# Big Data Analytics Using Cloud Computing Based Frameworks for Power Management Systems: Status, Constraints, and Future Recommendations

**Ahmed Hadi Ali AL-Jumaili** [1,2], **Ravie Chandren Muniyandi** [1], **Mohammad Kamrul Hasan** [1,*],
**Johnny Koh Siaw Paw** [3] and **Mandeep Jit Singh** [4]

[1]  Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia,
    Bangi 43600, Selangor, Malaysia
[2]  Computer Centre Department, University of Fallujah, Anbar 00964, Iraq
[3]  Department of Electronic & Communication Engineering, Universiti Tenaga Nasional, Km 7, Jalan
    Ikram-Uniten, Kajang 43009, Selangor, Malaysia
[4]  Department of Electrical, Electronic and System Engineering, Faculty of Engineering and Built Environment,
    Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia
*   Correspondence: mkhasan@ukm.edu.my

**Abstract:** Traditional parallel computing for power management systems has prime challenges such as execution time, computational complexity, and efficiency like process time and delays in power system condition monitoring, particularly consumer power consumption, weather data, and power generation for detecting and predicting data mining in the centralized parallel processing and diagnosis. Due to these constraints, data management has become a critical research consideration and bottleneck. To cope with these constraints, cloud computing-based methodologies have been introduced for managing data efficiently in power management systems. This paper reviews the concept of cloud computing architecture that can meet the multi-level real-time requirements to improve monitoring and performance which is designed for different application scenarios for power system monitoring. Then, cloud computing solutions are discussed under the background of big data, and emerging parallel programming models such as Hadoop, Spark, and Storm are briefly described to analyze the advancement, constraints, and innovations. The key performance metrics of cloud computing applications such as core data sampling, modeling, and analyzing the competitiveness of big data was modeled by applying related hypotheses. Finally, it introduces a new design concept with cloud computing and eventually some recommendations focusing on cloud computing infrastructure, and methods for managing real-time big data in the power management system that solve the data mining challenges.

**Keywords:** data mining; big data; cloud computing; parallel computing; power system

## 1. Introduction

In the early stage of the development of condition monitoring technology, most of the monitoring systems were developed for a certain type of equipment [1,2], and each system was scattered and isolated [3]. This was an information island without data sharing and interaction and was not conducive to the management and comprehensive analysis of monitoring data [4]. In addition, hardware resources such as network, computing, and storage of these monitoring systems are also difficult to share, resulting in a waste of IT facilities [5]. Therefore, an integrated management system built in the main control room has emerged, which can centrally process various monitoring data collected by different monitoring devices [6,7]. In order to integrate monitoring devices of different specifications into the centralized monitoring system, the State Grid Corporation of China issued a number of technical regulations and communication protocols, and established state monitoring centers for power transmission and transformer equipment in many

provincial grid companies [8]. However, a current limitation of the monitoring device is it can only upload the processed simplified data to the monitoring center, and the data collection frequency is not high [9,10]. However, with the popularization and application of a high-speed optical fiber network and wireless transmission technology in the power industry, the future power equipment condition monitoring center will be able to receive panoramic real-time condition monitoring data from a wider area and become a data center for data integration and information sharing [11]. Therefore, the amount of collected data in the monitoring center will be staggering in the future, and the information processing capability of the existing monitoring system will be insufficient to meet the storage and processing requirements of such massive data [12,13]. Obviously, the serial processing method has long been unable to meet the processing requirements of large amounts of data [14,15]. The conventional parallel computing paradigm based on high-performance computers has always been responsible for various computing problems encountered in scientific research and engineering practice [16]. There are also many deficiencies in the state data of electrical equipment [7]. In recent years, cloud computing technology sprouted from parallel computing and has developed rapidly, and its many advantages have brought new ideas for the establishment of monitoring center computing platforms, which have attracted the attention of scholars in the power industry [2,17], as shown in Figure 1. From the research status, most of the cloud computing platforms currently designed for monitoring centers are based on a single Hadoop framework, which has certain limitations. Hadoop is good at batch processing of big data, but it cannot meet the richer computing modes in power equipment condition monitoring, such as stream computing [18]. Especially under extreme conditions (such as severe weather), the condition monitoring alarms were activated for power equipment [19].
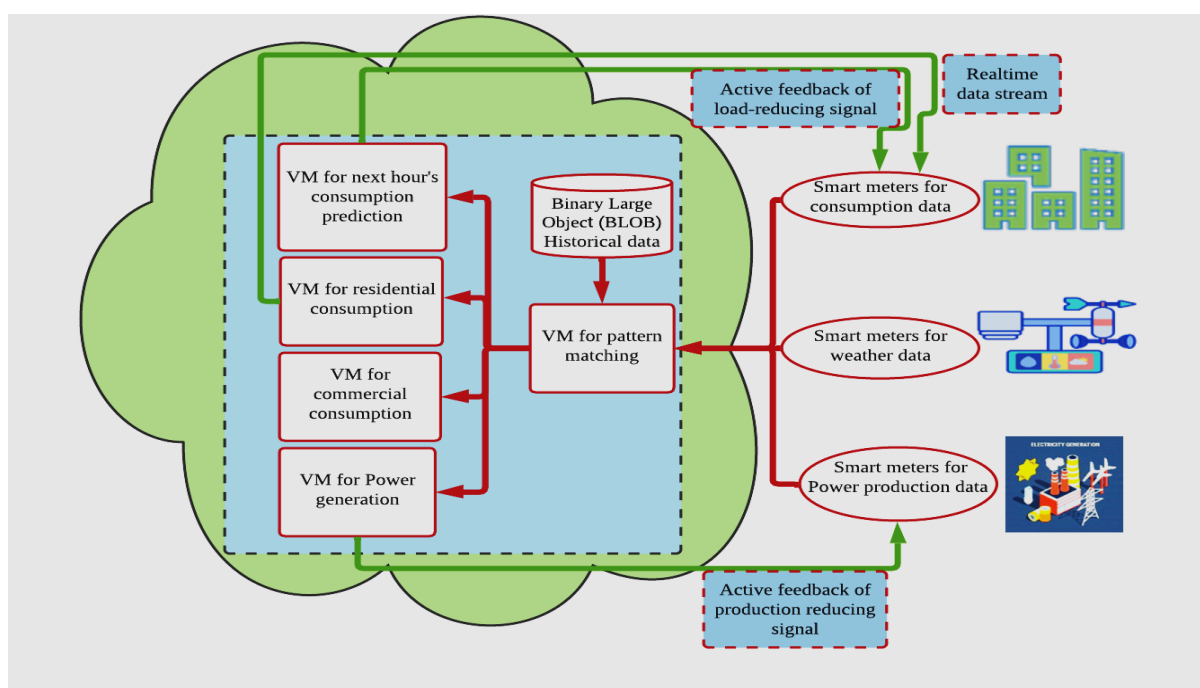


**Figure 1.** Efficient data processing for the power system.

## 1.1. Research Background

The power grid dispatch center is developing towards the promotion of online monitoring technology that integrates power regulation and control. The power equipment operating status data will be sent to the regulation center in the future. The existing Supervisory Control and Data Acquisition with power Management Systems (SCADA/EMS) are challenging for dealing with the massive amount of monitoring data [20]. In order to enable

monitoring devices of different manufacturers to be connected to a unified centralized monitoring system, a state monitoring center was established for power transmission and transformation equipment in many grid companies [2]. Although web service technology reduces the difficulty of data integration, challenges remain in meeting the real-time data requirements of electric power companies [14,15].

Therefore, the current monitoring system receives skilled data processed locally in the monitoring device. For example, the monitoring device must process high-voltage electrical equipment's partial discharge waveform signal into the number of discharges, the peak discharge volume, and the corresponding discharge phase before being uploaded [21–23]. Uploading "familiar data" instead of "raw data" can reduce network transmission costs and monitoring center storage costs [24]. However, a monitoring center that integrates data from different specifications of monitoring devices still faces a complex problem to in-depth state assessment and failure of the target device diagnosis [25,26].

Massive monitoring data will flood the remote power equipment monitoring center, resulting in heavy storage, processing, and analysis tasks. The condition monitoring data of power equipment has shown the main characteristics of big data: massive in volume, various in types, rapid in change, and low in value density. The monitoring data processing and analysis tasks will be transferred from distributed monitoring devices to centralized monitoring centers, which simplifies the hardware and software configuration of monitoring devices and facilitates the flexible expansion of the monitoring center infrastructure [27]. General Electric Company (GE), monitoring the operation of remote steam turbines, has shifted from the past data processing and uploading simplified data sets to the use of memory data grids and other technologies to receive and store raw sensor data [9]. The widespread application of high-speed optical fiber communication in the power industry provides solutions for massive data transmission [11].

Traditional data processing technology has encountered bottlenecks under the current explosive growth of data. It cannot meet the analysis needs of the power industry to obtain knowledge and information from massive data quickly. The power industry information is the research and application of power big data technology. It is an inevitable requirement for the development of technology and intelligence. Data mining can comprehensively use relevant algorithms to process a large amount of data and discover its hidden valuable information, which can realize the rapid conversion of data to knowledge and then value [28–32]. However, the current data volume is growing rapidly, and traditional algorithms based on single-node serial mining are no longer suitable for massive data requirements. As a distributed data processing platform, cloud computing can integrate many computer resources and increase technical capabilities considerably. It is more suitable for processing massive data than ordinary algorithms [14]. In addition, cloud computing models and cloud computing platforms do not have high requirements for network nodes, and regular computers can also participate in cloud computing, which reduces the complexity and cost of cloud platform construction to a certain extent [33–35].

In recent years, cloud computing has received more attention with a computing model that is integrated from the development of conventional computers and technological networks such as distributed computing, parallel computing, network storage, etc. [2,19,36]. The virtualization, distributed storage, and parallel computing technologies in cloud computing provide new ideas for constructing computing platforms for data centers' power equipment condition monitoring. It is possible to integrate the existing basic computing facilities of electric power enterprises and provide reliable, stable, and powerful storage and computing capacity support which is beneficial to monitor the online power equipment over a wide range [37,38].

Monitoring and information collection improve real-time analysis and intelligent diagnosis capabilities [19,39]. With the rapid development of cloud computing technology, various high-reliability and high-scalability, big data processing systems such as Hadoop, Spark, and Storm have emerged, providing favorable tools for the centralized processing of large-scale power equipment monitoring data [40–42]. Although these emerging com-

puting models all offer a unified programming interface and shield more low-level details than traditional parallel computing programming models, how can they be introduced into the data processing of the power equipment monitoring center? Combining specific professional backgrounds to solve practical problems and different high-level applications in the monitoring system is still a topic worthy of study [7,16]. The massive data characteristics in smart grid many studies are consistent with the widespread 5 V vast data with their 9 V's Characteristics paradigm as shown in the Figure 2 [43–46].
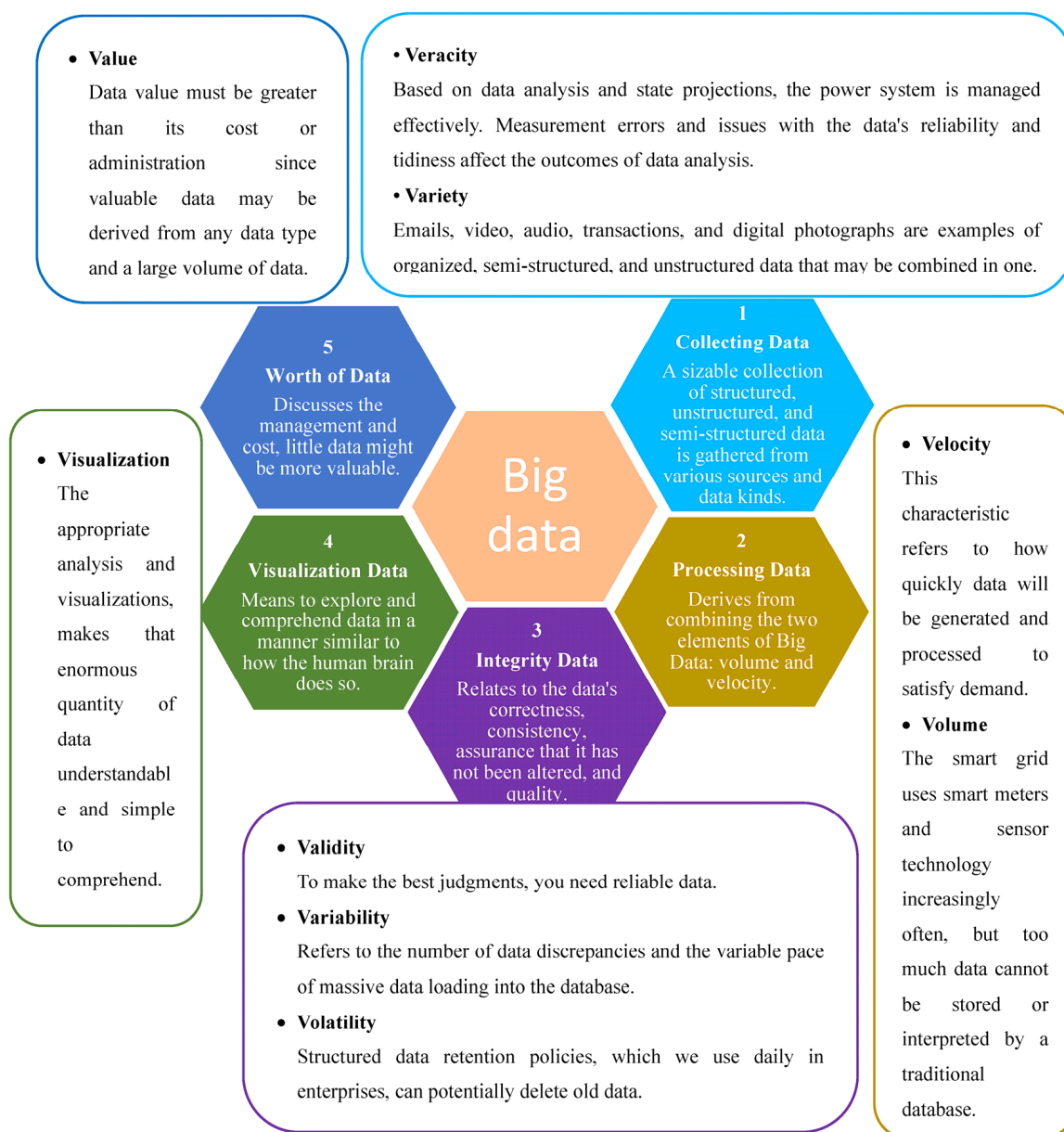


**Figure 2.** Five categories (collecting data, processing data, integrity data, visualization data, and worth of data) of big data with their 9 Vs characteristics.

Power System Condition Monitoring

Online monitoring refers to the continuous or periodic automatic collection of electrical, physical, chemical, and other state information of the monitored object with the help of monitoring devices installed on or near the monitored object without power failure. It can pass the field bus, Ethernet, wireless, and other communication methods transmit the status data to the remote monitoring system for centralized storage and processing [18,47],

as shown in Figure 3. The online monitoring system uses advanced information processing and diagnosis technology to process and comprehensively analyze the status data, which can predict the remaining life of the target equipment and provide a data basis and basis for the status maintenance [48].
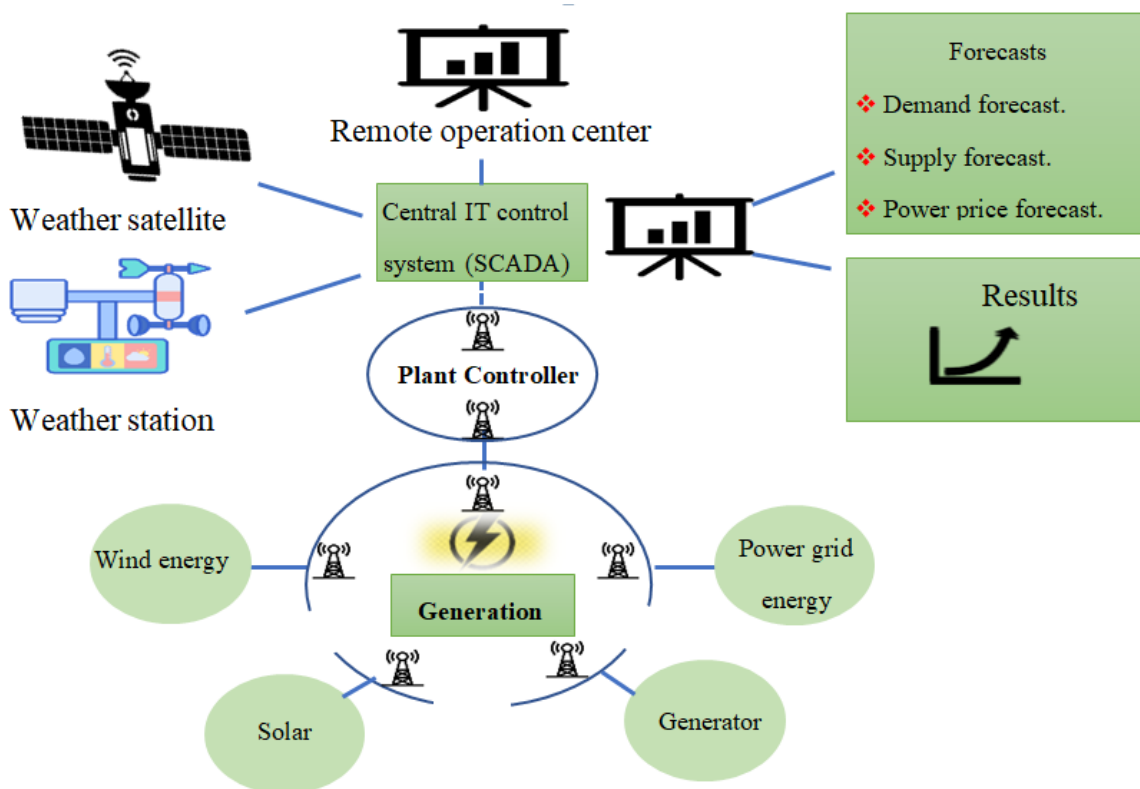


**Figure 3.** Power system controller monitoring.

At the same time, testing instruments capable of infrared, ultrasonic, and other non-electricity measurements have also appeared. Since the 1990s, sensor technology, computer technology, and digital signal acquisition technology have brought online monitoring into a new stage. A multifunctional microcomputer online monitoring system has appeared, which has realized the monitoring of more equipment parameters, such as dielectric loss factor, electrical capacity, leakage current, partial discharge, etc. This monitoring system has more complete functions and improved software and hardware configuration, which can realize data processing, analysis, diagnosis, alarms, and visualization of results [49]. Integration, automation, and intelligence have become the current development direction of online monitoring technology [50]. However, most of the existing power monitoring systems are in the stage of isolated operation, and the utilization of large amounts of data obtained is low [4].

In recent years, it can only allow partial monitoring subsystems to upload simplified "familiar data" and cannot transmit sampled data of high-frequency signals. Such as the discharge signal of electrical equipment, and these data are beneficial for fault diagnosis of equipment such as transformers and GIS switches [49]. The popularization and application of high-speed optical fiber communication technology will allow the transmission of large amounts of data [50,51]. The future power equipment monitoring system will receive panoramic real-time status monitoring data from a wide range of power equipment and become data that can realize data integration and information sharing center [52].

