Renewable and Sustainable Energy Reviews 79 (2017) 1099–1107

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

# Big data issues in smart grid – A review

Chunming Tu[a,b], Xi He[a,b,*], Zhikang Shuai[a,b], Fei Jiang[a,b]

[a] College of Electrical and Information Engineering, Hunan University, Changsha 401182, China
[b] National Electric Power Conversion and Control Engineering Technology Research Center, Hunan University, Changsha 410082, China

ARTICLE INFO

*Keywords:*
Big data
Smart grid
Big data analytical applications
Cloud platform
Data mining

ABSTRACT

There are both economic and environmental urges for transition from the current outdated power grid to a sensor-embedded smart grid that monitors system stability, integrates distributed energy and schedules energy consumption for household users. Especially with the proliferation of intelligent measurement devices, exponential growth of data empowers this transition and brings new tools for the development of different applications in power system. Under this context, this paper presents a holistically overview on the state-of-the-art of big data technology in smart grid integration. First, the features of smart grid and the multisource of energy data are discussed. Then, this paper comprehensive summarizes the applications leveraged by big data in smart grid, which also contains some brand new applications with the latest big data technologies. Furthermore, some mainstream platforms and knowledge extraction techniques are looked to promote the big data insights. Finally, challenges and opportunities are pointed out in this paper as well.

## 1. Introduction

Big data gains ever more attentions and is regarded as the intellectual “petroleum” that promotes modern socio-economic development. For information science, the big data is usually defined as a huge and complex data set, which is difficult for traditional tools to store, process and analyze [1]. In terms of energy, a revolutionary change is that the traditional unidirectional electricity grid has been gradually replaced by smart grid (SG), which is also called as ‘the next-generation power grid’. There are several trial projects of SG, such as the ENEL Telegestore project in Italy, which is regarded as the first attempt for SG construction in field [2]. Following that, several other SG projects have been implemented, including the Hydro One project in Canada [3], the InovGrid project in Portugal [4] and the Modellstadt Mannheim (Moma) project in Germany [5]. Compare to the conventional grid, the advantages of the SG are manifold, such as self-healing & recovering functionality, better incorporation with renewable energies, situational awareness and transient stability, etc., thanks to the deployment of smart meter devices and big data analytics, i.e. data mining or machine learning [6].

### 1.1. Sources of energy data in smart grid

The big data in smart grid is generated from various sources. Supervisory Control and Data Acquisition (SCADA) analog system, which with one sample per 2–4 s sampling rate, has been implemented in power grid for several decades. Due to the limited sampling rate, it is impossible to observe the transient stability and oscillation of the power system. Correspondingly, the Phasor Measurement Units (PMUs), which has much faster scan rate (30–60 samples per second), is able to produce direct time-stamped voltage/current magnitudes as well as the phase angles [7,8]. By the end of 2015, the total installed PMUs had reached more than 1380 granted by the American Recovery and Reinvestment Act (ARRA), covering nearly 100% of the US transmission system. In China, the State Grid and Southern Grid had installed 1717 PMUs by the end of 2013 [9,10]. In addition to the PMU, the advanced meter read (AMR) with 15-min read intervals has also been deployed to replace the traditional once a month reading meters. As a result, for each meter, it reads 96 data per day and carries out 2880 data per month, which means that 2880 times of kWh data has been increased just in the power metering field. The proliferation of PMUs, AMR and other advanced measurement devices such as Intelligent Electronic Devices (IEDs), Digital Fault Recorder (DFR), Sequence of Event Recorder (SER), etc. [123,124], have brought enormous data volume in power system for storage, curation, mining, sharing and visualization. As pointed out in [11], the numbers of global smart meter installation will triple from 10.3 million in 2011 to 29.9 million by the end of 2017. It shows that 100 PMUs with 20 measurements generate over 100 GB data per day at the rate of 60 Hz sampling rate [12].

* Corresponding author at: College of Electrical and Information Engineering, Hunan University, Changsha 401182, China.
*E-mail address:* forevermau@163.com (X. He).

http://dx.doi.org/10.1016/j.rser.2017.05.134
Received 11 July 2016; Received in revised form 20 March 2017; Accepted 19 May 2017
1364-0321/ 

### 1.2. Benefits from big data analytics in smart grid

It is known that, the big data has brought numerous tangible benefits to the utilities and electricity users, which can be systemically concluded as follows:

1) *Increasing System Stability & Reliability*: Safety is always placed in the first place in the priority of power grid and can be encompassed two major aspects: stability and reliability, which can be further sub-divided into some sub-aspects such as oscillation detection, voltage stability, event detection & restoration, islanding detection & restoration, post-event analysis, etc. [131]. Many of the above issues have been studied and used for decades. It is worthy pointing that, with the emergence of the big data and advanced data analytic technology, it is possible to explore some new capabilities or to improve the outdated monitoring and detection methods. For instance, as reported in [13], an oscillation in a wind farm was detected through PMU but unobservable by SCADA, this is a typical benefit from the big data analytics in the safety of smart grid.
2) *Increasing Asset Utilization & Efficiency*: In practice, the big data analytics can increase asset utilization and efficiency, especially in better understanding of the operating characteristics and physical limits of the assets, better validating and calibrating of the models, and better integrating the renewable resources with big data tools. For instance, in [14], the authors use the voltages measured by smart meters and Geo - Information system (GIS) data to conduct transformer fatigue analysis, which aware the operators to overhaul or change the transformer in advance. In addition, many works have also been carried out to investigate the usage of big data for model validation and calibration in [15–19].
3) *Better Customer Experience & Satisfaction*: Recently, significant progresses have been made for deploying smart meters at homes, offices or other premises worldwide [20,21]. The mass rollout enables easier billing, fraud detection, forewarning of blackouts, smart real-time pricing schemes, demand response and efficient energy utilization. However, all the above applications need high sampling rate by the meters and advanced data analytics, as well as the information communication technologies.

### 1.3. Main contributions

Up to this moment, the smart grid and the big data are usually reported separately [22–25] and [26,27] [125,126]. However, the analysis of the big data in smart grid is rarely reported. To the best of our knowledge, K Zhou et al. reviewed the big data driven smart energy management in [28], which mainly illustrates the architecture and industrial applied energy management tools. BA Schuelke-Leech et al. reviewed the opportunities of the big data in electric utilities [29]. However, this work cannot be extended to smart grid. Hu and Vasilakos presented a comprehensive review of the big data application in energy taxonomy and security [30]. In order to get the full utilization of the big data in smart grid, this paper presents a holistically review of the big data issues in smart grid.

In general, the contributions of this paper are manifold and can be summarized as follows:

- It provides a first comprehensive survey covering both smart grid and energy big data analytics. To the best of our knowledge, this is the first attempt to systematically look into the big data issues in smart grid.
- It points out some the latest applications or methods in power system empowered by big data technology. For instance, the data-based approach for fault detection and classification in Micro-grid, which results in a much better performance than the conventional model-based approach [31]; the measurement-based estimation of the power flow Jacobian matrix which is proved to be more efficient for inferring the pertinent information about the system topology in near real-time [32].
- Both implemented data-driven applications by utilities and practical used data analytic methods worldwide have been listed in this paper.

