Big data usage in electrical distribution systems: A review

Okan Şen[1] K. Çağatay Bayındır[2]

Received: 03/05/2019 Accepted: 06/07/2019
Online Published: 28/10/2019

Abstract

By this study, author aims for reader to get basic knowledge about relation between data and distribution systems and big-data analytics concept on energy distribution systems. Concept of intelligent energy systems, namely smart grid is described. Big-data basics is explained. Benefits of data-driven systems are searched. A detailed information about applications of big-data on distribution systems is given. Sources of energy big-data is mentioned. Finally, we point out the challenges of managing big data-driven distribution systems. Affect of big-data analytics of energy efficiency. A comprehensive review on big data driven smart energy management applications is presented.

Keywords: Distribution System, Dynamic Energy Management, Smart Grid, Data Analytics, Big-data

1. Introduction

Classically, data produced by an energy system was only used to see errors occurred while system is not observed. Works on recent years show that data collection is much more important than it is thought before. To enable system to be "smart" a huge amount of data is exchanged between components and the enterprise systems managing these components. Moreover, it is understood that data production is inevitable for energy systems.

Especially, after development of intelligent energy network systems, benefits about work on data analytics is seen clearly and on the age data size extracted is getting bigger every day.

Compared to classical energy systems, in the modern industrial systems data analytics is playing a more important role. With development of information and communication technology, Information layer is a part of a modern electrical transmission system. Such a system has wide installation of smart meters and sensors [1].

Smart grid is the brain of an intelligent distribution system which requires reliable real-time data for processing. As a result, energy systems are designed such to produce increasing amount of data to be used for advanced statistical machine learning applications.

2. Big Data

With the fast development of digital technology and cloud computing, an opportunity occurred about data generated by electrical distribution systems can be analyzed to bring effectivity, efficiency and benefit to our daily life may be to company activities.

[1] Department of Energy Systems Engineering, Yıldırım Beyazıt University, Ankara-Turkey, okansen1@gmail.com
[2] Electrical and Electronics Engineering Department, Yıldırım Beyazıt University, Ankara, Turkey

Processing and analysis method of this data is a challenging task. Because data collected is growing exponentially and the structure of it is also getting complexity by time.

Big data generated by smart grids are usually heterogeneous and asynchronous have varying resolution, and are stored in different formats. A typical smart meter data is collected every 15 min. One million smart meters installed in a utility result in nearly 3 TB of new energy consumption data every year [2].

According to another study about size of data, when the data is collected from 1 million devices every 15 minutes, 35.04 billion records are obtained with the volume of 2920 Tb data [3].

At the beginning of this century there occurred an opportunity about such a large and complicated data when concept of "big data" is declared [4].

Big data analytics is being focused by many areas ranging from sales and consumer analysis to fast-growing industries such as information technology. With huge amount of data measured in petabytes, the information technology industry should be leading to adopt big data algorithms and processes. Nowadays, most popular computer and web applications are powered by big data analytics algorithms.

2.1. Characteristics of Big Data

The characteristics of big data is defined by five major elements. These are **V**olume, **V**elocity, **V**ariety, **V**eracity and **V**alue. These are 5Vs of big data.

Volume: Big data has large volume of data. This is a consequence of data being built by large number of deployed smart metering devices and a wide-range level of measurement. This presents also new opportunities to detail analysis, but also a challenging task to store and processing of data .

Velocity: This refers to speed for collecting, processing and usage of big data of systems. Time intervals of data collection should be as short as possible to obtain data in details. Processing time should be short to decrease response time to an acceptable level.

Power systems are required to work in real-time or almost real-time manner for known reasons.

Variety: This stands for having to consider different types of data. Structured or unstructured data may be in consideration for a system due to different levels of recording and at different time. Since standardization is reduced, processing algorithms have to adapt to this situation.

Veracity: This represents uncertainity of data. Especially in power systems, sometimes generation, sometimes demand and mostly both of them is related to natural events. Overall system of processing of data must assure to regard this veracity. Fig 1. shows major elements of big data concept.

Value: This refers to be able to extract valuable information from data. Endless amount of data collected is useless unless it is turned into value. The opportunity of big data applications is directly related with the value hidden inside it.

