

A review: evolution of big data in developing country

Bayu Prasetyo^{a,1,*}, Faiz Syaikhoni Aziz^{a,2}, Kamil Faqih^{a,3}, Wahyu Primadi^{a,4}, Roni Herdianto^{b,5},
Wicaksono Febriantoro^{c,6}

^a Electrical Engineering Postgraduate, Electrical Engineering Department, Universitas Negeri Malang

^b Graduate School, Universitas Negeri Malang

^c Metrological Resources Development Centre, Ministry of Trade-Republic of Indonesia

¹ bayoe.30015@gmail.com*; ² faizsyaikhoni@gmail.com; ³ kamil.faqih201@yahoo.com; ⁴ pakwahyuprimadi@gmail.com; ⁵ roni.herdianto@um.ac.id;

⁶ wicaksono.febriantoro@kemendag.go.id

* corresponding author

ARTICLE INFO

ABSTRACT

Article history

Received January 7, 2019

Revised January 24, 2019

Accepted February 6, 2019

Keywords

Big data

Healthcare

Agriculture

Building

Transportation

Developing Country

The development of technology from year to year is increasingly rapid and diverse. All systems that exist in human life began to be designed with technology that requires large data storage. Big Data technology began to be developed to accommodate very large data volumes, rapid data changes, and very varied. Developing countries are starting to use Big Data a lot in developing their systems, such as healthcare, agriculture, building, transportation, and various other fields. In this paper, it explains the development of Big Data applied to the sectors previously mentioned in developing countries and also the challenges faced by developing countries in the process of developing their systems.

This is an open access article under the [CC-BY-SA](#) license.



1. Introduction

The need for data storage is increasing along with the progress of systems that begin to use data storage technology as their primary storage. Up to the present, some systems store data on their hard drives at the most, but the more data stored, the more storage and big data technology as a solution are required. Big Data and its analysis system lie in modern data centers. Data stored on Big Data is obtained from online transactions, E-Mail, picture media, audio video, log data, posts, search requests, health records, social networking interactions, science data, sensors and cellphones and their application [1], [2]. All data obtained is stored in databases that grow massively and begin to be difficult to capture, form, store, manage, share, analyze, and visualize through unique database software [3].

Utilization of Big data has begun to penetrate in many human's life aspects, for instance, Big Data in the context of Health Services. In the context of health, there are many medical imaging techniques to know certain structures or know what is in the human body. For example, visualizing the structure of blood vessels can be performed using Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Ultrasound and Photoacoustic Imaging [4]. The scanning process requires a large storage space. For example, microscopic scanning of high-resolution human brains requires 66TB of storage space [5]. Ordinary storage systems will not be able to accommodate that much data storage. Big Data technology, accordingly, is very necessary to meet the demands of storage in the health sector.

Besides the health sector, the utilization of Big Data technology has also begun to be utilized in the field of Agriculture. Smart Farming development began to be intensively implemented by utilizing Big Data technology as storage. Big data on agriculture is often used for further analysis, such as research on soil type, temperature, biodiversity, plants and so on. This analysis is used to determine the right techniques to produce better agricultural products [6].

Development of Big Data requires developing countries to be prepared in all their architectures, both communication networks, compatible devices, and construction costs. In various cases, there are several developing countries that have failed in technology development, especially in the information field. For example, the development of a health information system in South Africa has failed due to a system that spent a high cost of data or less consumer use which has caused a loss [7].

System development in developing countries often has problems due to incompatibility between the previous system and the future system to be developed. In addition, system development in developing countries is often constrained due to gaps in cultural, economic and systemic contexts with software designers [7], [8]. However, development must have progressed in the process. In this paper, the evolution of Big Data technology will be presented in various fields, such as healthcare, agriculture, building, and transportation.