Most traditional grid SCADA systems collect data through circular inquiry, which cannot collect continuous monitoring data of power equipment. It is challenging to meet real-time processing requirements and dynamic access of streaming big data for power

equipment monitoring [20]. The recent emergence of cloud computing technology may bring hope to difficulties in collecting, storing, and processing big data for power equipment monitoring [53]. Ref. [19] proposed information management and computing platform architecture based on cloud computing technology for smart grid status monitoring, in which the computing layer only uses the Hadoop framework. In [54] discussed the feasibility and necessity of the power system utilizing by cloud computing, explained the features and structure goals of the power system utilizing by cloud computing center, to propose system structure design by taking simulation computing as an example to meet the needs of future smart grids for the storage, sharing, and processing of massive panoramic information. Ref. [55] proposed computing platform provides a storage and computing environment for business applications in the data center. Ref. [18], combined the characteristics of intelligent substation state monitoring data, studied the online monitoring data processing platform of substation equipment based on Hadoop, and focused on the storage and fast query methods of monitoring data in the column storage distributed database "Apache HBase". Ref. [19] studied an intelligent diagnosis system for substation equipment based on a cloud platform, provided a variety of intelligent diagnosis methods, and realized collaborative fault diagnosis and hierarchical diagnosis through information fusion strategies.

### 1.2. Challenges

Online monitoring of intelligent power primary equipment and conventional power equipment has been dramatically developed and has become a trend [2,19,56]. Monitoring data is becoming more and more massive, and online monitoring system of power equipment faces huge technical challenges, like (real-time, rapid change, high precision, different application (various types), abnormal data brought by power data, large-scale data with complex structures, large dimensions, and massive data) [27]. Research work can be roughly divided into two kinds of detection methods based on traditional detection methods and detection methods based on data mining [14]. Traditional power load abnormal detection methods are generally based on human experience, state estimation, load curve, similarity and load change rate [57]. Due to the low efficiency, strong subjectivity, and many human factors of this method. This method is suitable just for abnormal data with large abrupt changes, and cannot detect abnormal data with unobvious changes [58]. With the continuous development of artificial intelligence and cluster analysis theory, the speed of data mining technology in the detection of abnormal power load data is also increasing [59]. where commonly can be roughly divided into two categories: firstly, the abnormal value detection method with supervised learning, which selects a part of the power system load data as the training sample and then uses the corresponding algorithm to make the selected sample data and the expected output meet the corresponding requirements based on support vector machines, detection methods based on artificial neural networks, and detection methods based on decision trees [60]. secondly, the unsupervised learning power load abnormal data detection method does not need to select part of the historical power load data as training samples [61]. This type of abnormal data detection method usually includes density analysis, cluster analysis and so on. The distance-based outlier detection algorithm belongs to the unsupervised algorithm. This method is easy to understand and explain, and it is one of the most representative methods in density-based outlier detection algorithms, and it has better results when dealing with medium and high-dimensional data [61]. The density-based abnormal load data detection method still has the problem that some parameters are specified by subjective factors such as human experience, which reduces the accuracy of abnormal data detection [62]. The abnormal value detection method based on the clustering algorithm can more accurately distinguish the normal data and abnormal data in the electric load data set according to the characteristics of the electric load data.

At present, cloud computing technology has achieved good results in the fields of medical data storage, traffic data real-time analysis, and weather data analysis [37,63]. The cost of cloud computing is low, and there are no complicated requirements for the servers in

the cluster establishment [64]. With the help of the large scale and fast calculation speed of cloud computing, the management and analysis of power monitoring data can be realized more effectively through integration with traditional data mining techniques [33,65]. In summary, the power load and abnormal data detection method based on data mining and cluster analysis has become a research hotspot in recent years because of its ability to dig deeper into the changing law of the load curve and effectively detect abnormal load data, and a certain result has been achieved [12,66]. In practice, the types of power load data are complex and diverse, and the contradiction between the increasing scale of power load data and the low efficiency of data mining algorithms has gradually become prominent. Each clustering algorithm has defects such as difficulty in optimizing the determination of initial parameters and high sensitivity [67]. Therefore, the related algorithms for the parallel detection of power load data need to be further studied. All the above explanation summarizes in Figure 4.
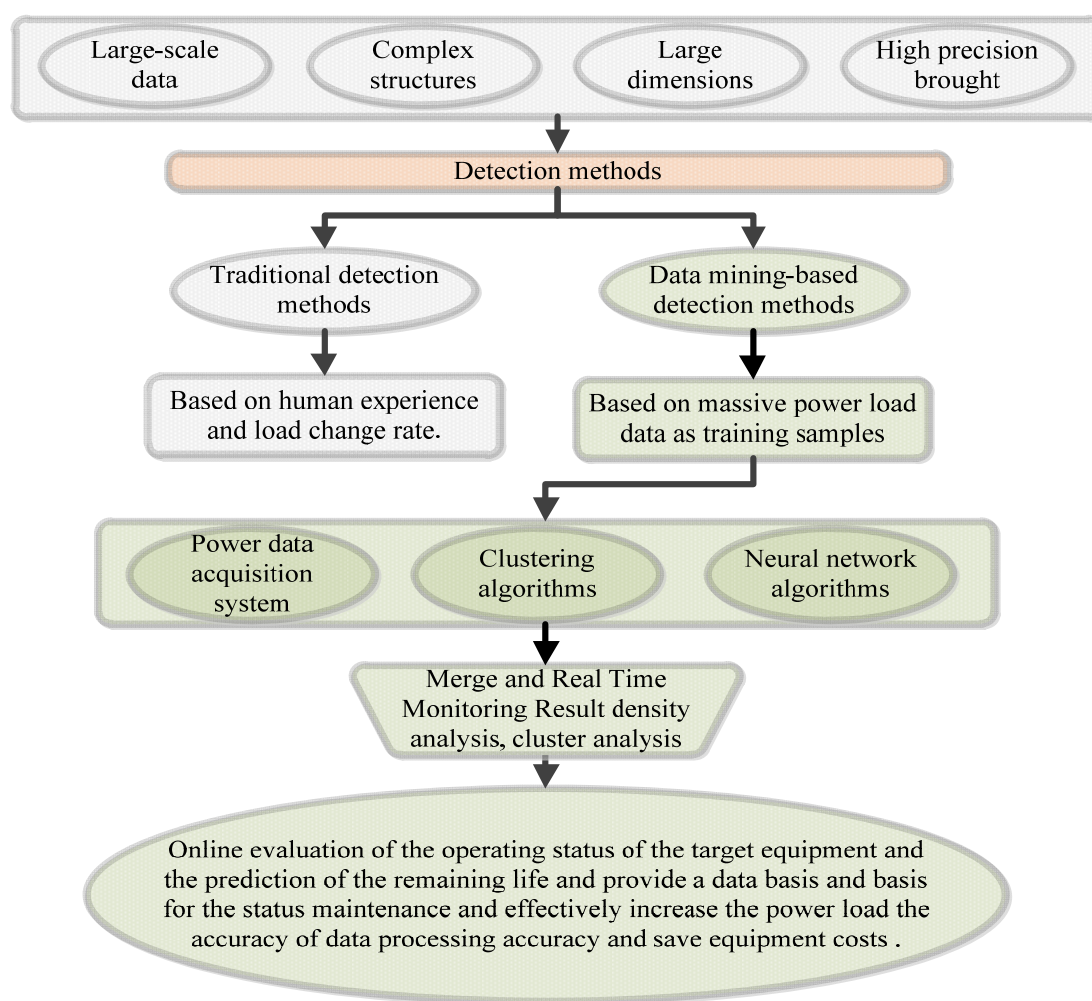


**Figure 4.** Techniques power system detection.

At present, the application research of cloud computing-based big data technology in the power industry is still in the exploratory stage [30]. Presently, most cloud computing platforms designed for monitoring centers are based on a single Hadoop framework and have certain limitations storing data before centralized processing that lead to processing time delay is too long [29], which cannot meet the requirements of online monitoring data [28,42]. Meet the data flow quickly and efficiently will become a trend of information processing in the future [12,68]. Despite several studies on velocity, volume, and variety, a full and efficient solution is currently unavailable on the market; the most popular method

is to use a database management system (DBMS) that may be incompatible with older systems [69]. The rapid processing of monitoring data and the problem of abnormal detection in real-time are becoming increasingly prominent [13]. The introduction of cloud computing technology into the power field has essential research significance, but the application in the power industry is still in the exploratory stage, and more in-depth research is needed to put it into power production [70]. Cloud computing technology has attracted much attention with high performance, but how to use cloud computing technology for large-scale real-time data processing has not been studied [71]. Additionally, multisource heterogeneous urban sensor access and data management technologies provide strong support for intelligent perception and scientific management at the city scale and can accelerate the construction of smart cities or digital twin cities with virtual reality features [72–74]. The difficulty of the processes has frequently increased with the single computing resources insufficient because of the complicated operation of the combined prediction model and meet in real-time for intelligent power systems [7,39,75]. The challenge to analyzing and studying the electricity consumption massive data lead to necessitate the development of clustering analysis algorithms combined with cloud computing [69]. In most of the studies of cloud computing, used as a local controlling system [76,77]. Therefore, there was a lack of studies that used cloud computing to benefit from the ability to share the systems that control the process more widely. Several studies proposed cloud computing products, like artificial intelligence (AI) [78], and the internet of things (IoT) [5,79–81] in their models, however, all of the studies were focusing only on the smart vehicles' batteries [77,78,82–84]. Therefor this research focuses on the problem of centralized parallel processing and diagnosis of power system condition data based on cloud computing and big data technology because difficult to meet the real-time requirements of the power system and construction of a safe, stable, cost-effective, green and environmentally friendly smart grid. The challenges of power management and monitoring have many problems; this work is proposed to target three different Problems Figure 5.

1. Meet the multiple real-time requirements of power system condition monitoring;
2. The weakness of traditional data mining based on single-node serial mining;
3. The insufficient algorithm that combines data mining and computing technology to deal with massive data.

Finally, cloud computing could be a great addition to an intelligent power system that is aimed at solving the challenge of massive data from large-scale areas of a smart power system. Contributions to this literature evaluation take into account the aforementioned issues. To give an overview of big data analytics using cloud computing frameworks for power management systems at the moment, we propose the first universally applicable framework for parallel optimization in power systems, which researchers can use to systematically describe their parallelization studies and place them in the landscape of parallel optimization without regard to the application domain, problem addressed, methodology parallelized, or technology used. The proposed approach, in particular, incorporates both algorithmic design and computational implementation challenges of parallel optimization, which are often dealt with separately in the literature. Second, we use the integrative framework to consolidate earlier research in the field of power systems on parallel optimization.

Furthermore, there are also difficulties managing information and data due to the millions of intelligent meters that need to be managed effectively. Cloud computing could offer a more affordable option for data analytics and storage., as shown in Figure 6, and Table 1 [85].
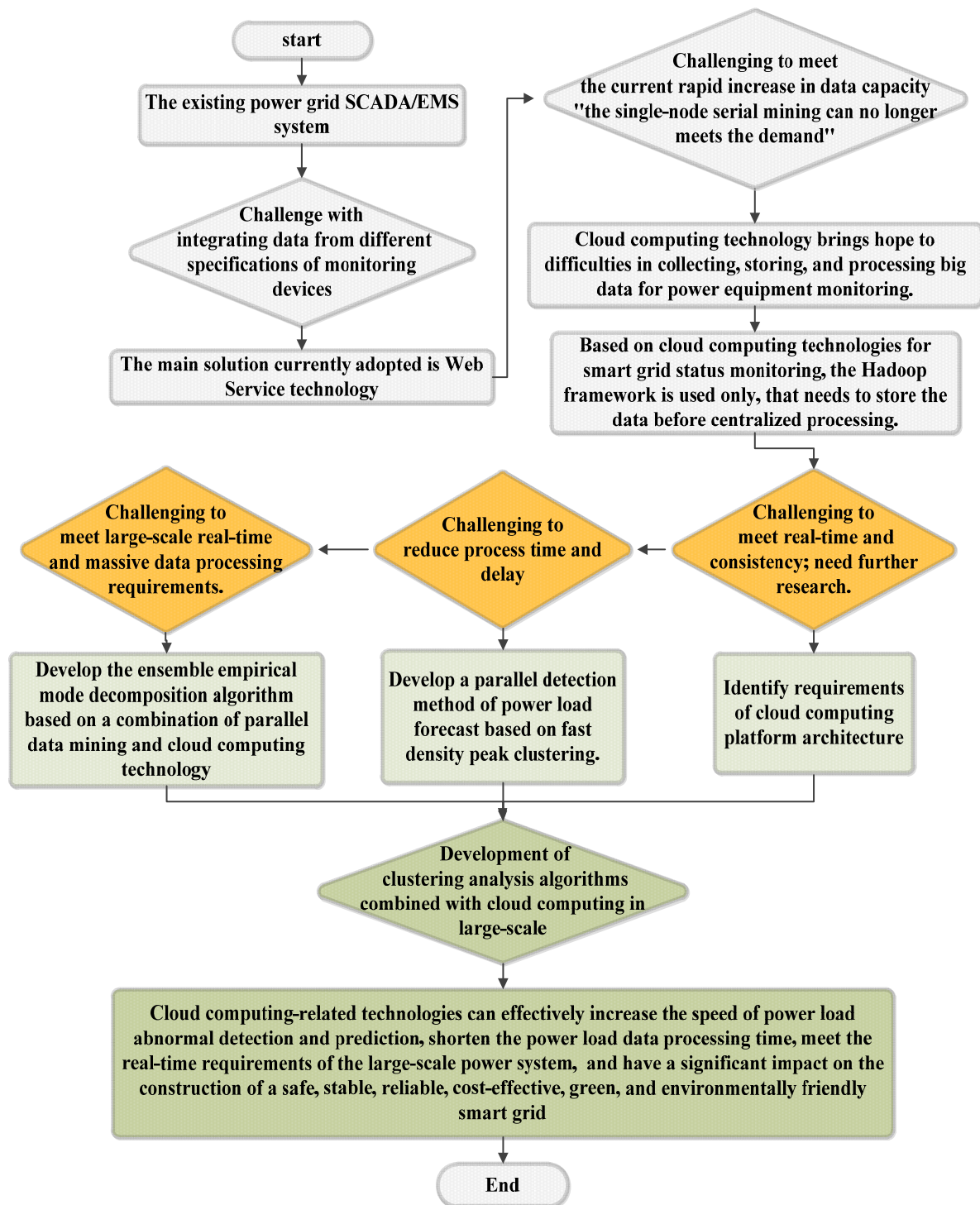
**Figure 5.** General flowchart for challenges and solutions the proposed system.

**Challenges**
- ✔ Multidisciplinary
- ✔ Inadequate techniques for storing energy
- ✔ Sharing energy storage with the cloud
- ✔ Managing and analyzing massive data

**Smart grid**

**Challenges**
- ✔ Capacity restriction of memory
- ✔ Stabilization
- ✔ Administration
- ✔ Inadequate techniques for dealing with enormous volumes of data
- ✔ Issues in data information management

**Big data**

**Cloud computing**

**Challenges**
- ✔ Exchanging data Difficulties
- ✔ Absence of cloud integration for several mobile brokers
- ✔ Reliability alone is insufficient
- ✔ Lack of platform implementation to provide long-term evolution
- ✔ Operation out of sync
- ✔ Distributed Denial of Service Threat
- ✔ The mechanism will cause failure by any mistake.
- ✔ Inadequate techniques for storing and analyzing data

**Figure 6.** Key characteristics of large data in smart grids.

**Table 1.** The difficulties managing information and data due to the millions of intelligent meters data that need to be managed effectively through cloud computing.

| Ref. | Existing Challenging | Proposed Solution | Challenging Proposed Solutions and Future Work |
|---|---|---|---|
| [20] | The problem with concurrent power transmission networks is the uneven temporal distribution and the growing number of fault occurrences that cause power outages or interruptions. | The suggested model incorporates and explicitly assesses seldom occurring environmental components, faults, and periods with fewer fault events, which improves the forecast performance of power transmission fault events. | challenging to deal with the massive amount of monitoring data |
| [24] | The problem of conventional clustering algorithms for Big Data Analytics. | Parallel algorithms of k-means and canopy are implemented using the Hadoop environment and Mahout to solve the problem of conventional clustering algorithms. | Process locally after storage data. |
| [14] | Problems with traditional data mining that is generated in single-node chain mining. | This work uses single-node serial mining to tackle the classic data mining problem in power systems. It has vast storage and processing capacities, and accuracy 87%. | Did not use cloud computing so it was hard to meet real-time and large scale |
| [2] | The limitations of centralised administration based on LAN design prevent broad-area monitoring and the resolution it's issues. | This study describes cloud-based power grid-wide-area monitoring architecture for parallel computing and big data mining to give intelligent grid decisions. | This paper's flaw is a lack of data exchange during processing. |
| [31] | Considering real-time application, the smart grid still needs to advance in terms of efficiency, power management, dependability, and value. | Using cloud computing architecture from any location and at any time, design remote real-time monitoring of substation power data in a safe, efficient, and effective manner. | The weakness of this work the power flow in the grid is continuously monitored using PLC and Energy Meter, it doesn't use cloud computing applications. |
| [19] | Large-scale data processing and analysis methods in a real-time panoramic grid are a challenge for smart grids. | This paper use data mining and integrated information technology platform to present a smart grid building a large multi-level data storage system to extract valuable knowledge to support grid scheduling decisions. | Dealing with redundant data and noise in data mining results remains a barrier for technology. It is also uncertain if the current cloud platform will get real-time smart grid monitoring data. |
| [70] | As smart grids spread, terminal devices like cutting-edge sensors and smart metres tend wide access to distribution networks, providing major challenges to the information perception, analysis, and processing capacities of the distribution automation system. | This paper aims at guiding to preserve CPU and memory resources and increase resource utilisation. through presents a configuration technique for computing resources for the microservice-based edge computing apparatus in the smart distribution transformer region. | The lack is the trade-off methods between robustness and economy in computing resource configuration problems and apply the achievement of this work to investigate the computing resource scheduling problem of the cloud-edge collaborative system in the smart grids. |
| [71] | It has become very difficult to process big amounts of real-time data in research and applications, and it hasn't been researched how to employ cloud computing technology for large-scale real-time data processing. | This research focuses on the big data processing architecture of the cloud computing platform. It creates a large data processing calculation mode and establishes the overall real-time big data processing architecture that acts as the foundation for the RTDP (Real-Time Data Processing) | The RTDP is a tough project, and many issues still need to be researched further: Choosing the most effective technique for calculating future design performance; Real-time data processing hardware must be implemented equally. |

