

# A Big Data Management Architecture for Smart Cities based on Fog-to-Cloud Data Management Architecture

Amir Sinaeepourfard <sup>[0000-0002-6922-7858]</sup>, John Krogstie <sup>[0000-0003-4830-1876]</sup>, and Sobah Abbas Petersen <sup>[0000-0001-6055-9285]</sup>

Department of Computer Science,  
Norwegian University of Science and Technology (NTNU), Trondheim, Norway  
{a.sinaee, john.krogstie, sobah.a.petersen}@ntnu.no

**Abstract.** A massive volume of data has been created and saved in their related data repositories in either structured or unstructured format (Big Data). Therefore, there are several ongoing studies (including in different environments, scenarios, and sciences) nowadays to propose the Big Data management architectures. Data management (including data acquisition, data preservation, or data processing) becomes a complex task in these studies during their entire life cycles. In a particular example of the Big Data environment, the smart city is an advanced technological option to increase citizens' quality of life through different smart services. It is evident that data is the most critical element in a smart city. Without relevant data, no smart services can be launched. Recently Fog-to-Cloud (F2C) data management combines the advantages of both the centralized (Cloud) and distributed data management in a smart city. The advantage of this hierarchical distributed data management is diverse, including reducing communication latencies for real-time or critical services, decreasing network data traffic, applying different policies (for example, for data filtering, data aggregation, data securities, etc.) and so on. In this paper, we developed our proposed hierarchical distributed data management architecture for Zero Emission Neighborhoods (ZEN) center in Norway. In the beginning, the hierarchical distributed architecture, that has the potential to organize all data life cycle stages, has been described (from creation to consumption). Afterward, it has been illustrated that the architecture can manage different obtained data types (including real-time, last-recent, and historical data) in each cross-layer (from IoT devices to cloud technologies) of the architecture.

**Keywords:** Smart Cities, IoT, Data Management, Fog-to-Cloud (F2C) Data Management, Big Data Management, Data LifeCycle (DLC), Vs Challenges.

## 1 Introduction and Motivation

A massive amount of data is being produced and stored rapidly, regularly and unlimitedly in their related distributed data repositories in the contemporary IT world. This massive data can be in structured, semi-structured or unstructured formats (Big Data). In addition, the above mentioned data have the possibility to be shared and openly accessible (Open Data) and used by the potential clients in either private or public sector

Copyright held by the author(s). NOBIDS 2018

(Open Government Data). However, it is widely accepted that sharing huge amounts of heterogeneous data imposes several challenges and difficulties for the data management systems. Several specific Data LifeCycle (DLC) models have been designed for particular scenarios and different fields [1-4], to name a few, but there is still an open discussion about how to organize and manage these complexities of data and their related challenges. This solution can contribute to efficient usage and exploitation of data during all steps of the data life cycle.

As an example of the Big Data scenario, smart cities are the available technological solutions to organize all the challenges and complexities of the increasing population and urban density. Data is one of the most valuable ingredients in the smart cities' environments. They provide a fertile ground for a city to be smart and creative. In fact, data create the required information for services to proceed according to contextual parameters, or some higher value knowledge extracted from complicated data analysis. Therefore, smart cities are an ideal scenario to create abundant data from many different sources in the city (such as sensors, smartphones, camera surveillance etc.) to combine with historical information. Consequently, there are many ongoing efforts and active studies in academia as well as in industry nowadays to manage these large data amounts.

In this research paper, we developed further our Big Data management architecture proposal for the ZEN research centers [5]. For more information, we recently proposed a ZEN data management architecture based on the F2C data management architecture in the smart cities [6]. Our main interest in this paper is to show that the ZEN data management architecture is able to cover all data life stages (from generation to usage) through cross-layers (from IoT devices to cloud technologies) with respect to the recent proposal for the Comprehensive Scenario Agnostic DLC (COSA-DLC) model in the smart cities [7]. Besides that, we depict that different data types in terms of the product time (including historical, last-recent, and real-time data) can be supported and managed through their data life cycle stages in our proposed Big Data management architecture. On top of that, we used the F2C technologies [8] to have both advantages of the centralized and distributed data management architectures for our proposed Big Data management architecture in the smart cities context.

The rest of this paper is structured as follows. Section 2 introduces background about the data management architectures (including cloud and F2C data management) in the smart city environments. In Section 3, we explain some main insights related to the Big Data contexts and concepts. Section 4 describes the essential concepts in a Big Data management architecture based on F2C data management in smart cities. Section 5 pertains to the different smart cities scenario through ZEN center in Norway. In addition, our focus is to extend our proposed ZEN data management architecture [6] with some additional Big Data management insights (including different data types, covering all data life stages and so on). Section 6 concludes this paper and introduces future research works.

