Chapter 71
Big Data, Privacy, and Healthcare

Nirav Bhatt
https://orcid.org/0000-0002-8105-3563
Charotar University of Science and Technology, India

Amit Thakkar
Charotar University of Science and Technology, India

ABSTRACT

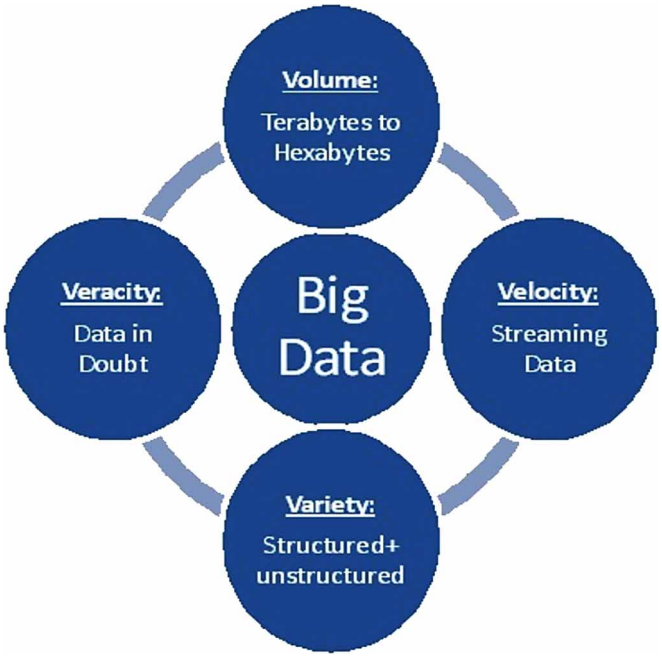
In the era of big data, large amounts of data are generated from different areas like education, business, stock market, healthcare, etc. Most of the available data from these areas are unstructured, which is large and complex. As healthcare industries become value-based from volume-based, there is a need to have specialized tools and methods to handle it. The traditional methods for data storage and retrieval can be used when data is structured in nature. Big data analytics provide technologies to store large amounts of complex healthcare data. It is believed that there is an enormous opportunity to improve lives by applying big data in the healthcare industry. No industry counts more than healthcare as it is a matter of life and death. Due to rapid development of big data tools and technologies, it is possible to improve disease diagnosis more efficiently than ever before, but security and privacy are two major issues when dealing with big data in the healthcare industry.

INTRODUCTION

Due to the growth of the healthcare industry, a lot of data is generating. There are massive problems presented in modern healthcare including high cost, high waste, and low quality. But the hope of big data is lead to better care and lower cost. There is massive amount of data which gives analytical algorithms or systems a lot to act on. Fig 1 shows characteristics of Big Data

DOI: 10.4018/978-1-7998-8954-0.ch071

Figure 1. Characteristics of Big Data



Characteristics of Big Data

1. **Volume:** There is a massive amount of data which gives analytics algorithms or systems a lot to act on. For example, for genomic data each human genome requires 200 gigabyte of raw data, or 125 megabyte if store only snips. For medical imaging data, single FMRI is about 300 gigabyte.
2. **Variety:** There is a variety of data which connect lots of information sources together. For example, clinical informatics such as patients diagnosis, procedure, meditation, lab results and clinical notes. Patient generated data, such as the information coming out of on-body sensors and other devices that patients wear. Real-time data sources such as blood pressure measures, temperature heart rate, drug dispensing levels at intensive care units.
3. **Velocity:** Often data is coming in live and need to be processed and analyze live.
4. **Veracity:** There is a lot of noise, missing data, errors and a lot of false alarms.

Challenges for Big Data and Health Analytics

1. Incorporating new information, such as biomedical data, and new technologies into electronic health records that store big data. Text data require special algorithms; generic data may be voluminous, and continuously monitored.
2. Harnessing the potential of unstructured data for analysis, such as medical imaging and text.
3. Building a culture of data sharing and architecture, including interoperability, to meet health system needs, including future meaningful use requirements.
4. Building data systems that meet requirements of accountable care organizations and other types of payment reforms.