**Table 1.** *Cont.*

| Ref. | Existing Challenging | Proposed Solution | Challenging Proposed Solutions and Future Work |
|------|---------------------|-------------------|-----------------------------------------------|
| [72] | The huge challenge of integrating and exchanging vast sensor information resources that differ widely in hardware design, connection protocols, formatting, conversational skills, sampling rate, and data accuracy. | This paper provides a deeper understanding of the needs, platforms, most current technical developments, and open research problems of urban sensor applications for academics and leaders in the IoT and smart cities sectors. | Relational databases usually struggle with scalability, availability, and concurrent reading and writing, especially for big data handling in wireless sensor networks. As IoT and sensor technology continues to progress, cloud computing will be used. |
| [86] | The ability to detect and analyse anomalies for huge data in real-time is a tough problem due use conventional detection methods of data processing. | An anomaly detection model based on Hadoop distributed processing method, cloud computing and MapReduce monitoring framework is presented using machine learning. | The challenge to Meeting the real-time and large scale |
| [16] | Data from networks and smart cities is increasing and it is becoming huge so it need to big data analysis (BDA) | BDA generated in the smart city (IoT) to turn the smart city toward safety, efficient data processing, and good governance. | The flaw is the system created for the study only offers offline batch analysis and prediction functions. |
| [87] | Smart grids (SGs) are utilizing massive data for operations and services. | Information and communication technologies (ICTs) play an important role, particularly in the computing model, which governs how data analytics in SG may be carried out. | The design of EC systems, EC-appropriate algorithms, resource management in the EC environment, and even hardware accelerations might all be improved. |
| [88] | Increasing renewable energy sources making the power system more complex. | This study focuses on using ICT data in smart grid decision-making to ensure systems are secure and reliably operate. | The SCADA issues caused by ICT integration continue to exist like interdependency analysis, and decision-making. |
| [89] | There are challenges to controlling MGs in a logical and coordinated way | In this study, control objectives are categorized in line to the hierarchical control layers in MGs, and the development approaches given by MGSC/EMS are summarized. | the challenging issue is the uncertainty about power production related to weather, load calculation times and response time brings more challenges to MGSC/EMS. |
| [21] | The challenge of extracting data value through the statistical analysis of an immense amount of data generated by cyber-physical systems. | The goal of this paper was not to give the solutions, but rather to name the problems. A major challenge is the changing nature of the technical systems | software-based devices change frequently due to bug fixing and software updates. Therefore, the data we collected is after time only partially valid. |
| [90] | The challenge of clustering techniques in Big Data context. | Provide a thorough analysis of the Big Data clustering problems and highlight the benefits of the key methods. | Data are too big, dynamic, and complex. Traditional data handling struggle to collect, store, and analyse data. |
| [28] | The execution of the Hadoop cluster when processing a high number of tiny files is the true problem businesses face. The solutions are restricted to NameNode memory | Some novel strategies have been put forth, such as combining tiny heterogeneous files in various formats in a quasirandom manner, which resolves the memory issue by drastically reducing the amount of metadata. | Hadoop cannot satisfy real-time demands because it stores data before processing. |
| [29] | Big Data poses difficulties for Digital Earth in terms of data mining, processing, and storage. Transforming big data's volume, velocity, and diversity into values is the main challenge. | Cloud computing provides fundamental support to address the challenges with shared computing resources including computing, storage, networking and analysis, that fostered Big Data advancements. | It is extremely difficult to achieve in real-time processing. |

**Table 1.** *Cont.*

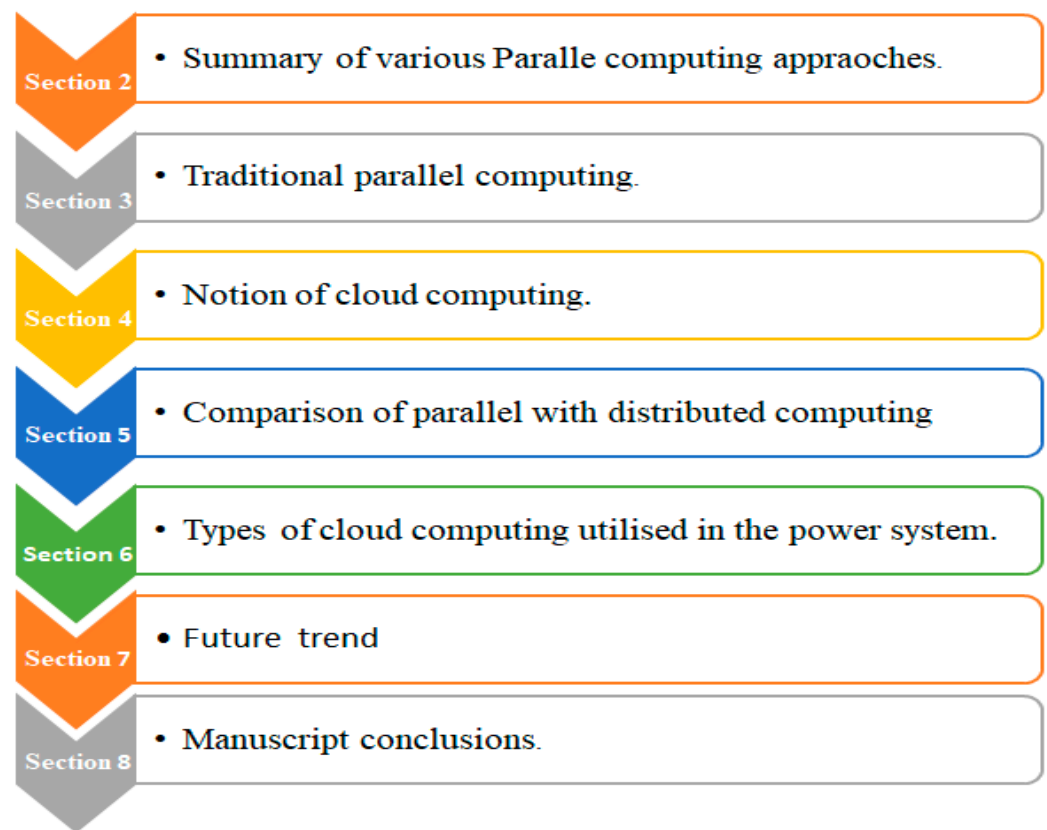| Ref. | Existing Challenging | Proposed Solution | Challenging Proposed Solutions and Future Work |
|------|---------------------|-------------------|-----------------------------------------------|
| [30] | Large data environments lack capabilities like support for massive data, high performance, high reliability, scalability, and high resource. | This paper studied features of popular NoSQL and NewSQL databases for unified storage management and quick data access. | It is extremely difficult to achieve in real-time processing. |
| [69] | Big data is currently the most difficult organisational problem due to the rapid generation of new data every second. Systems cannot be compatible with typical DBMS solutions. | In order to address diversity in greater detail, this article discusses current problems, possibilities, trends, and difficulties associated with big data. We'll talk about an effective fix for the huge data variety issue. | It is extremely difficult to achieve in real-time processing. |

### 1.3. Novelty

Recently, Ref. [7] introduced a framework that contains an EMS stored on the cloud computing service as a combination of monitoring different power sources as well as managing the charging and discharging process. This model is useful for handling the best-optimized system and for controlling the switches the power hub before the need to deal with the data that comes from the power system. Therefore, this study focuses on solving this problem by use of the first step of the optimum system that will be Big Data Analytics Using Cloud Computing Based Frameworks for Power Management Systems through an overview of big data analytics using current cloud computing frameworks for power management systems, and we propose the first universally applicable framework for parallel optimization in power systems, which researchers can use to systematically describe their parallelization studies and place them in the landscape of parallel optimization without regard to the application domain, problem addressed, methodology parallelized, or technology used. The proposed approach, in particular, incorporates both algorithmic design and computational implementation challenges of parallel optimization, which are often dealt with separately in the literature. Second, we use the integrative framework to consolidate earlier research in the field of power systems on parallel optimization. Cloud computing could be a great addition to any system aiming for an optimal solution, especially intelligent power systems that solve the challenge of massive data from large-scale areas of a smart power system. All these processes and data will be saved and controlled by a cloud computing framework using a cloudsim.

### 1.4. Organizing of Paper

The flowchart is demonstrates the layout and organization of the manuscript in graphical form as follows [91]:

- Section 2 • Summary of various Paralle computing appraoches.
- Section 3 • Traditional parallel computing.
- Section 4 • Notion of cloud computing.
- Section 5 • Comparison of parallel with distributed computing
- Section 6 • Types of cloud computing utilised in the power system.
- Section 7 • Future trend
- Section 8 • Manuscript conclusions.

## 2. Parallel Computing

### 2.1. Concept of Parallel Computing

Parallel computing is a computing model relative to serial computing [92]. In this mode, multiple computing processes are performed simultaneously, which is different from the conventional computing method that is performed sequentially [93]. As shown in Figure 7, users write a series of computer instructions for specific problems to form tasks. In serial computing mode, these instructions can only be executed one by one by a single processing element (PE). Decomposition into multiple subtask instruction sets that allow simultaneous solving is mapped to multiple PEs for execution. Where a PE refers to a single-core central processing unit (CPU) or a core of a multi-core CPU, and is the basic hardware unit for executing instructions [94].



a. Serial computing    b. Parallel computing
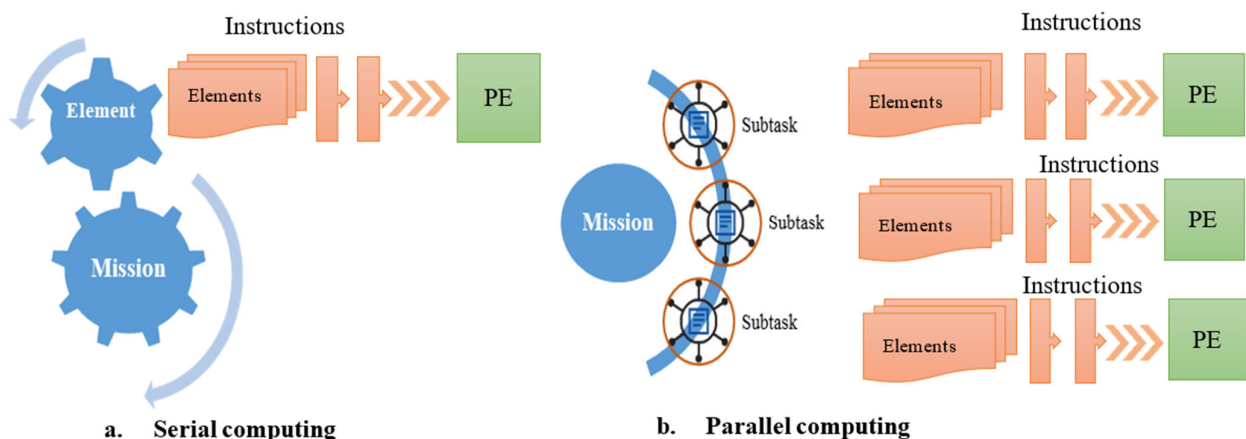
**Figure 7.** Sequential computing and parallel computing.

## 2.2. Classification of Parallel Computing Technology

There are many classification methods for parallel computing technology, and the following only introduces the classification methods related to this article [95].

Flynn Classification

The most classic Flynn classification is classification from the most basic instructions and data processing methods [96,97]. Figure 8 shows the four different types after classification.

(1)  SISD is a traditional serial computing method. Early computers fell into this category in a certain clock cycle, only one instruction is executed and only one data stream is processed;

(2)  SIMD is uses one instruction to process multiple data streams simultaneously in a certain clock cycle. Current single-core computers also fall into this category and are widely used in the fields of digital signal processing, image processing, and multimedia information processing;

(3)  MISD is uses multiple instruction streams to process a single data stream. Currently, it is only a theoretical model and has no application examples;

(4)  MIMD are currently the most popular. Multicore processors fall into this category which can execute multiple instruction streams on multiple different data streams at the same time.
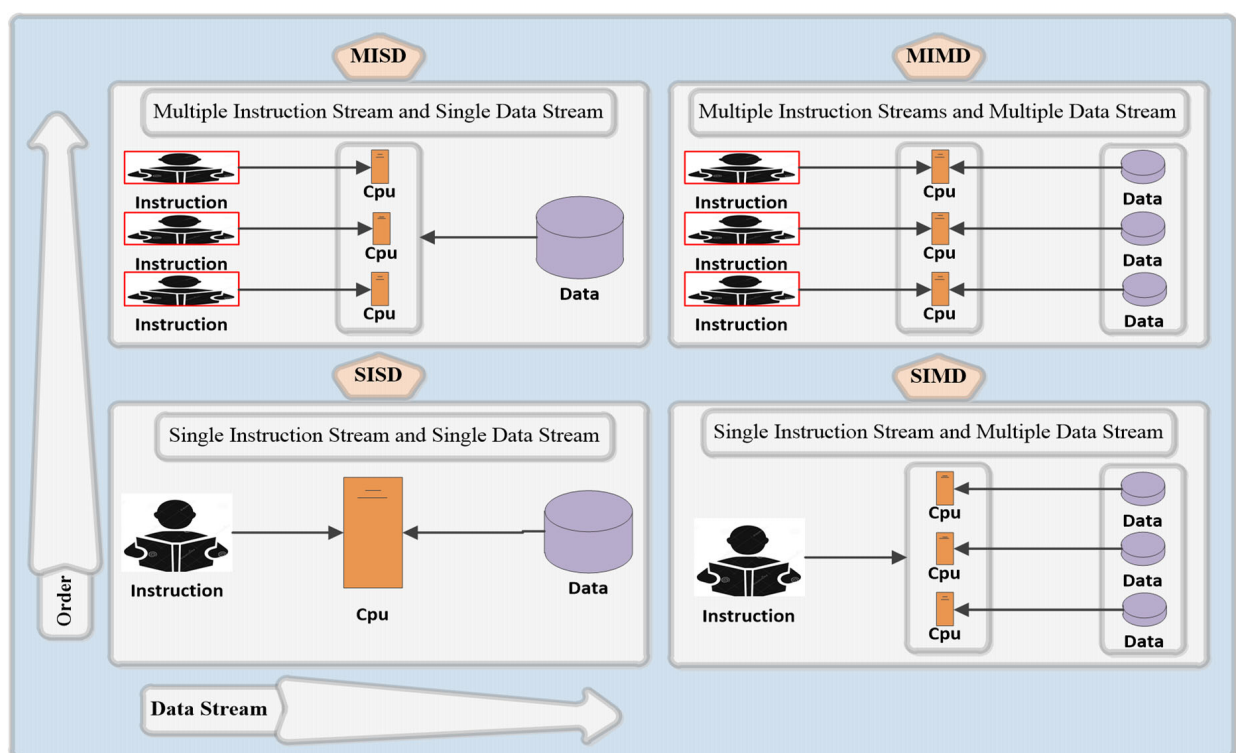


**Figure 8.** Classification by Flynn for computer systems.

## 2.3. Classification by Computational Characteristics of Applications

(1) Data-intensive applications: Such applications are faced with a huge amount of data with relatively simple calculations [98]. Usually, parallel computing is performed by dividing the data, which can also be called data parallelism [26,99].

(2) Computation-intensive applications: The amount of data processed by this type of application is not large, but the calculation is very complex [100]. Generally, this method of decomposing tasks is used for parallel computing, which can also be called task parallelism [101].

(3) Mixed-intensive applications: The amount of data and computation can use either data parallelism or task parallelism to solve the problem, or a combination of the two parallel methods can be used to solve the problem [37]. Often, partitioning data is easier to handle than breaking down tasks [102].

## 3. Shortcomings of Traditional Parallel Computing

### 3.1. Computational Complexity Issues

Real-world optimization issues in a various applications fields are often NP-hard, and even the development of (meta) heuristic optimization algorithms might necessitate significant computer resources [103]. The rising availability of powerful computer capabilities will be used to solve complex optimization issues with parallel algorithms in numerous industries, including finance, logistics, production, and design [70]. A lack of unifying frameworks for parallel optimization across techniques, application sectors, and challenges adds to the heterogeneity [37,38]. The concerns of a novel integrative framework for parallel computational optimization across optimization problems, techniques, and application domains. The framework combines algorithmic design and parallel optimization computational implementation viewpoints [104]. Unsurprisingly, the application of parallel optimization has been hesitant because (i) parallelizing algorithms is challenging in general from both the algorithmic and the computational perspective, and (ii) a viable alternative to parallelizing algorithms has been the exploitation of ongoing increases of clock speed of single CPUs of modern microprocessors [105].

Parallelization initiatives are now considerably more crucial than they were in the past due to progress. Fortunately, the need for parallelization has been realized, and parallel computing resources are now more widely available. The phenomenon: The rapid development of parallel hardware architectures and infrastructures, including multi-core CPUs and GPUs, local high-speed networks, and large data storage, as well as libraries and software frameworks for parallel programming, is the fundamental cause of this [106]. The study discovered a large number of published reviews on parallel optimization for specific issues, approaches, uses, fields of study, and technological advancements. The majority of the assessments we discovered center on parallel optimization with respect to specific approaches. Metaheuristics have dominated methodological literature evaluations, as indicated in Table 2. Several points that demand for a new literature review have been presented in earlier reviews. In the first place, we were unable to locate reviews of the most recent research on the parallelization of both precise and (meta)heuristic approaches that had been published in the years 2008–2017. Second, the categories that were employed to define and organize earlier literature were different. Aside from this heterogeneity, there aren't any frameworks that can be used to describe parallel optimization across techniques, application areas, and issues [77,78]. Finally, this has produced a fragmented overall view of what has been accomplished and what needs to be done in parallel optimization in operations research (OR). As a side effect, the heterogeneity with which parallelization studies in OR have been described in terms of algorithmic parallelization, computational parallelization, and parallelization performance is high. This is advantageous from a diversity perspective, but it also raises problems such as heterogeneity makes it often time-consuming and, in some cases, impossible for readers to identify the aforementioned parallelization characteristics of a study, to classify the study.

**Table 2.** Several parallel optimization methods and technologies use Metaheuristics.

| Algorithm | Reference |
| --- | --- |
| Tabu search (TS) | [107] |
| Simulated annealing (SA) | [108] |
| Variable neighbourhood search (VNS) | [109] |
| Greedy Randomized Adaptive Search Procedures | [110] |
| Swarm intelligence algorithms | [111] |
| Particle swarm optimization algorithms | [112] |
| Genetic algorithms (GAs) | [113] |
| Ant colony optimization algorithms | [114] |
| Scatter search | [115] |
| Several reviews have covered sets of Metaheuristics | [116] |
| Hybrid Metaheuristics | [117] |
| General-purpose computation on graphics processing units (GPC-GPU), in particular, are noteworthy parallelization approaches | [118,119] |

### 3.2. Multi-Source Heterogeneous Problem

Mostly in the domain of computers, heterogeneity typically refers to various instruction-set architectures (ISA), also called computer architecture, a machine that carries out the commands specified by that ISA, such as CPU [120]. The amount of heterogeneity in contemporary computing systems is progressively rising, with many new processors now featuring integrated logic for interacting with other devices (SATA, PCI, Ethernet, USB, RFID, radios, UARTs, and memory controllers), as well as programmable functional units and hardware accelerators (GPUs, cryptography co-processors, programmable network processors, A/V encoders/decoders, etc. [121]. When a system uses the same ISA but has a heterogeneous CPU topology, the speed of the cores varies [5]. In contrast to normal homogeneous systems, heterogeneous computing systems come with different issues [122]. All of the problems associated with homogeneous parallel processing systems are brought up by the presence of many processing units, and the degree of heterogeneity in the system can bring about non-uniformity in system development, programming techniques, and overall system capabilities [123]. Three categories of heterogeneity can be present: the first is instruction-set architectures (ISA): Binary incompatibility could result from differing instruction set designs in compute elements. The second is application binary interface (ABI): Different memory interpretations may be possible for compute elements [124]. This depends on the architecture and compiler being used and may include bendiness, calling convention, and memory layout. The third is application programming interface (API): It's possible that not all compute parts will have access to all OS and library services at once [125].