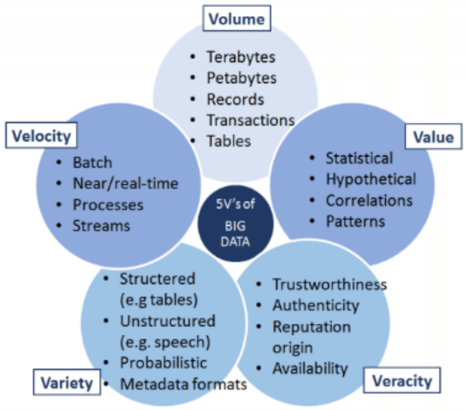

Fig. 1. *Major elements of big data concept*

According to [5] there are there are three major obstacles to the development and implementation of big data analytics in power distribution systems. (i) Lack of innovative use cases and application proposals that convert big data into valuable operational intelligence. (ii) Insufficient research on big data analytics system architecture design and advanced mathematics for this size of data. (iii) Risk of failing to adhere to data privacy and data protection standards.

2.2. Computation Techniques

There are some techniques which are used by modern technologies to analyze big data. Some of them adore to be explained briefly here. These are cloud computing, Internet of things, (IoT), granular computing, data center and quantum computing. [6]

Cloud computing: Cloud computing is delivery of applications and services through Internet. It is closely related to big data. Cloud computing aims to use distributed computing and storage resources at the same time to provide big data applications [7]. The occurrence of big data concept also accelerates the development of cloud computing. By the help of distributed storage technology and the parallel computing big data can be analyzed effectively. Cloud computing increased requirement of transferring and sharing data also.

Internet of Things (IoT): The Internet of Things (IoT) is a network of things by name. "Things" stands for physical objects implemented with electronics, software and sensors. By exchanging data inbetween “things” (IoT) achieves a greater service than sum of independent network sources can do. (IoT) consists of high number of sensors embedded in devices in network. Deployed sensors in different fields collect various kinds of data. Naturally, IoT enables advanced applications for smart grid, since different types of data can be collected, which has heterogeneity and variety. It is expected that by 2030, the IoT data will be the most important part of big data. [6]

Granular computing: Granular computing (GrC), assumes classes, clusters, subsets, groups and intervals as granules and are used to build computational models for complex applications. It offers a solution to uncertainity of big data which often contains a significant amount of unstructured, uncertain and imprecise data. GrC aims to find a suitable level of granularity of a problem level of which can be adjusted according to the degree of complexity of the problem.

Data center: Acquiring, managing & organizing data is performed by datacenters. Data centers are infrastructure and basic for data collection. Data centers should be modernized continuously, due to rapid growth of IT technologies. As grids gets smarter, explosive growth of data is experienced. As a consequence enhancement of infrastructure and related software of data center is a must. The development of data centers should be taken into consideration to improve the capacity of rapidly and effectively processing of big data.

Quantum computing: Quantum computing studies quantum computers that use quantum-mechanical b [8]. Unlike digital computers Quantum computers require data to be encoded into binary digits. These are called quantum bits. Actual quantum computers have newly begun to develop, but experiments have been carried out for developing. It is calculated that large-scale quantum computers will be able to solve problems much more quickly than classical computers.

3. Big Data in Electrical Distribution Systems

3.1. Smart Grid Concept

"The grid" alone stands for a network of transmission lines, substations, transformers and more. A grid is used to deliver electricity from power plant to clients. This delivery is performed by transmission lines. Smart grid is a grid which senses values along these transmission lines by means of smart meters and manages two way communication between utility and consumers by means of digital technology.

Smart grid enhances reliability, availability, and efficiency in distribution systems. During transmission of energy, carry out testing, technology improvements, consumer education, development of standards and regulations, and information sharing is critical.

Uninterrupted power supply, central management, integrated renewable energy resource as power supply, bidirectional electrical grid, consumption data driven pricing, giving change to grid operator to manage network, tracking power consumption data and analyzing history are some advantages of smart grids.

Benefits associated with the Smart Grid listed as following [9]

- More efficient transmission of electricity
- Quicker restoration of electricity after power disturbances
- Reduced operations and management costs for utilities, and ultimately lower power costs for consumers
- Reduced peak demand, which will also help lower electricity rates
- Increased integration of large-scale renewable energy systems
- Better integration of customer-owner power generation systems, including renewable energy systems
- Improved security

A smart grid not only collects data both from customer and utility side, and also analyze historical and immediate data to perform multi-dimensional and multi-level management to adopt changing demands of consumption.