## 2 Related Work

There is recently much ongoing research to build the Big Data Management architectures in different environments, scenarios and sciences [9-11], to name a few. Data management and organization considered as a complicated task during their entire life cycle (DLC) [12, 13]. The principal aim of data management is to give direction for the easy and safe access to the data sources and repositories, in order to discover any kind of new value from your Big Data sources. Because of the above mentioned reason, advanced data management and organization systems are core topic for the efficient value generation. Moreover, the available concepts of Relational Database Management Systems (RDBMS) and the recent Extract-Transform-Load (ETL) process have been suggested for modeling the typical data life cycles in data warehousing environments [4, 14] but these are not sufficient for Big Data paradigm. Therefore, Big Data concepts impose several new challenges to the existing data management and organization systems [4, 15].

In a specific example of the Big Data scenario, smart cities are composed of large amount of data (including unstructured, semi-structured, and structured). Therefore, there are two main references for the data management architecture in smart cities. On one side, the majority of the architectures designed with explicit data management schemes are centralized in one place. This highlights that even though data is obtained from distinct sources spread across the city (including sensors, surveillance cameras, third-party applications, external databases, etc.), data can be accessed from a centralized platform, normally using cloud technologies [16, 17]. On the other side, minor architectures go beyond a distributed schema for data management [6, 8, 18], using some advanced technologies such as Fog Computing [19] or F2C Computing [8].

To sum up, we can highlight that all the available data management architectures for smart cities have the following concerns and limitations:

- There are a few studies to contribute the distributed architecture especially in terms of the distributed data management architecture;
- There is no proposal for DLC model (including all data life stages from creation to conception) within ZEN data management architecture in smart city context.
- There is no previous work (based on F2C data management architecture) to handle all the data life stages through different smart city scenarios.

Therefore; because of the all above mentioned points, hierarchical distributed data management architecture [6] for ZEN center [5] to cover all concerns and limitations has been developed.

## 3 Big Data Contexts and Concepts

Coined a few years ago, the importance of the “Big Data” term has been discussed by several scientific communities. Nowadays, there are several technical issues and business impacts of Big Data but there is no global consensus on a uniform and highly accepted definition [20]. That said, we might suggest that Big Data definitions can be

assumed into three main categories depending on the main character used to formally establish the definition after a thorough reading process on the current related literature. Those main characters are “data quantity”, “challenges”, and “Big Data management complexities”. For each of the mentioned characteristics, Table 1 defines relevant references in the literature.

After revisiting all definitions about the Big Data contexts [1], it has recognized that there is a difference between using the Vs model for Big Data definition and Big Data challenges. In one side, the complete definition of Big Data can be suggested considering Variety, Volume, and Velocity, since these are the features that explain Big Data – Value might be also added considering Big Data has imagined fueled by the potential value among such massive data. On the other side, with respect to the Big Data challenges, it has not offered that Visualization, relating to a way of presenting data once processed [26], is one main challenge for Big Data technology. This is rooted in the fact that Visualization is an optional software programming aspect for end-users. Indeed, in [1, 6] offered that the 6Vs challenges model, including of Value, Volume, Variety, Velocity, Variability and Veracity.