In terms of interconnecting, the compute pieces may be interconnected in a variety of ways. While some parts of a heterogeneous system may be cache-coherent, maintaining consistency and coherency may be explicitly required for other parts of the system [126]. A heterogeneous system may include CPUs with equal architecture in terms of performance, but with subtle changes in microarchitecture that affect both performance and power consumption [127]. Performance predictability issues, particularly when dealing with mixed workloads, can occasionally be caused by asymmetries in capabilities combined with opaque programming models and operating system abstractions [128,129]. Regarding Data Segmentation While data splitting on homogeneous platforms is frequently easy, it has been demonstrated that the task is NP-Complete for the general heterogeneous situation [130]. It has been demonstrated that there are ideal partitioning's for small numbers of partitions that completely balance the load and reduce communication volume [131].

### 3.3. Data-Intensive Challenges

The parallel computing facility resources are usually a single high-performance computer or high-performance computing cluster, whether they are a shared storage architecture or a distributed storage architecture or they are still logically reflected as a single high-performance computer for parallel computing users [100,132].

The computing resources and storage resources are relatively isolated physically, and the computing unit accesses data on the storage unit through a data bus or a high-speed network. For compute-intensive applications with small amounts of data, this does not pose a problem [133]. When faced with increasingly prominent data-intensive application requirements, frequent read and write accesses between computing units and storage units will become the performance bottleneck of the entire parallel system. This problem is especially true in a shared storage architecture where data is centrally stored [134]. In order to reduce the amount of communication data between processes, each worker has a local backup of all the data to be analyzed, and only the metadata of the data is sent when assigning tasks [135]. However, this precondition is not easily satisfied in clusters of distributed storage architectures, so the parallel scheme is only suitable for multi-core or many-core computers when solving data-intensive problems [136].

### 3.4. Scalability

The high-performance computer (HPC) clusters consist of high-performance hardware and are very expensive [137]. The cluster expansion generally adopts a vertical expansion method, which improves the computing performance of the cluster by replacing CPUs, expanding memory, and adding disks. However, vertical scaling is very limited, it is easy to reach the limit and the upgrade is expensive. The scalability can be achieved by horizontal expansion, such as a cluster of workstations (COW), but there are still other problems in traditional parallel computing based on this hardware [138].

### 3.5. Usability

Although the existing traditional parallel computing programming models such as Message Passing Interface (MPI) and OpenMP have been encapsulated at the bottom, and that data storage management, data division, task allocation and scheduling, data synchronization and communication, fault tolerance, and many other technical details need to be handled by users themselves, which is still very cumbersome [139]. Users are entangled in many underlying technical details while considering the application problem itself, which makes parallel programming not easy [105].

## 4. Cloud Computing

### 4.1. Concept of Cloud Computing

The concept of cloud computing, is defined by the US National Institute of Standards and Technology (NIST): as a model that can achieve convenience for obtaining the required resources (including networks, servers, storage, applications, and services) through network access on demand, and the required resources can be quickly provided or released, with little management effort or little interaction with service providers interaction [140,141]. Cloud computing is generally considered to have the following characteristics:

(1) Virtualization: Virtualization is the core technology of cloud computing, and many other features that depend on it. The application of virtualization technology can integrate heterogeneous computing resources to form a resource pool for users to access [142].

(2) Service-oriented: Cloud computing provides three levels of services, namely Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). IaaS is the lowest-level service that directly provides compute, memory, and networking equipment. Users have the greatest degree of freedom and can build their own platforms and software. PaaS is one level higher than IaaS, providing a ready-made cloud platform, saving the work of developing the platform. SaaS provides

more convenient services; users can directly use the provided software without any development [143];

(3)　Elasticity and scalability: The cloud scale can be easily expanded without affecting the cloud services currently provided externally. Resources in the cloud are infinitely desirable to users and can be automatically provisioned and reclaimed quickly on demand [144];

(4)　Reliable and universal: Cloud computing technology provides a variety of fault-tolerant mechanisms to ensure high reliability of services [145]. Data is placed with multiple copies to prevent data loss due to hardware failure [146]. Compute services that were stopped due to hardware failures can still continue elsewhere through virtual machine migration. Virtualization makes cloud computing resources transparent to users and supports applications in different industries at the same time [147];

(5)　Economies of scale: The cloud computing platform does not have high requirements for hardware facilities, and a large number of idle ordinary computers can be integrated into the resource pool through virtualization [33]. For users, it saves hardware costs and daily management costs of self-built platforms [57]. For cloud service providers, the versatility of cloud computing has greatly improved the utilization of resources, and the scale has significantly increased economic benefits [148].

### 4.2. Cloud Computing Environment

A smart grid is a heterogeneous and complex environment containing different kinds of devices, networks, systems, and data. IEC 61970 and IEC 61850 were discovered via a study on the principal Smart Grid standards and are open-source platforms based on cloud computing. IEC 61970 specifies the application program interface (API), while IEC 61850 specifies the abstract communication services interface (ACSI) [149].

Compatibility with the IEC61970 and IEC61850 standards for cloud computing technologies such as Hadoop, Spark, and Storm involves careful analysis of existing systems, translation of data, adoption of standards-compliant data formats, integration of systems, and thorough testing to ensure that the systems are functioning as intended. Because these two standards defined data models and the interface separately, the models are not uniform, and seamless communication between the substation and the control center is not possible. Whereas IEC 61970 defines the power of an information model and is widely used in enterprise integration. IEC 61850 is restricted to data exchange within substation equipment. Research revealed that Hadoop might pool idle power system resources and provide "super-computing capability" for the smart grid's data integration platform. The grid dispatch automation system's support platform and application software should be upgraded in accordance with the component interface specification (CIS) and Common Information Models (CIM) standards. The data is integrated using the IEC61970 standard and connected to the smart grid via a platform for data sharing [150]. The main user of these standards will not be a person, but a computer and they have to be machine-readable. At the same time, they are very complex documents involving thousands of different items. The full series of IEC 61850 Standards is now available as a global package. They are issued with the available associated code components. The series includes no less than 35 documents, dealing with substation automation, DER integration or cyber security, to name but a few [151]. With the development of cloud computing technology and the needs of big data processing, a number of new computing models and programming models have emerged, providing users with a basic platform for parallel programming in the cloud environment, and shielding users as much as possible from the bottom layer. The details are presented to the user through a higher-level abstract interface [152].
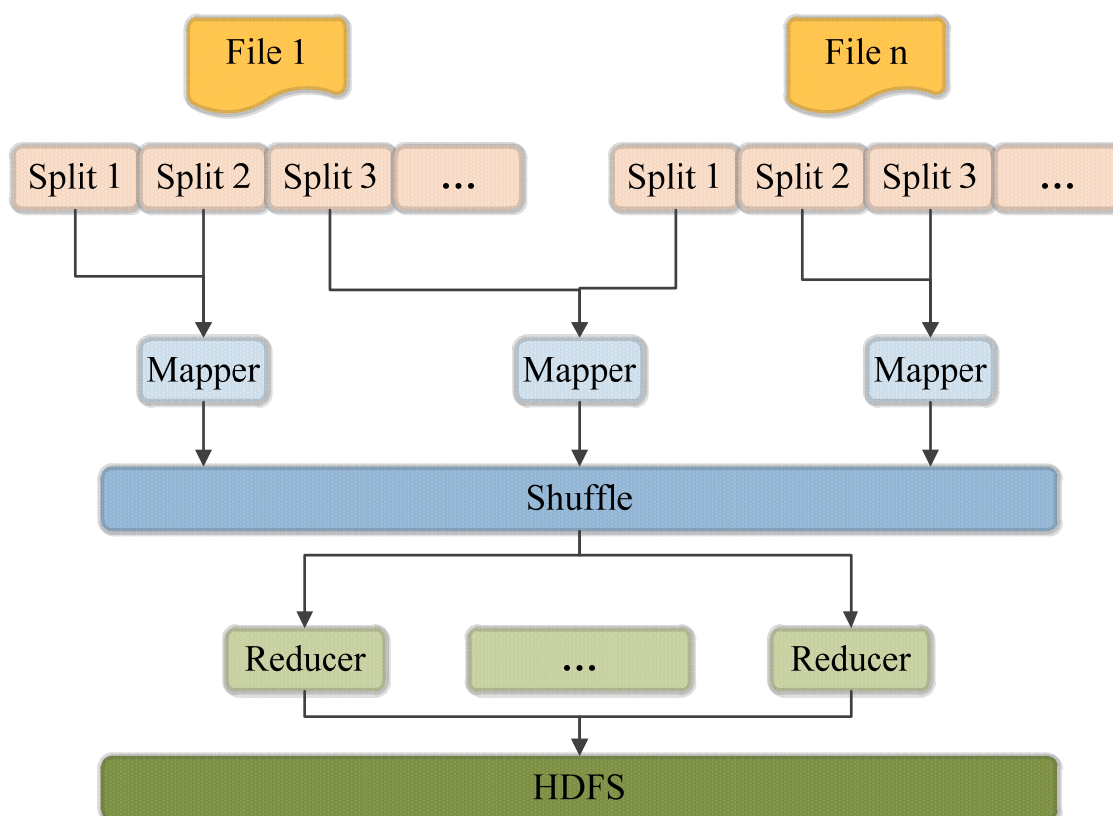
### 4.2.1. Hadoop Technology

Hadoop is a massive data processing system open sourced by the Apache Software Foundation. It includes many components for storage or processing such as the Hadoop Distributed File System (HDFS) and the MapReduce parallel computing model [153].

HDFS has the characteristics of distributed storage, high concurrent access, high fault tolerance, simple consistency and provides a reliable storage environment for parallel computing models [154]. It adopts a master-slave architecture and is built on a physical cluster connected by multiple computers through a network. The bottom layer is the local file system of the operating system. Its architecture is shown in Figure 9, shows HDFS includes a master node NameNode, a backup master node SecondaryNameNode and a set of DataNode slave nodes. The NameNode is responsible for managing the HDFS namespace, saving all metadata, and responding to client access requests. SecondaryNameNode is used to solve the single point of failure problem of Hadoop. The DataNode is responsible for the actual storage and management of redundant data blocks of files, and the data block files are actually stored on the local file system of each node [155].
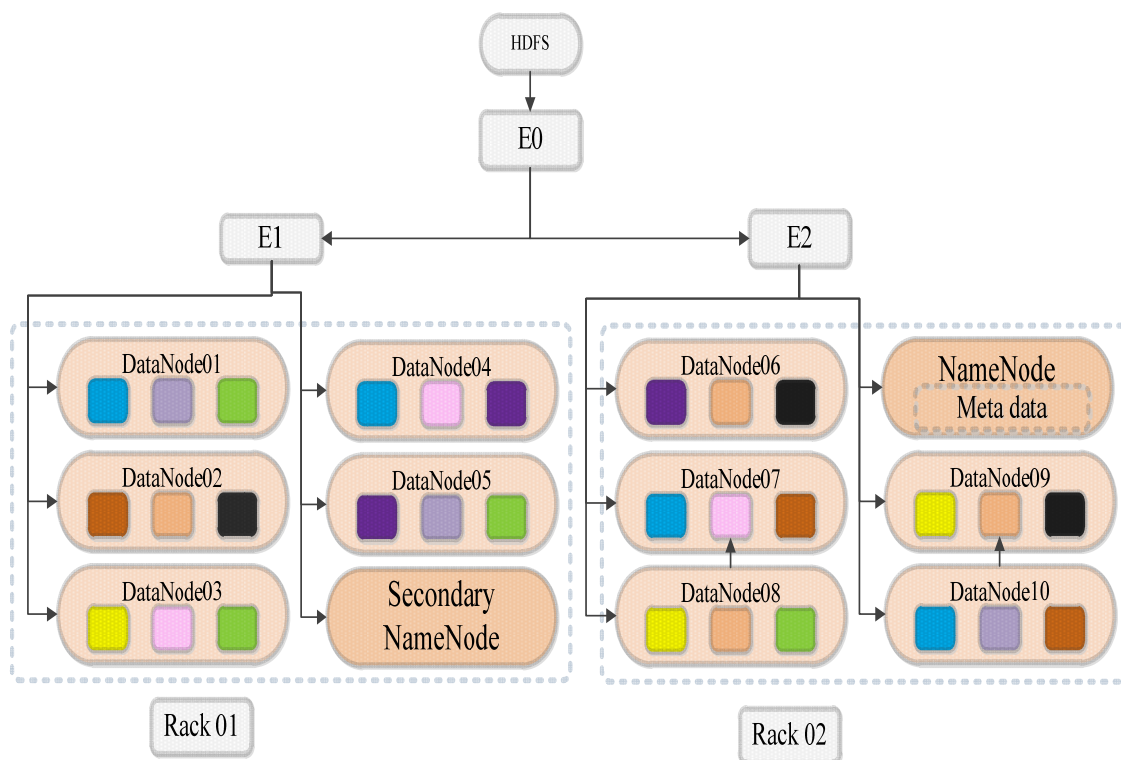


**Figure 9.** Architecture of HDFS.

The basic idea of MapReduce is basically the same as that of traditional parallel computing models such as MPI, which is to "divide and conquer" a large amount of data. The system provides two simpler interfaces, Map and Reduce which automatically completes many underlying functions such as task division and scheduling, communication, load balancing, and failure recovery [156]. Figure 10, shows the MapReduce programming mode [140]. The input file is divided into several slices (InputSplit) according to a specific format, and converted into <key, value> key-value pairs, which are input to Mapper for calculation, and the intermediate results are key–value pairs. The collection is aggregated by the Reducer after a shuffle phase, and the result is saved to HDFS. Hadoop MapReduce is mainly oriented to the batch mode of massive static data, and its real-time performance is not high [141]. The collection is aggregated by the Reducer after a shuffle phase and the result is saved to HDFS. Hadoop MapReduce is mainly oriented to the batch mode of massive static data, and its real-time performance is not high [142]. Therefore, it too quickly processes and analyze the massive historical data accumulated in the condition monitoring of power equipment, hoping to discover valuable knowledge from it [143,144].

**Figure 10.** MapReduce programming model.

4.2.2. Spark Technology

Spark is a Hadoop MapReduce-like general-purpose parallel computing framework that appeared in 2012 [157]. The difference is that it puts data (including some intermediate data) in memory for calculation, avoiding a large amount of disk I/O caused by frequent reading and writing of HDFS during the calculation process. Therefore, spark is suitable for iterative and interactive computing scenarios, and even in general application scenarios which is more efficient than Hadoop MapReduce [158]. Spark's in-memory computing features from its core abstraction, Resilient Distributed Dataset (RDD) [159]. RDDs are read-only, fault-tolerant, distributed computing, partitionable, coarse-grained transformations, and in-memory storage. Each partition of the new RDD generated during Spark's calculation process has a dependency relationship with the partition of its parent RDD due to the calculation, which is called lineage. The lost RDD partition can be regenerated from the ancestor RDD by tracing the lineage, so as to implement fault tolerance. At the same time, Spark divides the entire computing process into multiple stages according to the different dependencies between RDDs, each stage generates a job, and creates tasks in units of RDD partitions and distributes them in multiple stages. Figure 11, shows the process in parallel on two computing nodes, and the Spark computing model [160]. The computing model is richer and more flexible than the single MapReduce model of Hadoop, and is compatible with various data sources such as HDFS, HBase, and Hive [161]. At present, Spark has been widely used in Internet companies such as Amazon, Yahoo, and Taobao [162], and it is still in the research stage in the power industry, and research still needs to be carried out in combination with typical application scenarios [163]. Based on Spark, is incapable of how the complex signal processing algorithm can perform fast calculation when the amount of data is large, which makes up for the application scenarios that Hadoop MapReduce [164].
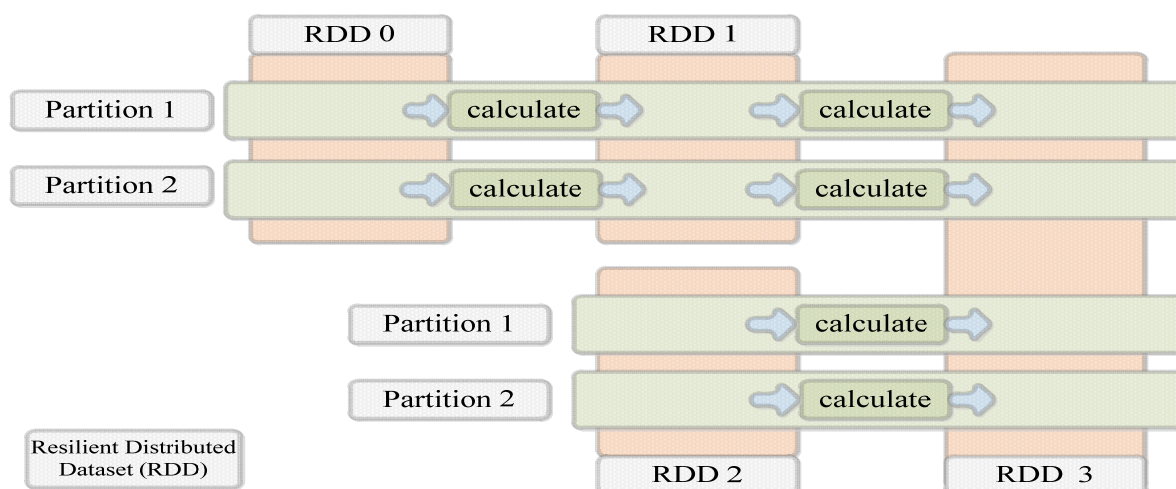
**Figure 11.** The computing model based on resilient distributed dataset (RDD).

