Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 172 (2017) 544 – 549

Procedia Engineering

www.elsevier.com/locate/procedia

Modern Building Materials, Structures and Techniques, MBMST 2016

Big data in building energy efficiency: understanding of big data and main challenges

Natalija Koseleva*, Guoda Ropaite

Faculty of Civil Engineering, Department of Construction Technology and Management, Vilnius Gediminas Technical University, Sauletekio al. 11, 10223, Vilnius

Abstract

Data generation has increased drastically over the past few years. Data management has also grown in importance because extracting the significant value out of a huge pile of raw data is of prime important thing to make different decisions. One of the important sectors nowadays is construction sector, especially building energy efficiency field. Collecting big amount of data, using different kinds of big data analysis can help to improve construction process from the energy efficiency perspective. This article reviews the understanding of Big Data, methods used for Big Data analysis and the main problems with Big Data in the field of energy.

Peer-review under responsibility of the organizing committee of MBMST 2016

Keywords: Energy Efficiency; Big Data; Characteristics of Big Data; Big Data analysis; Construction; Web of Science.

1. Introduction

The building sector is evolving to be the greatest energy consumer around the world, accounting for 40% of the global energy use and one third of the global greenhouse gas emissions. As a result, building energy efficiency has become one of the top concerns of a sustainable society and attracted increasing research and development efforts in recent years [1-2].With the rapid development of sensor technology, wireless transmission technology, network communication technology, cloud computing, and smart mobile devices, large amounts of data has been accumulated

* Corresponding author.
E-mail address: natalija.koseleva@vgtu.lt

Peer-review under responsibility of the organizing committee of MBMST 2016
doi:10.1016/j.proeng.2017.02.064

in almost every aspects of our lives. Moreover, the volume of data is growing rapidly with increasingly complex structures and forms. A research report of International Data Corporation (IDC) pointed out that 1.8ZB data were created and replicated in 2011 worldwide, and it is estimated that this figure will increase by 50 times by the year 2020. The energy big data provides a new way to analyze and understand individuals' energy consumption behavior, and thus to improve energy efficiency and promote energy conservation [3-5].

2. The understanding and the use of Big Data

Big data is a popular phenomenon that aims to provide an alternative to traditional solutions based on databases and data analysis. Big data is not just about storage or access to data; its solutions aim to analyze data in order to make sense of them and exploit their value. Big data refers to datasets that are terabytes to petabytes (and even exabytes) in size, and the massive sizes of these datasets extend beyond the ability of average database software tools to capture, store, manage, and analyze them effectively.

The concept of big data has been defined through the 3V model, which was defined in 2001 by Laney as: "high-volume, high velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making". More recently, in 2012, Gartner updated the definition as follows: "Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization". Both definitions refer to the three basic features of big data: Volume, Variety, and Velocity. Other organizations and big data practitioners have extended this 3V model to a 4V model by including a new "V": Value. This model can be even extended to 5Vs if the concept of Veracity is incorporated into the big data definition [6-9]. The concepts in the energy sector can be briefly described as follows (Fig. 1):

- *Volume*: refers to large amounts of any kind of data from any different sources, including mobile digital data creation devices and digital devices. The benefit from gathering, processing, and analyzing these large amounts of data generates a number of challenges in obtaining valuable knowledge for people and companies [6]. In energy sector, the wide deployment of smart metering devices (e.g., smart meters) created massive amounts of data. For instance, for a distribution network with 1 million smart meters deployed, the amount of electricity consumption data collected in one year is very large. According to Zhou ant other scientists, the amount of electricity consumption data collected by 1 million smart meters in a distribution network within one year, the amount of electricity consumption data collected once every 15 mins by 1 million smart meters within one year will be up to 2920 TB. This presents not only a storage problem, but an analytic problem of making sense of all that data.
- *Velocity*: refers to the speed of data transfers. The data's contents are constantly changing through the absorption of complementary data collections, the introduction of previous data or legacy collections, and the different forms of streamed data from multiple sources. From this point of view, new algorithms and methods are needed to adequately process and analyze the online and streaming data [6]. Velocity mainly means the speed of energy big data collection, processing and analysis. Different from traditional post-processing type business intelligence and data mining, the collection and processing of energy big data need surprising speed. To support the near real-time decision-makings in energy system, the speed of data collection and processing ranges from sub-second to 5 or 15 mins intervals [3].
- *Variety*: refers to different types of data collected via sensors, smartphones or social networks, such as videos, images, text, audio, data logs, and so on. Moreover, these data can be structured (such as data from relational databases) or unstructured in format [6]. Energy big data has a high degree of variety. Generally, it is a mix of structured (e.g., the energy consumption data), semi structured (e.g., data exchanged between smart energy management platform and third-party data aggregators using XML, Web services), and unstructured data (e.g., email or SMS notification about energy use, interactions of consumers on social media about their energy use). In addition, there are also some inter-industry data (e.g., electric vehicle-related data) and outside-industry data (e.g., weather data) in the energy big data. These different types of data all combined will result in a significant increase in the complexity of energy big data applications [6].
- *Value*: refers to the process of extracting valuable information from large sets of social data and it is usually referred to as big data analytics. Value is the most important characteristic of any big-data-based application,