**Table 1.** Big Data contexts and concepts [1].

<b><i>Big Data Classification and definition</i></b>	
<b><i>Data Quantity</i></b>	Those were among the first authors to define the term Big Data literally, related to the utilization of large quantities of scientific data for visualization [21].
	Big Data relates to datasets whose size is beyond the capability of typical database software tools to capture, store, manage, and analyze [22].
	Big Data is about the increasing number of challenges that organizations face. As they deal with large and fast-increasing sources of data or information which highlights a complex range of analysis [23].
<b><i>Challenges</i></b>	The “V” discussion is offered with 3Vs models. The term 3Vs (Volume, Velocity and/or Variety) were defined by Gartner to explain the main factors of Big Data challenges. [24-28].
	Then, the 3Vs model were moved to 4Vs. However, there is no consensus on the fourth V. This reference suggests veracity, showing the 4Vs as Volume, Velocity, Veracity, and Variety [29].
	Another option for the added V is Value. Then, it is considered the 4Vs as Volume, Velocity, Variety, and Value (instead of Veracity) [20].
	Next, the discussion went forward to 5Vs. Similar to 4Vs, several thoughts suggested characterizing the 5Vs. Some references mention that 5Vs model includes Volume, Velocity, Variety, Value, and Veracity [30-32].
	This works added variability (replacing veracity), so defining 5Vs as Volume, Velocity, Variety, Value and Variability [33].
	New school of thought suggested 7Vs as Volume, Variety, Velocity, Veracity, Value, Variability and Visualization [34, 35].
<b><i>Big Data Management Complexities</i></b>	Big Data shall define the datasets that are not able to be perceived, managed, and processed by traditional IT and software/hardware tools within a tolerable time [20].
	Big Data and data-intensive technologies are considered as a new technology trend in science, industry, and business [31].
	Big Data considered as a set of techniques and technologies that need new forms of integration to uncover large hidden values from large datasets that are diverse, complex, and of a massive scale [24].
	The process of handling Big Data includes with the collection, storage, transportation, and exploitation. It is obvious that the collection, storage, and transportation stages are necessary precursors for the main object of exploitation through data analytics, which is the core of the Big Data processing [29].
	Big Data shall assume that the traditional relational methods for data models face a problem with data volume, acquisition speed, or data representation. This problem is related to the effective analysis and deep processing of those data models. [36].
	Big Data is a term including distinct types of complicated and large datasets, all becoming hard to process with the traditional data processing systems [33].
Big Data is a massive amount of both structured and unstructured data that that is hard to be processed by using traditional database and software algorithms [37].	

This section is organized into three main subsections with respect to the Big Data definitions concepts as it is categorized in Table 1. First, we briefly mention the quantity of Big Data environments. Second, we discuss Big Data challenges. These challenges are mainly related to the Vs challenges generation (including 3Vs, 4Vs, 5Vs, 6Vs, etc.). Third, we describe the complexities of the Big Data management.

### 3.1 Big Data Quantity

Data is being produced in large amounts unstructured and structured formats in every moment. In addition, the produced data generated from different data sources, with distinct formats and so on. This freshly obtained data is often combined together with the archived historical data. This accumulated data builds initial ingredients for future knowledge purposes in many different fields of sciences and Big Data scenarios. As an example of the Big Data environment in the biological systems field [38], the data growth is from a rate of 1 KB per day in 1996 to a rate of 10 GB per day in 2011. Therefore, the size and number of experimental datasets are increasing exponentially in fact.

### 3.2 Big Data Challenges

The main challenges in Big Data environment have been widely discussed [1] through the 3Vs challenges (including Volume, Variety and Velocity) as proposed by Gartner [24]. This model has been suggested to the 5Vs challenges, which may be considered as 4 +1, or 4Vs parameters (including Value, Variety, Velocity and volume). The additional V challenge may consider either Variability as stated by [33, 39] or Veracity as to be read in [30-32]. Currently few authors showed that the challenges can be 7Vs [34, 35] (including both Variability and Veracity, and adding Visualization). Then, other authors added Volatility, Viscosity and Virality as additional challenges [40-54]. Indeed, after a thorough reading process on the current related literature, in [1, 6] suggested that the Big Data challenges can be defined as a 6Vs challenges (including Volume, Variety, Velocity, Variability, Veracity, and Value).

### 3.3 Big Data Management Complexities

The chronological evolution of digital data generation can be related to some years ago. Then, it is seems that the first digital data was produced by the first computer in the world (Data Creation). The produced digital data might be eventually stored in distinct types of media for the future purpose (Data Storing). Then, the stored data must be converted into information and useful knowledge through some specific processing (Data Processing). Then data analyzed to extract new values for the end users (Data Analysis). This initial roadmap (from data creation to data analysis through data storing and data processing) demonstrates a simple data life cycle.

DLC models have been proposed to set a high-level framework standing for a global data life view from the creation stage to the usage stage. The main purpose of a DLC

model is to organize data management in order to suggest the data products to the end-users, which is fitted to their requirements [1-4]. In addition, several DLC models exist in the literature as part of particular sciences and/or environments and/or data cycles management. In [7] the authors proposed the COSA-DLC model to manage and organize data in any scenario, science, and Big Data environment. On top of that, the model is able to fit easily to any Big Data scenario (including smart city, Scientific areas and so on).

The COSA-DLC model consist of three main blocks and is presented in [7] as shown in Fig. 1. Each block contains a set of sophisticated phases. These phases make the model comprehension, agnosticism, and adaption true. For more detailed information, the Data Acquisition block includes four phases. The Data Processing block offers in three phases. Finally, the Data Preservation block consists of four phases. The description of each phase (including all responsibilities and functionalities) together with the relationship among phases, is named the Data LifeCycle Management (DLM) and is presented in [7].