2. Big Data Technology

Big data is a term used to refer to a set of volumes of data that are difficult to store, process, and analyze and are efficient if only using simple database technology [9], [10]. Big Data technology is created since the need for data storage volumes is getting considerable and more complex with various types of media stored. Big data allows people to store various types of media ranging from online transactions, E-Mail, picture media, audio video, log data, posts, search requests, health records, social networking interactions, science data, sensors and cellphones and their application [1], [2]. The kinetic of Big Data is explained in four dimensions, namely:

- Volume (V1): Size of data collected for analysis.
- Velocity (V2): refers to the speed of data transfer. The contents of the data continue to change due to absorption complementary data collections, the introduction of previously archived data or legacy collections, and streaming data coming from various sources.
- Variety (V3): Diverse data sources that have different formats and from various disciplines and from several application domains.
- Veracity (V4): Quality, reliability, and potential data [6], [9].

Yet, there are some opinions that convey five dimensions. The fifth dimension is Valorization (V5): The ability to spread knowledge, appreciation, and innovation [6]. Although these five dimensions can describe Big Data, Big Data analysis does not need to fulfill all dimensions. Large data is generally known to be less accurate and stable due to compromising the 4th dimension. Other relevant dimensions can be visualization hence an informative data structure presentation is needed thus it is easy to understand [6], [9], [11], [12].

3. Evolution of Big Data

3.1. Big Data in Healthcare

Healthcare is a sector that plays an important role in a country. Evolution of Big Data in the healthcare sector has been carried out in developing countries, for example, the system of managing records in hospitals from manual to digital [2]. Big Data facilitates the identification, collection, and storage of data related to the healthcare sector [13]. In the healthcare sector, Big Data is summarized into three categories, namely traditional medical data originating from the health system (e.g personal and family health history, medical history, laboratory reports, pathology results), omics data referring to large-scale datasets in the biological and molecular fields (for example genomics, microbiomics, proteomics, and metabolomics) and data from social media [13].

3.2. Big Data in Agriculture

Agriculture is a sector that plays an important role in the economy of a country. A good agricultural sector will improve other aspects which follow it, such as employment, the country's food availability, and the supply of raw materials in the food industry. In addition, the agricultural sector is able to contribute more to the Gross Domestic Product.

Evolution in agriculture has happened since decades ago. Evolution in agriculture takes place in various aspects including aspects of pest management, planting techniques to produce quality and

quantity [14]. In the past 150 years, agricultural innovation has become an important means by which food and agricultural systems have increased productivity and increased world food availability [15], [16].

In the Big Data evolution of the agriculture sector, there are several datasets made so that Big Data Agriculture can be precise and can be used as a system algorithm parameter [13]. The types of data used include:

- **Historical data:** Includes, soil testing, crop patterns, field monitoring, monitoring of results, climate conditions, weather conditions, GIS data, and labor data.
- **Data on Agricultural Equipment and Sensors:** Includes data collected from remote sensing devices, GPS receivers based references, variable level fertilizers, soil moisture, temperature sensors, farmers call records and equipment logs.
- **Social and Web-Based Data:** These include, farmers and customer feedback, agricultural websites and blogs, social media groups, web pages, and data from search engines
- **Publications:** Includes agricultural research on cultural reference materials such as text-based practice guidelines for land and agricultural needs (e.g. pesticides, fertilizers, and equipment information).
- **Flowed Data:** This includes data from plant monitoring, mapping, drones, airplanes, wireless sensors, smartphones, and security surveillance.
- **Business, Industrial and External Data:** Data from billing and scheduling systems, agricultural departments and other agricultural equipment manufacturing companies.

Evolution in agriculture has now begun to involve technology both in monitoring, communication, and data storage technology. In 2015, processing techniques and program models for distributed computing were developed, namely, Map Reduce. This technique is used in Smart Agriculture for system decision making with several parameters as considerations, namely weather, soil conditions, and market conditions [13], [17].

Then in 2018, the Map-Reduce model was used as the basic algorithm in designing Big Data in the Smart Agriculture system developed. Big Data developed can be seen in Fig. 1. Smart Agriculture also allows for monitoring in the form of images and graphics since humans capture information in the form of images and graphics faster than using plain text [18].

