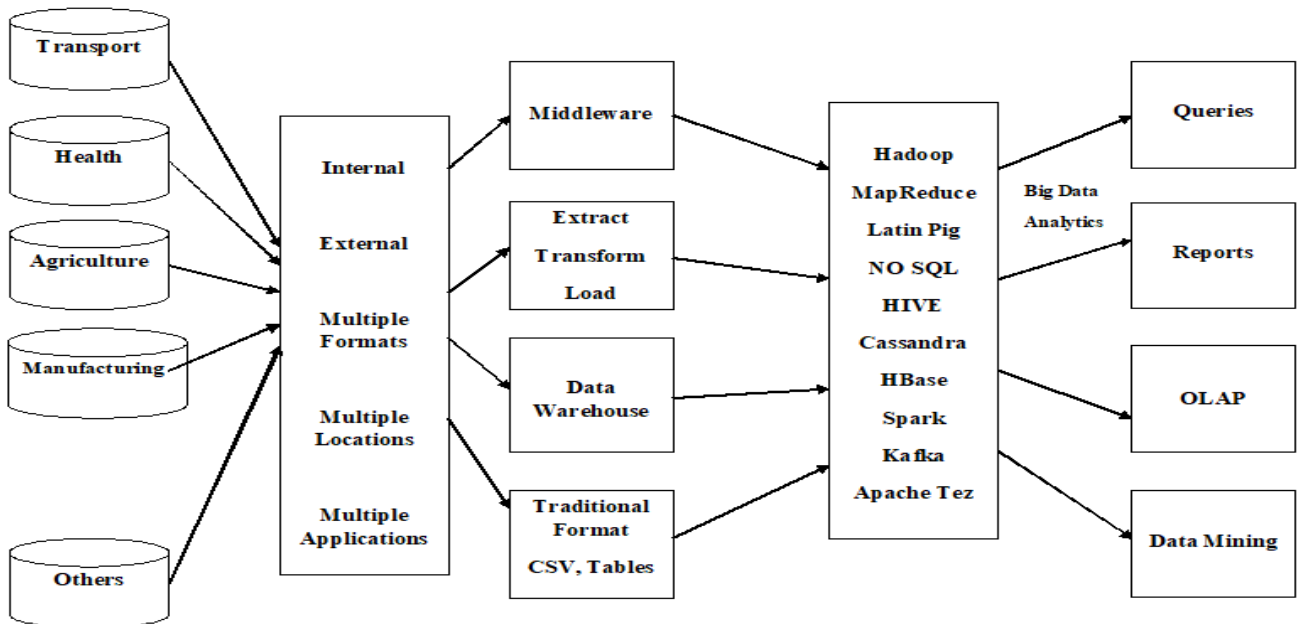


Fig 1: Spatiotemporal Framework for Big Data Tools

Table 1: Spatio-temporal Tools Used Industries Wise

Spatio-temporal Tools / Industries	Hadoop	Map Reduce	Pig	No SQL	Hive	Cassandra	Hbase	Spark	Kafka	Apache Tez
Transportation	Yes [35]	Yes [35]	No	Yes [36]	Yes [37]	Yes [38]	Yes [39]	Yes [40]	Yes [37]	No
Health / Medical	Yes [41]	Yes [41]	Yes [42]	Yes [43]	Yes [40]	Yes [44]	Yes [40]	No	No	No
Agriculture	Yes [45]	Yes [45]	No	Yes [46]	Yes [47]	Yes [48]	Yes [49]	Yes [49]	No	No
Manufacturing	Yes [50]	Yes [50]	No	No	No	No	No	Yes [50]	No	No