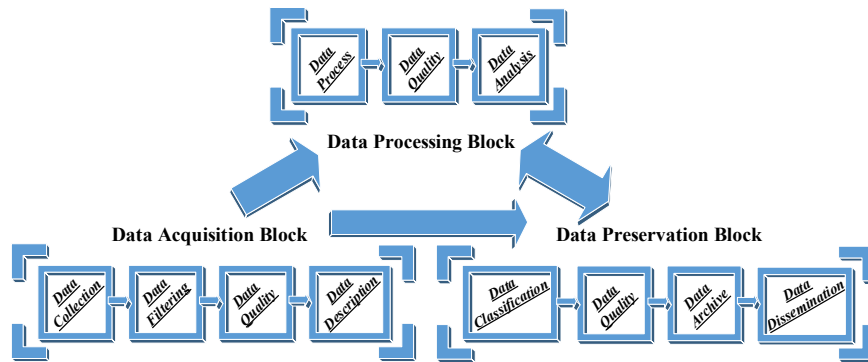


Fig. 1. The DLC model [1, 7]

#### 4 Big Data Management in Smart Cities

Smart cities include several technological challenges and need for a ubiquitous deployment of computing resources throughout the city (from IoT devices to advanced data centers). All data resources must be connected through several communication networks by many different network technologies (including wireless sensor networks, Bluetooth, 4G, Wi-Fi, etc.) and this scenario together should be organized by deploying advanced architectural approaches (including IoT, IoE etc [55]) to build the smart city idea. However, beyond all technologies, the most precious resource for a city to become smart is data. In addition, there is a huge number of data sources (including sensors, smart devices, 3<sup>rd</sup> party applications and so on) across smart cities in today's world. As a result, there is a huge concern to organize the big amount of produced data in the smart cities in academia and industries nowadays. Such a huge concern faces the following difficulties, to name a few:

- Data sources are distributed in a smart city.
- The data volume is growing exponentially in a smart city.
- Data generates by varieties sources, types, and formats (heterogeneous data).
- There are many redundant and dark (useless) data in related data storages.
- Smart services request for the use of both, real-time data (for fast access and critical services) and historical data (for deeper computing services).

In [7, 56] the authors defined a comprehensive DLC model in the context of a smart city, which combines the advantages of the centralized and distributed data management strategies. Therefore, if services or data stakeholders request particular (or critical) data, it will be collected from the distributed data sources (means that the real-time data comes from a close location). Although if in case more historical dataset is required, probably least recent, it is collected from upper levels (thus with higher level of the capabilities), more centralized nodes. The model is called the Smart City Comprehensive Data Lifecycle (SCC-DLC) model based on the F2C data management architecture (from distributed to centralized data management). . This model is able to manage all data life cycle stages (from data collection to usage). Moreover, the model covered all other important features, such as data quality and data security. The benefits of this model is that it mix the benefits of both the cloud and fog computing technologies. These benefits are using high performance capacities for computationally intensive applications, diminish communication latencies for real-time or critical services, decrease network data traffic and increase fault tolerance and security protection, to name a few.

## 5 Use Case for Big Data Management in Smart Cities

As an example of the smart cities use case (which can be assumed as a particular scenario for the Big Data environments), the ZEN center is located in Norway [5]. The ZEN center looks at groups of buildings, instead of a single building, which was the Zero Emission Buildings (ZEB) center’s objective [57]. A neighborhood is described as a group of interconnected buildings with related infrastructure [5]. Therefore, the final target moves to “smart cities” through the idea of the zero-emission neighborhoods [5]. The ZEN center has eight different pilot projects in distinct cities (Bodø, Trondheim, Steinkjer, Evenstad, Elverum, Oslo, Bærum and Bergen) in Norway as depicted in Fig. 2.

Our main effort in this paper is to develop our proposed data management architecture for the ZEN center [6] and their related pilots. Then, [6] designed the hierarchy architecture as shown in Fig.2. The base of our hierarchy architecture is established throughout the two main axes, Time and Location, has been depicted. Those axes helps us to depict our idea about Big Data Management in smart cities through the concept of the “data types”, “data management architecture” and “DLC models”.

There are several applications/services in the ZEN pilots, which are working with multiple data sources (including IoT sources, web services and so on). Then, it seems

that there are several complexities to discover suitable data sources regarding time-alignment (real-time, historical, and last-recent data). Therefore, in our Big Data management proposal, we categorized data according to its age, ranging from real-time to historical data. First, real-time data needs to be consumed for critical applications as it is generated. Such real-time data requires some implicit proximity constraints because these data faces some obstacles to be critical in remote services. Second, data can be considered historical (older data) as long it is accumulated and saved on the data repositories for the future purpose. In this situation, historical data can reside in a place (for instance cloud technologies) which is farther away of their own related data sources. It can result in the higher level of latencies that can be assumed to data access from the cloud. Finally, the last-recent data is in the middle layer of our time-alignment schedule assumption through the architecture. Then, the data will be received from all lower data sources in this layer. This layer has individual tasks (including processing and storage). Afterward the data will be sent to the upper layers (cloud technologies).