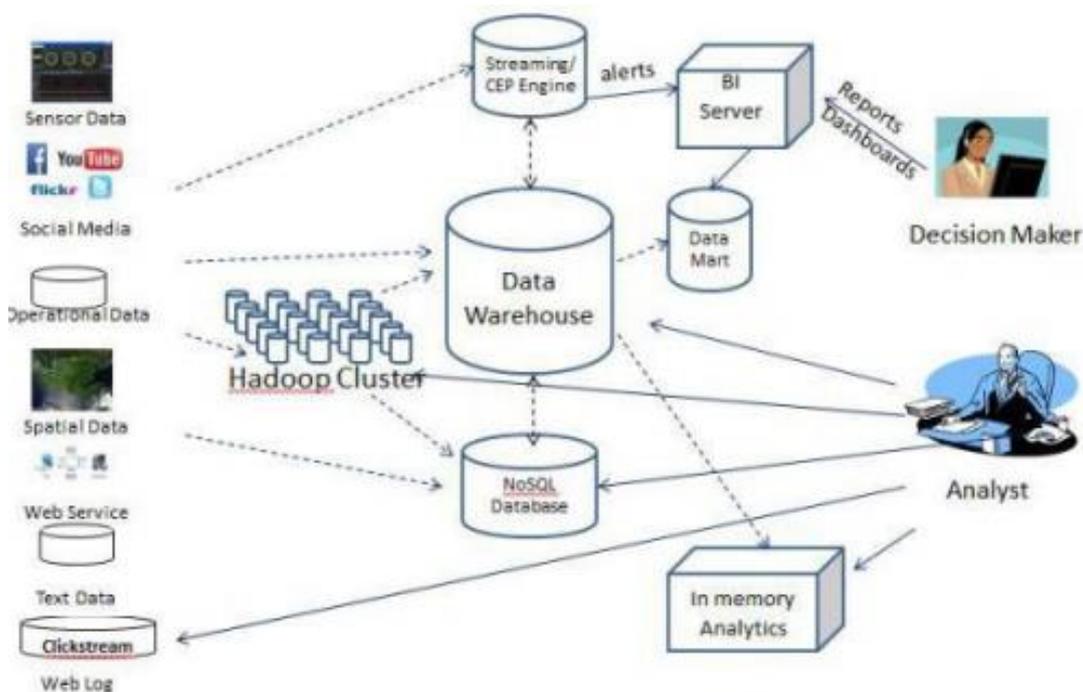


Fig. 1. Big data system of smart agriculture [17].

3.3. Big Data in Building Energy

Building energy efficiency has become one of the main concerns of the community in terms of sustainability and has attracted research and development efforts in recent years. Big Data Analysis can be one method used to analyze and understand individual energy consumption behaviors, help improve energy efficiency in the building sector and promote energy conservation [19]. Household energy consumption can be described in three dimensions, namely time dimensions, user dimensions and spatial dimensions as in Fig. 2.