## 6. REFERENCES

- [1] M. Coyle, "Data models in geographic information systems," *COMMUNICATIONS OF THE ACM*, vol. 40, no. 4, pp. 103-111, 1997.
- [2] R. T. Snodgrass, "Temporal Data Management," *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, vol. 11, no. 1, pp. 36-44, 1999.
- [3] M. ERWIG, "Spatio-Temporal Data Types: An Approach to," *GeoInformatica*, vol. 3, no. 3, pp. 269-296, 1999.
- [4] A. Gandomi, "Beyond the hype: Big data concepts, methods, and analytics," *International Journal of Information Management*, vol. 35, no. 1, pp. 137-144, 2015.
- [5] T. R. Soomro and A. G. S. Shoro, "Big Data Analysis: Ap Spark Perspective," *Global Journal of Computer Science and Technology: Software & Data Engineering*, vol. 15, no. 1, 2015.
- [6] T. R. Soomro and N. Thabet, "Big Data Challenges," *Journal of Computer Engineering and Information Technology*, vol. 4, no. 3, 2015.
- [7] H. Hajari and F. Hakimpour, "A SPATIAL DATA MODEL FOR MOVING," *International Journal of Database Management Systems*, p. Vol 6. No 1, 2014.
- [8] M. A. Cheema, X. Lin, W. Zhang and Y. Zhang, "A Safe Zone Based Approach for Monitoring," in *VLDB '05 Proceedings of the 31st international conference on Very large data bases*, Norway, 2005.
- [9] L. Wu, S. Hu, L. Yin, Y. Wang and Z. Chen, "Optimizing Cruising Routes for Taxi Drivers Using a Spatio-Temporal Trajectory Model," *International Journal of Geo-Information*, vol. 6, no. 12, p. 373, November 2017.
- [10] M. Vazirgiannis and O. Wolfson, "A Spatiotemporal Model and Language for Moving Objects on Road Networks," no. pp. 20-35, 2001.
- [11] E. I. Vlahogianni, M. G. Karlaftis and J. C. Golias, "Spatio-Temporal Short-Term Urban Traffic Volume Forecasting Using Genetically Optimized Modular Networks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 22, no. 5, pp. 317-325, 2007.
- [12] S.-C. Chen, M.-L. Shyu, S. Peeta and . C. Zhang, "Learning-Based Spatio-Temporal Vehicle Tracking and Indexing for Transportation Multimedia Database Systems," vol. 1, no. 11, 2002.
- [13] Z. Lijun, W. Huimin and S. Jianping, "Application Oriented Spatio-temporal Data Model Design for Transportation Planning," in *Proceedings of the IEEE ITSC 2006*, Toronto, 2006.
- [14] O. Wolfson and B. Xu, "Spatio-temporal Databases in Urban Transportation," 2010.
- [15] F. Jiang, K. Thilakarathna, M. A. Kaafar and F. Rosenbaum, "A Spatio-Temporal Analysis of Mobile Internet Traffic in Public Transportation Systems," 2015.
- [16] J. Librero, B. Ibañez, N. Martinez-Lizaga and S. Peiro, "Applying spatio-temporal models to assess variations across health care areas and regions: Lessons from the decentralized Spanish National Health System," vol. 12, no. 2.
- [17] Y.-W. Chiu, M.-Q. Wang, H.-Y. Chuang, C. Ed Hsu and E. T. Nkhoma, "A NEW APPLICATION OF SPATIOTEMPORAL ANALYSIS FOR DETECTING DEMOGRAPHIC VARIATIONS IN AIDS MORTALITY: AN EXAMPLE FROM FLORIDA," vol. 24, no. 11, 2008.
- [18] S. Hu and Z. Zhang, "Development and applications of spatial-temporal," 2010.
- [19] V. A. OZAKI, S. K. GHOSH, B. K. GOODWIN and R. SHIROTA, "SPATIO-TEMPORAL MODELING OF AGRICULTURAL YIELD DATA WITH AN APPLICATION TO PRICING CROP INSURANCE CONTRACTS," vol. 90, no. 4, 2008.
- [20] A. Nistor, R. J. Florax, J. Lowenberg-DeBoer and J. P. Brown, "SPATIOTEMPORAL MODELING OF AGRICULTURAL YIELD MONITOR DATA," vol. 08, no. 1, 2008.
- [21] P. P. Jayaraman, D. Palmer and A. Zaslavsky, "Do-it-Yourself Digital Agriculture applications with semantically enhanced IoT platform," in *2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, Singapore, 2015.
- [22] R. Kumar and A. Shankhan, "Spatio - Temporal Analysis of Industrial Hubs Development in," vol. 39, no. 1, 2017.
- [23] R. Y. Zhong, G. Q. Huang, S. Lan, Q. Dai and X. Chen, "A big data approach for logistics trajectory discovery from RFID-enabled production data," *International Journal of Production Economics*, vol. 165, no. C, pp. 260-272, 2015.
- [24] Y. R. Sagaerta, E.-H. Aghezzaf, N. Kourentzes and B. Desmet, "Temporal Big Data for Tactical Sales Forecasting in the Tire Industry," *Institute for Operations Research and the Management Sciences (INFORMS)*, no. ISSN 0092-2102, 2017.
- [25] V. T. Sharmila and S. R. Ratna, "An Efficient Bandwidth Aware Scheduling Algorithm in Hadoop," *International Journal of Computer Science and Mobile Computing*, vol. 6, no. 10, pp. 11-17, 2017.
- [26] K.Kavitha, "ASSESSMENT ON IMPROVISING FREQUENT ITEMSET MINING IN HUGE VOLUME OF DATASET," *International Journal of Computer Engineering and Applications*, vol. 9, no. 12, pp. 275-281, 2017.
- [27] C. Olston, B. Reed and U. Srivastava, "Pig latin: a not-so-foreign language for data processing," in *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, Vancouver, Canada , 2008.
- [28] S. Rao, S. S. N and S. M, "SECURITY SOLUTIONS