Smart grid also claimed to be automatic for some of its basic jobs like coordination, awareness, healing and reconfiguration. Additionally, integrating renewable energy resources, enhancing power generation efficiency, transmission and usage can be counted.

Fig. 2 shows smart grid model updated by NIST.

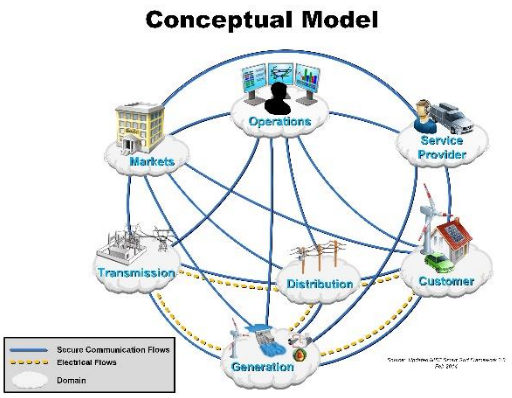

Fig. 2. *A typical Smart Grid Model*

Security is another concept that a smart grid should satisfy. Physical and cyber attacks should be foreseen or response times to these attacks should be shorten as much as possible.

3.2. Big Data in Smart Grids

The major goal of 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 [10]. Combined Cooling, Heating and Power systems (CCHP) is the main component of a multiform energy system including gas turbines, steam turbines, micro turbines and reciprocating internal combustion engines integrated with thermally activated cooling technologies.

Recently designed CCHP systems are examples of data driven models. They were developed as a supply way which is economical and reliable [11]. Since optimization and economic evaluation is critical for deployment of a CCHP system, new researches are being executed on its design, control, operation and planning.

A data-driven method is better for operating such a multiform energy system, when uncertainty on price and demand is considered. At this point power grids plays a basic role for collecting large amount of data from power systems. Smart grids are designed to contain huge number of intelligent devices which are connected to each other without human intervention. Then only analyzing of data is left to optimize a system. Fig. 3 shows a CCHP system controlled by a smart grid.

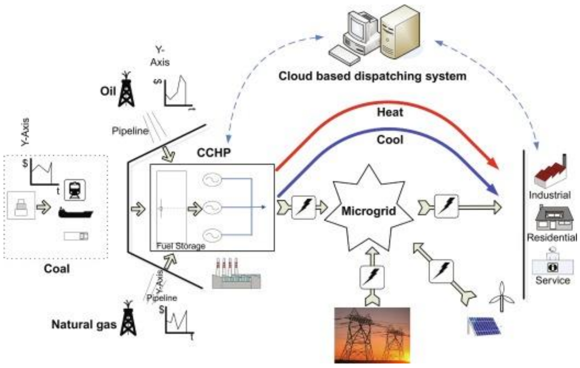

Fig. 3. *A Smart Grid controlled CCHP system*

As seen the multiform energy system employs more than one thermo-dynamic cycle. The outstanding electricity or energy demand is compansated by power grid.

Some researches on subject classifies big data applications on power distribution systems into two, namely short term and long term applications [12]. Energy fraud detection, EV detection and Rooftop Solar Integration, Develop More Granular Load Forecast, Develop More Granular Renewable Generation Forecast, Visualization in Power Distribution Systems, State Estimation and Equipment Diagnosis are short term applications. Model Customer Consumption Behavior under Various Incentive and Pricing Structures, Transformation of Distribution System Planning Process and as long term applications.

An algorithm for distribution system parameter estimation is presented [13]. It is demonstrated using a real utility feeder model with advanced metering infrastructure (AMI) and PV micro-inverters. Illustration in Fig. 4 shows the flows of Big Data for distribution system parameter estimation.

