BIG DATA APPLICATION IN POWER SYSTEMS – BRIEF OVERVIEW OF ISSUES AND PROPOSALS REGARDING THE USE OF THESE DATA FOR PREDICTION

Author: Anna Kamińska-Chuchmała

(“Rynek Energii” – 1/2019)

Key words: Big Data, Power Systems, Smart Grid, Satellite Data, Prediction

Summary. This study was designed to presents concise review of a novel subject regarding the use of large data sets (Big Data) which generates the functioning of the power system and their use to improve the operation and economic benefits of Smart Grids. Thanks to smart metering, we have current access to the data on the use of resources, which then using SCADA system and servers that support large data sets such as Apache Hadoop or Spark can be stored. Afterwards, these data are used for predictive calculations that are extremely important from an economic point of view. At the end of the paper, an interesting proposition of research is given by Author, namely to use, as ancillary information, the satellite data obtained from the Copernicus Programme provided by the European Space Agency ESA related for example with temperature to forecast energy consumption in electricity transmission and distribution networks.

1. INTRODUCTION

The issue of big data is quite fresh in power systems context. The actual rapidly developments in advanced metering infrastructure (AMI), monitoring, and sensor network provided fast increase of the Four V's Big Data (volume, variety, velocity and veracity) of measurement’s in electricity transmission and distribution networks. Furthermore, it could be noticed higher availability of open source platform for data analytics to transforming the power system area and turning utilities into data-driven enterprises with using machine learning (ML), database structure, advanced statistics and data mining [1].

The paper is structured in the following way. Firstly, variety of Big Data are introduced. Next, application and ML concerning high volume of data using in power systems are discussed. Subsequently applied Big Data to power systems and mixed model of prediction are presented and finally summarize ending this paper.

2. VARIETY OF BIG DATA AVAILABLE FROM POWER SYSTEMS

The fundamental of data-driven utilities is the information and communication technology (ICT) platform focus on two-way flow of data between customers and transmission and generation operators. The intensive development of utilities by adopting the distribution automation (DA) solutions followed by AMI platforms has led to the extensive innovation in grid operation.

Although considerable issues have been introduced to openly discuss and clarify the foundational ICT requirements to serve all functions of the electric grid in context to the data flood from DA and AMI in data-driven utilities.

Nowadays, where new technologies in transmission and gathered data (like high-speed Internet, distributed resources, smart phones, and cloud computing) are common on high advanced levels the new approach to the data-driven utility is inescapable. Moreover, the data generated by different resources are analysed by using advanced tools like ML, deep learning, and statistical inference. What is more, using advanced big data analytics it is possible to mixed models of estimation, prediction and diagnostics on basis historical and real-time data flows. Summarizing, the co-action of operators, customers, ICT networks, and grid components run the power system into a complex giant for ad hoc data-driven approaches and policies.

3. APPLICATION FOR BIG DATA USING IN POWER SYSTEMS

The aim of developing big data techniques in smart grid is to support complementary operation of multiform energy system in order to achieve energy cascading utilization and boosts the energy efficiency [2]. In the last years, distributed combined cooling heating and power (CCHP) system was developed as a novel energy supply way like an environment-friendly, economical, and reliable option for the future [3]. One of the main components in multiform energy system is CCHP for which typical primary movers are gas turbines, steam turbines, microturbines, and reciprocating internal combustion engines integrated with thermally activated cooling technologies. The purpose of this unique characteristics and complexities is to make new research focus on system design, control, operation, and planning. The most important issues of profitable implementation of CCHP system are optimization and economic evaluation techniques based on thermo-dynamics and thermo-economic rules. Moreover, here also to support functioning multiform energy system a data-driven method with a consideration of uncertainty on price and demand is valuable.

4. MACHINE LEARNING INTEGRATED WITH BIG DATA ANALYTICS IN POWER SYSTEMS

The data generated by rapid development: AMI, sensor networks and monitoring increase the variety, volume, and velocity of measurement data in electricity transmission and distribution networks. Exemplary, the advent of phasor measurement unit (PMU) has granted researches and operators the access to high quality system information that would otherwise be unobservable using traditional technologies. Thanks to the PMUs real-time measurement of three-phase voltage and current magnitude and phase angle with a high accuracy and a refined time resolution could be provided [4]. The time-synchronized observability of the system is enable by GPS time stamped PMU data. In a different manner would be unavailable with traditional measurement technologies. Figure 1 illustrates the integration of PMU network, machine learning (ML)-based event detection tools, and traditional power system infrastructures.