### 4.2.3. Storm Technology

Storm is primarily geared towards real-time analytics on large-scale, uninterrupted streams of data, unlike Hadoop which focuses on batch processing of massive amounts of data. Although Spark Streaming can also achieve the function of stream computing by decomposing batch jobs, its latency is longer than Storm [165]. Storm also adopts a master-slave architecture and uses ZooKeeper to coordinate the entire cluster. Where the master node is called the control node and runs the Nimbus daemon, which is responsible for publishing topology programs, distributing tasks and monitoring cluster status [166]. The slave node is called a worker node, running the Supervisor daemon, which is responsible for accepting the assigned tasks and starting the Java Virtual Machine (JVM) process worker to execute that shows in Figure 12a [167,168]. Figure 12b shows the topology where Spout is the data entry of the topology, connecting to an external data source and converting the data into tuples that are sent to Bolt. The processing logic for tuples is encapsulated in Bolts, and after the processing is completed, tuples can be transmitted to subsequent Bolts [169]. The Spout and Bolt components are linked by a stream grouping strategy and can be configured as multiple instances to achieve parallel processing. Each instance will eventually form a task to be scheduled for execution [170].
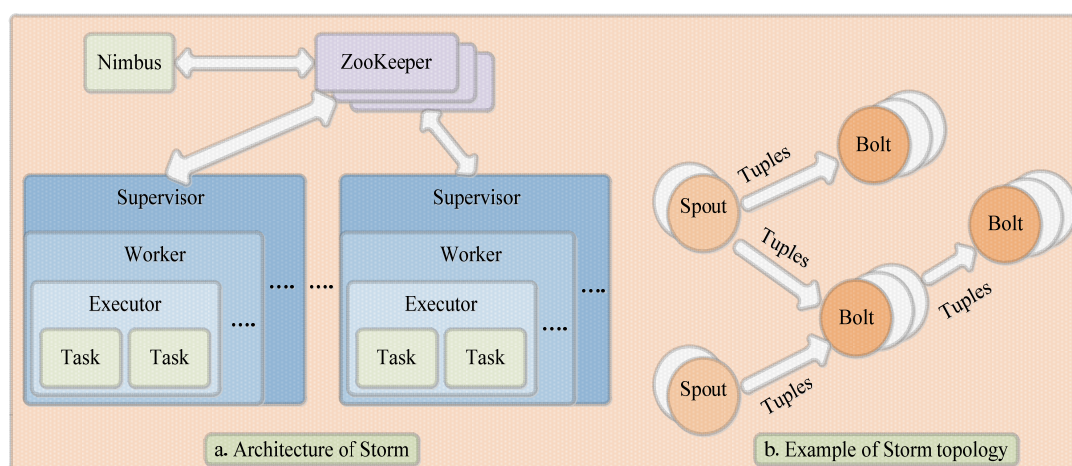


**Figure 12.** Architecture and topology of storm.

## 5. Comparison of Parallel Computing with Cloud Computing

Cloud computing is the fusion of many technologies, including parallel computing technology, so it is not equivalent to parallel computing, and its connotation richer [36,171]. The parallel computing usually refers to a specially designed parallel computer, while cloud computing combines multiple ordinary computers to achieve the purpose of improving computing performance [172,173]. In a broad sense, the computing technology it adopts also belongs to parallel computing. The technical point of parallel computing only focuses on computing and ignores data storage [174]. This is because traditional parallel programming models such as MPI are designed for high-performance computing, and small amount of data [174]. Cloud computing includes storage and computing, and the two cooperate with each other. For example, in the Hadoop system, data is stored in a distributed manner, and then the calculation is moved to the location of the data for execution, because mobile computing is more efficient than moving data [175]. From the perspective of applicable fields, parallel computing is suitable for the scientific computing field with high-performance computing requirements [2]. It is oriented to computing-intensive applications and requires users to have high professional quality in order to be able to deal with many low-level details [105]. The cloud computing provides services to users through three different levels, which is easier for users to use. The cloud computing is a key technology for big data processing and is suitable for data-intensive applications, but the large system management and maintenance costs make it not good at computing-intensive applications with a small amount of data [25]. To sum up, parallel computing and cloud computing technologies have a wide range of applications in various fields including the power industry [139,176]. The two are complementary rather than mutually exclusive, and each has different application scenarios [177].

*Distributed Cloud Computing and Parallel Computing*

A new method to connect data and applications supplied from several places is distributed cloud computing. A shared resource geographically dispersed among several users or systems is referred to as distributed, according to Table 2. The ability to execute several jobs simultaneously is a key characteristic of cloud computing, which also aims to reduce CPU consumption, cut down on switching times, reduce waiting times for data processing, increase server throughput, and enhance data processing and communication speed. Another feature makes using any cloud application to communicate with users in various places simple for users. The final crucial component is enhancing server performance since communication performance is crucial.

After reviewing each reference included in Table 3, it was determined that [178] was the finest work on distributed cloud computing since it covered the most features.

**Table 3.** Distributed Cloud Computing.

| Feature | [178] | [179] | [180] | [181] | [182] | [183] | [184] |
|---|---|---|---|---|---|---|---|
| Reduction in CPU use | ✓ | | | | | | |
| Reduce multiple-process tasks | ✓ | ✓ | | ✓ | | ✓ | ✓ |
| Reduce waiting times. | ✓ | | | | | | |
| enhanced use of resources | ✓ | | | | | | |
| Enhance server efficiency | ✓ | | | | | | ✓ |
| Increasing server performance | ✓ | | | | | | |
| Balance of loads | | | ✓ | | | | |
| performance in terms of costs | | | ✓ | | | | |
| lessen the demand on the memory | | ✓ | | ✓ | | | |
| Create a cloud architecture program | | | | ✓ | ✓ | | ✓ |
| enhance inter-humans communication | | | | | ✓ | | |
| Increasing safety | | | | | ✓ | | |
| Increasing effectiveness and creating a system expand | | | | | ✓ | | |
| Increase the scope of cloud computing | | | | | | ✓ | |
| Comparison of the benefits and drawbacks of MPI, oprnMPI, and MapReduce | | | | | | ✓ | |

The three different methods of parallel processing are distributed, shared, and hybrid memory systems. A distributed parallel computing focus identified certain key criteria, as indicated in Table 4. The important feature mentioned in more than one resource is improving performance using the load balancing technique through distributing the process and making a balance between servers for processing the jobs and improving the performance of our distributed system. Another feature is minimizing resource costs because when we divide the load among servers, we can minimize the resource cost such as CPU, memory, and storage. Since it is preferable to employ a system for handling user requests with a low response time, all of the references apply this principle by recommending a method for distributed parallel computing based on that factor. After looking through the references in this work, it was determined that Ref. [185] was superior since it provides a wide range of characteristics, such as load balancing, enhancing system performance, and lowering reaction time and resource cost.

**Table 4.** Distributed Parallel Processing.

| Features | [181] | [182] | [186] | [187] | [188] | [189] | [185] | [190] |
|---|---|---|---|---|---|---|---|---|
| Utilize load balancing to increase performance | ✓ | ✓ | ✓ | ✓ |  | ✓ | ✓ | ✓ |
| Requesting each node's status |  |  |  |  | ✓ |  |  |  |
| developed a reduce reaction time-based algorithm |  |  |  | ✓ | ✓ | ✓ |  |  |
| Decrease requests on resources that are available |  |  |  |  | ✓ |  |  |  |
| Minimize server-to-server interaction and processing |  |  |  |  | ✓ |  |  |  |
| Take every resource's load into account |  |  |  |  |  | ✓ |  |  |
| Optimize CPU throughput |  |  |  |  |  |  | ✓ |  |
| Reduce productivity |  |  |  |  |  |  | ✓ |  |
| Reduce reaction time. |  |  |  |  |  |  | ✓ |  |
| Reduce long waits |  |  |  |  |  |  | ✓ |  |
| lessen the cost of resources |  |  |  | ✓ |  |  | ✓ |  |
| Ensure error tolerance and QoS |  |  |  |  |  |  | ✓ |  |
| Effective implementation of parallelism | ✓ |  |  |  |  |  |  |  |
| Improving the way jobs are arranged |  |  |  | ✓ |  |  |  |  |
| Improve allocation of resources |  |  |  | ✓ |  |  |  |  |
| Faster performance with better outcomes |  |  |  | ✓ |  |  |  |  |

## 6. The Application Basis of Cloud Computing in a Power System

In the early days of the Internet, the cost of hardware was relatively high. With the continuous development of information and network technology, the data generated and processed by the Internet has grown exponentially. In order to cope with these changes, the investment in hardware equipment has to be increased input cost [38]. However, although the hardware investment cost is high, the scalability of the system is very poor, the information transmission efficiency is low, and the performance difference between devices is also the processing effect is not good [171]. The theoretical basis of cloud computing is to virtualize equipment and services, and then interconnect several distributed nodes on the cloud platform to realize the superposition and coupling of computing resources [191]. In order to better complete the processing of massive data the distribute a large number of processing tasks to virtual nodes in the resource pool [29,166]. Cloud is a very broad concept with many types, and it is usually classified according to different service objects and service types [192].

(1) Public cloud

As the name suggests, it is a cloud service that is open to the public. It is large-scale, low-cost, and the most popular cloud service for the public. The most typical application is Amazon Web Services ("AWS"). The app provides a complete set of infrastructure and cloud solutions to customers around the world. AWS provides users with a complete set of cloud computing services, which can help enterprises reduce IT investment costs and maintenance costs and easily migrate to the cloud [193].

(2) Private cloud

It is a cloud that does not provide services publicly and is used within a group or organization. Provide private cloud services to internal users. Because they cannot be used publicly, most firewalls are set up [194]. The typical representative of private cloud is the Blue Cloud plan launched by IBM. Blue Cloud is based on open standards and open source software powered by IBM software, systems technologies and services [195]. The Blue Cloud developed by more than 200 IBM researchers around the world, will help clients quickly and easily explore cloud computing infrastructure for extreme-scale computing [196].

(3) Hybrid cloud

That is, the combination of public cloud and private cloud is between private and public, such as Amazon's virtual private cloud (VPC) [54]. A VPC is a dynamically provisioned pool of public cloud computing resources that requires the use of encryption protocols, tunneling protocols, and other security procedures to transfer data between private enterprises and cloud service providers [197–200]. The services provided by each layer are as follows:

(1) **Application layer**

The application layer provides users with various application software and services required by a friendly user interface [201]. The application layer directly faces customer needs and provides enterprise customers with enterprise applications such as enterprise resource planning (ERP) and customer relationship management (CRM) [202], and office automation (OA) [203].

(2) **Platform layer**

The platform layer provides services for users who can use the platform to realize the value they want to achieve [204].

(3) **Infrastructure layer**

This layer provides infrastructure-level services, that is, the establishment of the cloud computing platform infrastructure is directly open to users, so that they can use the powerful storage and computing capabilities of cloud computing. Users can directly store files and run calculations in the cloud, and also the infrastructure can be allocated independently, which is equivalent to the user having a scalable computer with large storage space and supercomputing performance through the terminal [205].

In general, the advantage of cloud computing is that it can integrate various resources without special requirements for these computing resources and does not require a specific computer with powerful performance [33]. Users can easily obtain cloud services through ordinary terminals and utilize the capabilities of supercomputers. In the same way, an intelligent cloud system can also be built in the power system. Users can monitor and track the power system directly on the mobile terminal to improve the convenience of the power system [206–208].

At present, the smart grid construction the structure diagram shown in Figure 13. Compared with the previous computing model, cloud computing has made many qualitative breakthroughs. Its scale is large, but it has strong reliability [209,210]. The expansion type, which integrates these features together makes it have unparalleled advantages [211]. Google is the initiator of cloud computing ideas. Later, on the basis of Google's previous work, developed the significant Hadoop open-source cloud computing platform. The key

to Hadoop is to include two major systems: the distributed file system HDFS and the distributed computing framework [212,213].
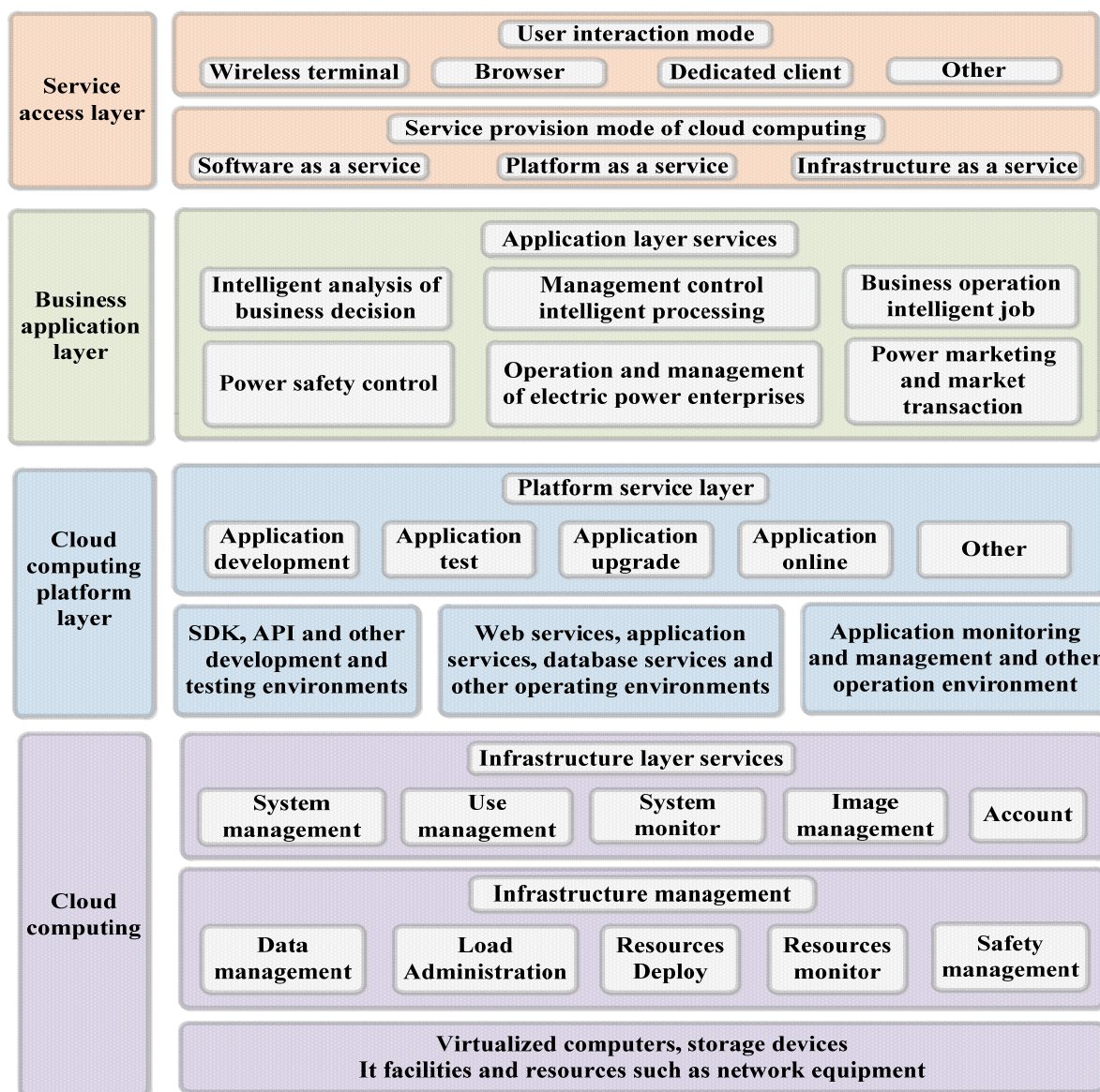
| Service access layer | **User interaction mode** |
| --- | --- |
| | Wireless terminal / Browser / Dedicated client / Other |
| | **Service provision mode of cloud computing** |
| | Software as a service / Platform as a service / Infrastructure as a service |

**Figure 13.** Structure of grid intelligent information platform.

## 7. Future Trend

Cloud-based tools and technologies for microgrid control are a set of software and hardware solutions that are designed to monitor and manage the power usage and distribution of microgrids. The architecture of cloud computing provides robust and efficient management of data, computing resources, and applications, with features like fault isolation cloud computing, which refers to the ability of the system to isolate a failure or malfunction in one part of the system so that it does not affect the entire system. Furthermore, system self-healing refers to the ability of the cloud computing system to automatically detect and recover from failures without human intervention. Where in a cloud computing environment, micro-grid control can be achieved by using cloud-based tools to monitor and control the power usage and distribution of the micro-grid. These systems can provide real-time data and insights into the power usage of individual devices and systems, allowing for more efficient and effective management of the microgrid. Local data interaction can be achieved through the use of APIs (Application Programming Interfaces) and other

communication protocols that allow for seamless integration between the cloud and local systems. APIs can be used to access cloud resources and services, as well as to exchange data between the cloud and local systems. This allows for more efficient and effective use of resources and provides a convenient way to access cloud services from local systems.

In the future, power system cloud computing will have broad applications, which can be described as "the future can be expected" as shown in Figure 14. Its effective application can provide a large number of high value-added services inside and outside the industry and has high practical value for the profitability and control cost level of power companies [214].
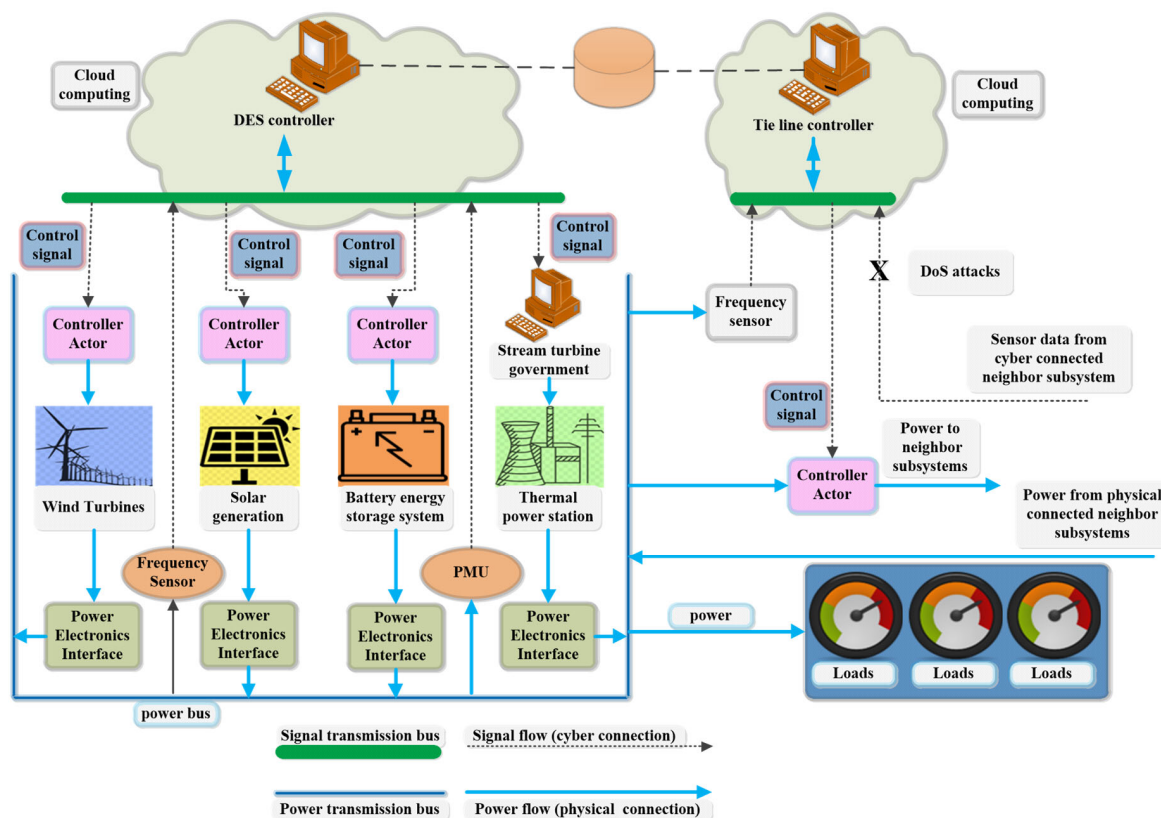


**Figure 14.** Schematic diagram of a subsystem.

The subsystem model has long-term benefits for national defense, military, and medical care. The application value in the power grid alone is immeasurable.