There are two types of the data management architecture in our proposed Big Data architecture. One is centralized (Cloud) and the other one is distributed data management architecture. In this proposed architecture, we used the advantage of using both data management architectures to handle the Big Data complexities and challenges in our scenario. On one hand, the distributed data management architecture positioned to the closest layers of the data sources in the city. In addition, the distributed data management architecture uses the potential of the fog technologies to overcome the Big Data challenges in the distributed schema. On the other hand, the centralized data management architecture is located to the top place in our architecture. Sometimes, the physical place can be considered in different cities or continent that is far away from the data sources. Traditionally, the centralized data management model uses the cloud technologies to organize all historical (centralized) data.

As we discussed in Section 4, our proposed Big Data management architecture fits the SCC-DLC model to organize all data life stages (from creation to consumption) in our scenario. In addition, we explained our vision of data lifecycle in terms of core steps and data flow. We proposed three main blocks (which are namely Data Acquisition, Data Processing, and Data Preservation). Data Acquisition block plans to collect all available data from our city pilots. Therefore, the Data Acquisition block is mainly responsible for collection, classification, quality check, description tag of data and prepare them for the next purpose. Then, the data can be processed or preserved. The Data Processing block is mainly converting data into valuable information through several analysis or analytical processes. This processed data can be utilized by the users or saved for the next purpose. Indeed, the Data Preservation block is mainly handling the data storing tasks (received in either the Data Acquisition or Data Processing blocks), and make them ready for publication, or for advanced processing purposes.



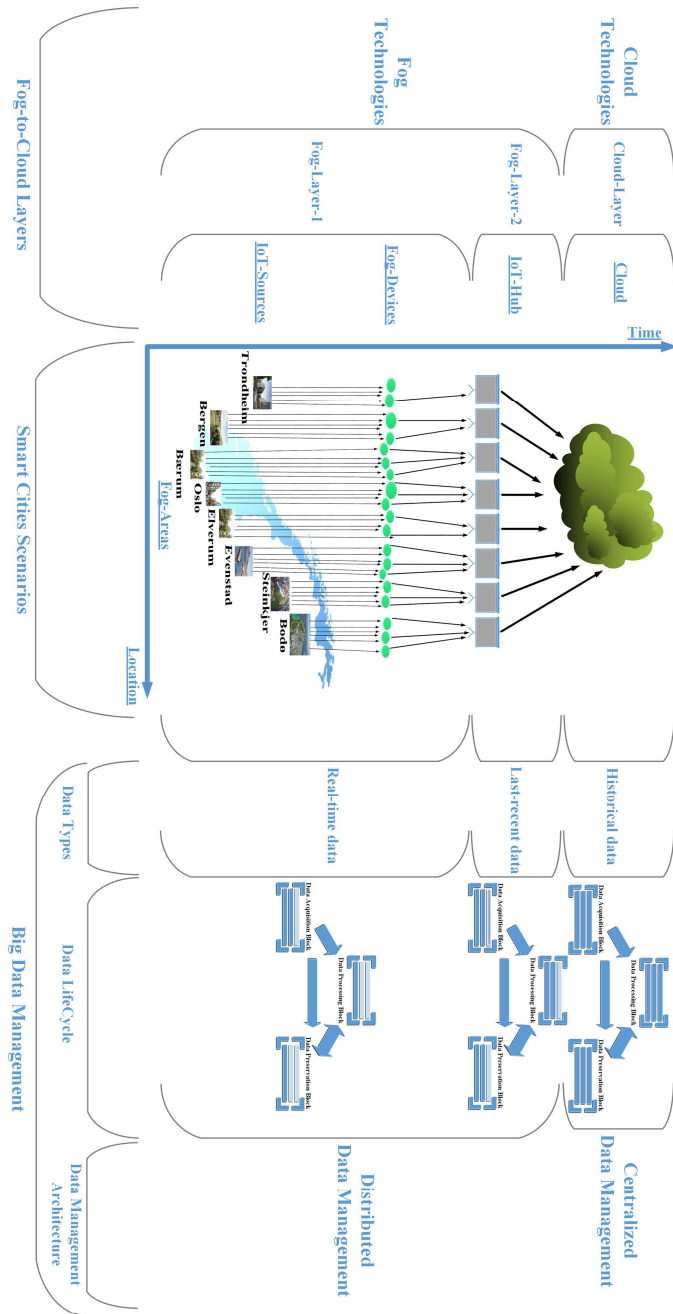


Fig. 2. The ZEN Big Data Management architecture based on F2C data management

As shown in Fig.2, a proposed ZEN data management architecture consists of the following three layers architecture: (Fog-Layer-1, Fog-Layer-2, and Cloud layer) and described below.