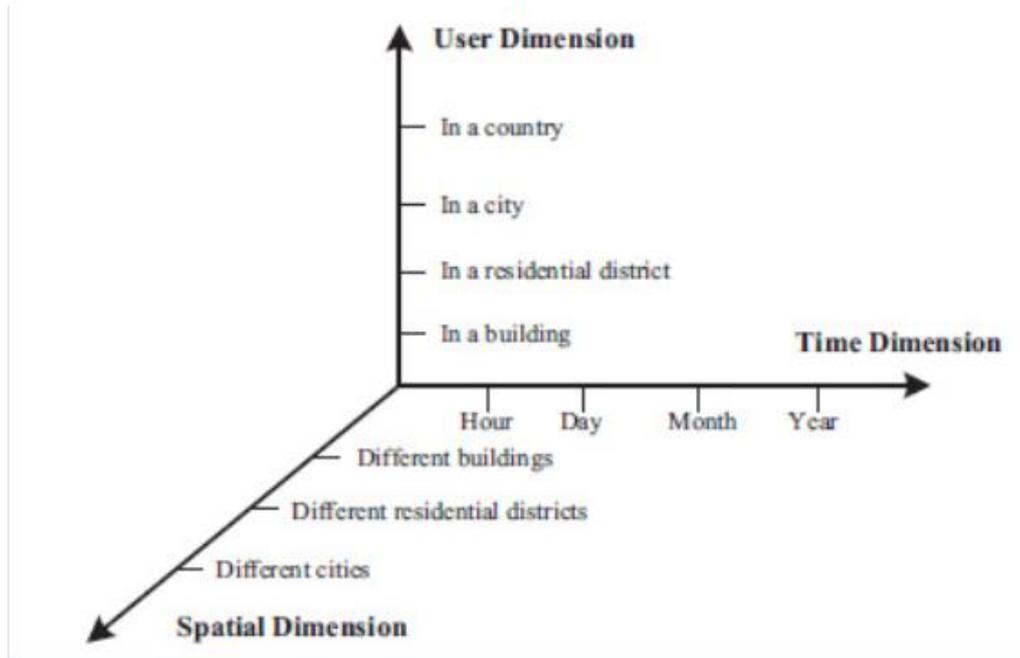


Fig. 2. Energy consumption dimensions [19].

The parameters for measuring household energy consumption can be in one hour, a day, a month or even a year. Household energy consumption in a day often shows several different differences in time of day. Monthly and annual energy consumption is usually influenced by many external factors [20]–[22].

Different household energy consumption also varies greatly. Individual energy use is generally influenced by various factors, including internal factors such as the use of basic needs and external factors such as building characteristics and building location [19]. Household energy use is often influenced by the geographical environment, level of economic development, climate characteristics and other factors.

The amount of data in the energy sector is growing at any time. Another big challenge for data analysis is exemplified by applications with limits on size. Occasionally, the limits are relatively arbitrary; About 256 columns, 65,536 rows are bound to worksheet sizes in all versions of Microsoft Excel, yet when Microsoft Excel was updated since 2007, 16,384 columns and one million rows can be collected [19].

3.4. Big Data in Transportation

Urban traffic has become a concern for many people and gathers increasing interest as cities become bigger, crowded, and “smart” [23]. Many people use Big Data analysis in various fields and have achieved great success [24]. With the successful Big Data analysis application in various fields, Intelligent Transportation Systems (ITS) also began to see Big Data with great interest [25].

The evolution of Intelligent Transportation Systems (ITS) was developed since the early 1970s, initially using traditional inefficient data processing systems. Intelligent Transportation Systems is the future direction of the transportation system. ITS combines advanced technology that includes electronic sensor technology, data transmission technology, and intelligent control technology into

the transportation system [26]. The aim of ITS is to provide better services for drivers and motorists in the transportation system [26]–[28].

Intelligent Transportation Systems (ITS) data can be obtained from various sources, such as smart cards, GPS, sensors, video detectors, social media, and so on [29], [30]. With the development of ITS, the amount of data generated at ITS expanded from the Trillion bytes to Petabyte level.

With ITS monitoring devices deployed along selected main roads in the downtown area, a large amount of traffic data can be a useful resource to help traffic operations, transportation design, planning, management, performance measurement, and research by identifying the main dynamic properties of the road which varies [23]. Big Data Analysis offers ITS a new technical method. ITS can obtain benefit from the Big Data analysis as follows [31], [32]:

- Big Data Analysis has solved three problems: data storage, data analysis, and data management. Big Data platforms like Apache Hadoop and Spark are able to process large amounts of data, and they have been widely used in academic setting and industry.
- Big Data Analysis can improve the efficiency of ITS operations and the traffic management department can predict traffic flow in real time. Big Data Analysis from transport developers can help users to reach their destination on the most suitable route and with the shortest possible time.
- Big Data Analysis can increase the level of safety of ITS. Using advanced sensors and detection techniques, the amount of transportation information in real time can be obtained. Through Big Data analysis, we can effectively predict traffic accidents.

The architecture of the Big Data analysis of Intelligent Transportation Systems (ITS) is shown in Fig. 3. This can be divided into three layers, namely the data collection layer, data analysis layer, and data application layer [26].