- First, it will help grid companies to carry out grid operation and maintenance monitoring and improve response sensitivity [215]. Use the data collected from the power system to monitor, control, or adjust the power generation, load, and fault status in the network, and respond accordingly when there is an error or an upgrade in the power grid [209,216].
- Secondly, it will help grid companies conduct special analysis on equipment maintenance, operation and maintenance, improve system reliability, power supply qualification rate, reduce costs, and reduce power outages [217]. In the field of power grid maintenance, operation and maintenance, through the selection of key indicators of power equipment from the three aspects of safety, benefit, and cost, analysis of the mutual influence of "safety", "benefit" and "cost" in maintenance management, coordination of the three these factors are comprehensively optimized, and at the same time, real-time online monitoring of the maintenance indicators of power grid enterprises is realized, providing guidance and services for the company's maintenance strategy formulation [218–220].

Some points that need improvement due to the limitations of previous work are personal ability and data source:

➢ The data capacity is still not large enough, and the data dimension change interval is small [221]; the electricity data comes from residential electricity consumption and does not include electricity consumption data in other areas such as industrial electricity [222];

➢ The coverage of parallel algorithm design is relatively narrow, and the application range in power system data processing is not wide enough. However, with the increasing informatization of the power system and the continuous quantification of power data, the application scope of data mining technology continues to expand. Parallel algorithms can be designed in more aspects to enhance the data processing effect and carry out all around power system production and dispatching [223,224];

➢ To design a comprehensive cloud computing platform architecture based on cloud computing and big data processing technology, which provides a reference plan for the construction of the computing platform of the power equipment monitoring center [225,226];

➢ To improve the real-time response speed of online monitoring data of power equipment, a real-time processing framework for streaming data based on Storm is needed, and an incremental variable prediction model classification method [227–229].

## 8. Conclusions

This review article briefly expounds on the basic concepts and programming models of parallel computing and analyzes the shortcomings of traditional parallel computing in the face of current big data scenarios with a specific application example. Then, cloud computing solutions are introduced for the field of power equipment condition monitoring under the background of big data, the concepts and characteristics of cloud computing are introduced, and emerging parallel programming models, such as Hadoop, Spark, and Storm, are briefly described. Finally, for different application scenarios in the field of power equipment condition monitoring, a comprehensive cloud computing platform architecture is designed to meet the multi-level real-time requirements by drawing on the experience of the Internet field.

There is a need to improve the parallel detection algorithm for power load anomaly detection efficiency. The detection of the abnormal value is only the first step in the analysis of the abnormal value of the power. After the abnormal value is detected, the cause of the abnormal data can be quickly analyzed, such as economic reasons, regular maintenance, equipment failure, illegal electricity stealing, temperature, or sudden environmental changes. Forming a power load abnormal detection response model to provide guiding opinions for demand-side power supply is one direction of future research. For power load data prediction, the further parallelization of the power load data prediction algorithm is another key direction of the next work. As an important part of power data analysis, research on power load anomaly detection and prediction methods in the cloud computing environment still has huge research opportunities waiting for us to study.

**Author Contributions:** A.H.A.A.-J.: drafted the original manuscript, conceptualization, literature analysis. R.C.M. and M.K.H.: conceptualization and methodology, scientific advice, supervision, drafting, review, and editing. J.K.S.P.: scientific advice, supervision, M.J.S.: scientific advice, supervision. All authors have read and agreed to the published version of the manuscript.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Lin, W.; Wu, G.; Wang, X.; Li, K. An artificial neural network approach to power consumption model construction for servers in cloud data centers. *IEEE Trans. Sustain. Comput.* **2019**, *5*, 329–340. [CrossRef]
2. Liu, Y.; Liang, S.; He, C.; Zhou, Z.; Fang, W.; Li, Y.; Wang, Y. A Cloud-computing and big data based wide area monitoring of power grids strategy. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *677*, 042055. [CrossRef]
3. Raza, M.Q.; Khosravi, A. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renew. Sustain. Energy Rev.* **2015**, *50*, 1352–1372. [CrossRef]
4. Santos, M.A.G.; Munoz, R.; Olivares, R.; Filho, P.P.R.; Del Ser, J.; de Albuquerque, V.H.C. Online heart monitoring systems on the internet of health things environments: A survey, a reference model and an outlook. *Inf. Fusion* **2020**, *53*, 222–239. [CrossRef]
5. Sharma, S.; Kotturu, P.K.; Narooka, P.C. Implication of IoT components and energy management monitoring. In *Swarm Intelligence Optimization: Algorithms and Applications*; John Wiley & Sons, Inc.: New York, NY, USA, 2020; pp. 49–65. [CrossRef]
6. Xu, Y.; Liu, C.C.; Schneider, K.P.; Tuffner, F.K.; Ton, D.T. Microgrids for service restoration to critical load in a resilient distribution system. *IEEE Trans. Smart Grid* **2018**, *9*, 426–437. [CrossRef]
7. AL-Jumaili, A.H.A.; Al Mashhadany, Y.I.; Sulaiman, R.; Alyasseri, Z.A.A. A Conceptual and Systematics for Intelligent Power Management System-Based Cloud Computing: Prospects, and Challenges. *Appl. Sci.* **2021**, *11*, 9820. [CrossRef]
8. Miu, W.; Zhang, Z.; Wang, X.; Hou, J.; Sun, Y. A Real-time Detection Framework for Abnormal Devices in the Power Internet of Things. *J. Phys.* **2022**, *2166*, 012057. [CrossRef]
9. Jain, S.; Chandrasekaran, K. Industrial automation using internet of things. In *Research Anthology on Cross-Disciplinary Designs and Applications of Automation*; IGI Global: New York, NY, USA, 2022; pp. 355–383.
10. Zaheeruddin; Singh, K.; Amir, M. Intelligent Fuzzy TIDF-II Controller for Load Frequency Control in Hybrid Energy System. *IETE Tech. Rev.* **2021**, *39*, 1355–1371. [CrossRef]
11. Manimegalai, C.T.; Kalimuthu, K.; Gauni, S. Toward integrating bidirectional multiband data and power transmission using double clad optical fibers for the next generation disaster resilient managing communication systems. *Microw. Opt. Technol. Lett.* **2022**, *64*, 816–820. [CrossRef]
12. Jia, X. Research on network abnormal data flow mining based on improved cluster analysis. *Distrib. Parallel. Databases* **2021**, *40*, 797–813. [CrossRef]
13. Ahamed, R.; Habeeb, A.; Nasaruddin, F.; Gani, A.; Abaker, I.; Hashem, T.; Ahmed, E.; Imran, M. International Journal of Information Management Real-time big data processing for anomaly detection: A Survey. *Int. J. Inf. Manag.* **2018**, *45*, 289–307.
14. Lin, Z.; Xiangping, L.; Wenzhong, C.; Haoyue, P. Computer aided analysis and control of power system based on data mining technology. In Proceedings of the 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers, Dalian, China, 14–16 April 2021; pp. 1258–1261. [CrossRef]
15. Malik, M.I. Cloud Computing-Technologies. *Int. J. Adv. Res. Comput. Sci.* **2018**, *9*, 379–384. [CrossRef]
16. Li, X.; Liu, H.; Wang, W.; Zheng, Y.; Lv, H.; Lv, Z. Big data analysis of the internet of things in the digital twins of smart city based on deep learning. *Futur. Gener. Comput. Syst.* **2022**, *128*, 167–177. [CrossRef]
17. Gharehpasha, S.; Masdari, M.; Jafarian, A. Power efficient virtual machine placement in cloud data centers with a discrete and chaotic hybrid optimization algorithm. *Cluster Comput.* **2021**, *24*, 1293–1315. [CrossRef]
18. Li, Y.; Shi, F.; Zhang, H. Panoramic synchronous measurement system for wide-area power system based on the cloud computing. In Proceedings of the 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, China, 31 May–2 June 2018; pp. 764–768. [CrossRef]
19. Wang, Y.; Chen, S. An approach to smart grid online data mining based on cloud computing. *Int. J. Simul. Syst. Sci. Technol.* **2016**, *17*, 17.1–17.5. [CrossRef]
20. Sun, C.; Wang, X.; Zheng, Y.; Zhang, F. A framework for dynamic prediction of reliability weaknesses in power transmission systems based on imbalanced data. *Int. J. Electr. Power Energy Syst.* **2020**, *117*, 105718. [CrossRef]
21. Sliwa, J. Assessing complex evolving cyber-physical systems (case study: Smart medical devices). *Int. J. High Perform. Comput. Netw.* **2019**, *13*, 294. [CrossRef]
22. Urooj, S.; Singh, T.; Amir, M.; Tariq, M. Optimal Design of Power Transformer with Advance Core Material using ANSYS Technique. *Eur. J. Electr. Eng. Comput. Sci.* **2020**, *4*, 1–17. [CrossRef]
23. Bekemeier, B.; Park, S.; Whitman, G. Challenges and lessons learned in promoting adoption of standardized local public health service delivery data through the application of the Public Health Activities and Services Tracking model. *J. Am. Med. Inform. Assoc.* **2019**, *26*, 1660–1663. [CrossRef]
24. Sagheer, N.S.; Yousif, S. Canopy with &-means Clustering Algorithm for Big Data Analytics. *AIP Conf. Proc.* **2021**, *2334*, 070006. [CrossRef]
25. Tao, D.; Lin, Z.; Wang, B. Load feedback-based resource scheduling and dynamic migration-based data locality for virtual hadoop clusters in openstack-based clouds. *Tsinghua Sci. Technol.* **2017**, *22*, 149–159. [CrossRef]

26. Dong, F.; Guo, X.; Zhou, P.; Shen, D. Task-aware flow scheduling with heterogeneous utility characteristics for data center networks. *Tsinghua Sci. Technol.* **2019**, *24*, 400–411. [CrossRef]

27. Javed, A.; Larijani, H.; Ahmadinia, A.; Gibson, D. Smart Random Neural Network Controller for HVAC Using Cloud Computing Technology. *IEEE Trans. Ind. Inform.* **2017**, *13*, 351–360. [CrossRef]

28. Eddoujaji, M.; Samadi, H.; Bohorma, M. Data Processing on Distributed Systems Storage Challenges. In *Smart Innovation, Systems and Technologies*; Springer: Berlin, Germany, 2022; Volume 237, pp. 795–811.

29. Yang, C.; Huang, Q.; Li, Z.; Liu, K.; Hu, F. Big Data and cloud computing: Innovation opportunities and challenges. *Int. J. Digit. Earth* **2017**, *10*, 13–53. [CrossRef]

30. Zhang, Y.; Lang, Y.; Yang, S.; Zhao, K.; Liu, P.; Han, F. Big Data Storage Technology suitable for the Operation and Maintenance of New Generation Power Grid Dispatching Control System Operation. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *300*, 042084. [CrossRef]

31. Kulkarni, N.; Lalitha, S.V.N.L.; Deokar, S.A. Real time control and monitoring of grid power systems using cloud computing. *Int. J. Electr. Comput. Eng.* **2019**, *9*, 941. [CrossRef]

32. Ahmed, M.H.; Tiun, S.; Omar, N.; Sani, N.S. Short Text Clustering Algorithms, Application and Challenges: A Survey. *Appl. Sci.* **2022**, *13*, 342. [CrossRef]

33. Alarifi, A.; Dubey, K.; Amoon, M.; Altameem, T.; El-Samie, F.E.A.F.E.A.; Altameem, A.; Sharma, S.C.C.; Nasr, A.A.A. Energy-Efficient Hybrid Framework for Green Cloud Computing. *IEEE Access* **2020**, *8*, 115356–115369. [CrossRef]

34. Ali, A.H.; Mohammad, O.K.J. Impacting of the E-Platforms on the 4.0th Industrial Educational Revolution. In Proceedings of the Pervasive Health: Pervasive Computing Technologies for Healthcare, Cairo, Egypt, 24–26 March 2019.

35. Al-Nuaymy, A.I.; Hamed, W.A.; Ali, A.H. The Role of Social Media in Enhancing the Learning Process of Iraq Students. *ARPN J. Eng. Appl. Sci.* **2019**, *14*, 7142–7153. [CrossRef]

36. Li, X.; Zhuang, W.; Zhang, H. Short-term Power Load Forecasting Based on Gate Recurrent Unit Network and Cloud Computing Platform. *Pervasive Health Pervasive Comput. Technol. Healthc.* **2020**, *29*, 1–6. [CrossRef]

37. Yuan, J. An Anomaly Data Mining Method for Mass Sensor Networks Using Improved PSO Algorithm Based on Spark Parallel Framework. *J. Grid Comput.* **2020**, *18*, 251–261. [CrossRef]

38. Deng, C.; Liu, J.; Liu, Y.; Yu, Z. Cloud computing based high-performance platform in enabling scalable services in power system. In Proceedings of the 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Changsha, China, 13–15 August 2016; pp. 2200–2203. [CrossRef]

39. Hu, L.; Yang, C.; Meng, X.; Pang, P. Short-Term Power Load Forecasting Method Based on Cloud Computing and Intelligent Algorithms. In *Lecture Notes on Data Engineering and Communications Technologies*; Springer: Berlin, Germany, 2022; Volume 98, pp. 518–525.

40. Litchfield, A.T.; Althouse, J. A systematic review of cloud computing, big data and databases on the cloud. In Proceedings of the Americas Conference on Information Systems, Savannah, GA, USA, 7–9 August 2014.

41. Wan, L.; Zhang, G.; Li, H.; Li, C. A Novel Bearing Fault Diagnosis Method Using Spark-Based Parallel ACO-K-Means Clustering Algorithm. *IEEE Access* **2021**, *9*, 28753–28768. [CrossRef]

42. Aziz, K.; Zaidouni, D.; Bellafkih, M. Real-time data analysis using Spark and Hadoop. In Proceedings of the 2018 International Conference on Optimization and Applications, ICOA 2018, Mohammedia, Morocco, 26–27 April 2018; pp. 1–6. [CrossRef]

43. Zhu, T.; Xiao, S.; Zhang, Q.; Gu, Y.; Yi, P.; Li, Y. Emergent Technologies in Big Data Sensing: A Survey. *Int. J. Distrib. Sens. Networks* **2015**, *2015*. [CrossRef]

44. Jiang, H.; Wang, K.; Wang, Y.; Gao, M.; Zhang, Y. Energy big data: A survey. *IEEE Access* **2016**, *4*, 3844–3861. [CrossRef]

45. Sami, S.; Sael, N. Extract Five Categories CPIVW from the 9V's Characteristics of the Big Data. *Int. J. Adv. Comput. Sci. Appl.* **2016**, *7*, 254–258. [CrossRef]

46. Jayaprakash, J.S.; Balasubramanian, K.; Sulaiman, R.; Hasan, M.K.; Parameshachari, B.D.; Iwendi, C. Cloud data encryption and authentication based on enhanced Merkle hash tree method. *Comput. Mater. Contin.* **2022**, *72*, 519–534. [CrossRef]

47. Lin, C.-H.; Amir, M.; Tariq, M.; Shahvez, M.; Alamri, B.; Alahmadi, A.; Siddiqui, M.; Beig, A.R. Comprehensive Analysis of IPT v/s CPT for Wireless EV Charging and Effect of Capacitor Plate Shape and Foreign Particle on CPT. *Processes* **2021**, *9*, 1619. [CrossRef]

48. Jimeno, J.; Anduaga, J.; Oyarzabal, J.; De Muro, A.G. Architecture of a microgrid energy management system. *Eur. Trans. Electr. Power* **2011**, *21*, 1142–1158. [CrossRef]

49. Nizeyimana, E. *Remote Sensing and GIS Integration*; McGraw-Hill Professional Publishing: New York, NY, USA, 2020.

50. Sharma, B.S.S.P.M.; Kamath, H.R.; Brahmaiah Rama, V.S. Modelling of cloud based online access system for solar charge controller. *Int. J. Eng. Technol.* **2018**, *7*, 58–61. [CrossRef]

51. Kong, X.; Zheng, Y.; Ouyang, M.; Li, X.; Lu, L.; Li, J.; Zhang, Z. Signal synchronization for massive data storage in modular battery management system with controller area network. *Appl. Energy* **2017**, *197*, 52–62. [CrossRef]

52. Kabalcı, E.; Kabalcı, Y.; Siano, P. Design and implementation of a smart metering infrastructure for low voltage microgrids. *Int. J. Electr. Power Energy Syst.* **2022**, *134*, 107375. [CrossRef]

53. Amir, M.; Zaheeruddin. ANN Based Approach for the Estimation and Enhancement of Power Transfer Capability. In Proceedings of the 2019 International Conference on Power Electronics, Control and Automation (ICPECA), New Delhi, India, 16–17 November 2019; pp. 1–6.

54. Yuan, H.; Bi, J.; Zhou, M. Spatiotemporal Task Scheduling for Heterogeneous Delay-Tolerant Applications in Distributed Green Data Centers. *IEEE Trans. Autom. Sci. Eng.* **2019**, *16*, 1686–1697. [CrossRef]

55. Bitzer, B.; Gebretsadik, E.S. Ensuring future clean electrical energy supply through cloud computing. In Proceedings of the 2015 International Conference on Clean Electrical Power (ICCEP), Taormina, Italy, 16–18 June 2015; 155–159. [CrossRef]

56. Ma, F.; Luo, X.; Litvinov, E. Cloud Computing for Power System Simulations at ISO New England-Experiences and Challenges. *IEEE Trans. Smart Grid* **2016**, *7*, 2596–2603. [CrossRef]

57. Hassen, H.B.; Ayari, N.; Hamdi, B. A home hospitalization system based on the Internet of things, Fog computing and cloud computing. *Inform. Med. Unlocked* **2020**, *20*, 100368. [CrossRef] [PubMed]

58. Rao, S.N.V.B.; Yellapragada, V.P.K.; Padma, K.; Pradeep, D.J.; Reddy, C.P.; Amir, M.; Refaat, S.S. Day-Ahead Load Demand Forecasting in Urban Community Cluster Microgrids Using Machine Learning Methods. *Energies* **2022**, *15*, 6124. [CrossRef]

59. Xia, D.; Ning, F.; He, W. Research on Parallel Adaptive Canopy-K-Means Clustering Algorithm for Big Data Mining Based on Cloud Platform. *J. Grid Comput.* **2020**, *18*, 263–273. [CrossRef]

60. Wang, T.; Zhang, W.; Ye, C.; Wei, J.; Zhong, H.; Huang, T. FD4C: Automatic Fault Diagnosis Framework for Web Applications in Cloud Computing. *IEEE Trans. Syst. Man, Cybern. Syst.* **2016**, *46*, 61–75. [CrossRef]

61. Wang, H.; Sun, J. Quantitative analysis of data mining application and sports industry financing mechanism based on cloud computing. *Int. J. Grid Distrib. Comput.* **2016**, *9*, 233–244. [CrossRef]

62. De Assis, M.V.O.; Novaes, M.P.; Zerbini, C.B.; Carvalho, L.F.; Abrao, T.; Proenca, M.L. Fast Defense System Against Attacks in Software Defined Networks. *IEEE Access* **2018**, *6*, 69620–69639. [CrossRef]

63. Henao, N.; Agbossou, K.; Kelouwani, S.; Dube, Y.; Fournier, M. Approach in Nonintrusive Type i Load Monitoring Using Subtractive Clustering. *IEEE Trans. Smart Grid* **2017**, *8*, 812–821. [CrossRef]

64. Zhang, Q.; Chen, H.; Shen, Y.; Ma, S.; Lu, H. Optimization of virtual resource management for cloud applications to cope with traffic burst. *Futur. Gener. Comput. Syst.* **2016**, *58*, 42–55. [CrossRef]

65. Ngah Nasaruddin, A.; Tee, B.T.; Mohd Tahir, M.; Md Jasman, M.E.S. Data Assessment on the relationship between typical weather data and electricity consumption of academic building in Melaka. *Data Br.* **2021**, *35*, 106797. [CrossRef] [PubMed]

66. Gharehpasha, S.; Masdari, M.; Jafarian, A. Virtual machine placement in cloud data centers using a hybrid multi-verse optimization algorithm. *Artif. Intell. Rev.* **2021**, *54*, 2221–2257. [CrossRef]

67. Amir, M.; Zaheeruddin; Haque, A. Integration of EVs Aggregator with Microgrid and Impact of V2G Power on Peak Regulation. In Proceedings of the 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), Kuala Lumpur, Malaysia, 24–26 September 2021; pp. 1–6.