- Fog-Layer-1: is an adjacent layer to the end-users and IoT devices in the pilot. This layer constitutes many different IoT-Sources (including sensors, smart-phones and so on) in the Fog-Areas (including various types of the building and their neighborhoods) and Fog-Device (is the most robust node for the processing and storage among the IoT-Sources). This layer can handle several Big Data management tasks as shown below.
  - Data Type: Fog-Layer-1 is saving the real-time data.
  - Data Management architecture: this layer is considered as a part of the distributed data management architecture.
  - DLC model: on one hand, this layer is mainly responsible for the duties of the Data Acquisition block (darker color level) because the majority of the data sources are positioned in this layer. On the other hand, this layer provides the processing and storage capacities (lighter color level).
- Fog-Layer-2 is a middle layer between Cloud and Fog-Layer-1 layers. This layer includes with IoT-Hub (is the strongest node for processing and storage to handle all obtained data). Moreover, this layer is somewhere in the city of the pilot, but it is not as close as to IoT devices like Fog-Layer-1. This layer can manage several Big Data management tasks as shown below.
  - Data Type: Fog-Layer-2 is a place for storing the last-recent data.
  - Data Management architecture: again, this layer assumes under consideration of the distributed data management architecture.
  - DLC model: IoT-Hub is responsible for high-level tasks under Data Preservation and Data Processing blocks (medium color level). However, the cloud is responsible for the advanced level of the processing and storage tasks. Besides that, the Data Acquisition block has less responsibility in this layer (lighter color level) than the lower layer because the data sources are fewer than the lower layer.
- Cloud layer is in the prominent position. The cloud accumulates with the most potent resources concerning processing and storage. Data Type: Fog-Layer-1 provides facility to store all historical data. This layer can organize several Big Data management tasks as shown below.
  - Data Type: cloud is responsible for keeping the historical data.
  - Data Management architecture: this layer is the central place for centralized data management.
  - DLC model: the cloud technology has almost unlimited resources for all enquires of data. Then, all related tasks will be done in the cloud environment (darker color level).

## 6 Conclusion

In this paper, we developed our proposed data management architecture for ZEN center based on a distributed hierarchical F2C data management. The main points of this development are the following:

- We illustrated the upsides of using both data management architectures (including distributed and centralized) in the context of the smart cities to handle Big Data complexities and challenges;
- We explained that our proposed architecture has an interesting facility to organize all different data types (including real-time, last-recent, and historical data) from IoT devices to cloud technologies in each cross-layer of our architecture;
- We described that F2C data management (from distributed to centralized) has a great possibility to handle all data life stages (from creation to conception) with respect to the DLC concepts;
- We contributed to different smart city scenarios to demonstrate our proposed Big Data architecture for the smart cities.

As a part of our future work, we will discover more options related to developing our ZEN data management architecture, such as extending other data sources and external third-party applications.

## Acknowledgment

This paper has been written within the Research Centre on Zero Emission Neighborhoods in smart cities (FME ZEN). The authors gratefully acknowledge the support from the ZEN partners and the Research Council of Norway. In addition, the first author would like to express my very great appreciation to the Advanced Network Architecture Lab (<https://craax.upc.edu/>) in UPC university of Barcelona, Spain because of their support for his Ph.D. thesis under the FI-DGR scholarship 2015FI\_B100186 (<https://upcommons.upc.edu/handle/2117/114435>).