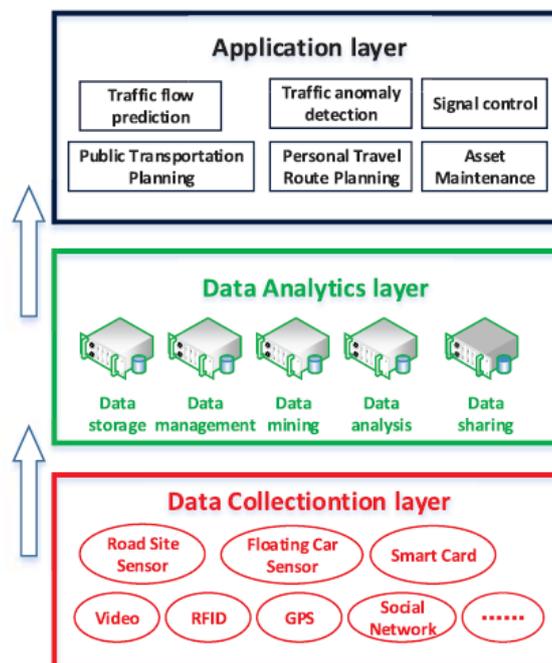


Fig. 3. Analysis architecture of ITS big data [26].

Using advanced data collection techniques, layer data collection monitors people, vehicles, roads, and the environment. Original traffic data which includes structured data, semi-structured data and mixtures are transmitted to layer analysis data via wired or wireless communication. After the layer analysis data receives original traffic data, first classifies the data, deletes duplicate data, cleanses the data and distributes useful and accurate data distributed [26].

4. Big Data Development Challenges in Developing Countries

The development of Big Data implementation in developing countries has faced numerous challenges. To develop a big data, it requires a strong physical infrastructure for its operations [4]. On the operation of big data, it requires a server architecture consisting of thousands of nodes with multiple processors and disks connected by high-speed networks working in a distributed manner [27]. Internet companies such as Google, Microsoft, Yahoo, and Amazon use this architecture with centers scattered throughout the world offering their services yet costing a lot [28]. Many developing countries cannot afford architectures that support big data [29]. In addition, apart from the server architecture, it also requires additional components that are needed by software and a reliable workforce [26]. Many developing countries lack the storage and communication infrastructure needed to regulate and integrate the amount of information generated in Big Data. Not only countries that lack resources, but they do not have computing capacity, electricity networks, and telecommunications networks [30]–[32]. After identifying the challenges of big data in developing countries, we discussed the challenges of big data in the sectors in Healthcare, Agriculture, Building, and Transportation.

4.1. Big Data Development Challenges in Healthcare

The Big Data Development Challenge in Healthcare is divided into two main categories, namely fiscal and technology [7]. In fiscal challenges, health practitioners interact without face-to-face but have risks about payment. The biggest technological challenge is the state of health data [7].

4.2. Big Data Development Challenges in Agriculture

Basically, agriculture requires a complex system with several types of data variables taken. An example is data regarding the weather. In smart agriculture, there is often a weather forecasting system. Numerical Weather Prediction or (NWP) has several problems, such as requiring large volumes, complex calculations, and real-time operations. This will also have an impact on large energy consumption as well [33], [34]. In addition, modeling in weather forecasting is limited and insufficient therefore this is a challenge in the development of agriculture [35].

4.3. Big Data Development Challenges in Building Energy

The amount of data in the energy sector is a challenge in the development of Big Data in building energy. Where data in the energy sector grows every time. Another big challenge for data analysis is exemplified by applications with limits on size. The limit is relatively arbitrary; About 256 columns, 65,536 lines are bound to worksheet sizes in all versions of Microsoft Excel. According to Adam Jacobs, Excel is not targeted at users who deploy very large data sets [19].

4.4. Big Data Development Challenges in Transportation

Big Data analysis has indeed made great achievements on Intelligent Transportation Systems (ITS), but there are still open challenges that need to be addressed in future work. Some open challenges to the use of Big Data analysis in ITS are, data collection, data privacy, data storage, data processing, and data opening [36]. Big Data Analysis will have a profound impact on intelligent system transportation design, and make it safer, more efficient and profitable [37].