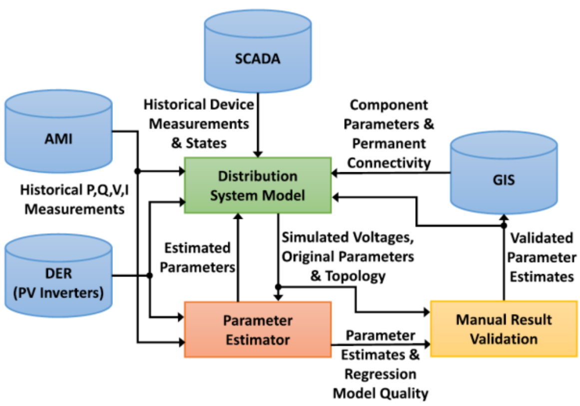

Fig. 4. *Big Data for distribution system parameter estimation*

There exists a proposal of a consumer partitioning methodology at the distribution level spite of probable disturbances and subsequent connectivity issues in the network induced by renewable energy resources[14]. K-means and a fuzzy logic engine is used on Big data environment.

Reliable and Feasible micro energy generated due to Dynamic Energy Management (DEM) and the electricity market including consumers and suppliers. Smart grid aims to reduce the cost of power by the help of their enhanced operations such as data analytics, high performance estimation, adequate data network management and cloud computing. Some studies focuses on the issues in big data and difficulties of it experienced by the DEM handled in Smart Grid [15]. Data processing techniques that are mostly implemented are explained in details, a brief description of the most commonly used data processing methods is given and recommended proposes a directional future research in the field.

Cost efficiency is a big task to come through. Usage of big data analytics to forecast electricity price is another area of big data searches [16]. There are some obstacles on the way to this kind of a solution. First of all there is a huge price data to process. There would be unavoidably redundancy from feature selection and a lack of integrated infrastructure for coordinating the procedures in electricity price. To solve this a modal proposal implemented by integrating three modules is done. Random Forest (RF) and Relief-F algorithms merged on Grey Correlation Analysis (GCA) to be a hybrid feature selector. An integration of Kernel function and Principle Component Analysis (KPCA) is used for feature extraction process to realize the dimensionality reduction. And to forecast price classification, a differential evolution (DE) based Support Vector Machine (SVM) classifier is suggested.

By using an intelligent big data platform, some other researches are done also. There are some considerations on analytics of an extremely large dataset of smart grid electricity price and load, which is very difficult to process with conventional computational models [17]. A deep learning based model is proposed for the forecast of price and demand on big data using deeper Long Short-Term Memory (LSTM). Due to adaptive and automatic feature learning of DNNs, processing of big data is easier with LSTM, compared to purely data driven methods. The proposed model is evaluated using a well-known real electricity markets' data.

From Utility, industry, and research points of view presents a review of big data analytics and its applications in smart grids, and also identifies challenges and opportunities. For future research research gaps and insights are considered on directions to integrate big data analytics into power systems [18].

An architecture using Random matrix theory (RMT) is proposed to motivate data-driven tools to perceive the complex grids in high-dimension [19]. The architecture performed a high-dimensional analysis and compares the findings with RMT predictions to conduct anomaly detections. To reflect the correlations of system data in different dimensions Mean spectral radius (MSR), as a statistical indicator, is defined. A management mode perspective is discussed to make advanced big data analyses possible. It is proved that proposed architecture, as a data-driven solution, is sensitive to system situation awareness, and practical for real large scale interconnected systems. Validation of the designed architecture in various fields of power systems accomplished by case studies.

Smart grids bring deeper changes in the information systems that drive them. A multi player, multi dimensional system is in consideration. Information flows from the electricity grid, new age producers of renewable energies. Usage of electric vehicles and connected houses. New communicating equipments like smart meters, sensors and remote control points. This will cause a flow of excess data that the energy companies will have to face. Big Data technologies offers suitable solutions for utilities. Some studies [20] provide an overview of data management for smart grids, summarize the added value of Big Data technologies for this kind of data, and discuss the technical requirements, the tools and the main steps to implement Big Data solutions in the smart grid context.

On-line solution optimization problems is crucial for a reliable, secure and economic smart grid operation. Reducing the effect of perturbations or adopting the grid state to new load conditions, data acquired by sensors should be continuously analyzed to identify proper control actions. A case based reasoning extracting useful knowledge from historical data is suggested to overcome these high number of computations. A solution is presented aimed at searching k most similar cases based on a paradigm which they call MapReduce paradigm [21]. It allows to distribute data and jobs over a compute cluster of commodity hardware. Experiments are done under different conditions, such as the dataset, size, the number of available nodes/jobs, the block size.

