

Received March 15, 2016, accepted March 28, 2016, date of current version August 15, 2016.

Digital Object Identifier 10.1109/ACCESS.2016.2580581

Energy Big Data: A Survey

HUI JIANG¹, KUN WANG¹, (Member, IEEE), YIHUI WANG¹, MIN GAO²,
AND YAN ZHANG³, (Senior Member, IEEE)

¹Key Laboratory of Broadband Wireless Communication and Sensor Network Technology, Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

²Department of Electrical Engineering, University of California at Los Angeles, Los Angeles, CA 90095, USA

³Simula Research Laboratory and University of Oslo, Norway

Corresponding author: Y. Zhang (yanzhang@ieee.org)

This work was supported in part by the National Natural Science Foundation of China under Grant 61572262, in part by the National Science Foundation of Jiangsu Province under Grant BK20141427, in part by the Nanjing University of Posts and Telecommunications under Grant NY214097, in part by the Open Research Fund within the Key Laboratory of Broadband Wireless Communication and Sensor Network Technology, Ministry of Education, Nanjing University of Posts and Telecommunications under Grant NYKL201507, in part by the Qinlan Project of Jiangsu Province, and in part by the Research Council of Norway under Project 240079/F20.

ABSTRACT Energy is one of the most important parts in human life. As a significant application of energy, smart grid is a complicated interconnected power grid that involves sensors, deployment strategies, smart meters, and real-time data processing. It continuously generates data with large volume, high velocity, and diverse variety. In this paper, we first give a brief introduction on big data, smart grid, and big data application in the smart grid scenario. Then, recent studies and developments are summarized in the context of integrated architecture and key enabling technologies. Meanwhile, security issues are specifically addressed. Finally, we introduce several typical big data applications and point out future challenges in the energy domain.

INDEX TERMS Big data, energy, smart grids, data processing, data mining, energy internet, survey.

I. INTRODUCTION

Smart grid is a major research and development direction in today's energy industry. It refers to the next generation power grid which integrates conventional power grid and management of advanced communication networks and pervasive computing capabilities to improve control, efficiency, reliability and safety of the grid [1]. Smart grid delivers electricity between suppliers and consumers so as to form a bidirectional electricity and information flow infrastructure. It fulfills the demands of each stakeholder, functionality coordination in electric power generation, terminal electricity consuming and power market. It also improves the efficiency in each part of the system operation and reduces the cost and environmental impact. At the same time, the system reliability, self-healing ability and stability are improved [2].

Integrating telecommunication, automation and electric network control, smart grid requires reliable real-time data processing. To support this requirement and benefit future analysis and decision making, huge amount of historic data should be well fetched and stored in a reasonable time budget. In addition, there are various sources that huge amount of data can be generated through diverse measurements acquired by Intelligent Electronic Devices (IEDs) in the smart grid:

- Data from power utilization habits of users;
- Data from Phasor Measurement Units (PMUs) for situation awareness;
- Data from energy consumption measured by the widespread smart meters;
- Data from energy market pricing and bidding collected by Automated Revenue Metering (ARM) system;
- Data from management, control and maintenance of device and equipment in the electric power generation, transmission and distribution in the grid;
- Data from operating utilities, like financial data and large data sets which are not directly obtained through the network measurement.

All aforementioned sources increase the grid's data volume. According to the investigation in [3], by 2009 the amount of data in electric utilities' system has already reached the level of TeraBytes (TBs). With every passing day, more TBs of data are emerging at infrastructure company data centers. This fact accelerates the pace of big data technology applying to the smart grid area.

To obtain a better understanding of the big data application on smart grid, we give an overview of energy big data (big data applied in the energy domain) covering the