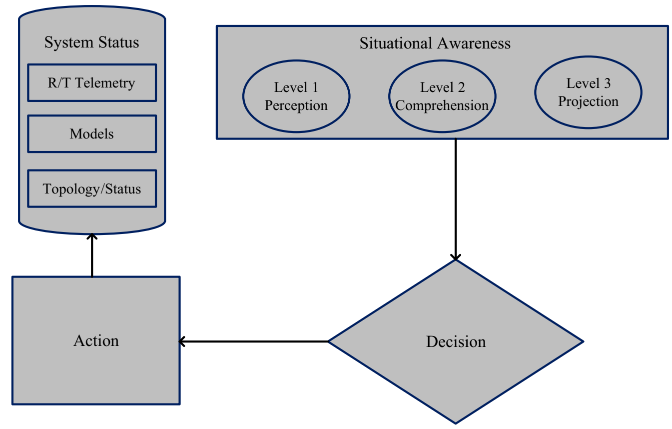


**Fig. 1.** WASA in the decision-making process [33].

The rest of this paper is organized as follows: In Section 2, the mainstream applications empowered by big data have been discussed. In Section 3, the big data platforms and data mining methods are presented and compared. We argue some future research directions and challenges in Section 4. Section 5 is devoted to the conclusions.

## 2. Big data applications in smart grid

### 2.1. Wide area situational awareness

The concept of "situational awareness" (SA) first appears in aviation industry. Fig. 1 shows the process flow of SA in a general system. It could be seen that a complete SA process has three steps, i.e. perception, comprehension and projection. The first step is to perceive heterogeneous data which may come from the conventional SCADA system or the newly installed IEDs and PMUs. The second step is to comprehend what the perceived data mean in relation to system oscillation or instability. This step is actually very resource-demanding and requires good knowledge of information extraction. The last step projection means the understanding of future behavior of the system through the above two steps. By constant and proper projection, the control room operators have enough time to respond to the events, which contributes to the prevention of cascading catastrophe.

In the real-world scenario of wide area situational awareness application, there are two issues need to be solved: the limitation of the installed PMUs and the latency brought by the decision making algorithms.

Due to the expensive cost and complicated factors of deploying PMUs in the grid, the number of the synchrophasor sensors is limited and need to be optimally placed. Many optimal PMU placement (OPP) methods are raised, such as mixed-integer programming [34], model-based OPP [35], zero-injection bused for further reducing [36], generic algorithm [37], etc. [38–40]. Sodhi et al. [41] proposed an improved OPP framework using five applications viz., improving state estimation, assessing voltage /angular stability, monitoring tie-line oscillations and the availability of communication infrastructure, to assess the potential PMU sites.

For transient fault, the reaction time is normally within 100 ms with which automatic protection devices take action without human decision; while for long term stability, control room operators have enough time to learn the situation by simulations or experiences. However, for the situation between those two scenarios, the operator's

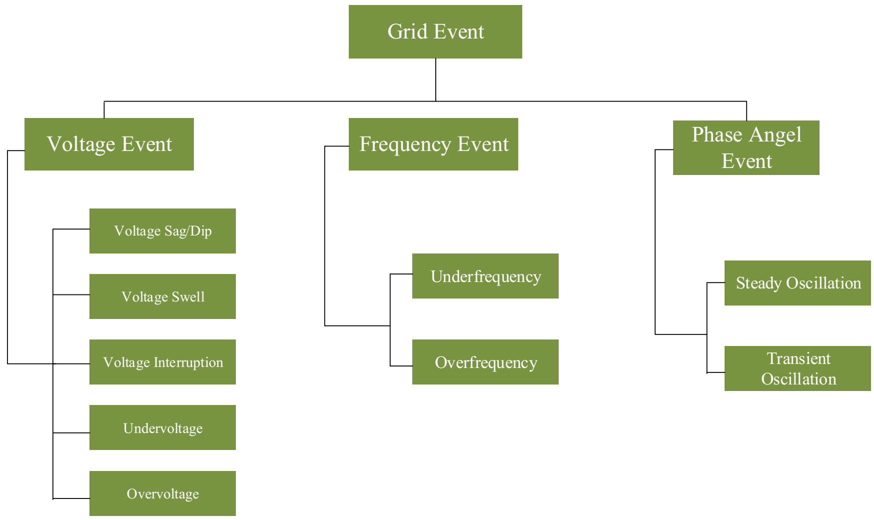


**Fig. 2.** Hierarchical event classification.

decision in a relatively short time is essential. Despite that the batching processing artificial intelligence (AI) could aid the decision making process, the latency suffered by mathematical calculations is intolerable. Decision tree looks promising in dealing with moderate data process, while steam mining is fit for big data process for decision making. A decision tree built from data stream using Hoeffding bound was proposed by Domingos and Hulten [42]. A main tree classifier and a cache-based classifier which can handle high-speed data streams are used to facilitate the intelligent decision making. The stream mining techniques require no model information but achieve on-line SA with reasonable accuracy, processing time and computational resources [43–45].

There are some examples of WASA application. The situational system SMDA (ver5.0) was used for wide-area monitoring and event detection in Hydro-Quebec [46]. NYISO used the real-time and off-line data to display the information on the dashboard which alerts operators of anomalies including voltage drop, transient oscillation, line tripping [7]. Peppanen et al. [47] developed a distribution system state estimation (DSSE) and situational awareness system to monitor Georgia Tech campus distribution system and deployed the 3-D graphical user interface to enhance situational awareness. The data collected by Oklahoma Gas & Electric via PMUs to conduct WASA in Oklahoma and western Arkansas [9].

### 2.2. State estimation

Since initial introduction and pioneering work by Schweppe and Wildes [48] [127,128], power system state estimation (PSSE) has become an essential part in power system automation. Conventional state estimation problem is solved iteratively due to the nonlinear measurements by SCADA system. This is quite inefficient and bad data intolerable. Motivated by the advent of big data and smart grid, new algorithms and techniques have been raised and deployed. A decoupled linear estimation method using the time sequence data measured by PMUs is proposed in [129], which decouples the problem into two independent smaller problems for accelerating the estimation process. A PMU based robust state estimation method (PRSEM) [49] that adapts weight assignment function to eliminate the unwanted disturbed data is developed in order to increase algorithm robustness. Actually there are two major problems in conducting PSSE, one is the bad data filtering and the other is dimensionality reduction of big data.

In the real field, there are many causes for “bad data” (BD), such as metering devices failure and electromagnetic interferences. The state-of-the-art techniques for BD detection can be classified into two categories: pre-estimation and post estimation [50].

The pre-estimation approach uses normalized residuals test and re-estimate the state, in which case the BD is part of the iterative process. However BD post estimation is more reliable, faster and non-iterative compared to BD pre-estimation, which makes it suitable for bad data detection in PSSE applications.

Dimensionality reduction of big data in smart grid is essential for state estimation. Principal component analysis (PCA) is one of the most used method for dimension reduction for its good performance in preserving original data, as well as its fast computation feature [51].

There are several SE projects implemented in US [52]. ATC performed a high-level validation of its state estimator model using PMU voltage angle data. Duke Energy used energy data to improve state estimation, system model accuracy, and the time required for post-event analysis. ISO-NE conducted a feasibility demonstration of a data-based state estimation for the 345 kV network and evaluating a hybrid PMU/SCADA state estimator.

### 2.3. Event classification & detection

Disturbance in power system involves many types, such as faults, line trip, load shedding, generation loss, oscillation, etc. Conventional event detection is model/topology based post-event analysis. In smart grid, the big amount of data and information make it feasible to use data-driven approach for real-time event classification and detection.