## References

1. Sinaeepourfard, A., Garcia, J., Masip-Bruin, X.: Hierarchical distributed fog-to-cloud data management in smart cities. Departament d'Arquitectura de Computadors, vol. Doctoral thesis. Universitat Politècnica de Catalunya (UPC), Barcelona, Spain (2017)
2. Levitin, A., Redman, T.: A model of the data (life) cycles with application to quality. *Journal of Information and Software Technology on Elsevier* 35, 217-223 (1993)

3. Rüegg, J., Gries, C., Bond-Lamberty, B., Bowen, G.J., Felzer, B.S., McIntyre, N.E., Soranno, P.A., Vanderbilt, K.L., Weathers, K.C.: Completing the Data Life Cycle: using information management in macrosystems ecology research. *Journal of Frontiers in Ecology and the Environment* 12, 24-30 (2014)
4. Hu, H., Wen, Y., Chua, T.-S., Li, X.: Toward scalable systems for big data analytics: A technology tutorial. *Journals & Magazines on IEEE Access* 2, 652-687 (2014)
5. <https://fimezen.no/>
6. Sinaeepourfard, A., Krogstie, J., Petersen, S.A., Gustavsen, A.: A Zero Emission Neighbourhoods Data Management Architecture for Smart City Scenarios: Discussions toward 6Vs challenges. *International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE (2018)
7. Sinaeepourfard, A., Garcia, J., Masip-Bruin, X., Marin-Torder, E.: Towards a comprehensive data lifecycle model for big data environments. *Proceedings of the 3rd IEEE/ACM International Conference on Big Data Computing, Applications and Technologies*, pp. 100-106. ACM (2016)
8. Sinaeepourfard, A., Garcia, J., Masip-Bruin, X., Marin-Torder, E.: Data Preservation through Fog-to-Cloud (F2C) Data Management in Smart Cities. *2nd International Conference on Fog and Edge Computing (ICFEC)*, 2018 IEEE, pp. 1-9. IEEE (2018)
9. Siddiqi, A., Hashem, I.A.T., Yaqoob, I., Marjani, M., Shamsirband, S., Gani, A., Nasaruddin, F.J.J.o.N., Applications, C.: A survey of big data management: Taxonomy and state-of-the-art. 71, 151-166 (2016)
10. Assunção, M.D., Calheiros, R.N., Bianchi, S., Netto, M.A., Buyya, R.J.J.o.P., Computing, D.: Big Data computing and clouds: Trends and future directions. 79, 3-15 (2015)
11. Storey, V.C., Song, I.-Y.J.D., Engineering, K.: Big data technologies and management: What conceptual modeling can do. 108, 50-67 (2017)
12. Demchenko, Y., Zhao, Z., Grosso, P., Wibisono, A., De Laat, C.: Addressing big data challenges for scientific data infrastructure. In: *IEEE 4th International Conference on Cloud Computing Technology and Science (CloudCom)*, pp. 614-617. IEEE, (2012)
13. Grunzke, R., Aguilera, A., Nagel, W.E., Herres-Pawlis, S., Hoffmann, A., Krüger, J., Gesing, S.: Managing complexity in distributed Data Life Cycles enhancing scientific discovery. In: *IEEE 11th International Conference on E-Science (e-Science)*, pp. 371-380. IEEE, (2015)
14. Henry, S., Hoon, S., Hwang, M., Lee, D., DeVore, M.D.: Engineering trade study: extract, transform, load tools for data migration. In: *IEEE Conference on Design Symposium, Systems and Information Engineering*, pp. 1-8. (2005)
15. Almeida, F.L.F., Calistru, C.: The main challenges and issues of big data management. *International Journal of Research Studies in Computing* 2, (2012)
16. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.J.F.g.c.s.: Internet of Things (IoT): A vision, architectural elements, and future directions. 29, 1645-1660 (2013)
17. Jin, J., Gubbi, J., Marusic, S., Palaniswami, M.: An information framework for creating a Smart City through Internet of Things. *Journal of Internet of Things Journal on IEEE* 1, 112-121 (2014)
18. Sinaeepourfard, A., Garcia, J., Masip-Bruin, X., Marin-Torder, E.: A Novel Architecture for Efficient Fog to Cloud Data Management in Smart Cities. In: *IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, pp. 2622-2623. IEEE, (2017)