An optimized smart electricity grid faciliates power flow in terms of economic efficiency, reliability and sustainability. This is a two-way flow between suppliers and consumers. In this situation consumers and producers take active role in market and is called the dynamic energy management. A smart grid turns users' participation to power cost reduction. This infrastructure permits the consumers and the micro-energy producers to take a more active role in the electricity market and the dynamic energy management (DEM). The most important challenge in a smart grid (SG) is how to take advantage of the users' participation in order to reduce the cost of power. In order to achieve this intelligent methods and solutions for the real-time exploitation of large volumes of data generated by the vast amount of smart meters is required. A research on subject aimed to highlight the big data issues and challenges faced by the DEM employed in SG network and to provide a brief description of the most commonly used data processing methods used [22].

3.3. Challenges and future perspectives

Still big data can be used to reveal demands, automatic decision making, forecasting and etc. However, big data technology applied to power system is thought to be in its infancy stage. On the way to setup advanced systems there inevitably would be challenges in the future. Today, these difficulties are analyzed and classified to be a base for future researches [23].

Data integration and storage: Classical data analysis usually considers data collected from single source, it is essential to integrate dataset collected from multisource, which has different standards, formats, and representations. Although some storage systems are designed to collect big data, it still needs to be developed and revised in order to accommodate big data collected from power grid.

Data processing: Fault detection and oscillation detection are urgent applications for a smart grid for which the reaction time should be minimized. Even in the cloud system which is able to provide fast computation service, a complicated algorithm with a the massive amount of data is not so easy to process.

Data compression: Data compression technique is not applicable in Wide Area Monitoring

System (WAMS). In order to achieve required high compression ratio to detect disturbances, some special compression methods are also needed.

Data visualization: Visual graphs and charts are much more effective to warn operators with abnormal values of voltage or frequency. Finding and representing the correlations or trends between multi-source data may become a big challenge. Other challenges lie in visualization algorithms, information extraction & presentation and image synthesis technology [24].

Data security: Smart grids are interconnected in a SCADA system but cyber-attack prevention is not in the considerations of a SCADA system design. A power system could be threatened by meta-data spoofing, wrapping and phishing attacks. Interoperation through APIs prevent grid versus these risks [25].

Data privacy: On the customer side, risk is the increase in private data as a result of smart meters data collection for measurements of household energy consuming. Private data leakage would be a disaster [26].

4. Conclusion

In this article, big data applications in smart grid which is the intelligent center members of electrical distribution systems have been reviewed and discussed. The large scale data which contains valuable information are collected from smart meters installed in the power systems, electricity market, meteorological system and other systems related to electrical distribution system. Given applications in this paper processes big data to solve physical problems, eliminate fault risks, forecast future realizations, compansate external parameters that affect transmission and demand oriented to every state of distribution systems like operation, maintenance, supply/demand projection and protection.

Advanced information and communication technology and developed big data processing techniques are used in concerned systems. Desired features are extracted from raw information and they are used as a feedback critical information again for tha system. This process goes on to obtain an effective, cost reduced enhanced system.

Despite developed technological abilities of information technology, performing fast and efficient data analytics on a huge data is still a challenging bussiness. Sometimes it takes much effort to provide cyber security and to satisfy privacy principles. In such a multiparameter systems there would be difficulties for identifying consumers' behaviour analysis. Data analytics in smart grids is a complicated area of work. It involves information tecnologies, artificial intelligence and mathematics, electrical engineering, meteorological science and some others. So, multidisciplinary teamwork is needed to design and implement big data based electrical distribution systems of future.

References

[1] Zhang, Y., Huang, T., Bompard, E. F., "Big data analytics in smart grids a review", Energy Informatics, pp. 1-8, 2018

[2] Abhisek, U., Zivanovic, R., "Automated Analysis of Power Systems Disturbance Records: Smart Grid Big Data Perspective", IEEE Innovative Smart Grid Technologies, 2015.

[3] Sagiroglu, S., Terzi, R., Canbay, Y., Colak, I., "Big Data Issues in Smart Grid Systems", IEEE International Conference on Renewable Energy Research and Applications (ICRERA), 2016.