Event classification/categorization is the preparation for event detection and location. Fig. 2 shows a hierarchical approach for disturbance event classification. For example, the event “Voltage sag/dip” means voltage reduction of more than 10% (up to 30%) of nominal voltage for more than 8 millisecond (up to 1 min). This is the functional specification for a certain disturbance event and all the voltage and frequency events could be categorized either oscillatory or non-oscillatory. However, it is worth noting that this hierarchical categorization only considers the most frequently happened events in power system. A comprehensive unsupervised clustering method (hierarchical, partitioning and density-based approach) is proposed to classify 2226 disturbances stored in Public Service Company of New Mexico (PNM) from 2007 to 2010 [53]. Chen et al. [54] propose a scatter-plot-based event classification (SPEC) algorithm to conduct the classification. The dots scatter outside the core subspace indicate a disturbance and the topological shape decides the types of the events.

Current PMU standard IEEE C37.118 does not define the transient or dynamic events which is vital for power system stability [55,56]. Detrended fluctuation analysis approach is suitable for dynamic events detection using the big amount of energy data. A refined parallel detrended fluctuation analysis (PDFA) algorithm is proposed for fast detection with parallel MapReduce architecture [57]. Neural networks and fuzzy logic are exploited for transient event detection in smart grid due to its inter-neuron connection strength and synaptic weights. Supervised learning and unsupervised learning are two ways for conducting neural networks. A combined supervised/unsupervised scheme is proposed and has proven to be more efficient and capable than either single approach [58]. For multiple event recognition and detection, the successive disturbances might be overshadowed and overlooked. Event unmixing conception and nonnegative sparse event unmixing (NSEU) algorithm have been proved as an effective way for multiple disturbance events separation and detection [59].

The transparent data mining algorithm Decision Tree (DT) with a Radom Forest (RF) based black box model is proposed to detect and classify faults in Micro-grid. A large data set of 3860 samples is established to train the DT and results in 97% fault detection accuracy compared to 56% by the existing over current relay method [31].

There are nine projects relate to event detection, management & restoration under ARRA [7]. Western Electric Coordinating Council (WECC) used correlation algorithm leveraged by big data for event detection and decision support. ATC uses PMU data to provide a check on the operation of protection devices, determining if faults are cleared properly and if all three phases open at the same time.

### 2.4. Other applications

#### 2.4.1. Power plant models validation and calibration

Power plant models validation and calibration perplexed the utilities and operators for a long time. The performance could only be conducted by staged testing and requires the plants to be shut down, costing utilities approximately US$15,000 to US$35,000 every time [15]. Leveraged by the big measurement data metered by the PMUs, IEDs, FDRs, new data-driven approach has been developed to verify the power plant models. The measured disturbance recording could be used to compare with the simulations, thus supplement the baseline test to adjust the models.

#### 2.4.2. Short-term load forecasting

Big data based short-term load forecasting methods have been proposed in recent years [60–62]. The core technique of this method is to classify load patterns using association and clustering analysis based on the smart metering data in addition to historical load data and ambient data, such as temperature, humidity, and rainfall data. Due to the influx of load forecasting models, the conventional "abstract metrics", such as mean absolute error (MAE) and root mean square error (RMSE), are not accurate enough to evaluate the residual errors between predicted values and actual values. With the fine spatial and temporal granular data, more sophisticated techniques such as regression tree learning [63] and artificial neural networks [64] can be used to solve this evaluation problem.

#### 2.4.3. Distribution network verification

Because of the low accuracy GIS inputs data, it is important to verify the network connectivity regularly. Big data analysis assists in verifying the distribution network topology in smart grid, especially for the underground feeders which are difficult to check [16,64]. This is a typical correlation statistical algorithm use-case leveraged by power big data. Similar applications including secondary modeling [14,65], transformer identification [66], and electricity theft [130] are developed based on the same algorithm.

#### 2.4.4. Big data driven demand response

Demand response management is an effective way to reduce the load burden during peak hour, however the conventional approaches are to cut off the predetermined loads, which result in inflexibility. In [67], 66,434,179 load profile data per 24 h, which is collected by more than 200,000 smart meters in Pacific Gas and Electric Company (PG & E), is utilized to conduct the customer target for Demand Response (DR). To solve the stochastic knapsack problem (SKP) for optimal customer selection, both the efficient heuristic and greedy algorithm are applied.

#### 2.4.5. Parameter estimation for distribution system

Generally, automated parameter estimation (PE) is only suitable for the transmission system other than the distribution power system due to its intricate radial topology and lack of measurements. However, with the massive implementations of sensors in smart grid, new parameter estimation methods of the distribution system secondary network are proposed. The big data from advanced metering infrastructure (AMI) and other sensors offer the possibility to the secondary system to realize the line impedances calibration. For instance, Peppanen et al. utilized the data collected by AMI and some PV measurements to improve the calibration accuracy of the Georgia Tech campus distribution secondary circuit parameters [68].

#### 2.4.6. System security and protection

Cyber-attack is regarded as one of the biggest threat in smart grid scheme due to the interconnection and interoperation of the components and networks. Intrusion detection system (IDS) as a countermeasure traditionally is a knowledge-intensive host-based system, thus it has its limitation in term of scalability and flexibility. It might be a good thought to take multiple data source into account to build a specification-based hybrid IDS for comprehensive system monitoring and protection [69]. Compared to the conventional cryptography in cyber security community, real-time and tight cyber-physical couplings bring the difference to the security concept of smart grid. In [30], it is pointed out that smart grid is vulnerable to cyber-attacks due to the openness and autonomy feature. Although many security solutions have been made for smart grid, most of them are not based on big data. Currently, there are three typical achievements of big data security and privacy: (i) big data oriented cryptosystems, (ii) big data oriented anomaly detection, and (iii) big data oriented intelligent applications.

For sure, there are other applications brought or improved by big data analytics, such as islanding detection [70], oscillation detection [71], real-time rotor angle monitoring [72], et al. Table 1 lists some practical implemented applications empowered by big data in smart grid worldwide.

## 3. Techniques used for big data analysis in smart grid

### 3.1. The platforms for big data analysis

In order to realize the functionalities illustrated above, it is essential to find a suitable platform or model that accommodates these applications effectively and efficiently. Cloud storage and computing, a newly emerged technology, gains lots of attentions and earns huge popularity due to its advantages in many aspects. Chang et al. [79] compared the cloud and non-cloud big data in the term of data storage, and the result shows that on cloud basis the actual execution time is lower, while consistency and efficiency are higher compared to the non-cloud basis.

Fig. 3 depicts a ubiquitous model of data cloud model for smart grid. Heterogeneous data are metered by sensors and then transmitted to the data cloud sub-centers via application programming interface (API) and transmission network. All the data stored in the cloud can be queried by the legitimate actors (e.g. control room operators, 3rd party service providers) for specific applications. The merits of this schema are manifold, such as cost sharing, interoperability and extensibility, high efficiency of parallel computation, distributed management, as well as high fault-tolerance and data security.

