Proceedings of the 3rd International Conference on Green Energy, Environment and Sustainable Development (GEESD2022), X. Zhang et al. (Eds.)

doi:10.3233/ATDE220322

Research on Fragmentation Problems of Big Data Analytics in Environmental Management

Haijing WANG[a,1] and Yaping WANG[b]
[a] *Law School, Wenzhou University, Wenzhou 325035, Zhejiang, China*
[b] *Wenzhounese Economy Research Institute, Wenzhou 325035, Zhejiang, China*

Abstract. Fragmentation is a common but critical problem of big data analytics. Although big data analytics have the 5-v benefit, including volume, variety, veracity, velocity and value, fragmentation may make such benefit illusory. The benefit of big data analytics is based on the complete coverage of data collection and the sharing of data among multiple levels and multiple branches of the organization. A field investigation is conducted to test the big data analytic systems of an environmental protection agency of a province in China at the level of province, prefecture, and county. This field investigation reveals that the data collection coverage fails to cover some essential aspects, particularly when such data are generated by other governmental branches. The field investigation also reveals that the data are not shared among different levels or different branches of the government, which leads to the problems of fragmentation. Further investigation reveals that this fragmentation mainly lies in the insufficiency of prediction in the design of database framework and the fragmental purchase of big data analytic services. This paper recommends predictive designing in the initial design of the framework of big data analytic systems so that it may better accommodate changes in the future.

Keywords. Big data analytics, fragmentation in database analytics, predictive design

1. Introduction

Fragmentation is a common but critical problem to big data analytics. Fragmentation means the compartmentalized data collection and barriers preventing or hindering data sharing. “Big Data” is big because of the volume, velocity, and variety of data [1]. Fragmentation makes a seemingly “big” database actually “small”. Fragmentation may render such benefit illusory, although big data analytics (BDA) were developed to overcome the insufficiency of manual computing capacity [2-3], have been specially made for particular fields [4], and have the 5-v benefit, including volume, variety, veracity, velocity and value [5, 6]. The benefit of big data analytics is stressed in literature [5, 6], however, whether the benefits are actually achieved in practice is still a big question.

This paper examines the use of big data analytics actually used by governmental agencies on issues related to the environment. This paper chooses the environmental big

[1] Corresponding Author, Haijing WANG, Law School, Wenzhou University, Wenzhou, Zhejiang, China; Email: wang.haijing@wzu.edu.cn.

data analytic systems as the case for study because environmental issues call for the collaboration and sharing of data among governmental agencies. Currently, environmental issues are related to one primary governmental agency, i.e., the environmental protection bureau, and various other supporting governmental agencies. Therefore, this case may shed light on the collaboration and sharing of data necessary for big data analytics.

2. The Case Examined: Big Data Analytic Systems of Environmental Protection Agencies in a Province in China

2.1. Multiplicity of Governmental Branches in Data Collection and Data Sharing

Various ministries and departments of the central government have authorities on various aspects of work related to environmental protection. Key ministries at the national level include the Ministry of Ecology and Environment (MEE), the National Development and Reform Commission (NDRC), the Ministry of Housing and Urban-Rural Development (MOHURD), and the Ministry of Agriculture and Rural Affairs (MOA). The primary agency is the MEE but others all have authorities and responsibilities on environmental protection. They all have powers on issues related to environmental protection and also have resources to generate data related to environmental protection [7, 8]. In one word, they all benefit from big data analytic systems as well as may contribute to the big data analytic systems. All these ministries at the national level have their subordinate corresponding governmental branches in local governments.

2.2. Multiplicity of Levels of Governments

Below the national government, there are the provincial government, the prefecture government, and the county government. The lower government reports to the higher government while the higher government supervises the work of lower governments. They all have governmental branches corresponding to the national government.

2.3. The Particular Environmental Issues Studied

This paper chooses the pilot projects as cases for study because pilot projects have new big data analytic systems and draws more governmental as well as non-governmental attention. Governments at the provincial level choose some cities as pilot cities for various types of pilot projects. Currently the pilot project cities are mainly in Table 1.

2.4. Components of Big Databases

The big data analytic system of the pilot projects generally includes the basic database, the sectoral database, and the integration process [7, 8].

The basic databases include environmental pollution, energy consumption, climate, geographic, and demographic information. They are accessible by all governmental agencies, although the demographic information does not include individual-level information, which is only available to the police.