- FOR BIG DATA ANALYTICS IN HEALTHCARE," in Second International Conference on Advances in Computing and Communication Engineering, Rohtak, Dehradun, 2015.
- [29] A. Thusoo, J. Sen Sarma and N. Jain, "Hive - a petabyte scale data warehouse using Hadoop," in IEEE 26th International Conference on Data Engineering (ICDE 2010), Long Beach, CA, USA, 2010.
- [30] I. A. TargioHashem, I. Yaqoob and N. BadrulAnuar, "The rise of "big data" on cloud computing: Review and open research issues," *Informations Systems*, Elsevier, vol. 47, pp. 98-115, 2015.
- [31] M. N. Vora, "Hadoop-HBase for large-scale data," in Proceedings of 2011 International Conference on Computer Science and Network Technology, Harbin, China, 2011.
- [32] M. Zaharia, R. S. Xin, P. Wendell and T. Das, "Apache Spark: a unified engine for big data processing," *Communications of the ACM*, vol. 59, no. 11, pp. 56-65, 2016.
- [33] K. M. M. THEIN, "Apache Kafka: Next Generation Distributed Messaging System," *International Journal of Scientific Engineering and Technology Research*, vol. 1, no. 47, pp. 9478-9483, 2014.
- [34] B. Saha, H. Shah and S. Seth, "Apache Tez: A Unifying Framework for Modeling and Building Data Processing Applications," in Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Victoria, Australia , 2015.
- [35] X.-Y. Chen, H.-K. Pao and Y.-J. Lee, "Efficient Traffic Speed Forecasting Based on Massive Heterogenous Historical Data," in IEEE International Conference on Big Data, Washington, DC, 2014.
- [36] W. Q. Wang, X. Zhang and J. Zhang, "Smart Traffic Cloud: An Infrastructure for Traffic Applications," in Parallel and Distributed Systems (ICPADS), 2012 IEEE 18th International Conference, Singapore, 2012.
- [37] H. Poonawala, V. Kolar, S. Blandin and L. Wynter, "Singapore in Motion: Insights on Public Transport Service Level Through Farecard and Mobile Data Analytics," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, 2016.
- [38] L. Zhou, N. Chen and Z. Chen, "Efficient Streaming Mass Spatio-Temporal Vehicle Data Access in Urban Sensor Networks Based on Apache Storm," *Sensors*, vol. 17, no. 4, p. 815, 2017.
- [39] C. Zhang, X. Chen and X. Feng, "Storing and Querying Semi-structured Spatio-Temporal Data in HBase," in International Conference on Web-Age Information Management, Springer International Publishing AG 2016, 2016.
- [40] A. Eldawy and M. F. Mokbel, "The Era of Big Spatial Data," in Data Engineering Workshops (ICDEW), 2015 31st IEEE International Conference, Seoul, 2015.
- [41] F. Wang, A. Aji and G. Teodoro, "Medical Image Dataset Processing over Cloud/MapReduce with Heterogeneous Architectures," Atlanta, GA, 2017.
- [42] A. Samuel, A. Ghafoor and E. Bertino, "Context-Aware Adaptation of Access-Control Policies," *IEEE Internet Computing*, vol. 12, no. 1, 2008.
- [43] J. R. Lourenço, V. Abramova and B. Cabral, "No SQL in Practice: A Write-Heavy Enterprise Application," in BIGDATA CONGRESS '15 Proceedings of the 2015 IEEE International Congress on Big Data, Washington, 2015.
- [44] A. Mohan, D. Bauer and D. M. Blough, "A Patient-centric, Attribute-based, Source-verifiable Framework for Health Record Sharing," Georgia Institute of Technology, Atlanta, GA, Atlanta, 2009.
- [45] A. Manjula and G. Narsimha, "SPATIAL TEMPORAL DATA MINING FOR CROP YIELD PREDICTION," *INTERNATIONAL JOURNAL OF ADVANCED RESEARCH*, pp. 848-859, 2016.
- [46] S. C. Satapathy, V. K. Prasad and B. P. Rani, Proceedings of the First International Conference on Computational Intelligence and Informatics, Hyderabad: Springerlink, 2016.
- [47] M. P. McGuire and M. C. Roberge, "The Design of a Collaborative Social Network for Watershed Science," *Geo-Informatics in Resource Management and Sustainable Ecosystem*, Springerlink, vol. 482, pp. 95-106, 2015.
- [48] S. Sharma and S. Gadia, "Expanding ParaSQL for spatio-temporal (big) data," *The Journal of Supercomputing*, Springer US, pp. 1-20, 2017.
- [49] A. Karmas, A. Tzotsos and K. Karantzalos, *Geospatial Big Data for Environmental and Agricultural Applications*, Springerlink, 2016, pp. 353-390.
- [50] S. Suma, R. Mehmood and N. Albugami, "Enabling Next Generation Logistics and Planning for Smarter Societies," *Procedia Computer Science*, Elsevier, vol. 109, pp. 1122-1127, 2017.
- [51] W. Korres, T. Reichenau, P. Fiener and C. Koyama, "Spatio-temporal soil moisture patterns – A meta-analysis using plot to catchment scale data," *Elsevier, Journal of Hydrology*, vol. 520, pp. 326-341, 2015.