In fact, there are many industrial cloud platforms that have been used for smart grid big data. Microsoft Azure is a flexible and interoperable platform hosting cloud based applications and conducting data processing as well [80]. Holm [81], a private household energy

**Table 1**
Smart grid applications empowered by big data.

| Applications | Software name | Developer | Description | Refs |
|---|---|---|---|---|
| Situational awareness system | FNET/GridEye | Yilu Liu( Leader) | A variety of applications were developed, including real-time event detection and location estimation, oscillation detection etc. | [73,74,93] |
| Wide area situational awareness | SMDA (ver5.0) | Hydro-Quebec | Collects the wide-area phasor data in real time and monitors inter-area oscillations, covering about 25% of the 735 kV substation. | [46,75] |
| Event detection & Alarm Management | e-terra3.0 | Alstom | Present and visualize disturbances and navigate to the relevant diagnostic information | [76] |
| Power plant models validation | CERTS | BPA & CERTS | BPA engineers calibrate the Columbia Generating Station (CGS) model without off-line the generators | [15] |
| Oscillation Detection and Mitigation | GRID-3P platform | Electric Power Group | Oscillations are not observable using SCADA technology but are observed by fine granular PMU data. | [7] |
| Renewable Resource Integration | DEMS | Siemens | A data-driven system for monitoring, managing and integrating distributed generation and renewable energy into the bulk power system | [77] |
| Transient stability and Intrude Detection | WARMAP5000 | NARI Technology | Combine real-time monitoring data and simulated data for wide-area transient stability control and cyber-attack prevention | [78] |

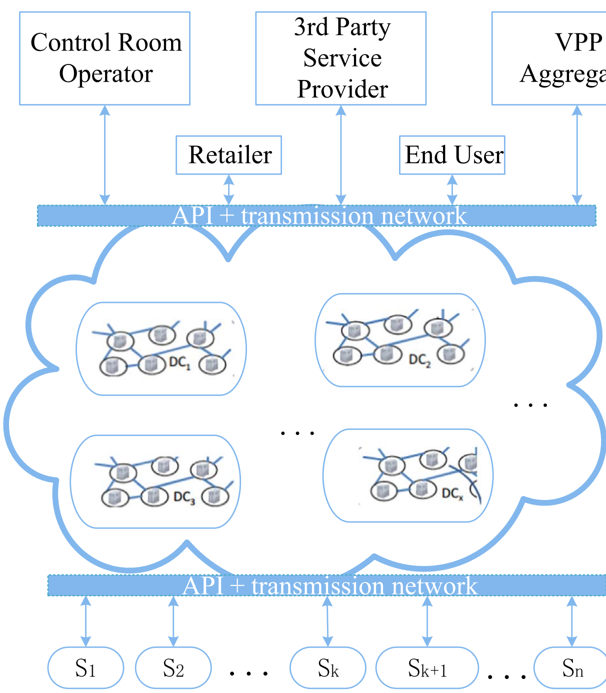

**Fig. 3.** Cloud architecture for smart grid big data.

management tool, is the first application in this category that based on the cloud platform Azure. Google PowerMerer [82], an application to track the household energy consumption is also based on the cloud platform Google APP Engine. InterPSS [82], short for An Internet Technology-based Open-source Power System Simulation, aims at developing a cloud data based simulation platform for newly applications in smart grid. In [83], Smart-Frame, a cloud computing based big data information management framework designed for smart grid, is proposed. The framework consists of three hierarchical levels: top level, regional level, and end-user level. This framework is utilized in a prototype on Eucalyptus, which is a popular open source cloud computing based platform. Other platforms or cloud computing modules are studied or have been applied including MapReduce, Chord, Dynamo, Zookeeper, Chubby, etc. [84–87].

In the following part, the two most frequently used platforms in smart grid nowadays will be discussed.

#### 3.1.1. Hadoop MapReduce platform

MapReduce, initially developed by Google in 2004, is the most prevalent programming model for large scale data processing. It has several implementations such as Hadoop, Mars, Phoenix, Dryad and Sector/Sphere. Hadoop is regarded as the most promising implementation due to its scalability, fault-tolerance and automatically failure restoration techniques. Firstly developed by Doug Cutting and Mike Cafarella in 2005 [88], it is widely used among big technology giants such as Google, Yahoo, Facebook, YouTube, IBM and Microsoft. Fig. 4 depicts the architecture of Hadoop MapReduce framework. There are two major parts in this framework: MapReduce and Hadoop Distributed File System (HDFS).

MapReduce, initially developed by Google in 2004, is the most prevalent programming model for large scale data processing. It has several implementations such as Hadoop, Mars, Phoenix, Dryad and Sector/Sphere. Hadoop is regarded as the most promising implementation due to its scalability, fault-tolerance and automatically failure restoration techniques. Firstly developed by Doug Cutting and Mike Cafarella in 2005 [88], it is widely used among big technology giants such as Google, Yahoo, Facebook, YouTube, IBM and Microsoft. Fig. 4 depicts the architecture of Hadoop MapReduce framework. There are two major parts in this framework: MapReduce and Hadoop Distributed File System (HDFS).

Due to the unique characteristic of energy big data, the information science originated platform needs to be modified and to be used in power system. In [89], Zhang et al. proposed a new incremental MapReduce and named it as $i^2$MapReduce to perform key-value pair level incremental processing and to support more sophisticated iterative computation in power system. In [90], Xing et al. proposed a platform Petuum for the wide range machine learning. The efficacy of this general-purpose framework was verified in the experiment by comparing with MapReduce as well.

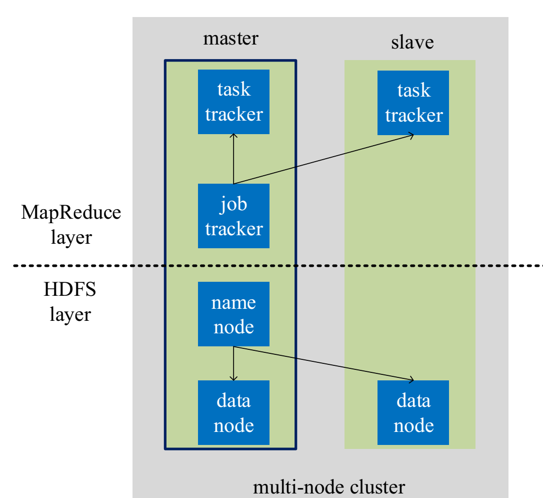

**Fig. 4.** Architecture of Hadoop MapReduce.

**Table 2**
Data analysis algorithms applied for the applications in smart grid.

| Algorithms | Applications | Refs |
|---|---|---|
| Principal Component Analysis (PCA) | Dimensionality reduction | [51,102,103] |
| Artificial Neural Networks (ANNs), K-means, Fuzzy c-means, eXtended Classifier System for clustering (XCSc) | Load classification | [104–107] |
| ANNs, Empirical Mode Decomposition, Extended Kalman Filter, Extreme Learning with Kernel, Decision Tree | Short-term load forecasting(STLF) | [108–111] |
| Statistical Relational Learning (SRL) | Knowledge Graphs | [112,113] |
| Random Matrix Theory | Anomaly detections | [114,115] |
| Deep Neural Networks, Multi-view Learning, Matrix Factorization | Cross-Domain Data Fusion | [33,116,117] |
| Lyapunov Exponents | Rotor Angle Monitoring | [72] |
| Playback Process, Sensitivity Analysis | Generating Unit Model Validation | [18,118] |
| Decision Tree (DT) & Radom Forest (RF) | Fault Detection & Classification | [31] |
| Additive Quantile Regression | Individual Smart Meter Data Estimation | [100] |
| K-means Clustering and Principal Component Analysis | Estimation Invisible Solar Sites Power Generation | [98,99] |

#### *3.1.2. Apache spark*

There are mainly three ways to tackle the big data processing, i.e. batch, stream and iterative. In [91], it is pointed out that, Hadoop MapReduce is suitable for the analysis of empirical and static data, but not suitable for real-time and stream data analysis. Compared to Hadoop, Apache spark is able to process on-line and streaming data. Apache Spark is an open source framework for big data computing, initially developed at the Berkeley's AMPLab, University of California. Spark runs programs up to 100x faster than Hadoop MapReduce in memory, or 10x faster on disk [92].