5. Conclusion

The rapid development of technology with the amount of data which needs to be increasingly stored encourages the need for a system that is able to accommodate the entire data that must be stored. Big Data technology is one of the solutions for storing data on a large scale with increasingly complex computing. Some sectors have started using Big Data for storage and computing media for example Healthcare. The development of Big Data in Healthcare offers an easy approach to administer and store health data from medical devices or medical methods. In addition, in the field of Agriculture, Building Energy and Transportation also utilize Big Data to store or compute data for control based on that data.

In the development of Big Data, there are several challenges that are generally caused by financial and capital conditions. In the future, it is expected that the Big Data system will be more efficient and economical which will be fulfilled by several developments on Low-Cost Computing.

References

- [1] P. Zikopoulos and C. Eaton, *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. McGraw-Hill Osborne Media, 2011.
 - [2] R. D. Schneider, *Hadoop for Dummies Special Edition*. Canada: John Wiley & Sons, 2012.
 - [3] S. Sagioglu and D. Sinanc, "Big Data: A Review," in *2013 International Conference on Collaboration Technologies and Systems (CTS)*, 2013, pp. 42–47.
 - [4] R. C. Gessner, C. B. Frederick, F. S. Foster, and P. A. Dayton, "Acoustic angiography: a new imaging modality for assessing microvasculature architecture," *J. Biomed. Imaging*, vol. 2013, no. 14, 2013.
 - [5] I. Scholl, T. Aach, T. M. Deserno, and T. Kuhlen, "Challenges of medical image processing," *Comput. Sci. Dev.*, vol. 26, no. 1–2, pp. 5–13, 2011.
 - [6] A. Kamilaris, A. Kartakoullis, and F. X. Prenafeta-Boldú, "A review on the practice of big data analysis in agriculture," *Comput. Electron. Agric.*, vol. 143, pp. 23–37, 2017.
 - [7] D. Dada, "The failure of e-government in developing countries: a literature review," *Electron. J. Inf. Syst. Dev. Ctries.*, vol. 26, no. 1, pp. 1–10, 2006.
 - [8] R. Heeks, "Information systems and developing countries: Failure, success, and local improvisations," *Inf. Soc.*, vol. 18, no. 2, pp. 101–112, 2002.
 - [9] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, "The rise of 'big data' on cloud computing: Review and open research issues," *Inf. Syst.*, vol. 47, pp. 98–115, 2015.
 - [10] J. Manyika *et al.*, "Big data: The next frontier for innovation, competition, and productivity," 2011.
 - [11] D. Rodriguez, P. de Voil, M. C. Rufino, M. Odendo, and M. T. van Wijk, "To mulch or to munch? Big modelling of big data," *Agric. Syst.*, vol. 152, pp. 32–42, 2017.
 - [12] A. Karmas, A. Tzotsos, and K. Karantzalos, "Geospatial big data for environmental and agricultural applications," in *Big Data Concepts, Theories, and Applications*, S. Yu and S. Guo, Eds. Springer, Cham, 2016, pp. 353–390.
 - [13] M. R. Bendre, R. C. Thool, and V. R. Thool, "Big data in precision agriculture: Weather forecasting for future farming," in *2015 1st International Conference on Next Generation Computing Technologies (NGCT)*, 2015, pp. 744–750.
 - [14] S. Sonka, "Big data: fueling the next evolution of agricultural innovation," *J. Innov. Manag.*, vol. 4, no. 1, pp. 114–136, 2016.
 - [15] N. E. Borlaug, "Ending world hunger. The promise of biotechnology and the threat of antiscience zealotry," *Plant Physiol.*, vol. 124, no. 2, pp. 487–490, 2000.
 - [16] S. Chakraborty and A. C. Newton, "Climate change, plant diseases and food security: an overview," *Plant Pathol.*, vol. 60, no. 1, pp. 2–14, 2011.
 - [17] M. Kumar and M. Nagar, "Big data analytics in agriculture and distribution channel," in *2017 International Conference on Computing Methodologies and Communication (ICCMC)*, 2017, pp. 384–387.
 - [18] O. Kumar and A. Goyal, "Visualization: a novel approach for big data analytics," in *2016 Second International Conference on Computational Intelligence & Communication Technology (CICT)*, 2016, pp. 121–124.
 - [19] N. Koseleva and G. Ropaite, "Big data in building energy efficiency: understanding of big data and main challenges," *Procedia Eng.*, vol. 172, pp. 544–549, 2017.
 - [20] K. Zhou, C. Fu, and S. Yang, "Big data driven smart energy management: From big data to big insights," *Renew. Sustain. Energy Rev.*, vol. 56, pp. 215–225, 2016.
 - [21] L. Mashayekhy, M. M. Nejad, D. Grosu, Q. Zhang, and W. Shi, "Energy-aware scheduling of mapreduce jobs for big data applications," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 10, pp. 2720–2733, 2014.
 - [22] J. Cooper, M. Noon, C. Jones, E. Kahn, and P. Arbuckle, "Big Data in Life Cycle Assessment," *J. Ind. Ecol.*, vol. 17, no. 6, pp. 796–799, 2013.
-