[4] Günther, W. A., Mehrizi, M. H. R., Huysman, M., Feldberg, F., "Debating big data: a literature review on realizing value from big data", Journal of Strategic Information Systems, 26:191–209, 2017.

[5] Yu, N., Shah, S., Johnson, R., Sherick, R., Hong, M., Loparo, K., "Big Data Analytics in Power Distribution Systems" IEEE Power & Energy Society Innovative Smart Grid Technologies Conference, 2016.

[6] Lai, C. S., Lai, L. L., "Application of Big Data in Smart Grid", IEEE International Conference on Systems, Man, and Cybernetics, 2015.

[7] "Big data in the cloud: Converging technologies", IT Center, Intel, 2015.

[8] http://en.wikipedia.org/wiki/Quantum_computing, 2015.

[9] https://www.smartgrid.gov/the_smart_grid/smart_gri
d.html, 2019.

[10] Feng, G., "Frameworks for Big Data Integration, Warehousing, and Analytics", chapter 4 in "Big Data Application in Power Systems", edited by Arghandeh, R. and Zhou, Y., Elsevier, 2018.

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

[12] Yu, N., Shah, S., Johnson, R., Sherick, R., Hong, M. & Loparo, K., "Big Data Analytics in Power Distribution Systems", IEEE Power & Energy Society Innovative Smart Grid Technologies Conference, 2016.

[13] Peppanen, J., Reno, M. J., Broderick, R. J., Grijalva, S., "Distribution System Model Calibration With Big Data From AMI and PV Inverters", IEEE Transactions on Smart Grid, Vol 7, No 5, 2016.

[14] Nasiakou, A., Alamaniotis, M., Tsoukala, L. H., "Power distribution network partitioning in big data environment using k-means and fuzzy logic", Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion, 2016.

[15] Lydia, E. L., Kumar, B. P., Ramya, D., "Generation of dynamic energy management using data mining techniques basing on big data analytics issues in smart grids", International Journal of Engineering & Technology, pp. 85-89, 2018.

[16] Flick, D., Kuschicke, F., Schweikert, M., Thiele, T., Herrmann, C., "Ascertainment of Energy Consumption Information in the Age of Industrial Big Data", Procedia CIRP, V.72, pp. 202-208, 2018.

[17] Mujeeb, S., Javaid, N., Khalid, R., Nazeer, O., Sha, I. & Khan, M., "Big Data Analytics for Price and Load Forecasting in Smart Grids", 13th International Conference on Broad-

Band Wireless Computing, Communication and Applications (BWCCA), 2018.

[18] Ukil, A., Zivanovic, R., “Automated Analysis of Power Systems Disturbance Records: Smart Grid Big Data Perspective”, IEEE Innovative Smart Grid Technologies, 2015.

[19] Park, E., “Positive or negative? Public perceptions of nuclear energy in South Korea: Evidence from Big Data”, Nuclear Engineering and Technology, V.51, pp. 626-630, 2019.

[20] Chen, J., Chen, H. Y., Wen, M. & Sun, S., “Industrial power demand forecasting based on big data technology orienting to energy internet: A case study of Hunan Province”, IOP Conference Series: Earth and Environmental Science 188, 2018.

[21] Batra, N., Singh, A., Whitehouse, K., “Neighbourhood NILM: A Big-data Approach to Household Energy Disaggregation”, Computer Science Machine Learning, 2015.

[22] Wu, W., Lin, W., Hsu, C. H., He, L., “Energy-efficient hadoop for big data analytics and computing: A systematic review and research insights”, Future Generation Computer Systems, cilt 86, pp. 1351-1367, 2018.

[23] Tu, C., He, Xi., Shuai, Z., Jiang, F., “Big data issues in smart grid – A review”, Renewable and Sustainable Energy Reviews, V.79, pp. 1099-1107, 2017.

[24] Kennedy, S., J., “Transforming big data into knowledge: experimental techniques in dynamic visualization”, Massachusetts Institute of Technology, 2012.

[25] Cuzzocrea, A., “Privacy and security of big data: current challenges and future research perspectives”, Proceedings of the international workshop on privacy & security of big data, p. 45–7, 2014.

[26] Bertino, E., “Big data – security and privacy”, Proceedings of the IEEE international congress on big data, 2015.