Energy Reports 4 (2018) 91–100

Contents lists available at ScienceDirect

Energy Reports

journal homepage: www.elsevier.com/locate/egyr

Research paper

Power systems big data analytics: An assessment of paradigm shift barriers and prospects

Hossein Akhavan-Hejazi, Hamed Mohsenian-Rad *

Department of Electrical Engineering, University of California, Riverside, CA, USA

ARTICLE INFO

Article history:
Received 5 April 2017
Received in revised form 22 November 2017
Accepted 30 November 2017

Keywords:
Energy
Big data analytics
Internet of energy
Smart grid

ABSTRACT

Electric power systems are taking drastic advances in deployment of information and communication technologies; numerous new measurement devices are installed in forms of advanced metering infrastructure, distributed energy resources (DER) monitoring systems, high frequency synchronized wide-area awareness systems that with great speed are generating immense volume of energy data. However, it is still questioned that whether the today's power system data, the structures and the tools being developed are indeed aligned with the pillars of the big data science. Further, several requirements and especial features of power systems and energy big data call for customized methods and platforms. This paper provides an assessment of the distinguished aspects in big data analytics developments in the domain of power systems. We perform several taxonomy of the existing and the missing elements in the structures and methods associated with big data analytics in power systems. We also provide a holistic outline, classifications, and concise discussions on the technical approaches, research opportunities, and application areas for energy big data analytics.

1. Introduction

Started in the information technology (IT), Big Data Analytics (BDA) has now found extensive applications in many areas of technology and business intelligence (Chen et al., 2012). Those serving mass consumers are particularly interested in using such tools to understand the current state of their business and track the still-evolving aspects. The electric power industry, interacting with one of the largest customer-serving critical networks is going through some drastic, rapid changes in both business and technical paradigms (Bui et al., 2012; Jaradat et al., 2015; Aiello and Pagani, 2014). Thus, naturally it is presenting limitless opportunities for BDA. Power system Big Data (BD) brings new opportunities such as providing an otherwise non-existing feedback loop, taking actions to correct and enhance planning, and enabling accurate realization of the system states, leading to more informed operations.

In this paper, we aim to overview some fundamental concepts and characterizations of BD and BDA, in the domain of power systems. We address questions such as: What are the attributes of energy data and whether they constitute BD? What are the distinct concepts in BDA related to power systems? What are the challenges in generation, communications, management and analysis of BD? What are the new core theories that furnish BDA in power system domain? What are the barriers to adopt the existing generic BDA tools and platforms for BDA in power systems?

2. Energy big data characterization

Although the term "Big Data" is self-explanatory, it still can be a source of confusion or controversy. For example, what an electric utility may consider BD could be seen as moderate size data for data-centric enterprises. The relativity of BD to the systems that operate based on those data is recognized even within the IT community (Chen et al., 2012; Russom et al., 2011). Nevertheless, a definition often used for BD is a "high-volume, high-velocity and high-variety information asset that requires and demands cost-effective, innovative forms of information collection, storage, and processing for enhanced insight and decision making" (De Mauro et al., 2016). From this definition, the volume, i.e., size of data is not the only factor, as there are other factors too. The so called "three Vs" of BD, see Fig. 1, are described as follows:

- Volume: Many IT-related organizations define BD in terabytes—sometimes petabytes (Cohen et al., 2009; Huang et al., 2014). For instance, the data warehouse of Fox Audience Network (a large advertisement network) holds over 200 terabytes of production data (Cohen et al., 2009). The scope of BD also affects its quantification. For example, as shown in Fig. 2, one smart meter, with resolution of seconds to minutes generates much fewer data than one phasor

* Corresponding author.
E-mail address: hamed@ece.ucr.edu (H. Mohsenian-Rad).

