Electric Power Systems Research 151 (2017) 369–380

Contents lists available at ScienceDirect

Electric Power Systems Research

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

Big data framework for analytics in smart grids

Amr A. Munshi [a,b,*], Yasser A.-R. I. Mohamed [a]

[a] Electrical and Computer Engineering Department, University of Alberta, Edmonton, Alberta, Canada
[b] Computer Engineering Department, Umm Al-Qura University, Makkah, Saudi Arabia

ARTICLE INFO

Article history:
Received 27 March 2017
Received in revised form 15 May 2017
Accepted 9 June 2017

Keywords:
Smart grid
Smart meters
Big data
Data management
Dynamic demand response

ABSTRACT

Smart meters are being deployed replacing conventional meters worldwide and to enable automated collection of energy consumption data. However, the massive amounts of data evolving from smart grid meters used for monitoring and control purposes need to be sufficiently managed to increase the efficiency, reliability and sustainability of the smart grid. Interestingly, the nature of smart grids can be considered as a big data challenge that requires advanced informatics techniques and cyber-infrastructure to deal with huge amounts of data and their analytics. For that, this unprecedented smart grid data require an effective platform that takes the smart grid a step forward in the big data era. This paper presents a framework that can be a start for innovative research and take smart grids a step forward. An implementation of the framework on a secure cloud-based platform is presented. Furthermore, the framework has been applied on two scenarios to visualize the energy, for a single-house and a smart grid that contains over 6000 smart meters. The application of the two scenarios to visualize the grid status and enable dynamic demand response, suggests that the framework is feasible in performing further smart grid data analytics.

1. Introduction

Ever growing volume of data production is the reality we are living in. The recent technological advancements have led to a deluge of data from various domains, such as social networks, scientific sensors, smart cities, and the Internet. The global data volume from 2005 to 2020 is predicted to grow by a factor of 300, from 130 exabytes to 40,000 exabytes, representing a double growth every two years [1]. To cope with the volume, velocity and variety of data produced, the term "big data" was brought up to capture the meaning of this evolving trend of data.

Big data are becoming a new technology focus in science and engineering domains. Big data systems include a set of tools and mechanisms to acquire, store, and process disparate data while leveraging the massively parallel processing power to perform complex transformations and analysis. However, designing and deploying a big data framework system for a specific application is not a trivial or straightforward task [2,3]. This is due to the fact that data comes from multiple, heterogeneous and autonomous sources with complex and evolving relationships, and keeps on growing. Moreover, the rise of big data applications where data collection has grown tremendously is beyond the capability of current commonly used hardware and software platforms to manage, store and process within a tolerable amount of time [2].

Many utilities are transferring to smart meters and smart grids as part of long range planning to improve the reliability of power supply, incorporate distributed generation resources, develop storage solutions, use the power plants efficiently, and enable customers to participate in controlling their energy use. To accomplish this, many utilities are deploying smart meter systems as a first step. This leads to incorporate other challenges. For example, going from a system that reads the meter once a month to a smart meter that can provide meter readings every few minute leads to millions of reads per hour. The result is a massive increase in data that is overwhelming if not managed properly. This generated data if managed efficiently can provide better understanding of customer behavior and assist in defining electric tariffs. For example, time of use pricing encourages customers to operate certain higher voltage appliances at off-peak periods. Consequently, customers save money and less power is generated.

Smart grids are a dynamic field of research and development and are currently observing a phase where research ideas are of interest. Technology changes are starting to permeate through the entire

* Corresponding author at: Electrical and Computer Engineering Department, University of Alberta, Edmonton, Alberta, Canada.
E-mail address: aamunshi@ualberta.ca (A.A. Munshi).

http://dx.doi.org/10.1016/j.epsr.2017.06.006
0378-7796/

smart grid, from generation to transmission and distribution. For example, renewable power resources such as wind and solar power are being included in the generation mix, not just by the power generation utilities, but also by consumers through micro-grids. Also, vehicle-to-grid (V2G) and grid-to-vehicle (G2V) technologies can provide power flow from vehicles to power lines and back. Also, smart meters are being deployed at consumer premises to monitor near real-time energy consumption data and securely communicate it back to the utility over communication networks. These meters can also receive data from the utility with information on dynamic power usage and incentives for reducing load during peak periods. Streaming various data from thousands of smart meters, sampled every few minutes, must be collected and correlated for monitoring, controlling, and research purposes. Also, such data can be processed and shared between utilities and customers to enable dynamic demand response (DDR) for near real-time monitoring and response. Smart grids are characterized by concepts of sustainability, interoperability, and controllability. The contemporary developments toward the future smart grid will require acquiring and analyzing data of integrated devices, such as distributed storage, intelligent loads, and distributed energy resources [4]. However, the massive amounts of data evolving from smart grids needs to be sufficiently managed to increase the efficiency, reliability and sustainability of the grid. This is a big data challenge that requires advanced informatics techniques and cyber-infrastructure to deal with huge data and their analytics. Interestingly, big data reflects the true nature of the smart grids. For that, the unprecedented data volumes require an effective platform that takes the smart grid a step forward in the big data era.