The spark framework powers a stack of libraries including structured query language (SQL) & data frames, MLlib for machine learning, GraphX, and Spark streaming. The North American Power Grid Frequency Monitoring Network (FNET/GridEye) [93], which is led by Yilu Liu, is a typical example of the spark implementation in smart grid. This system deploys 150 frequency disturbance recorders in United States and about 50 ones in the world, whose architecture includes the openPDC for real-time applications, the distributed analytics cluster for near-real-time apps and the Apache Spark for post event & statistical analysis. Thanks to the high-speed, widely monitored data and distributed data analytics platform, some power system events, such as the generator shedding at James A. Fitzpatrick power plant during 2012 Hurricane Sandy [93], can be detected promptly.

### *3.2. Data analytics for energy big data*

Collecting the data alone is useless, it is the information extracted from the big dataset that matters. Data mining which identifies insights and patterns in data is considered as the one of the most useful techniques for knowledge extraction. Data mining is not new in power system, however the technique used over the past few decades is primarily based on SQL database or even spreadsheets statistics. In the context of smart grid, it is truly demanding for new efficient and effective algorithms and tools to deal with large influx of data.

The initial data mining methods are pretty primitive, such as static knowledge and single-source mining methods [94]. They are not suitable for smart grid scenario which encompasses massive heterogeneous and stream data. Multi-source mining mechanism and dynamic data mining methods are put forward to solve this problem. A local pattern analysis approach was firstly proposed by Wu et al. [95], laying a foundation for multi-source mining mechanism and paving the way for the divide-and-conquer method in data mining territory. The conventional centralized approach for data processing and analysis is inefficient, resource-demanding and costly. An alternative way is distributed computation and it has been applied in many fields such as Geo, climate and environment analysis, Human Genome Project, Dark Energy Survey (DES) project etc. [96]. It is well grounded that distributed data analysis is a good option for massive multi-source dataset mining in smart grid. There are already some literatures looking into this edge data processing method [93,97].

With the drastically increasing computation ability and lowering hardware cost, some new information extraction methods have been proposed and developed. Machine learning is among one of them. There are nine most frequently used machine learning algorithms, including k-means, linear support vector machines (LSVM), logistic regression (LR), locally weighted linear regression (LWLR), Gaussian discriminant analysis (GDA), back-propagation neural network (BPNN), expectation maximization (EM), naive Bayes (NB), and the independent variable analysis (IVA). Each of the algorithms has its own characteristic and can be used in different scenarios. In [98,99], a method, which combines hybrid k-means clustering with principal component analysis, was utilized to carry out the data dimension reduction and estimation mapping to estimate the power generation of invisible solar sites. Besides, 405 PV-sites data across California was employed to prove the concept [99]. In [100], the authors proposed an additive quantile regression to conduct the probabilistic forecasting for the disaggregated smart metering. It is based on 3639 households meters installed by Commission for Energy Regulation (CER) in Ireland, which is also a novel attempt for individual smart meter data estimation other than the traditional aggregated load forecast.

Shallow learning models (e.g. k-means, LSVM, LR) have been proven to be useful in the terms of well-constrained simple problems optimization. However, for complicated application cases, such as natural language problem, it yields results far from satisfaction. Instead, a refined and much more complicated approach is proposed, i.e. deep learning [101]. Other than shallow learning, deep learning exploits multi-layers of the hierarchical structure and it can be divided into supervised deep learning and unsupervised deep learning. Table 2 lists some proposed or practical implemented algorithms applied for the applications in smart grid.

## 4. Challenges and future prospects

The concept of big data is not new, it can be used to create transparency, reveal demands, and replace manual decision making and so on. However, big data technology applied in power system is currently in its infancy stage and there is a long way to go. We point out some challenges lying ahead in the terms of smart grid big data technology.

(1) ***Multisource data integration and storage.*** Traditional data analysis usually deals with data from single domain, it is essential to find a fusion method for multisource dataset, which has different modalities, formats, and representations. In the terms of big data storage, although some of the system such as Hadoop distribute file system (HDFS) seems to be feasible, it still needs to be tailored and modified in order to accommodate power grid big

data.

(2) ***Real-time data processing technology.*** For some urgent applications such as fault detection and transient oscillation detection, the reaction time scale is milliseconds. Although the cloud system is able to provide fast computation service, the network congestion, complicated algorithm, combined with the massive amount of data still result in a latency. Database based on memory seems to be a feasible way to tackle this problem, and the memory based database HANA developed by SAP was used to deal with massive kilo-watt meter data in order to better distribute power flow [119].

(3) ***Data compression.*** Data compression technique is indispensable in Wide Area Monitoring System (WAMS). It should have its own characteristics to meet the high fidelity requirements. Besides, in order to detect the transient disturbance while achieving high compression ratio (CR), some special compression methods are also needed.

(4) ***Big data visualization technology.*** Visualized graphs and charts can present operators with granular and explicit changes of the voltage and frequency. However, how to effectively find and represent the correlations or trends between multi-source data is a big challenge. Other challenges lie in visualization algorithms, information extraction & presentation and image synthesis technology [120].

(5) ***Data privacy and security.*** Legacy SCADA system will coexist with newly AMI and IT systems in the foreseeable future. Cyber-attack prevention is not in the consideration of SCADA system design. The legacy systems and interoperation through APIs expose the grid to dangerous scenes, such as meta-data spoofing, wrapping and phishing attacks [121]. On the customer side, ever-increasing smart meters for household energy consuming create ever-increasing individual information [122]. As data are shared between different entities, private data leakage could be a disaster and leads to cascading troubles.

## 5. Conclusion

Big data technology is regarded as the major factor for smart grid construction. In this paper, the big data issues of smart grid have been surveyed from the following aspects: the energy big data sources; the benefits of data-based approaches in smart grid; the theoretical and practical implemented applications empowered by big data; and the current platforms and techniques for big data analysis.

In smart grid, the latest deployed smart meters, such as PMUs, AMR, DFR etc., and the legacy power grid field devices together constitute the big data scenario for utilities. Actually, this construction form not only brings multifold advantages, but also means many challenges. In this paper, theory and practical implemented applications in power grid, which empowered by big data analytics, have been thoroughly discussed. It is worthy pointing out that some of the discussed applications are novel and effective compared to the conventional non data-driven approaches. In addition, since the platforms and methods for big data analysis is originally come from information/computer science and need to be revised and tailored, they are also covered in this paper. Furthermore, both the challenges and the future prospects for big data utilization in smart grid are carefully summarized at the end of this paper.