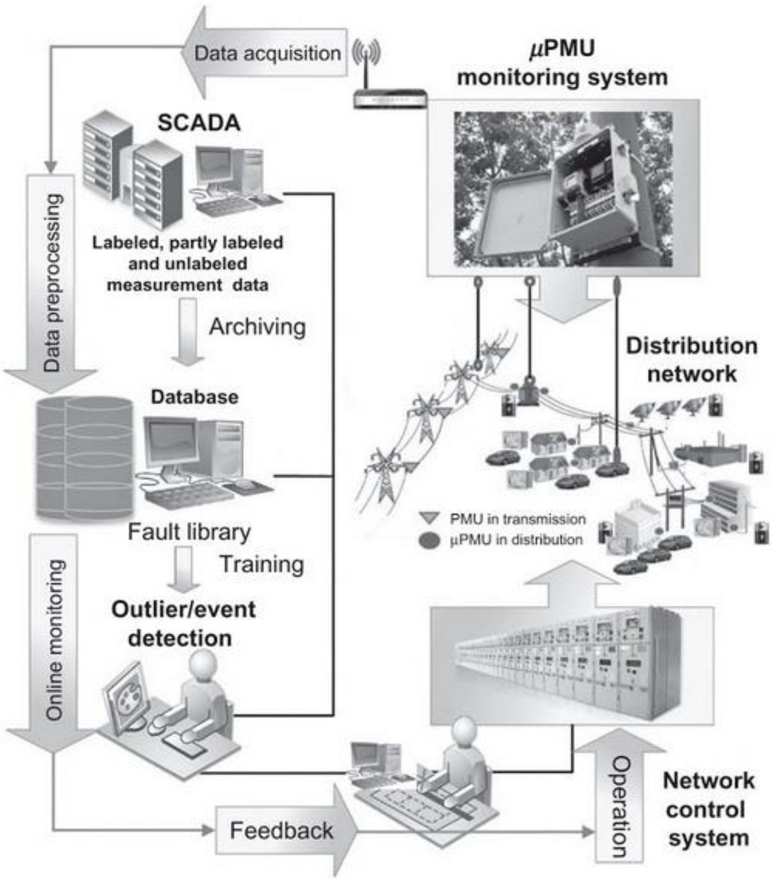

Fig. 1. The integration of ML-based event detection tool with PMU monitoring and SCADA system [5]

Overall, contemporary power systems characteristic a mixed of measurement technology, networking, system dynamics, and computational techniques like ML and control are highly integrated with the demand and behaviour of their users [6].

5. APPLIED BIG DATA INTO POWER SYSTEMS

The specificity of data-driven requires special approaches to the design, operation, and planning of utilities. Furthermore, data-driven utilities need new business models for knowledge extraction from data. Hitherto, there are exist some examples analysis of the demand response (DR) potential of grid users, big data preprocessing from grid sensors, large-scale simulation of electricity markets, and predictive maintenance of electrical equipment [7, 8]. Unquestionably high business value for utilities' stakeholders and customers have a prediction of real-time and day ahead market price, load, and renewable generation time series (TS). Generally, in the distribution and transmission networks the big data applications are driven by two objectives: first to increase the monitoring and situational awareness capability and develop fast decision-making methods for operators, and second to implement predictive active management strategies that take advantage of flexibility from various technologies in the electricity supply and demand such distributed energy resources, DR, and energy storage.

Nevertheless, utilize the full potential of big data in utilities is challenged by deficiency of statistics and data analytics knowledge in utilities workforce. Ready to deploy industry-level ML tools and solutions are not commonly available to utilities which may increase the learning curve and utilities' modernization time. It should be pay attention on modern data-driven solutions such as distributed learning and optimization, spatial-temporal modeling of TS, data reduction, assimilation, and visualization methods for classic power system problems including topology detection, fault detection, load disaggregation, and state estimation [9].

