Studi On Big Data Analytics Framework in Smart City Context

Rahmawati Dinda*, Arief Assaf and Do Abdullah Saiful Saiful

Smart City Islands Research Group Laboratory of Faculty of Engineering, Universitas Khairun, Ternate, Indonesia

Abstract. The issue of global urbanization, which is a separate problem faced by the government, is the very rapid growth of population density in cities. To face this challenge, the government launched a smart city project by targeting sustainable economic growth and improving the quality of life. Information and Communication Technology governance is the key to realizing a smart city. However, each of these I.C.T. tools produce large amounts of data known as Big Data. Data processing with the Big Data approach is becoming a trend in information systems to provide better public services and provide references in the policy-making process. However, to obtain important information in the scope of big data, a Big Data Analytics process is needed, also known as Big Data Value Chain. Extracting knowledge from the related literature can identify the characteristics of the big data analytic framework for smart cities. This paper reviews several big data analytic frameworks applied to smart cities. This paper is to find the advantages and disadvantages of each framework so that it can be a direction for future research

Keywords: *Big Data Analytic, Framework, Smart City, Big Data Value Chain*

1 Introduction

The issue of urbanization and global population growth that is increasingly uncontrollable is causing problems in urban areas. The government launched a smart city project by targeting sustainable economic growth with a better quality of life. In Indonesia, the Smart City concept was initiated by an expert from ITB, Suhono S. Supangkat. A smart city is a city's government that is the fastest and most accurate in providing solutions to its citizens. Smart cities have been initiated by European countries that the smart city concept consists of supporting components, namely: smart economy, smart people, smart governance, smart government, smart mobility, smart environment, and smart living [1]. The smart city is not only a process of making each of these domains "smart", but also how the integration process between domains forms a unified system. This integration process requires the sharing of information between domains [2]-[3].

The use of digital technology from various aspects and interactions between humans, between machines and humans with machines produce a very large volume of data which is called Big Data [4,5]. A term where there is complex and rapidly growing data. So to get the right information, a precise and accurate analysis process is needed. This process is called Big Data Analytic BDA or Big Data Value Chain (B.D.V.) [6]-[7].

Reviewing related literature on BDA can provide knowledge about how to identify the characteristics of BDA and B.D.V. framework for smart city

implementation [7]. This paper reviews the big data analytic framework applied to smart cities. The purpose of this review is to find the advantages and disadvantages of each of these frameworks.

2 Big Data Analytics Concept

The term Big Data is a natural product of digital artefacts and their applications. Cell phones, sensors and social media. Increasing human-to-human, human-to-machine and machine-to-machine interactions to a more advanced level. Issues related to Big Data are volume, velocity, variety, and veracity [8].

Big Data Analytics BDA is the entire process and tool needed to find the right information to support the decision-making process. The process consists of data extraction, transformation to knowledge, and analysis. Involves BDA specific tools, techniques and methods. Because this process is not like ordinary traditional information extraction, this creates opportunities for researchers and technology providers to develop reliable models, platforms, frameworks and algorithms for BDA [6].

Big Data management platform scalability is divided into two approaches, namely: vertical approach and horizontal approach, namely Vertical scaling, which accommodates increased data volume by increasing computer capabilities (memory, CPU, etc.). Horizontal scaling is dividing the workload evenly in parallel to several computers at once.

* Corresponding author : dinda26kc@gmail.com

There are two approaches, both from perspective advantages and disadvantages. Vertical scaling has a drawback in terms of the maximum limit of a computer being upgraded. While horizontal scaling requires complicated handling when using many computers with different operating systems. High-Performance Computing (HPC) and Apache Hadoop are examples of vertical and horizontal scaling platforms[8]. Cost issues, limitations in computer upgrades and dynamics of multi-domain datasets make horizontal scaling more widely used as a reference in designing big data analytic platforms such as Apache Hadoop.

3 Smart City Concept

The term Smart city is the concept of a system with the use of technology by using sophisticated data processing with the aim to make the governance of the city be more efficient, people are satisfied, and promising business environment more secure and convenient [9]. Three factors influence the formation of a smart city: technological aspects (hardware and software infrastructure), human aspects (creativity, diversity and education) and institutional aspects (government and regulations)[2].