## References

[1] Chen HH, Chen S, Lan Y. Attaining a sustainable competitive advantage in the smart grid industry of China using suitable open innovation intermediaries. Renew Sustain Energy Rev 2016;62:1083–91.
[2] Botte B, Cannatelli V, Rogai S. The telegestore project in Enel's metering system[C]. In: Proceedings of the international conference and exhibition on electricity distribution. IET; 2005. p. 1–4.
[3] National energy technology laboratory for the U.S. Department of energy, modern grid benefits, white paper; 2007.
[4] Inovgrid. [Online]; 2010. Available: ⟨http://www.inovcity.pt/en/pages/inovgrid.aspx⟩.
[5] Moma. [Online]; 2012. Available: ⟨http://www.modellstadt-mannheim.de⟩.
[6] Farhangi H. The path of the smart grid. Power Energy Mag IEEE 2010;8(1):18–28.
[7] DOE. Advancement of synchrophasor technology in ARRA Projects; 2016. ⟨https://www.smartgrid.gov/recovery_act/program_publications.html⟩.
[8] Yang B, Yamazaki J. Big data analytic empowered grid applications- Is PMU a big data issue? In: Proceedings of the 2015 12th international conference on the European energy market (EEM). IEEE; 2015. p.1–4.
[9] DOE. Summary of the North American synchrophasor initiative (NASPI) activity area; 2015. ⟨https://www.naspi.org/documents⟩.
[10] Yuan J, Shen J, Pan L, et al. Smart grids in China. Renew Sustain Energy Rev 2014;37(3):896–906.
[11] Alahakoon D, Yu X. Smart electricity meter data intelligence for future energy systems: a survey. IEEE Trans Ind Inf 2016;12(1):425–36.
[12] Klump R, Agarwal P, Tate JE, Khurana H, Lossless compression of synchronized phasor measurements. In Proceedings of the IEEE power and energy society general meeting. Minneapolis, MN, USA; Jul. 2010. p. 1–7.
[13] Electric Power Group. Wind farm oscillation detection and mitigation. ⟨https://www.smartgrid.gov/recovery_act/project_information.html⟩.
[14] Short TA. Advanced metering for phase identification, transformer identification, and secondary modeling. IEEE Trans Smart Grid 2013;4(2):651–8.
[15] Overholt P, Kosterev D, Eto J, et al. Improving reliability through better models: using synchrophasor data to validate power plant models. IEEE Power Energy Mag 2014;12(3):44–51.
[16] Luan W, Peng J, Maras M, et al. Smart meter data analytics for distribution network connectivity verification. IEEE Trans Smart Grid 2015;6(4), [1–1].
[17] Berrisford AJ. A tale of two transformers: an algorithm for estimating distribution secondary electric parameters using smart meter data. Electrical and computer engineering (CCECE). In: Proceedings of the 2013 26th annual IEEE Canadian conference. IEEE; 2013. p. 1–6.
[18] Hajnoroozi AA, Aminifar F, Ayoubzadeh H. Generating unit model validation and calibration through synchrophasor measurements. IEEE Trans Smart Grid 2015;6(1):441–9.
[19] Bolognani S, Bof N, Michelotti D. et al. Identification of power distribution network topology via voltage correlation analysis. In: Proceedings of the IEEE conference on decision and control; 2013. p. 1659–64.
[20] Erlinghagen S, Lichtensteiger B, Markard J. Smart meter communication standards in Europe – a comparison. Renew Sustain Energy Rev 2015;43:1249–62.
[21] Zhou K, Yang S. Understanding household energy consumption behavior: the contribution of energy big data analytics. Renew Sustain Energy Rev 2016;56:810–9.
[22] Tuballa ML, Abundo ML. A review of the development of Smart Grid technologies. Renew Sustain Energy Rev 2016;59:710–25.
[23] Haider HT, See OH, Elmenreich W. A review of residential demand response of smart grid. Renew Sustain Energy Rev 2016;59:166–78.
[24] Moretti M, Djomo SN, Azadi H, et al. A systematic review of environmental and economic impacts of smart grids. Renew Sustain Energy Rev 2017:68.
[25] Ali SM, Jawad M, Khan B, et al. Wide area smart grid architectural model and control: a survey. Renew Sustain Energy Rev 2016;64:311–28.
[26] Manyika J, Chui M, Brown B, et al. Big data: the next frontier for innovation, competition, and productivity. Analytics 2011.
[27] Naimi AI, Westreich DJ. Big data: a revolution that will transform how we live, work, and think. Am J Epidemiol 2014;17(9):181–3.
[28] Zhou K, Fu C, Yang S. Big data driven smart energy management: from big data to big insights. Renew Sustain Energy Rev 2016;56:215–25.
[29] Schuelke-Leech BA, Barry B, Muratori M, et al. Big data issues and opportunities for electric utilities. Renew Sustain Energy Rev 2015;52(2):937–47.
[30] Hu J, Vasilakos A. Energy big data analytics and security: challenges and opportunities. IEEE Trans Smart Grid 2016;7(5):2423–36, [2016, 7(5):1-1].
[31] Mishra DP, Samantaray SR, Joos G. A combined wavelet and data-mining based intelligent protection scheme for microgrid. IEEE Trans Smart Grid 2016;7(5):2295–304.
[32] Chen YC, Wang J, Dominguez-Garcia AD, et al. Measurement-based estimation of the power flow Jacobian matrix. IEEE Trans Smart Grid 2016;7(5):2507–15.
[33] Endsley MR, Connors ES. Situation awareness: State of the art[C]. In: Proceedings of the IEEE power & energy society general meeting-conversion & delivery of electrical energy in century. IEEE; 2008. p. 1–4.
[34] Aminifar F, Fotuhi-Firuzabad M, Shahidehpour , et al. A probabilistic multistage PMU placement in electric power systems. IEEE Trans Power Deliv 2011;26(2):841–9.
[35] Gopakumar P, Chandra G, Reddy M, et al. Optimal placement of PMUs for the smart grid implementation in Indian power grid – a case study. Front Energy 2013;7(3):358–72.
[36] Xu B, Abur A. Observability analysis and measurement placement for systems with PMUs. In: Proceedings of the IEEE PES power systems conference and exposition. Vol. 2; October 2004. p. 943–6.
[37] Milosevic B, Begovic M. Nondominated sorting genetic algorithm for optimal phasor measurement placement. IEEE Trans Power Syst 2003;18(1):69–75.
[38] Western interconnection synchrophasor program; June 2014. ⟨https://www.smartgrid.gov/project/western⟩.
[39] Huang J, Wu N, Ruschmann M. Data-availability-constrained placement of PMUs and communication links in a power system. IEEE Syst J 2014;8(2):483–92.
[40] Li Q, Cui T, Weng Y, Negi R, et al. An information-theoretic approach to PMU