19. Bonomi, F., Milito, R., Zhu, J., Addepalli, S.: Fog computing and its role in the internet of things. Proceedings of the first edition of the MCC workshop on Mobile cloud computing, pp. 13-16. ACM (2012)
20. Chen, M., Mao, S., Liu, Y.: Big data: A survey. Journal on Mobile Networks and Applications 19, 171-209 (2014)
21. Cox, M., Ellsworth, D.: Managing big data for scientific visualization. In: ACM Siggraph, pp. 21. (1997)
22. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H., Institute, M.G.: Big data: The next frontier for innovation, competition, and productivity. (2011)
23. Villars, R.L., Olofson, C.W., Eastwood, M.: Big data: What it is and why you should care. White Paper, IDC (2011)
24. Hashem, I.A.T., Yaqoob, I., Anuar, N.B., Mokhtar, S., Gani, A., Khan, S.U.: The rise of “big data” on cloud computing: Review and open research issues. Journal of Information Systems on Elsevier 47, 98-115 (2015)
25. Zikopoulos, P., Parasuraman, K., Deutsch, T., Giles, J., Corrigan, D.: Harness the Power of Big Data The IBM Big Data Platform. McGraw Hill Professional (2012)
26. Berman, J.J.: Principles of big data: preparing, sharing, and analyzing complex information. Newnes (2013)
27. Laney, D.: 3D data management: Controlling data volume, velocity and variety. META Group Research Note 6, 70 (2001)
28. O’Leary, D.E.: Artificial intelligence and big data. Journal of IEEE Intelligent Systems 28, 0096-0099 (2013)
29. Zhi-Hua, Z., Chawla, N.V., Yaochu, J., Williams, G.J.: Big Data Opportunities and Challenges: Discussions from Data Analytics Perspectives [Discussion Forum]. Computational Intelligence Magazine, IEEE 9, 62-74 (2014)
30. Rossi, R., Hirama, K.: Characterizing Big Data Management. Journal on Issues in Informing Science and Information Technology (IISIT) 12, (2015)
31. Gordijenko, D.: Big Security for Big Data: Addressing Security Challenges for the Big Data Infrastructure. In: Secure Data Management: 10th VLDB Workshop, SDM 2013, Trento, Italy, August 30, 2013, Proceedings, pp. 76. Springer, (2014)
32. Narasimhan, R., Bhuvaneshwari, T.: Big Data—A Brief Study.
33. Samuel, S.J., Koundinya, R., Sashidhar, K., Bharathi, C.: A SURVEY ON BIG DATA AND ITS RESEARCH CHALLENGES. (2006)
34. <http://dataconomy.com/seven-vs-big-data/>
35. <http://www.optimusinfo.com/blog/understanding-the-7-vs-of-big-data/>
36. Micheni, E.M.: Diffusion of Big Data and Analytics in Developing Countries. (2015)
37. Bloomberg, J.: The Big Data Long Tail. (2013)
38. Jagadish, H., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J.M., Ramakrishnan, R., Shahabi, C.: Big data and its technical challenges. Communications of the ACM 57, 86-94 (2014)
39. Chen, C.P., Zhang, C.-Y.: Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences 275, 314-347 (2014)
40. PWG, N.B.D.: Draft NIST Big Data Interoperability Framework: Volume 4, Security and Privacy. (2015)
41. Struijs, P., Braaksma, B., Daas, P.J.: Official statistics and Big Data. Big Data & Society 1, 2053951714538417 (2014)

42. Sahafizadeh, E., Nematbakhsh, M.A.: A Survey on Security Issues in Big Data and NoSQL. (2015)
43. <http://www.klarity-analytics.com/392-dimensions-of-big-data.html>
44. <http://insidebigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/>
45. C. Desouza, K., L. Smith, K.: Big Data for Social Innovation. Stanford Social Innovation Review (2014)
46. <http://www.dummies.com/how-to/content/how-to-ensure-the-validity-veracity-and-volatility.html>
47. Ali-ud-din Khan, M., Uddin, M.F., Gupta, N.: Seven V's of Big Data understanding Big Data to extract value. In: American Society for Engineering Education (ASEE Zone 1), 2014 Zone 1 Conference of the, pp. 1-5. IEEE, (2014)
48. <http://www.bigdata-alliance.org/what-is-big-data/>
49. <http://data-magnum.com/how-many-vs-in-big-data-the-characteristics-that-define-big-data/>
50. <http://cloudonomic.blogspot.com.es/2014/11/big-data-characterized.html>
51. <http://www.datamensional.com/big-data/>
52. <http://blog.softwareinsider.org/2012/02/27/mondays-musings-beyond-the-three-vs-of-big-data-viscosity-and-virality/>
53. Desouza, K.: Realizing the Promise of Big Data. Washington, DC: IBM Center for the Business of Government (2014)
54. Krishnan, K.: Data warehousing in the age of big data. Newnes (2013)
55. Kyriazopoulou, C.: Smart city technologies and architectures: A literature review. In: Smart Cities and Green ICT Systems (SMARTGREENS), 2015 International Conference on, pp. 1-12. IEEE, (2015)
56. Sinaeepourfard, A., Garcia, J., Masip-Bruin, X., Marin-Tordera, E., Yin, X., Wang, C.: A data lifeCycle model for smart cities. In: International Conference on Information and Communication Technology Convergence (ICTC), pp. 400-405. IEEE, (2016)
57. <https://www.zeb.no/>