68. Zhou, Y.; Zhou, Y.; Luo, Q.; Abdel-Basset, M. A simplex method-based social spider optimization algorithm for clustering analysis. *Eng. Appl. Artif. Intell.* **2017**, *64*, 67–82. [CrossRef]

69. Sui, X.; Liu, D.; Li, L.; Wang, H.; Yang, H. Virtual machine scheduling strategy based on machine learning algorithms for load balancing. *Eurasip J. Wirel. Commun. Netw.* **2019**, *2019*, 160. [CrossRef]

70. Hasan, M.K.; Hosain, M.S.; Saha, T.; Islam, S.; Paul, L.C.; Khatak, S.; Alkhassawneh, H.M.; Kariri, E.; Ahmed, E.; Hassan, R. Energy efficient data detection with low complexity for an uplink multi-user massive MIMO system. *Comput. Electr. Eng.* **2022**, *101*, 108045. [CrossRef]

71. Naeem, M.; Jamal, T.; Diaz-Martinez, J.; Butt, S.A.; Montesano, N.; Tariq, M.I.; De-la-Hoz-Franco, E.; De-La-Hoz-Valdiris, E. Trends and future perspective challenges in big data. In Proceedings of the Advances in Intelligent Data Analysis and Applications, Arad, Romania, 15–18 October 2019; pp. 309–325.

72. Cen, B.; Hu, C.; Cai, Z.; Wu, Z.; Zhang, Y.; Liu, J.; Su, Z. A configuration method of computing resources for microservice-based edge computing apparatus in smart distribution transformer area. *Int. J. Electr. Power Energy Syst.* **2022**, *138*, 107935. [CrossRef]

73. Zheng, Z.; Wang, P.; Liu, J.; Sun, S. Real-time big data processing framework: Challenges and solutions. *Appl. Math. Inf. Sci.* **2015**, *9*, 3169–3190. [CrossRef]

74. Yang, F.; Hua, Y.; Li, X.; Yang, Z.; Yu, X.; Fei, T. A survey on multisource heterogeneous urban sensor access and data management technologies. *Meas. Sensors* **2022**, *19*, 100061. [CrossRef]

75. Chen, Y. Integrated and Intelligent Manufacturing: Perspectives and Enablers. *Engineering* **2017**, *3*, 588–595. [CrossRef]

76. Niden, H.; Spriggs, T. How smart, connected products are transforming companies: Interaction. *Harv. Bus. Rev.* **2016**, *94*, 4.

77. Nosratabadi, S.M.; Hooshmand, R.A.; Gholipour, E. A comprehensive review on microgrid and virtual power plant concepts employed for distributed energy resources scheduling in power systems. *Renew. Sustain. Energy Rev.* **2017**, *67*, 341–363. [CrossRef]

78. Edpuganti, A.; Khadkikar, V.; Zeineldin, H.; El Moursi, M.S.E.M.S.; Al Hosani, M.; Al Hosani, M. Comparison of Peak Power Tracking Based Electric Power System Architectures for CubeSats. *IEEE Trans. Ind. Appl.* **2021**, *57*, 2758–2768. [CrossRef]

79. Wu, B.; Widanage, W.D.D.; Yang, S.; Liu, X. Battery digital twins: Perspectives on the fusion of models, data and artificial intelligence for smart battery management systems. *Energy AI* **2020**, *1*, 100016. [CrossRef]

80. Anusha, P.; Shabanabegum, S.K.K.; Pavaiyarkarasi, R.; Seethalakshmi, E.; Vadivukkarasi, K.; Vijayakumar, P. Smart internet of vehicle maintenance system. *Mater. Today Proc.* **2021**. [CrossRef]

81. Iqbal, A.; Amir, M.; Kumar, V.; Alam, A.; Umair, M. Integration of Next Generation IIoT with Blockchain for the Development of Smart Industries. *Emerg. Sci. J.* **2020**, *4*, 1–17. [CrossRef]

82. Zhang, Q.; Lin, M.; Yang, L.T.; Chen, Z.; Khan, S.U.; Li, P. A Double Deep Q-Learning Model for Energy-Efficient Edge Scheduling. *IEEE Trans. Serv. Comput.* **2019**, *12*, 739–749. [CrossRef]

83. Hasan, M.K.; Alkhalifah, A.; Islam, S.; Babiker, N.B.M.; Habib, A.K.M.A.; Aman, A.H.M.; Hossain, M.A. Blockchain Technology on Smart Grid, Energy Trading, and Big Data: Security Issues, Challenges, and Recommendations. *Wirel. Commun. Mob. Comput.* **2022**, 2022. [CrossRef]

84. Amir, M.; Haque, A.; Kurukuru, V.S.B.; Bakhsh, F.; Ahmad, A. Agent based online learning approach for power flow control of electric vehicle fast charging station integrated with smart microgrid. *IET Renew. Power Gener.* **2022**. [CrossRef]

85. Shariff, S.M.; Alam, M.S.; Faraz, S.; Khan, M.A.; Abbas, A.; Amir, M. Economic approach to design of a level 2 residential electric vehicle supply equipment. In *Advances in Power and Control Engineering: Proceedings of GUCON*; Springer: Singapore, 2020; pp. 25–40.

86. Hasan, M.K.; Habib, A.K.M.A.; Islam, S.; Balfaqih, M.; Alfawaz, K.M.; Singh, D. Smart Grid Communication Networks for Electric Vehicles Empowering Distributed Energy Generation: Constraints, Challenges, and Recommendations. *Energies* **2023**, *16*, 1140. [CrossRef]

87. Bera, S.; Misra, S.; Rodrigues, J.J.P.C. Cloud Computing Applications for Smart Grid: A Survey. *IEEE Trans. Parallel Distrib. Syst.* **2014**, *26*, 1477–1494. [CrossRef]

88. Cui, B.; He, S. Anomaly detection model based on hadoop platform and weka interface. In Proceedings of the 2016 10th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, IMIS, Fukuoka, Japan, 6–8 July 2016; pp. 84–89. [CrossRef]

89. Feng, C.; Wang, Y.; Chen, Q.; Ding, Y.; Strbac, G.; Kang, C. Smart grid encounters edge computing: Opportunities and applications. *Adv. Appl. Energy* **2021**, *1*, 100006. [CrossRef]

90. Narayan, A.; Krüger, C.; Göring, A.; Babazadeh, D.; Harre, M.C.; Wortelen, B.; Luedtke, A.; Lehnhoff, S. Towards future SCADA systems for ICT-reliant energy systems. In *ETG-Kongress 2019-Das Gesamtsystem im Fokusder Energiewende*; VDE: Esslingen am Neckar, Germany, 2019; pp. 364–370.

91. Meng, L.; Sanseverino, E.R.; Luna, A.; Dragicevic, T.; Vasquez, J.C.; Guerrero, J.M. Microgrid supervisory controllers and energy management systems: A literature review. *Renew. Sustain. Energy Rev.* **2016**, *60*, 1263–1273. [CrossRef]

92. Suganya, R.; Pavithra, M.; Nandhini, P. Algorithms and Challenges in Big Data Clustering. *Int. J. Eng. Tech.* **2018**, *4*, 40–47.

93. Narayanan, V.; Kavitha, R.; Srikanth, R. Performance Evaluation of Brahmagupta-Bhaskara Equation Based Algorithm Using OpenMP BT. *Proc. Data Anal. Manag.* **2022**, *90*, 21–28.

94. Dafir, Z.; Lamari, Y.; Slaoui, S.C. *A Survey on Parallel Clustering Algorithms for Big Data*; Springer: Amsterdam, The Netherlands, 2021; Volume 54.

95. Carro, M.; Zhao, J. Serial computing vs. parallel computing: A comparative study using MATLAB. *Int. J. Comput. Sci. Mob. Comput.* **2014**, *3*, 815–820.

96. Thakur, V. Perspective Study and Analysis of Parallel Architecture. *Int. J. Comput. Appl.* **2016**, *148*, 21–25. [CrossRef]

97. Navarro, A.; Hitschfeld-kahler, N.; Mateu, L. A Survey on Parallel Computing and its Applications in Data-Parallel Problems Using GPU Architectures. *Commun. Comput. Phys.* **2014**, *15*, 285–329. [CrossRef]

98. Yazici, A.; Mishra, A.; Karakaya, Z. Teaching Parallel Computing Concepts Using Real-Life Applications*. *Int. J. Eng. Educ.* **2016**, *32*, 772–781.

99. Meng, F. Performance Modeling on the Basis of Application Type in Virtualized Environments. *J. Softw.* **2013**, *8*, 2847–2854. [CrossRef]

100. Chen, C.L.P.; Zhang, C. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Inf. Sci.* **2014**, *275*, 314–347. [CrossRef]

101. Chen, Y.; Zhang, N.; Zhang, Y.; Chen, X. Dynamic computation offloading in edge computing for internet of things. *IEEE Internet Things J.* **2019**, *6*, 4242–4251. [CrossRef]

102. Ferry, D.; Li, J.; Mahadevan, M.; Agrawal, K.; Gill, C.; Lu, C. *A Real-Time Scheduling Service for Parallel Tasks*; IEEE: Piscataway, NJ, USA, 2013; pp. 261–272.

103. Guo, Y.; Rao, J.; Cheng, D.; Member, S. iShuffle: Improving Hadoop Performance with Shuffle-on-Write. *IEEE Trans. Parallel Distrib. Syst.* **2017**, *28*, 1649–1662. [CrossRef]

104. Juan, A.A.; Faulin, J.; Grasman, S.E.; Rabe, M.; Figueira, G. A review of simheuristics: Extending metaheuristics to deal with stochastic combinatorial optimization problems. *Oper. Res. Perspect.* **2015**, *2*, 62–72. [CrossRef]

105. Schryen, G. Parallel computational optimization in operations research: A new integrative framework, literature review and research directions. *Eur. J. Oper. Res.* **2020**, *287*, 1–18. [CrossRef]

106. Diaz, J.; Muñoz-Caro, C.; Niño, A. A survey of parallel programming models and tools in the multi and many-core era. *IEEE Trans. Parallel Distrib. Syst.* **2012**, *23*, 1369–1386. [CrossRef]

107. Schulz, C.; Hasle, G. GPU computing in discrete optimization. Part I: Introduction to the GPU. *EURO J. Transp. Logist.* **2013**, *2*, 129–157. [CrossRef]

108. Zohrabad, Z.M. Application of Hybrid HS and Tabu Search Algorithm for Optimal Location of FACTS Devices to Reduce Power Losses in Power Systems. *Eng. Technol. Appl. Sci. Res.* **2016**, *6*, 1217–1220. [CrossRef]

109. Wang, B.; Xu, J.; Cao, B.; Ning, B. Adaptive mode switch strategy based on simulated annealing optimization of a multi-mode hybrid energy storage system for electric vehicles. *Appl. Energy* **2017**, *194*, 596–608. [CrossRef]

110. Bouchekara, H.R.E.H.; Shahriar, M.S.; Javaid, M.S.; Sha, Y.A.; Zellagui, M.; Bentouati, B. A variable neighborhood search algorithm for optimal protection coordination of power systems. *Soft Comput.* **2021**, *4*, 10863–10883. [CrossRef]

111. Demesure, G.; Charpentier, P.; Siadat, A.; Mouayni, E.; El, I.; Mouayni, E. Jobs scheduling within Industry 4.0 with consideration of worker's fatigue and reliability using Greedy Randomized Adaptive Search Procedure. *Isc. FAC Pap.* **2019**, *52*, 85–90. [CrossRef]
112. Tan, Y.; Member, S.; Ding, K. A Survey on GPU-Based Implementation of Swarm Intelligence Algorithms. *IEEE Trans. Cybern.* **2016**, *46*, 2028–2041. [CrossRef] [PubMed]
113. Hasan, M.K.; Ismail, A.F.; Islam, S.; Hashim, W.; Pandey, B. Dynamic spectrum allocation scheme for heterogeneous network. *Wirel. Pers. Commun.* **2017**, *95*, 299–315. [CrossRef]
114. Hadji, S.; Gaubert, J.; Krim, F. Real-Time Genetic Algorithms-Based MPPT: Study and Comparison (Theoretical an Experimental) with Conventional Methods. *Energies* **2018**, *11*, 459. [CrossRef]
115. Asgher, U.; Rasheed, M.B.; Al-Sumaiti, A.S.; Rahman, A.U.; Ali, I.; Alzaidi, A.; Alamri, A. Smart energy optimization using heuristic algorithm in smart grid with integration of solar energy sources. *Energies* **2018**, *11*, 3494. [CrossRef]
116. Search, S. *Scatter Search*; Springer: Berlin, Germany, 2018.
117. Crainic, T.G. Parallel metaheuristics and cooperative search. In *Handbook of Metaheuristics*; Springer: Berlin, Germany, 2019.
118. Saber, T.; Marques-Silva, J.; Thorburn, J.; Ventresque, A. Exact and Hybrid Solutions for the Multi-Objective VM Reassignment Problem. *Int. J. Artif. Intell. Tools* **2017**, *26*, 1760004. [CrossRef]
119. Boyer, V.; El Baz, D. Recent advances on GPU computing in operations research. In Proceedings of the 2013 IEEE 27th International Symposium on Parallel & Distributed Processing Workshops and PhD Forum, Cambridge, MA, USA, 20–24 May 2013. [CrossRef]
120. Brodtkorb, R.; Schulz, C.; Hasle, G.; Hagen, T.R. GPU computing in discrete optimization. Part II: Survey focused on routing problems. *EURO J. Transp. Logist.* **2013**, *2*, 159–186. [CrossRef]
121. Cong, J.; Fang, Z.; Huang, M.; Wei, P.; Wu, D.; Yu, C.H. Customizable Computing—From Single Chip to Datacenters. *Proc. IEEE* **2019**, *107*, 185–203. [CrossRef]
122. Wang, K.; Mi, J.; Xu, C.; Zhu, Q.; Shu, L.; Deng, D.-J. Real-Time load reduction in multimedia big data for mobile internet. *ACM Trans. Multimed. Comput. Commun. Appl.* **2016**, *12*, 1–20. [CrossRef]
123. Chien, S.F.; Zarakovitis, C.C.; Ni, Q.; Xiao, P. Stochastic Asymmetric Blotto Game Approach for Wireless Resource Allocation Strategies. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 5511–5528. [CrossRef]
124. Cheng, D.; Rao, J.; Guo, Y.; Jiang, C.; Zhou, X. Improving Performance of Heterogeneous MapReduce Clusters with Adaptive Task Tuning. *IEEE Trans. Parallel Distrib. Syst.* **2017**, *28*, 774–786. [CrossRef]
125. Gai, K.; Qiu, M.; Zhao, H. Cost-Aware Multimedia Data Allocation for Heterogeneous Memory Using Genetic Algorithm in Cloud Computing. *IEEE Trans. Cloud Comput.* **2016**, *1*, 1212–1222. [CrossRef]
126. Agron, J.; Andrews, D. Hardware microkernels for heterogeneous manycore systems. In Proceedings of the International Conference on Parallel Processing Workshops Hardware, Vienna, Austria, 22–25 September 2009. [CrossRef]
127. Wang, W.; Jiang, K.; Tan, Y.; Wu, Q. Dominant fairness fairness: Hierarchical scheduling for multiple resources in heterogeneous datacenters. *IEICE Trans. Inf. Syst.* **2016**, *E99D*, 1678–1681. [CrossRef]
128. Hwang, E.; Kim, S.; Kim, J.-S.; Hwang, S.; Choi, Y.-R. On the role of application and resource characterizations in heterogeneous distributed computing systems. *Cluster Comput.* **2016**, *19*, 2225–2240. [CrossRef]
129. Gomatheeshwari, B.; Selvakumar, J. Appropriate allocation of workloads on performance asymmetric multicore architectures via deep learning algorithms. *Microprocess. Microsyst.* **2020**, *73*, 102996. [CrossRef]
130. Akhtaruzzaman, M.; Hasan, M.K.; Kabir, S.R.; Abdullah, S.N.H.S.; Sadeq, M.J.; Hossain, E. HSIC Bottleneck Based Distributed Deep Learning Model for Load Forecasting in Smart Grid with a Comprehensive Survey. *IEEE Access* **2020**, *8*, 222977–223008. [CrossRef]
131. Gou, C.; Benoit, A.; Marchal, L. Memory-aware tree partitioning on homogeneous platforms. In Proceedings of the 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), Cambridge, UK, 21–23 March 2018; pp. 321–324. [CrossRef]
132. Beaumont, O.; Becker, B.A.; Deflumere, A.; Eyraud-Dubois, L.; Lambert, T.; Lastovetsky, A. Recent Advances in Matrix Partitioning for Parallel Computing on Heterogeneous Platforms. *IEEE Trans. Parallel Distrib. Syst.* **2019**, *30*, 218–229. [CrossRef]
133. Yin, F.; Shi, F. *A Comparative Survey of Big Data Computing and HPC: From a Parallel Programming Model to a Cluster Architecture*; Springer: Berlin, Germany, 2022; Volume 50.
134. Nguyen, G.; Šipková, V.; Dlugolinsky, S.; Nguyen, B.M.; Tran, V.; Hluchý, L. A comparative study of operational engineering for environmental and compute-intensive applications. *Array* **2021**, *12*, 100096. [CrossRef]
135. Mutlu, O.; Ghose, S.; Gómez-Luna, J.; Ausavarungnirun, R. Processing data where it makes sense: Enabling in-memory computation. *Microprocess. Microsyst.* **2019**, *67*, 28–41. [CrossRef]
136. Zheng, X.; Mukkamala, R.R.; Vatrapu, R.; Ordieres-Mere, J. Blockchain-based personal health data sharing system using cloud storage. In Proceedings of the 2018 IEEE 20th international conference on e-health networking, applications and services (Healthcom), Ostrava, Czech Republic, 17–20 September 2018. [CrossRef]
137. Zhou, Y.; He, F.; Hou, N.; Qiu, Y. Parallel ant colony optimization on multi-core SIMD CPUs. *Futur. Gener. Comput. Syst.* **2018**, *79*, 473–487. [CrossRef]
138. Schmidt, B.; Hildebrandt, A. Next-generation sequencing: Big data meets high performance computing. *Drug Discov. Today* **2017**, *22*, 712–717. [CrossRef] [PubMed]
139. Gupta, A.; Christie, R.; Manjula, R. Scalability in Internet of Things: Features, Techniques and Research Challenges. *Int. J. Comput. Intell. Res.* **2017**, *13*, 1617–1627.