placement in electric power systems. IEEE Trans Smart Grid 2013;4(1):446–56.
[41] Sodhi R, Srivastava SC, Singh SN. Multi-criteria decision-making approach for multistage optimal placement of phasor measurement units. IET Gener Transm Distrib 2011;5(2):181–90.
[42] Domingos P, Hulten G. Mining high-speed data streams. In: Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. 2000. p. 71–80.
[43] Dahal N, Abuomar O, King R, et al. Event stream processing for improved situational awareness in the smart grid. Expert Syst Appl 2015;42(20):6853–63.
[44] Ma S, Fang S, Yuan D, et al. The design of power security system in smart home based on the stream data mining. Adv Data Min Appl 2014:716–24.
[45] Omitaomu OA. A control chart approach for representing and mining data streams with shape based similarity; 2014.
[46] Basu C, Agrawal A, Hazra J. et al. Understanding events for wide-area situational awareness. In: Proceedings of the innovative smart grid technologies conference. IEEE; 2014. p. 1–5.
[47] Peppanen J, Reno MJ, Thakkar M, et al. Leveraging AMI data for distribution system model calibration and situational awareness. IEEE Trans Smart Grid 2015;6(4):2050–9.
[48] Schweppe FC, Wildes J. Power system static-state estimation,part I: exact model. IEEE Trans Power Appl Syst . 1970;PAS-89(1):120–5.
[49] Zhao J, Zhang G, Das K, et al. Power system real-time monitoring by using PMU-based robust state estimation method. IEEE Trans Smart Grid 2015;7(1):1.
[50] Pignati M, Zanni L, Sarri S. et al. A pre-estimation filtering process of bad data for linear power systems state estimators using PMUs. In: Proceedings of the power systems computation conference; 2014. p. 1–8.
[51] Xie L, Chen Y, Kumar PR. Dimensionality reduction of synchrophasor data for early event detection: linearized analysis. IEEE Trans Power Syst 2014;29(6):2784–94.
[52] DOE. Project information; 2016. ⟨https://www.smartgrid.gov/recovery_act/project_information.html⟩.
[53] Dahal OP, Brahma SM, Cao H. Comprehensive clustering of disturbance events recorded by phasor measurement units. IEEE Trans Power Deliv 2014;29(3):1390–7.
[54] Chen Y, Xie L, Kumar PR. Power system event classification via dimensionality reduction of synchrophasor data. In: Proceedings of the IEEE, sensor array and multichannel signal processing workshop; 2014. p. 57–60.
[55] IEEE Std C37.118.1™–2011. IEEE standard for synchrophasor measurements for power systems; 2011. p. 1–61.
[56] IEEE Std C37.118.2™–2011. IEEE standard for synchrophasor measurements for power systems; 2011. p. 1–53.
[57] Khan M, Li M, Ashton P. et al. Big data analytics on PMU measurements. In: Proceedings of the international conference on fuzzy systems and knowledge discovery; 2014.
[58] Kezunovic M, Xie L, Grijalva S. The role of big data in improving power system operation and protection. In: Proceedings of the bulk power system dynamics and control – IX optimization, security and control of the emerging power grid (IREP), 2013 IREP symposium; 2013. p. 1–9.
[59] Meier R, Cotilla-Sanchez E, Mccamish B, et al. Power system data management and analysis using synchrophasor data. Technol Sustain IEEE 2014:225–31.
[60] Khan AR, Mahmood A, Safdar A, et al. Load forecasting, dynamic pricing and DSM in smart grid: a review. Renew Sustain Energy Rev 2016;54:1311–22.
[61] Selakov A, Cvijetinović D, Milović L, et al. Hybrid PSO–SVM method for short-term load forecasting during periods with significant temperature variations in city of Burbank. Appl Soft Comput 2014;16(3):80–8.
[62] Deihimi A, Orang O, Showkati H. Short-term electric load and temperature forecasting using wavelet echo state networks with neural reconstruction. Energy 2013;57(3):382–401.
[63] Zhu L, Wu QH, Li MS. et al. Support vector regression-based short-term wind power prediction with false neighbors filtered. In: Proceedings of the international conference on renewable energy research and applications. IEEE; 2013. p. 740–4.
[64] Rocio Cogollo M, Velasquez JD. Methodological advances in artificial neural networks for time series forecasting. IEEE Lat Am Trans 2014;12(4):764–71.
[65] Huang SC, Lu CN, Lo YL. Evaluation of AMI and SCADA data synergy for distribution feeder modeling. IEEE Trans Smart Grid 2015;6(4), [1–1].
[66] Berrisford A J. A tale of two transformers: an algorithm for estimating distribution secondary electric parameters using smart meter data; 2013. p. 1–6.
[67] Kwac J, Rajagopal R. Data-driven targeting of customers for demand response. IEEE Trans Smart Grid, IEEE Trans Smart Grid 2016;7(5):2199–207.
[68] Peppanen J, Reno MJ, Broderick RJ, et al. Distribution system model calibration with big data from AMI and PV inverters. IEEE Trans Smart Grid 2016;7(5):2497–506.
[69] Pan S, Morris T, Adhikari U. Developing a hybrid intrusion detection system using data mining for power systems. IEEE Trans Smart Grid 2015;6(6):134–43.
[70] Bayrak G, Kabalci E. Implementation of a new remote islanding detection method for wind–solar hybrid power plants. Renew Sustain Energy Rev 2016;58:1–15.
[71] Li C, Higuma K, Watanabe M, et al. Monitoring and estimation of interarea power oscillation mode Based on application of CampusWAMS. Turk J Agric - Food Sci Technol 2008;3(1).
[72] Dasgupta S, Paramasivam M, Vaidya U, et al. PMU-based model-free approach for real-time rotor angle monitoring. IEEE Trans Power Syst 2015;30(5):2818–9.
[73] Liu Y, Zhan L, Zhang Y, et al. Wide-area measurement system development at the distribution level: an FNET/GridEye example. IEEE Trans Power Deliv 2015;31(2), [1–1].
[74] Chai J, Liu Y, Guo J. et al. Wide-area measurement data analytics using FNET/GridEye: a review[C]. In: Proceedings of the power systems computation conference; 2016. p. 1–6.
[75] Kamwa J, Beland G, Trudel. et al. Wide-area monitoring and control at hydro-québec: past, present and future. In: Proceedings of the power engineering society general Meeting; 2006.
[76] e-terraphasorpoint 3.0. ⟨http://www.alstom.com/Global/Grid/Resources/Documents/Automation/NMS/e-terraphasorpoint%203.0.pdf⟩.
[77] Siemens. ⟨http://www.engerati.com/resources/efficient-network-integration-renewable-energy-resources-distribution-level⟩; 2015.
[78] NARI. ⟨http://www.naritech.cn/html/jie166.shtm⟩; 2015.
[79] Chang V, Wills G. A model to compare cloud and non-cloud storage of big data. Future Gener Comput Syst 2016;57:56–76.
[80] Singh M, Ali A. Big data analytics with Microsoft HDInsight in 24h, Sams Teach Yourself: Big Data, Hadoop, and Microsoft Azure for better business intelligence; 2016.
[81] Rusitschka S, Eger K, Gerdes C. Smart grid data cloud: a model for utilizing cloud computing in the smart grid domain. In: Proceedings of the IEEE international conference on smart grid communications; 2010. p. 483–8.
[82] Birman KP, Ganesh L, Renesse RV. et al. Running smart grid control software on cloud computing architectures. In: Proceedings of the the workshop on computational needs for the next generation electric grid; 2012.
[83] Baek J, Vu QH, Liu JK, et al. A secure cloud computing based framework for big data information management of smart grid. IEEE Trans Cloud Comput 2015;3(2):233–44.
[84] Simmhan Y, Prasanna V, Aman S, et al. Cloud-based software platform for big data analytics in smart grids. Comput Sci Eng 2013;15(4):38–47.
[85] Khan Mukhtaj. Hadoop performance modeling and job optimization for big data analytics. London: Brunel University; 2015.
[86] Markovic DS, Zivkovic D, Branovic I, et al. Smart power grid and cloud computing. Renew Sustain Energy Rev 2013;24(1):566–77.
[87] Fang B, Yin X, Tan Y, et al. The contributions of cloud technologies to smart grid. Renew Sustain Energy Rev 2016;59:1326–31.
[88] Apache Hadoop. ⟨http://hadoop.apache.org/⟩; 2016.
[89] Zhang Y, Chen S, Wang Q, et al. $i^2$MapReduce: incremental MapReduce for mining evolving big data. IEEE Trans Knowl Data Eng 2015;27(7):1.
[90] Xing EP, Ho Q, Dai W. et al. Petuum: a new platform for distributed machine learning on big data. In: Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining. ACM; 2015. p. 1–1.
[91] Shyam R, Bharathi GHB, Sachin KS, et al. Apache spark a big data analytics platform for smart grid. 21; 2015. p. 171–8.
[92] Apache Spark. ⟨http://spark.apache.org/⟩; 2016.
[93] Zhou D, Guo J, Zhang Y, et al. Distributed data analytics platform for wide-area synchrophasor measurement systems. IEEE Trans Smart Grid 2016;7(5):2397–405.
[94] Li D. Machine learning aided decision making and adaptive stochastic control in a hierarchical interactive smart grid. Dissertations & theses - gradworks; 2014.
[95] Wu X, Zhu X, Wu GQ, et al. Data mining with big data. IEEE Trans Knowl Data Eng 2014;26(1):97–107.
[96] Al-Jarrah OY, Yoo PD, Muhaidat S, et al. Efficient machine learning for big data: a review. Big Data Res 2015;2(3):87–93.
[97] Khazaei J, Fan L, Jiang W, et al. Distributed prony analysis for real-world PMU data. Electr Power Syst Res 2016;133:113–20.
[98] Shaker H, Zareipour H, Wood D. A data-driven approach for estimating the power generation of invisible solar sites. IEEE Trans Smart Grid 2016;7(5):2466–76.
[99] Shaker H, Zareipour H, Wood D. Estimating power generation of invisible solar sites using publicly available data. IEEE Trans Smart Grid 2016;7(5):2456–65.
[100] Taieb SB, Huser R, Hyndman RJ, et al. Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression. IEEE Trans Smart Grid 2016;7(5):2448–55.
[101] Green RC, Wang L, Alam M. Applications and trends of high performance computing for electric power systems: focusing on smart grid. IEEE Trans Smart Grid 2013;4(2):922–3.
[102] Partridge M, Calvo RA. Fast dimensionality reduction and simple PCA. Intell Data Anal 1998;2(1–4):203–14.
[103] Poekaew P, Champrasert P. Adaptive-PCA: an event-based data aggregation using principal component analysis for WSNs. In: Proceedings of the international conference on smart sensors and application. IEEE; 2015.
[104] Nuchprayoon S. Electricity load classification using K-means clustering algorithm. In: Proceedings of the Brunei international conference on engineering and technology. IET; 2014. p. 1–5.
[105] Yang H, Zhang J, Qiu J, et al. A practical pricing approach to smart grid demand response based on load classification. IEEE Trans Smart Grid 2016, [1–1].
[106] Mu FL, Li HY. Power load classification based on spectral clustering of dual-scale. In: Proceedings of the IEEE international conference on control science and systems engineering. IEEE; 2014.
[107] Ye M, Liu Y, Xiong W. et al. Voltage stability research of receiving-end network based on real-time classification load model. In: Proceedings of the IEEE international conference on mechatronics and automation; 2014. p. 1866-70.
[108] Che JX, Wang JZ. Short-term load forecasting using a kernel-based support vector regression combination model. Appl Energy 2014;132(11):602–9.
[109] Liu N, Tang Q, Zhang J, et al. A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids. Appl Energy 2014;129:336–45.
[110] Rahman S. Formation and analysis of a rule-based short-term load forecasting algorithm. In: Proceedings of the IEEE (Institute of electrical and electronics engineers), (USA). 78:5(5); 2015. p. 805–16.
[111] Hernández L, Baladrón C, Aguiar JM, et al. Artificial neural networks for short-