-
- [23] A. Artikis, M. Weidlich, A. Gal, V. Kalogeraki, and D. Gunopulos, "Self-adaptive event recognition for intelligent transport management," in *2013 IEEE International Conference on Big Data*, 2013, pp. 319–325.
- [24] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mob. networks Appl.*, vol. 19, no. 2, pp. 171–209, 2014.
- [25] C. R. Berger and E. Smith, "Intelligent transportation systems provide operational benefits for New York metropolitan area roadways: a systems engineering approach," in *2007 IEEE Long Island Systems, Applications and Technology Conference*, 2007, pp. 1–8.
- [26] L. Qi, "Research on intelligent transportation system technologies and applications," in *2008 Workshop on Power Electronics and Intelligent Transportation System*, 2008, pp. 529–531.
- [27] S. H. An, B. H. Lee, and D. R. Shin, "A survey of intelligent transportation systems," in *2011 Third International Conference on Computational Intelligence, Communication Systems and Networks*, 2011, pp. 323–337.
- [28] N. E. El Faouzi, H. Leung, and A. Kurian, "Data fusion in intelligent transportation systems: Progress and challenges—A survey," *Inf. Fusion*, vol. 12, no. 1, pp. 4–10, 2011.
- [29] J. Zhang, F. Y. Wang, K. Wang, W. H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1624–1639, 2011.
- [30] Q. Shi and M. Abdel-Aty, "Big data applications in real-time traffic operation and safety monitoring and improvement on urban expressways," *Transp. Res. Part C Emerg. Technol.*, vol. 58, pp. 380–394, 2015.
- [31] N. Mohamed and J. Al-Jaroodi, "Real-time big data analytics: Applications and challenges," in *2014 international conference on high performance computing & simulation (HPCS)*, 2014, pp. 305–310.
- [32] X. Lin, P. Wang, and B. Wu, "Log analysis in cloud computing environment with Hadoop and Spark," in *2013 5th IEEE International Conference on Broadband Network & Multimedia Technology*, 2013, pp. 273–276.
- [33] M. Zaharia *et al.*, "Fast and interactive analytics over Hadoop data with Spark," *Usenix Login*, vol. 37, no. 4, pp. 45–51, 2012.
- [34] H. Jain and R. Jain, "Big data in weather forecasting: Applications and challenges," in *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, 2017, pp. 138–142.
- [35] S. E. Haupt and B. Kosovic, "Big data and machine learning for applied weather forecasts: Forecasting solar power for utility operations," in *2015 IEEE Symposium Series on Computational Intelligence*, 2015, pp. 496–501.
- [36] J. Mayes, "From Observations to Forecasts – concluding article (Part 15): Opportunities and challenges for today's operational weather forecasters," *Weather*, vol. 67, no. 4, pp. 100–107, 2012.
- [37] M. Smith, C. Szongott, B. Henne, and G. Von Voigt, "Big data privacy issues in public social media," in *2012 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*, 2012, pp. 1–6.