https://doi.org/10.1016/j.egyr.2017.11.002

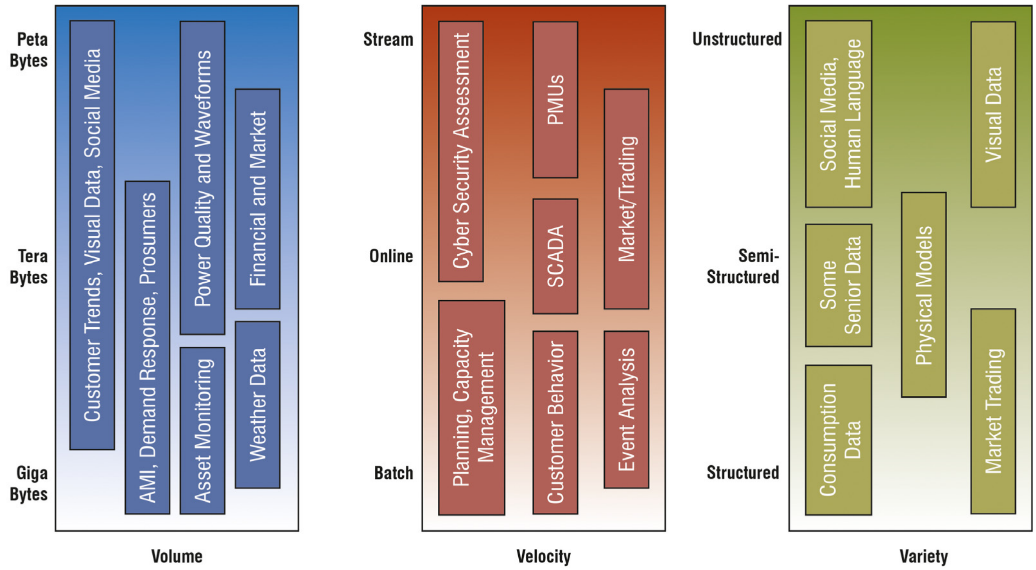

Fig. 1. The 3 V attributes of BD and some examples in the context of power systems.

measurement unit (PMU), with resolution of milliseconds; yet an advanced metering infrastructure (AMI) may generate large volume of data coming from millions of customers, e.g., the AMI in the New York State with seconds-resolution produces roughly 127.1 terabytes of consumption data per day (Huang et al., 2014).

- Variety: Data now comes from a much greater variety of sources compared to traditional data systems. The so-called structured data (e.g., tables and other data structures of relational databases, record formats of most applications, and the character-delimited rows of many flat files) which still form a majority of data, is now joined by *unstructured data* (e.g., text, voice, and video) and *semi-structured data* (e.g., XML, JSON, RSS feeds, and hierarchical data) (Russom et al., 2011). Examples of energy data are shown in Fig. 2.
- Velocity: The frequency of data generation or the frequency of data delivery is a key attribute. An example is a continuous stream of data, as opposed to once-in-a-while event-triggered data from a sensor. Although the majority of power system sensors are event-triggered, there are also sensors, e.g., PMUs both at transmission and distribution level, that produce data streams at high rates (Shand et al., 2015; Stewart et al., 2014b).

Do we currently face BD in the electric power sector? The answer does not seem to be a clear 'yes' or 'no'. Indeed, the *volume* of data being generated in the power sector has grown tremendously over the past few years due to the deployment of smart grid technologies, e.g., 45.8 million smart meters installed in US by 2013 (Alejandro et al., 2014). The energy data now come from *variety* of sources that span a wide range of locations, types and applications. Additionally, many forms of grid data are generated at high *velocities*. Yet, we may still have reservations to answer 'yes' to the above question, as we explain next.

A key reservation is that for many power system sensors, including many emerging and state-of-the-art sensors, the majority of data is either *not logged*, or they are *overwritten* very quickly. For example, in most protection relays and related sensors, the data collected is discarded shortly after internal use. Additionally, almost all state-of-the-art power quality sensors[1] tend to store the voltage or current waveforms only a few cycles before and after an *event* is detected (von Meier and McEachern, 2012). If a pre-programmed event is not detected, then no waveform data is automatically stored. Furthermore, the majority of measured data in power systems are intended to be used close to the point that they are generated, and were never intended to be carried to an enterprise data center, as opposed to the BD mentality in the IT sector (Stimmel, 2014). This is partly due to the different design requirements in the power sector, because traditionally, no centralized data storage model satisfied the needs of very low-latency controller systems in many practical power applications. Accordingly, while many of the recently deployed or emerging power data measurement systems lie in the description of BD, the way that they are currently managed does not exactly match the *soul* and *purpose* of BD. Once such hidden data is collected, managed, and analyzed, they will constitute the real BD in power systems. So far, we may have seen only the "Tip of the Iceberg" of the BD in power systems!

The type of data that can eventually form BD in power systems can be classified into *domain* data and *off-domain* data, see Fig. 3. The domain data can be further categorized by their sources. Here are few examples: (1) *Telemetry and SCADA data:* enable continuous flow of measurements on grid equipment status and parameters and other grid variables. The SCADA data can have various sources such as renewable energy resources which generate huge amount of data, such as real-time production, and equipment status. For instance the data from condition monitoring systems of many wind turbines can be utilized in predictive maintenance strategies (Qiu et al., 2016, 2012; Feng et al., 2013; Qiu et al., 2017; Long et al., 2015). (2) *Oscillographic and Synchrophasor data:* make up of voltage and current waveform samples in time or frequency domains that can create a graphical record. (3) *Consumption data:* is most often the smart meter data. (4) *Asynchronous event data:* often come from devices with embedded processors generating messages under a variety of normal and abnormal conditions. (5) *Metadata:* is any data that can describe other data. Grid metadata is highly diverse and may include internal sensor data, calibration data, and other device-specific information. (6) *Financial data:* may include day-ahead and real-time market bids and price data, bilateral transactions, and retail rates.

Traditionally, power grid operation relies also on different forms of *off-domain* data, i.e., the data that is not specific to or necessarily intended for the power sector. For instance, the weather

[1] For example refer to PQube sensors at http://www.powersensorsltd.com/PQube.