The consensus view seems to be that in the modern world the study of Big Data has become an important aspect of assess, storage and analysis of such enormous amount of data, which is still increasing. Recently, there has been growing interest of availability of variety servers, tools and platforms to streaming, storage and analytics of Big Data. Figure 2 presents possibility to use open-source Apache Spark [11] with plenty of choice environments, data source and application with which are compatible.

Another technology, also very popular, is an Apache Hadoop [12] and similar to Apache Spark accessible for free. This software library is a framework that acquiesce for the distributed processing of Big Data sets across clusters of computers using uncomplicated programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer, so delivering a highly-available service on top of a cluster of computers, each of which may be prone to failures.

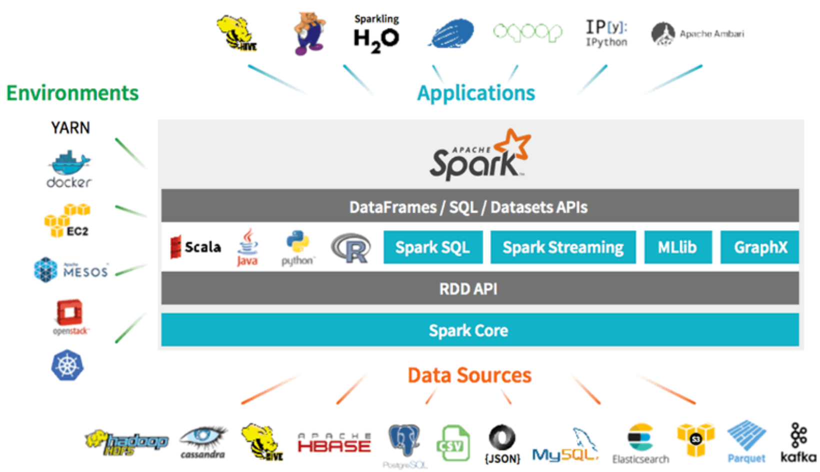

Fig. 2. Spark for big data analytics [10]

Obtain data is one challenge, but another even more complicated is to change data to concrete value which we can use to analytics models and draw valuable conclusions. The mentioned process is presented in fig. 3.

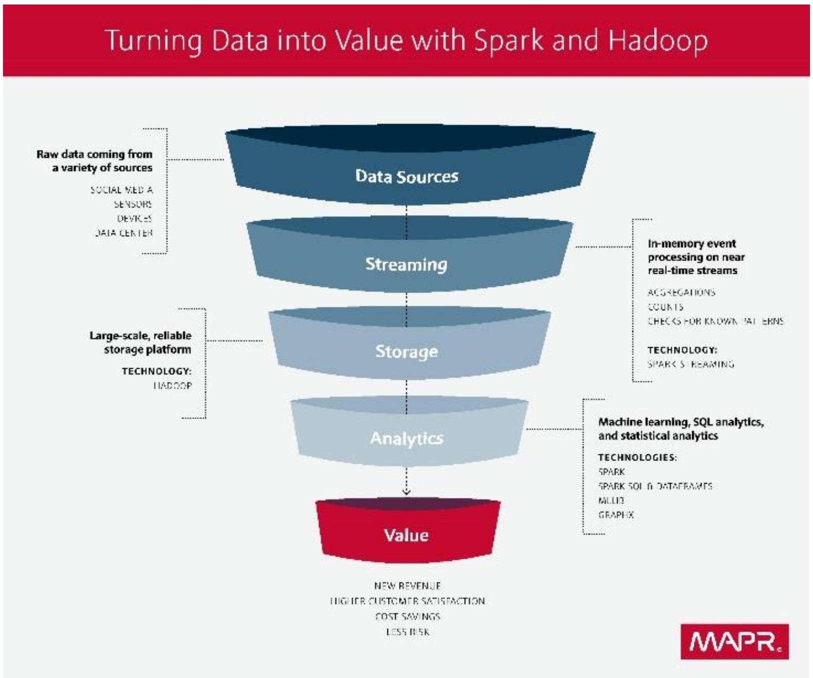

Fig. 3. Spark for big data analytics [13]

The exemplary architecture with description is presented in figure 4.