In Refs. [2,5–9] various challenges and issues in adapting big data technology were discussed. From their research, it was concluded that refining a unified framework suitable for every module is not straight forward due to the diversity of applications. For that big data frameworks for smart grids are of research interest. In Refs. [7,9] several main elements of big data and data base technologies that are beneficial within the utility ecosystem are scratched on, but a comprehensive idea on how big data elements can construct a framework to deal with smart grid data has not been presented. The authors of Ref. [8] designed architecture for smart grids based on mathematical foundations and statistical procedures. Ref. [3] analyzed several challenges of big data and suggested that high-performance computing platforms are required to unleash the power of big data. In Ref. [10], statistical learning algorithms for big data were presented. Also, Ref. [11] presented an unsupervised method for early event detection in smart grids with big data. A big data value chain in Ref. [2] was presented; it decomposed big data into four sequential modules, namely data generation, data acquisition, data storage, and data analytics. More in detail, numerous approaches for each phase were highlighted, and a prevalent framework for addressing big data challenges was suggested. Ref. [12] demonstrates a cloud-based DDR platform project being deployed in the University of Southern California (USC) campus micro-grid as a testbed for transforming Los Angeles utility into a smart grid in the future. Works in big data analytics mainly describe possible theoretical frameworks and challenges, and lack practical implementation. Specially, handling real smart grid big data is becoming an area of research interest for that, this paper focuses on the development of a comprehensive big data framework for analytics in smart grids. Furthermore, an implementation and application of the framework is applied to test its feasibility.

The contributions of this paper to the research field are:

- A comprehensive big data framework for smart grids.
- The utilization of open source state-of-the-art prevalent Hadoop platform for addressing smart grids big data challenges.
- The adoption of open source tools to provide an easy and cost-effective development environment for practicing engineering to develop similar tools for their demanding smart grid applications.
- Practical implementation of the framework, including necessary configuration and coding, and application of the framework on real smart grid data.
- The development of a secure cloud-based DDR platform.
- Outlining potential research directions that can be applied using the framework.
- The ability to perform various data analytical applications on top of the framework.

The remainder of this paper is structured as follows. Section 2 briefly discusses the main characteristics of big data. The big data core components used in the framework for smart grids and the features of using the Hadoop platform in the smart grid environment are highlighted in Section 3. Furthermore, the big data framework for smart grids including the lifecycle of smart grid data from data generation to data analytics is introduced. An implementation with relevant source code on a secured cloud platform is presented in Section 4. The application of the framework on two scenarios is presented in Section 5. Finally, the conclusions are drawn in Section 6.

2. Characteristics of big data

Concurrently, there has been much discussion about what big data actually means [2,12,13]. However, the most common definition in literature is the “Vs” definition [2,14–19] which contains several characteristics of big data beginning with the letter “V”. In this paper, we limit ourselves to the “4Vs” definition [19] (volume, variety, velocity and veracity) which are discussed next.

- **Volume**: big data implies enormous volumes of data. Concurrently, that data is generated by machines, networks, and social media. Thus, the volume of data to be analyzed is massive.
- **Variety**: refers to the various sources and formats of data (structured, semi- structured and unstructured). As data comes in the form of photos, videos, logs, sensor devices, etc., this variety of data formats creates challenges for storage, mining and analyzing.
- **Velocity**: the velocity of data is the rate at which data arrives. This also includes the time that it takes to process and understand the acquired data to help in decision making.
- **Veracity**: is a term that refers to the quality or trustworthiness of the data. Tools that help handle big data’s veracity discard noise and abnormality in data and transform the data into trustworthy insights.

3. Framework’s big data core components for smart grids

This section highlights the big data core components used in the framework for smart grids. Also, the features of using Hadoop’s big data platform in smart grids environment are highlighted. Further, the framework that covers the lifecycle of smart grid data from data generation to data analytics is covered.

3.1. Big data core components—state-of-the-art

Fig. 1 illustrates the hierarchical architecture of the core components of the suggested framework for smart grid big data. In the following sub-sections, the state-of-the-art components of the data acquisition, data storing and processing, data querying and, data analytics components are introduced. The data storing and processing components are included in the same subsection as they fall under the same platform (Hadoop).