

# Chapter 40

## A Service–Oriented Foundation for Big Data

Zhaohao Sun

 <https://orcid.org/0000-0003-0780-3271>

Papua New Guinea University of Technology, Papua New Guinea

### ABSTRACT

*This paper provides a service-oriented foundation for big data. The foundation has two parts. Part 1 reveals 10 big characteristics of big data. Part 2 presents a service-oriented framework for big data. The framework has fundamental, technological, and socio-economic levels. The fundamental level has four big fundamental characteristics of big data: big volume, big velocity, big variety, and big veracity. The technological level consists of three big technological characteristics of big data: Big intelligence, big analytics, big infrastructure. The socioeconomic level has three big socioeconomic characteristics of big data: big service, big value, and big market. The article looks at each level of the proposed framework from a service-oriented perspective. The multi-level framework will help organizations and researchers understand how the 10 big characteristics relate to big opportunities, big challenges, and big impacts arising from big data. The proposed approach in this paper might facilitate the research and development of big data, big data analytics, business intelligence, and business analytics.*

### 1. INTRODUCTION

Big data has become one of the most important frontiers for innovation, research and development in computing (Kumar B., 2015; McKinsey, 2014; Manyika, Chui, & Bughin, 2011) (Sun, Strang, & Li, 2018), industry and business (Chen & Zhang, 2014) (Sun Z., 2019). Big data has become a strategic asset for organizations, industries, enterprises, businesses, and individuals as well as national security. Big data has also been a key enabler of exploring business insights and economics of services. These have drawn an unprecedented interest in industries, universities, governments and organizations (Gartner, 2020) (Sun, Strang, & Li, 2018). Big data and its emerging technologies including big data analytics and intelligent analytics (Coronel & Morris, 2015; Tableau, 2020) have been not only making big changes in the way the business operate but also making traditional business analytics bring about new

DOI: 10.4018/978-1-6684-3662-2.ch040

big opportunities for academia and enterprises (Sun, Strang, & Yearwood, 2014; Manyika, Chui, & Bughin, 2011). Big data analytics and intelligent analytics has big market and investment opportunities (Sun Z., 2019). For example, International Data Corporation (IDC) forecasts that big data and business analytics (BDA) revenue will be \$274.3 billion by 2022 with a five-year compound annual growth rate (CAGR) of 13.2% from 2018 to 2022 (IDC, 2019). However, some fundamental issues are still open for comprehending big data. For example:

- Why is big important for big data?
- What are the characteristics of big data?
- How do we understand the relationships among the characteristics of big data?

This article will address these three issues. It addresses the first issue taking into account a mathematical thinking. To address the second issue, different from the existing literature on big data (Borne, 2014; Gandomi & Haider, 2015; McAfee & Brynjolfsson, 2012), this article identifies and reveals 10 Bigs as the ten big characteristics of big data based on our early research (Sun, Strang, & Li, 2018). The 10 Bigs include big volume, big velocity, big variety, big veracity, big intelligence, big analytics, big infrastructure, big service, big value, and big market. Then this article presents a service-oriented framework through exploring the interrelationships among these 10 Bigs. This framework reveals that the 4 Bigs are fundamental characteristics of big data; another 3 Bigs are technological characteristics of big data; the remaining 3 Bigs are socioeconomic characteristics of big data.

The remainder of this article is organized as follows. Section 2 overviews the characteristics of big data from an evolutionary Perspective. Section 3 looks at why big is important in a big data world. Section 4 identifies and reveals 10 Bigs as the ten big characteristics of big data. Section 5 presents a service-oriented framework and looks at service providers, brokers and requestors for fundamental, technological, socioeconomic level of the framework respectively. The final sections discuss the related work, implications, and end this paper with some concluding remarks and future work.

## **2. CHARACTERISTICS OF BIG DATA: AN EVOLUTIONARY PERSPECTIVE**

This section looks at the characteristics of big data from an evolutionary perspective.

The characteristics of big data have been scattered in a significant number of publications. From an evolutionary perspective, Doug Laney of the META Group (now Gartner) uses 3 Ds: data volume, data velocity, and data variety to represent the characteristics of data in e-commerce in 2001. These 3 Ds should be controlled in data management using novel techniques (Laney, 2001). Late these 3 Ds have been changed into 3 Vs (volume, velocity, and variety) which have been explained as three characteristics of big data (Tsai, Lai, Chao, & Vasilakos, 2015; McAfee & Brynjolfsson, 2012; Gandomi & Haider, 2015). These 3 Vs have been extended first to 4 Vs as four characteristics of big data by adding veracity (Sathi, 2013; IBM, 2015), and then to 5 Vs as five characteristics of big data (volume, variety, velocity, veracity, value) (DataCom, 2015; Wang, 2012), finally to 10 Vs by adding another 5 Vs: validity, variability, venue, vocabulary and vagueness (Borne, 2014).

With the development of big data as a computing, big data computing, more Vs might be proposed to extend these 10 Vs, because case based reasoning (CBR) researchers have extended the R model of CBR from 4 Rs, 5 Rs (Sun & Finnie, 2004; 2010) to 11 Rs (Bridge, 2005). To our knowledge, however, these

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/a-service-oriented-foundation-for-big-data/291015](http://www.igi-global.com/chapter/a-service-oriented-foundation-for-big-data/291015)

## Related Content

---

### Internet of Things: Concepts, Applications, and Challenges

Varsha Sharma, Vivek Sharma and Nishchol Mishra (2018). *Exploring the Convergence of Big Data and the Internet of Things* (pp. 73-95).

[www.irma-international.org/chapter/internet-of-things/187894](http://www.irma-international.org/chapter/internet-of-things/187894)

### A Predictive Analytics Framework for Blood Donor Classification

Kavita Pabreja and Akanksha Bhasin (2021). *International Journal of Big Data and Analytics in Healthcare* (pp. 1-14).

[www.irma-international.org/article/a-predictive-analytics-framework-for-blood-donor-classification/277644](http://www.irma-international.org/article/a-predictive-analytics-framework-for-blood-donor-classification/277644)

### Measuring the Quality of Healthcare Services in Bangladesh

Fahima Khanam and Nayem Rahman (2019). *International Journal of Big Data and Analytics in Healthcare* (pp. 15-31).

[www.irma-international.org/article/measuring-the-quality-of-healthcare-services-in-bangladesh/232323](http://www.irma-international.org/article/measuring-the-quality-of-healthcare-services-in-bangladesh/232323)

### Combining Clustering and Factor Analysis as Complementary Techniques

Lakshmi Prayaga, Krishna Devulapalli and Chandra Prayaga (2020). *International Journal of Data Analytics* (pp. 48-57).

[www.irma-international.org/article/combining-clustering-and-factor-analysis-as-complementary-techniques/258920](http://www.irma-international.org/article/combining-clustering-and-factor-analysis-as-complementary-techniques/258920)

### On Estimating Population Means of Two-Sensitive Variables With Ranked Set Sampling Design

Shivacharan Rao Chitneni, Stephen A. Sedory and Sarjinder Singh (2022). *Ranked Set Sampling Models and Methods* (pp. 104-140).

[www.irma-international.org/chapter/on-estimating-population-means-of-two-sensitive-variables-with-ranked-set-sampling-design/291281](http://www.irma-international.org/chapter/on-estimating-population-means-of-two-sensitive-variables-with-ranked-set-sampling-design/291281)