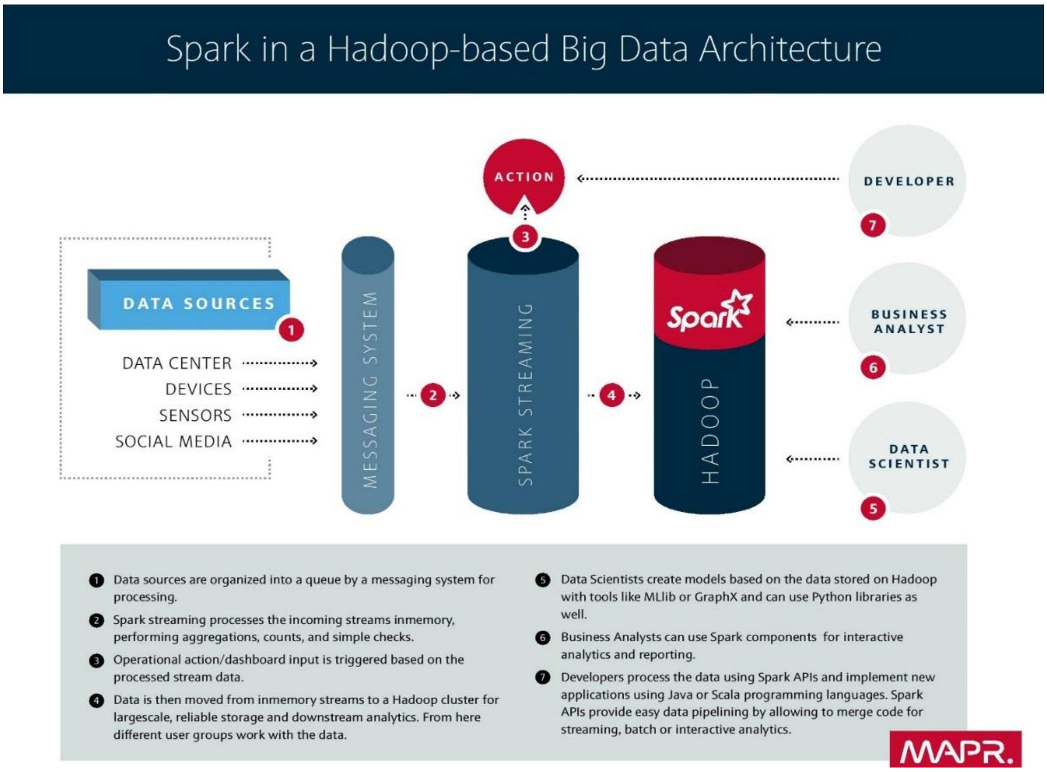

Fig. 4. Exemplary Big Data Architecture [14]

6. APPLIED BIG DATA TO PREDICTION MODELS

The potential of possibility to usage the enormous amount of data is huge. Excellent, for researchers, possibility to have access to Big Data is a Copernicus Programme provided by the ESA [15] which coordinates the delivery of data from upwards of 30 satellites. Dependent on kind of satellites information could be obtained, which would be useful in many domains. Further research in this area may include contamination stationary data and satellites in common model of analysis and prediction. Exemplary proposition of space-time (4D) model of prediction in power system given by the author of this paper is using real data from power system connection with temperature from Satellites images. On basis such mixed 4D model it could be created prediction of usage energy in concrete places in critical moments like in winter.

Similar proposition of usage mixed model Author applied with success in different domain – in hydrology [16]. This research presents a combined methodology of blending satellite and ground observations of rainfall for improved spatiotemporal mapping and analysis. The methodology is based on geostatistical and machine learning principles extended to the Bayesian framework. The case study is the island of Crete, Greece for ten years. Rainfall data from over 50 stations were used in combination with satellite images for the reference period and results were compared with the outcome of Regional Climate Models.

7. SUMMARIZE

To summarise, the data obtained from satellites could be significant useful in many kinds of domains. It is worth to pay attention for power energy companies to possibility of create such mixed models of power energy usage. Moreover, the new applications integrated with modern systems rest on energy consumption should be implemented as a new tools deal with issue of Big Data analysis.