Perspectives of I.C.T., industry and academia create a layered approach to model smart city. Yin Chuan Tai, et al. make 4 layers models: the acquisition and transmission of data, data virtualization, data services, and applications.

Fig.1. Find knowledge in the database

There are many applications for implementing BDA in various smart city domains, such as for planning [10], traffic control and transportation [11]. Analysis of crime[12], energy sector[13] and environment[14]. In order to design software that is in accordance with the objectives of a smart city, it is necessary to identify aspects of functional requirements and non-function requirements related to data sources and their management. As an example of the diversity of data sources (I.O.T. Data, Social Media Data, data and medical records) requires integration, scalability of the system, the security and privacy aspects [15].

Related research by [16], the aspect of non-functional requirements such as sustainability, availability, privacy, social aspects and flexibility/extensibility (scalability). Meanwhile, functional requirements aspects such as object interoperability, real-time monitoring, historical data, mobility, service composition and integrated urban management. Similar research by [17] mentions eight functional requirements, namely: data management, application run-time, Wireless Sensor Network W.S.N. management, data processing, external data access, service management, software engineering tools and definition of smart city model. It is coupled with eight

non-functional requirements, namely: interoperability, scalability, security, privacy, context awareness, adaptation, extensibility and configurability.

Can we conclude that there are seven aspects of functional requirements: interoperability, real-time analysis, historical analysis of data, mobility, iterative processing, data integration, and aggregation models. It has Nine aspects of non-functional requirements, such as scalability, security, privacy, context awareness, adaptation, extensibility, sustainability, availability, and configurability.

4 Results and Discussion of BDA Framework

Based on the results of reviewing and evaluating several related articles, there are 3 popular architectural frameworks related to BDA to be applied in the smart city context, namely the framework: BASIS[18], SWIFT[19] and RADICAL[20].

4.1 BASIS

BASIS is a 3 layer big data architecture for smart city. This layer consists of: conceptual, technology and infrastructure. The conceptual layer packs BASIS internal and external functionality which includes data capturing, integration, storage and APIs for web streams and data analysis. The technology layer is realized with the Hadoop platform. The infrastructure layer manages the hardware interconnection with the outside world. This infrastructure layer is the interface between the external technology layers. BASIS has design principles including:

- Open data services[21], facilitating the realization of new services developed by the government, organizations or communities.
- It is providing storage and processing of data that is distributed or centralized.
- Focus on security and privacyPartitioning data to manage data cycle
- Focus on services and the freedom of users
- Use of open-source technology, making it more open.

BASIS has been validated in case studies to find flight delay profiles in several American cities. The profiles were extracted using the K-Means algorithm which was run on 13 computer clusters using a dataset containing 3.5 million flight records in America.

4.2 SWIFT

SWIFT, which stands for Smart Wireless Sensor Network (W.S.N.)-based Infrastructural Framework for smart Transactions, is a three-layer architectural framework that supports the integration of various devices. The most basic layer is the Smart Wireless Sensor Network (S-WSN). The second layer is Smart Wireless-based Pervasive Edifice (SWIPE). Then the third layer is Smart Decision & Control Enabler (SDCE).

The framework has S-WSN layer functions as a SWIFT detector that looks for information in hundreds of wireless sensors that are spread out. Several sensors are grouped in a cluster and managed by a smart cluster head (SCH) which is distributed in various places. SCH collect and combine data from nearby sensors. In addition, the sensor also sends information about the identity of the sensor and its battery status. SCH has the ability to sound an alarm or activate an emergency signal etc. Data from SCH is sent to Smart Fusion Nodes (SFN), which are in the SWIPE layer. SWIPE is at the core of SWIFT. SFN has the function of classifying and merging data to get interpretation results from sensor data. Meanwhile, SDCE provides cloud-based services to all domains in the smart city.