Table 1. Cities engaged in pilot projects.

Ministries and departments	Types	Cities
MOHURD	Sponge cities	16 cities of the first group, 14 cities of the second group
NDRC	Low-carbon cities	8 cities and 5 provinces for the first group 18 cities and 1 province for the second group 14 cities for the third group
NDRC and MOHURD	Cities accommodated to climate change	28 cities
MEE	Ecological civilization pilot cities	Cities of Fujian Province, Jiangxi Province and Guizhou Province

The sectoral databases are established by various governmental agencies. For example, the police department has the data of individual citizens, the water affairs agency has the data on lakes, rivers, and underground water, the agricultural agency has the data on agriculture. These data are collected in the daily work by staff of these agencies.

The integration process is the key to big data analytics. It is the integration process that integrates the basic database and various sectoral databases. The differences between various pilot project lie in the integration processes.

3. Fragmentation Problems Revealed in Field Study

The pilot projects revealed some concerns with big data analysis as well as provides some insight into big data analytics.

3.1. Sufficiency and Accuracy of Data

The availability and accuracy of data are the bases of big data analysis [9, 10]. The data collection with big data technology is more accurate than the data generated with previous techniques. Traditionally, the cities require different agencies to file different tables. The data is fragmental and sometimes conflicting. Ideally, the conflicting data should be addressed at the earliest time but they are not addressed and cured because they are scattered and hard to detect. The use of big data collection techniques seemingly addresses this problem. However, our field study reveals that data collectors feel overwhelmed by the data and the suppliers of data may engage in unintended cheating when they are asked for overwhelming data. This affects the accuracy of data.

3.2. Sharing of Data

Our field study reveals severe problems in data sharing, far more problematic than the sufficiency and accuracy of data in the data collection. Vertically, the provincial system can have access to most but not all systems of subordinate prefectures and counties, while the subordinate prefectures and counties have no access to prefecture systems. Horizontally, governmental agencies have severe difficulty to have access to the database of other governmental agencies.

3.3. Sources of the Problem

The field investigation further reveals that this fragmentation mainly lies in the insufficiency of prediction in the initial design of the framework of the big data analytic systems. The initial design failed to predict that other different levels and different branches of government may need to participate in the collection of data and the sharing of data.

4. Conclusion

This study finds that theoretically speaking big data analytics have great benefit but in practice the benefit is not sufficiently generated because of the fragmentation problem. The designing of big data analytics systems should consider the problem of fragmentation. This paper recommends predictive designing in the initial design of the framework of big data analytic systems so that it may better accommodate changes in the future.

References

[1] Manogaran G, Lopez D. Spatial cumulative sum algorithm with big data analytics for climate change detection. Computers and Electrical Engineering. 2018;65:207-221.
[2] Grover V, Lindberg A, Benbasat I. The perils and promises of big data research in information systems. Journal of the Association for Information Systems. 2020;21:9-293.
[3] Kar AK, Dwivedi YK. Theory building with big data-driven research—moving away from the "what" towards the "why" International Journal of Information Management. 2020;54:102205.
[4] Power DJ, Kaparthi S. Building web-based decision support systems studies in informatics and control. Studies in Informatics and Control. 2020;11(4):291-302.
[5] Addo-Tenkorang R, Helo PT. Big data applications in operations/supply-chain management: A literature review. Computers & Industrial Engineering. 2016;101:528-543.
[6] Papadopoulos T, Gunasekaran A, Dubey R, Altay N, Childe SJ, Wamba SF. The role of big data in explaining disaster resilience in supply chains for sustainability. Journal of Cleaner Production. 2017;142(2):1108-1118.
[7] Nerini F, Sovacool B, Hughes N, et al. Connecting climate action with other sustainable goals nature. Sustainability. 2019;2:674-680.
[8] NDRC. National Strategy on Energy Production and Consumption Revolution (2016-2030). 2016 http://www.ndrc.gov.cn/fzgggz/fzgh/ghwb/gjjgh/201705/W020170517397451808076.pdf (in Chinese)
[9] Guo H. Big Data on the Earth for Sustainable Development of China. Beijing: Science Press; 2020. p. 178-193.
[10] Deng X, Wang J. Smart city climate chart based on big data times Shanghai urban management. 2014;23(04):33-36.