REFERENCES

[1] *Big Data Application in Power Systems*, Edited by Reza Arghandeh and Yuxun Zhou, Elsevier, 2018

[2] Gao Feng, *Frameworks for Big Data Integration, Warehousing, and Analytics*, chapter 4 in *Big Data Application in Power Systems*, Edited by Reza Arghandeh and Yuxun Zhou, Elsevier, 2018

[3] Liu Q., Cui L., Chen H., *Key technologies and applications of internet of things*, Computer Science 37 (6), 2010

[4] Von Meier A., Culler D., McEachern A., Arghandeh R., *Micro-synchrophasors for distribution systems*, in: Innovative Smart Grid Technologies Conference (ISGT), IEEE, 2014, pp. 1-5

[5] Yuxun Zhou, and Reza Arghandeh, *Moving Toward Agile Machine Learning for Data Analytics in Power Systems*, chapter 5 in *Big Data Application in Power Systems*, Edited by Reza Arghandeh and Yuxun Zhou, Elsevier, 2018
[6] Jordan M.I., Mitchell T.M., *Machine learning: trends, perspectives, and prospects*, Science 349 (6245), 2015
[7] Siano P., *Demand response and smart grids – a survey*, Renewable and Sustainable Energy Reviews 30, 2014
[8] Macedo M.N.Q., Galo J.J.M., de Almeida L.A.L., Lima A.D.C., *Demand side management using artificial neural networks in a smart grid environment*, Renewable and Sustainable Energy Reviews 41, 2015
[9] Paterakis N.G., Tascikaraoglu A., Erdinc O., Bakirtzis A.G., Catalao J.P., *Assessment of demand-response-driven load pattern elasticity using a combined approach for smart house-holds,* IEEE Transactions on Industrial Informatics 12 (4), 2016
[10] https://www.imagenesmi.com/im%C3%A1genes/spark-data-f3.html, available and checked on 11.01.2019
[11] https://spark.apache.org/, available and checked on 11.01.2019
[12] https://hadoop.apache.org/, available and checked on 11.01.2019
[13] https://dataconomy.com/2016/03/hadoop-and-spark-a-match-made-in-big-data-heaven/, available and checked on 11.01.2019
[14] https://mapr.com/blog/apache-spark-hadoop-based-big-data-architecture-infographic/assets/blogimages/mapr_ebook_graphic_architecture2.pdf, available and checked on 11.01.2019
[15] https://www.copernicus.eu, available and checked on 11.01.2019
[16] Kamińska-Chuchmała Anna, Varouchakis Emmanouil, and Karatzas George P.: *Blending satellite and ground precipitation observations using geostatistics and machine learning*, EGU General Assembly 2019, publication in review.

ZASTOSOWANIE DUŻYCH ZBIORÓW DANYCH DO SYSTEMÓW ELEKTROENERGETYCZNYCH – PRZEGLĄD ZAGADNIEŃ ORAZ PROPOZYCJE WYKORZYSTANIA TYCH DANYCH DO PREDYKCJI

Słowa kluczowe: duże zbiory danych, systemy elektronenergetyczne, dane satelitarne, sieci inteligentne

Streszczenie. Praca ta zawiera zwięzły przegląd bardzo świeżej tematyki dotyczącej zagadnień wykorzystania dużych zbiorów danych (Big Data) jakie generuje funkcjonowanie systemu elektroenergetycznego i użycie ich do ulepszania działania i ekonomicznych korzyści w tychże systemach typu Smart Grids. Dzięki inteligentnemu opomiarowaniu mamy bieżący dostęp do danych dotyczących wykorzystania zasobów, które następnie za pomocą systemu SCADA oraz serwerów obsługujących duże zbiory danych jak np. Apache Hadoop czy Spark mogą zostać składowane i następnie wykorzystane do obliczeń predykcyjnych niezmiernie istotnych chociażby z ekonomicznego punktu widzenia. Ponadto ciekawą propozycją Autora jest wykorzystanie jako informacji pomocniczych danych satelitarnych z Programu Copernicus udostępnianych przez Europejską Agencję Kosmiczną ESA związanych przykładowo z temperaturą do prognoz zużycia energii w sieci energetycznej.

Anna Kamińska-Chuchmała, Ph.D. Eng., Assistant Professor, Faculty of Computer Science and Management, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland, e-mail: anna.kaminska-chuchmala@pwr.edu.pl