4.3 RADICAL

A platform called Rapid Deployment on Intelligent Cities And Living (RADICAL) is a Service Oriented Architecture (S.O.A.) based platform that collects and analyzes IoT (Internet of Things) and Social Network (S.N.) data into value-added services for smart cities[22]. Data related to IoT is stored in the form of observations and measurements. As an example of measuring CO2 levels. Meanwhile, the S.N. data is accessed in real-time using the network API.

On top of the main platform, RADICAL provides a set of application management tools that facilitate users to maximize the use of the platform.

IoT data is collected in a RADICAL repository (MySQL Database) using an API (Application Programming Interface).

5 Result & Discussion

Table 1. Framework comparison table

Framework	Strengths	Weaknesses
BASIS	Shape 3 layers (3 Layers)	Between the technology and infrastructure layers, the difference is less clear. For example, in the case of the infrastructure layer that manages the design of hardware capable of interacting with the external environment, this can be the same as the additional functionality of the existing technology layer.
	The shape of an independent inter-domain	BASIS is comprehensive and has functions between independent domains, but there is no special design for smart cities

	Having the ability to batch data analysis and data streams	The linkage case of validation (flight delay profile) with a smart city is still questionable. We studied it from the point of view of a smart city (transportation, energy, traffic, etc.).
	The ability of service-oriented analysis	The difference between batch data analysis for offline applications is less obvious when compared to data stream analysis.
	Data collection and analysis is carried out using Hadoop, better known as the horizontally scalable platform	
	SWIFT	Relevant to the scope of the smart city (surveillance and traffic control)
	Able to accommodate a variety of sensors	Only use data from sensors. Data from other sources do not.
	The ability to create location-aware data	It cannot be applied to batch data analysis for offline applications.
	Run with the Online application or near online	SWIFT performance untested
	Although SWIFT is intended for the scope of traffic monitoring and control, it can also be applied to other areas of online analytics	
	RADICAL	A framework that combination of data collection and analysis from multiple sources (IoT and S.N.). This is the key link between domains in a smart city.
	Using S.O.A. to build smart city services	Dependence on the MySQL database as the main repository is a limiting factor
		The results of the analysis have not been able to be used for reference in further analysis