term load forecasting in microgrids environment. Energy 2014;75(C):252–64.
[112] Nickel M, Murphy K, Tresp V, et al. A review of relational machine learning for knowledge graphs. Proc IEEE 2015;104(1):11–33.
[113] Drumond LR, Diaz-Aviles E, Schmidt-Thieme L. et al. Optimizing multi-relational factorization models for multiple target relations. In: Proceedings of the ACM international conference; 2014. p. 191–200.
[114] Schweppe F, Wildes J. Power system static-state estimation, part I, II, III. IEEE Trans Power Appl Syst . 1970;PAS-89(1):120–35.
[115] Xu X, He X, Ai Q, et al. A correlation analysis method for power systems based on random matrix theory. IEEE Trans Smart Grid 2015.
[116] Yang Q. Cross-domain data fusion. Computer 2016;49(4), [18–18].
[117] Diou C, Stephanopoulos G, Panagiotopoulos P, et al. Large-scale concept detection in multimedia data using small training sets and cross-domain concept fusion. IEEE Trans Circuits Syst Video Technol 2011;20(12):1808–21.
[118] Tsai CC, Chang-Chien LR, Chen IJ, et al. Practical considerations to calibrate generator model parameters using phasor measurements. IEEE Trans Smart Grid 2016:1–11.
[119] Hana S. SAP AG, in-memory database, computer appliance. Vertpress; 2012.
[120] Kennedy , Stephen James MCP. M.C.P. Massachusetts Institute of TechnologyTransforming big data into knowledge: experimental techniques in dynamic visualization. Massachusetts Institute of Technology; 2012.
[121] Cuzzocrea A. Privacy and security of big data: current challenges and future research perspectives. In: Proceedings of the international workshop on privacy & secuirty of big data. ACM; 2014. p. 45–7.
[122] Bertino E. Big data – security and privacy. In: Proceedings of the IEEE international congress on big data. IEEE; 2015.
[123] Depuru SSSR, Wang L, Devabhaktuni V. Smart meters for power grid: challenges, issues, advantages and status. Renew Sustain Energy Rev 2011;15(6):2736–42.
[124] Kabalci Y. A survey on smart metering and smart grid communication. Renew Sustain Energy Rev 2016;57:302–18.
[125] Hashem IAT, Yaqoob I, Anuar NB, et al. The rise of “big data” on cloud computing: review and open research issues. Inf Syst 2015;47(47):98–115.
[126] Jagadish HV, Gehrke J, Labrinidis A, et al. Big data and its technical challenges. Commun Acm 2014;57(7):86–94.
[127] Schweppe FC, Rom DB. Power system static-state estimation, part II: approximate model. IEEE Trans Power Appl Syst 1970;PAS-89(1):125–30.
[128] Schweppe FC. Power system static-state estimation, part III: implementation. IEEE Trans Power Appl Syst 1970;PAS-89(1):130–5.
[129] Gol M, Abur A. A fast decoupled state estimator for systems measured by PMUs. IEEE Trans Power Syst 2015;30(5):2766–71.
[130] Jokar P, Arianpoo N, Leung VCM. Electricity theft detection in AMI using customers' consumption patterns. IEEE Trans Smart Grid 2016;7(1):216–26.
[131] Shuai Z, Sun Y, Shen Z, et al. Microgrid stability: classification and a review. Renew Sustain Energy Rev 2016;58:167–79.