References

- Giffinger; Fertner; Kramar; Klasek; Pichler; Majers *Smart cities Ranking of European medium-sized cities*; (2007); Vol. **16**;
- Nam, T.; Pardo, T.A. Conceptualizing smart city with dimensions of technology, people, and institutions. In Proceedings of the Proceedings of the 12th Annual International Digital Government Research Conference on Digital Government Innovation in Challenging Times - dg.o '11; A.C.M.: College Park, MD, U.S.A., (2011); p. 282.
- Fu'Adi, D.K.; Arief, A.; Sensuse, D.I.; Syahrizal, A. Conceptualizing smart government implementation in smart city context: A systematic review. In Proceedings of the 2020 5th International Conference on Informatics and Computing, I.C.I.C. 2020; (2020).
- Mergel, I.; Rethemeyer, R.K.; Isett, K. Big Data in Public Affairs. *Public Adm. Rev.* (2016), **76**, 928–937, doi:10.1111/puar.12625.
- Wong, J.S. Driverless Government: Speculation, Citizenship and Collective Civic Intelligence. *Archit. Cult.* (2019), doi:10.1080/20507828.2019.1647960.
- Tsai, C.W.; Lai, C.F.; Chao, H.C.; Vasilakos, A. V. Big data analytics: a survey. *J. Big Data* (2015), **2**, 1–32, doi:10.1186/s40537-015-0030-3.
- Hashem, I.A.T.; Chang, V.; Anuar, N.B.; Adewole, K.; Yaqoob, I.; Gani, A.; Ahmed, E.; Chiroma, H. The role of big data in smart city. *Int. J. Inf. Manage.* (2016), **36**, 748–758, doi:10.1016/j.ijinfomgt.2016.05.002.
- Chen, M.; Mao, S.; Liu, Y. Big data: A survey. *Mob. Networks Appl.* (2014), **19**, 171–209, doi:10.1007/s11036-013-0489-0.
- Yin, C.T.; Xiong, Z.; Chen, H.; Wang, J.Y.; Cooper, D.; David, B. A literature survey on smart cities. *Sci. China Inf. Sci.* (2015), **58**, 1–18, doi:10.1007/s11432-015-5397-4.
- Souza, A.; Figueredo, M.; Cacho, N.; Araújo, D.; Prolo, C.A. Using Big Data and Real-Time Analytics to Support Smart City Initiatives. *IFAC-PapersOnLine* (2016), **49**, 257–262, doi:10.1016/j.ifacol.2016.11.121.
- Singh, D.; Vishnu, C.; Mohan, C.K. Visual Big Data Analytics for Traffic Monitoring in Smart City. *Proc. 2016 IEEE 1st Int. Conf. Data Stream Min. Process. DSMP 2016* (2017), 886–891, doi:10.1109/icmla.2016.0159.
- Ghosh, D.; Ae Chun, S.; Shafiq, B.; Adam, N.R. Big Data-based Smart City Platform. (2016), 58–66, doi:10.1145/2912160.2912205.
- Horban, V. A multifaceted approach to smart energy city concept through using big data analytics. *Proc. 2016 IEEE 1st Int. Conf. Data Stream Min. Process. DSMP*, (2016), 392–396, doi:10.1109/DSMP.2016.7583585.
- Iancu, V.; Stegaru, S.C.; Tudose, D.S. A Smart City Fighting Pollution, by Efficiently Managing and Processing Big Data from Sensor Networks. (2016), 489–513, doi:10.1007/978-3-319-44881-7_23.
- Al Nuaimi, E.; Al Neyadi, H.; Mohamed, N.; Al-Jaroodi, J. Applications of big data to smart cities. *J. Internet Serv. Appl.* (2015), **6**, 1–15, doi:10.1186/s13174-015-0041-5.
- Da Silva, W.M.; Tomas, G.H.R.P.; Dias, K.L.; Alvaro, A.; Afonso, R.A.; Garcia, V.C. Smart cities software architectures: A survey. *Proc. A.C.M. Symp. Appl. Comput.* (2013), 1722–1727, doi:10.1145/2480362.2480688.
- Zambom Santana, E.F.; Chaves, A.P.; Gerosa, M.A.; Kon, F.; Milojicic, D.S. Software Platforms for Smart Cities: Concepts, Requirements, Challenges, and a Unified Reference Architecture. *A.C.M. Comput. Surv.* (2018), **50**, [1-37], doi:10.1145/3124391.
- Costa, C.; Santos, M.Y. BASIS: A big data architecture for smart cities. *Proc. 2016 S.A.I. Comput. Conf. S.A.I. 2016* (2016), 1247–1256, doi:10.1109/SAI.2016.7556139.
- Nandury, S. V Strategies to Handle Big Data for Traffic Management in Smart Cities. (2016), 356–364.
- Psomakelis, E.; Aisopos, F.; Litke, A.; Tserpes, K.; Kardara, M.; Martínez Campo, P. Big IoT and social networking data for smart cities: Algorithmic improvements on big data analysis in the context of RADICAL city applications. *CLOSER 2016 - Proc. 6th Int. Conf. Cloud Comput. Serv. Sci.* (2016), **1**, 396–405, doi:10.5220/0005934503960405.
- Rahmatika, M.; Krismawati, D.; Rahmawati, S.D.S.D.; Arief, A.; Sensuse, D.I.D.I.; Dzulfikar, M.F.M.F. An open government data maturity model: A case study in BPS-statistics Indonesia. *2019 7th Int. Conf. Inf. Commun. Technol. ICoICT*, (2019), 1–7, doi:10.1109/ICoICT.2019.8835352.
- Psomakelis, E.; Aisopos, F.; Litke, A.; Tserpes, K.; Kardara, M.; Campo, P.M. Big IoT and Social Networking Data for Smart Cities. *Proc. 6th Int. Conf. Cloud Comput. Serv. Sci. - Vol. 12*, (2016), **1**, 396–405.