140. Xiong, Z.; Xu, K. Lightweight job submission and file sharing schemes for a teaching ecosystem for parallel computing courses. *J. Ambient Intell. Humaniz. Comput.* **2020**. [CrossRef]
141. Badger, L.; Patt-corner, R.; Voas, J. Cloud Computing Synopsis and Recommendations Recommendations of the National Institute of Standards and Technology. *Nist Spec. Publ.* **2012**, *800*, 81.
142. Hogan, M.; Liu, F.; Sokol, A.; Tong, J. NIST Cloud Computing Standards Roadmap. *Commun. Comput. Inf. Sci.* **2015**, *3*, 1–6.
143. Javadpour, A. Improving Resources Management in Network Virtualization by Utilizing a Software-Based Network. *Wirel. Pers. Commun.* **2019**, *106*, 505–519. [CrossRef]
144. Al-Janabi, S.; Al-Shourbaji, I.; Shojafar, M.; Abdelhag, M. Mobile Cloud Computing: Challenges and Future Research Directions. In Proceedings of the 2017 10th International Conference on Developments in eSystems Engineering (DeSE), Paris, France, 14–16 June 2017; pp. 62–67. [CrossRef]
145. Lehrig, S.; Sanders, R.; Brataas, G.; Cecowski, M.; Ivanšek, S.; Polutnik, J. CloudStore—towards scalability, elasticity, and efficiency benchmarking and analysis in Cloud computing. *Futur. Gener. Comput. Syst.* **2018**, *78*, 115–126. [CrossRef]
146. Kurte, K.; Sanyal, J.; Berres, A.; Lunga, D.; Coletti, M.; Yang, H.L.; Graves, D.; Liebersohn, B.; Rose, A. Performance analysis and optimization for scalable deployment of deep learning models for country-scale settlement mapping on Titan supercomputer. *Concurr. Comput.* **2019**, *31*, e5305. [CrossRef]
147. Kaseb, M.R.; Khafagy, M.H.; Ali, I.A.; Saad, E.M. An improved technique for increasing availability in Big Data replication. *Futur. Gener. Comput. Syst.* **2019**, *91*, 493–505. [CrossRef]
148. Neaime, J.; Dhaini, A.R. Resource management in cloud and tactile-capable next-generation optical access networks. *IEEE/OSA J. Opt. Commun. Netw.* **2018**, *10*, 902–914. [CrossRef]
149. Pedreno-Manresa, J.; Khodashenas, P.S.; Siddiqui, M.S.; Pavon-Marino, P. On the Need of Joint Bandwidth and NFV Resource Orchestration: A Realistic 5G Access Network Use Case. *IEEE Commun. Lett.* **2018**, *22*, 145–148. [CrossRef]
150. Daki, H.; El Hannani, A.; Aqqal, A.; Haidine, A.; Dahbi, A. Big Data management in smart grid: Concepts, requirements and implementation. *J. Big Data* **2017**, *4*, 13. [CrossRef]
151. Lin, S.; Maalla, A.; Wu, G. Research on Real-Time Database Recovery Method of Smart Grid System Based on IEC61970 Standard. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *242*, 022046. [CrossRef]
152. Santodomingo, R.; Uslar, M.; Specht, M.; Rohjans, S.; Taylor, G.; Pantea, S.; Bradley, M.; McMorran, A. IEC 61970 for Energy Management System Integration. *Smart Grid Handb.* **2016**, 1–29. [CrossRef]
153. Lordan, F.; Badia, R.M. COMPSs-Mobile: Parallel Programming for Mobile Cloud Computing. *J. Grid Comput.* **2017**, *15*, 357–378. [CrossRef]
154. Sweetline Priya, E.; Suseendran, G. Cloud computing and big data: A comprehensive analysis. *J. Crit. Rev.* **2020**, *7*, 185–189. [CrossRef]
155. Berlian, M.H.; Sahputra, T.E.R.; Ardi, B.J.W.; Dzatmika, L.W.; Besari, A.R.A.; Sudibyo, R.W.; Sukaridhoto, S. Design and implementation of smart environment monitoring and analytics in real-time system framework based on internet of underwater things and big data. In Proceedings of the 2016 International Electronics Symposium (IES) IES 2016, Denpasar, Indonesia, 29–30 September 2016; pp. 403–408. [CrossRef]
156. Xu, X.; Cao, L.; Wang, X. Adaptive Task Scheduling Strategy Based on Dynamic Workload Adjustment for Heterogeneous Hadoop Clusters. *IEEE Syst. J.* **2016**, *10*, 471–482. [CrossRef]
157. Guo, Y.; Rao, J.; Jiang, C.; Zhou, X. Moving Hadoop into the Cloud with Flexible Slot Management and Speculative Execution. *IEEE Trans. Parallel Distrib. Syst.* **2017**, *28*, 798–812. [CrossRef]
158. Zaharia, M.; Chowdhury, M.; Das, T.; Dave, A.; Ma, J.; McCauley, M.; Franklin, M.J.; Shenker, S.; Stoica, I. Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing. In Proceedings of the 9th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 12), USENIX, San Jose, CA, USA, 25–27 April 2012; pp. 15–28.
159. Gopalani, S.; Arora, R. Comparing Apache Spark and Map Reduce with Performance Analysis using K-Means. *Int. J. Comput. Appl.* **2015**, *113*, 8–11. [CrossRef]
160. Hosseini, B.; Kiani, K. A big data driven distributed density based hesitant fuzzy clustering using Apache spark with application to gene expression microarray. *Eng. Appl. Artif. Intell.* **2019**, *79*, 100–113. [CrossRef]
161. Liu, X.; Nielsen, P.S. Regression-based Online Anomaly Detection for Smart Grid Data. *arXiv* **2016**, arXiv:1606.05781v1.
162. Nguyen, V.-Q.; Nguyen, S.N.; Kim, K. Design of a Platform for Collecting and Analyzing Agricultural Big Data. *J. Digit. Contents Soc.* **2017**, *18*, 149–158. [CrossRef]
163. Na, C.; Xin, C. The research of large scale data processing platform based on the spark. In Proceedings of the 2016 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Changsha, China, 17–18 December 2016. [CrossRef]
164. Tan, J.; Meng, S.; Meng, X.; Zhang, L. Improving ReduceTask data locality for sequential MapReduce jobs. In Proceedings of the IEEE INFOCOM, Turin, Italy, 14–19 April 2013; pp. 1627–1635.
165. Niu, S. Research on the application of machine learning big data mining algorithms in digital signal processing. In Proceedings of the 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, 14–16 April 2021; pp. 776–779. [CrossRef]
166. Peng, B.; Hosseini, M.; Hong, Z.; Farivar, R.; Campbell, R. R-storm: Resource-aware scheduling in storm. In Proceedings of the 16th annual middleware conference, Vancouver, BC, Canada, 7–11 December 2015; pp. 149–161. [CrossRef]

167. Soualhia, M.; Khomh, F.; Tahar, S. Task Scheduling in Big Data Platforms: A Systematic Literature Review. *J. Syst. Softw.* **2017**, *134*, 170–189. [CrossRef]

168. Simão, J.; Esteves, S.; Pires, A.; Veiga, L. GC-Wise: A Self-adaptive approach for memory-performance efficiency in Java VMs. *Futur. Gener. Comput. Syst.* **2019**, *100*, 674–688. [CrossRef]

169. Bobroff, N.; Westerink, P.; Fong, L. Active control of memory for java virtual machines and applications. In Proceedings of the 11th International Conference on Autonomic Computing ({ICAC} 14), Philadelphia, PA, USA, 18–20 June 2014; pp. 97–103.

170. Cugola, G.; Margara, A. Processing flows of information: From data stream to complex event processing. *ACM Comput. Surv.* **2012**, *44*, 1–62. [CrossRef]

171. Li, C.; Zhang, J.; Luo, Y. Real-time scheduling based on optimized topology and communication traffic in distributed real-time computation platform of storm. *J. Netw. Comput. Appl.* **2017**, *87*, 100–115. [CrossRef]

172. Cheng, J.Y.; Hung, M.H.; Lin, S.S.; Cheng, F.T. New remote monitoring and control system architectures based on cloud computing. *Adv. Mater. Res.* **2012**, *579*, 312–329. [CrossRef]

173. Shetty, S.D.; Vadivel, S.; Gandhi, N. A cloud enabled online quiz application deployed on Google cloud. In Proceedings of the 2012 International Conference on Cloud Computing Technologies, Applications and Management, ICCCTAM, Dubai, United Arab Emirates, 8–10 December 2012; pp. 108–113. [CrossRef]

174. Rashid, Z.N.; Zebari, S.R.M.; Sharif, K.H.; Jacksi, K. Distributed Cloud Computing and Distributed Parallel Computing: A Review. In Proceedings of the ICOASE 2018-International Conference on Advanced Science and Engineering, Duhok, Iraq, 9–11 October 2018; pp. 167–172. [CrossRef]

175. Yue, S.; Ma, Y.; Chen, L.; Wang, Y.; Song, W. Dynamic DAG scheduling for many-task computing of distributed eco-hydrological model. *J. Supercomput.* **2019**, *75*, 510–532. [CrossRef]

176. Jafarnejad Ghomi, E.; Rahmani, A.M.; Qader, N.N. Applying queue theory for modeling of cloud computing: A systematic review. *Concurr. Comput. Pract. Exp.* **2019**, *31*, e5186. [CrossRef]

177. Cao, Y.; Li, P.; Zhang, Y. Parallel processing algorithm for railway signal fault diagnosis data based on cloud computing. *Futur. Gener. Comput. Syst.* **2018**, *88*, 279–283. [CrossRef]

178. Alzakholi, O.; Haji, L.; Shukur, H.; Zebari, R.; Abas, S.; Sadeeq, M. Comparison Among Cloud Technologies and Cloud Performance. *J. Appl. Sci. Technol. Trends* **2020**, *1*, 40–47. [CrossRef]

179. Tripathy, L.; Ranjan Patra, R. Scheduling in Cloud Computing. *Int. J. Cloud Comput. Serv. Archit.* **2014**, *4*, 21–27. [CrossRef]

180. Sobie, R. Distributed cloud computing in high energy physics. In Proceedings of the 2014 ACM SIGCOMM workshop on Distributed cloud computing, Chicago, IL, USA, 18 August 2014; pp. 1–4. [CrossRef]

181. Pawade, P.A.; Gaikwad, P.V.T. Semi-Distributed Cloud Computing System with Load Balancing Algorithm. *Int. J. Comput. Sci. Inf. Technol.* **2014**, *5*, 2942–2947.

182. Lv, X.; Cheng, X.; Lv, Y.X. Distributed cloud storage and parallel topology processing of power network. In Proceedings of the 2016 Third International Conference on Trustworthy Systems and their Applications (TSA), Wuhan, China, 18–22 September 2016; pp. 142–146. [CrossRef]

183. Varghese, B.; Buyya, R. Next generation cloud computing: New trends and research directions. *Futur. Gener. Comput. Syst.* **2018**, *79*, 849–861. [CrossRef]

184. Peng, Z.; Gong, Q.; Duan, Y.; Wang, Y. The Research of the Parallel Computing Development from the Angle of Cloud Computing. *J. Phys. Conf. Ser.* **2017**, *910*, 012002. [CrossRef]

185. Ali, M.F.; Khan, R.Z. Distributed Computing: An Overview. *Int. J. Adv. Netw. Appl.* **2015**, *7*, 2630–2635.

186. Kapur, R. A workload balanced approach for resource scheduling in cloud computing. In Proceedings of the 2015 8th International Conference on Contemporary Computing (IC3), Noida, India, 20–22 August 2015; pp. 36–41. [CrossRef]

187. Li, X.; Jia, S.; Wang, K.; Yin, X. Distributed parallel processing of mobile robot PF-slam. *IET Conf. Publ.* **2012**, *2012*, 927–930. [CrossRef]

188. Khiyaita, A.; El Bakkali, H.; Zbakh, M.; Kettani, D. El Load balancing cloud computing: State of art. In Proceedings of the 2012 National Days of Network Security and Systems, Marrakech, Morocco, 20–21 April 2012; pp. 106–109. [CrossRef]

189. Sun, Y.; Zhu, Z.; Fan, Z. Distributed Caching in Wireless Cellular Networks Incorporating Parallel Processing. *IEEE Internet Comput.* **2018**, *22*, 52–61. [CrossRef]

190. Rao, P.S.; Rao, V.P.C.; Govardhan, A. Dynamic Load Balancing with Central Monitoring of Distributed Job Processing System. *Int. J. Comput. Appl.* **2013**, *65*, 43–47.

191. Sharma, A.; Peddoju, S.K. Response time based load balancing in cloud computing. In Proceedings of the 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), Kanyakumari, India, 10–11 July 2014; pp. 1287–1293. [CrossRef]

192. Chen, H.; Wang, F.; Helian, N.; Akanmu, G. User-priority guided min-min scheduling algorithm for load balancing in cloud computing. In Proceedings of the 2013 National Conference on Parallel Computing Technologies (PARCOMPTECH), Bangalore, India, 21–23 February 2013. [CrossRef]

193. Mondal, B.; Choudhury, A. Simulated Annealing (SA) based Load Balancing Strategy for Cloud Computing. *Int. J. Comput. Sci. Inf. Technol.* **2015**, *6*, 3307–3312.

194. Chen, H.; Liu, G.; Yin, S.; Liu, X.; Qiu, D. ERECT: Energy-efficient reactive scheduling for real-time tasks in heterogeneous virtualized clouds. *J. Comput. Sci.* **2018**, *28*, 416–425. [CrossRef]

195. Chunlin, L.; LaYuan, L. Optimal scheduling across public and private clouds in complex hybrid cloud environment. *Inf. Syst. Front.* **2017**, *19*, 1–12. [CrossRef]

196. Kim, J.; Lee, K. FunctionBench: A suite of workloads for serverless cloud function service. In Proceedings of the 2019 IEEE 12th International Conference on Cloud Computing (CLOUD), Milan, Italy, 8–13 July 2019; pp. 502–504. [CrossRef]

197. Wendy; Wang, G. Measuring information security and cybersecurity on private cloud computing. *J. Theor. Appl. Inf. Technol.* **2019**, *97*, 156–168.

198. Yu, J.; Zhao, Y.; Zhu, S.; Xu, B.; Li, S.; Zhang, M. Research on development of cloud computing based on patent analysis. In *2018 3rd Joint International Information Technology, Mechanical and Electronic Engineering Conference (JIMEC 2018)*; Atlantis Press: Berlin, Germany, 2018. [CrossRef]

199. Yu, J.; Yang, Z.; Zhu, S.; Xu, B.; Li, S.; Zhang, M. A bibliometric analysis of cloud computing technology research. In Proceedings of the 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 12–14 October 2018; pp. 2353–2358. [CrossRef]

200. Mane, A.S.; Ainapure, B.S. *Private Cloud Configuration Using Amazon Web Services BT-Information and Communication Technology for Competitive Strategies (ICTCS 2020)*; Springer: Singapore, 2021; pp. 839–847.

201. Duan, J.; Yang, Y. A Load Balancing and Multi-Tenancy Oriented Data Center Virtualization Framework. *IEEE Trans. Parallel Distrib. Syst.* **2017**, *28*, 2131–2144. [CrossRef]

202. AL-Jumaili, A.H.A.; Muniyandi, R.C.; Hasan, M.K.; Singh, M.J.; Paw, J.K.S. Analytical survey on the security framework of cyber-physical systems for smart power system networks. In Proceedings of the 2022 International Conference on Cyber Resilience (ICCR), Dubai, United Arab Emirates, 6–7 October 2022; pp. 1–8.

203. Hasan, M.K.; Habib, A.K.M.A.; Shukur, Z.; Ibrahim, F.; Islam, S.; Razzaque, M.A. Review on cyber-physical and cyber-security system in smart grid: Standards, protocols, constraints, and recommendations. *J. Netw. Comput. Appl.* **2022**, *209*, 103540. [CrossRef]

204. Satoh, F.; Yanagisawa, H.; Takahashi, H.; Kushida, T. Total energy management system for Cloud Computing. In Proceedings of the 2013 IEEE International Conference on Cloud Engineering (IC2E), San Francisco, CA, USA, 25–27 March 2013; pp. 233–240. [CrossRef]

205. Ruivo, P.; Oliveira, T.; Mestre, A. Enterprise resource planning and customer relationship management value. *Ind. Manag. Data Syst.* **2017**, *117*, 1612–1631. [CrossRef]

206. Li, B.; Li, Z. Office Automation Sub-Summary of the Work and the Project Management System. *Adv. Eng. Res. (AER)* **2017**, *61*, 346–348.

207. Chamoso, P.; González-Briones, A.; Rodríguez, S.; Corchado, J.M. Tendencies of Technologies and Platforms in Smart Cities: A State-of-the-Art Review. *Wirel. Commun. Mob. Comput.* **2018**, *2018*, 3086854. [CrossRef]

208. Alghofaili, Y.; Albattah, A.; Alrajeh, N.; Rassam, M.A.; Al-Rimy, B.A.S. Secure cloud infrastructure: A survey on issues, current solutions, and open challenges. *Appl. Sci.* **2021**, *11*, 9005. [CrossRef]

209. Byun, J.; Hong, I.; Park, S. Intelligent cloud home energy management system using household appliance priority based scheduling based on prediction of renewable energy capability. *IEEE Trans. Consum. Electron.* **2012**, *58*, 1194–1201. [CrossRef]

210. Fang, B.; Yin, X.; Tan, Y.; Li, C.; Gao, Y.; Cao, Y.; Li, J. The contributions of cloud technologies to smart grid. *Renew. Sustain. Energy Rev.* **2016**, *59*, 1326–1331. [CrossRef]

211. Singh, K.; Amir, M.; Ahmad, F.; Refaat, S.S. Enhancement of Frequency Control for Stand-Alone Multi-Microgrids. *IEEE Access* **2021**, *9*, 79128–79142. [CrossRef]

212. Cao, Z.; Lin, J.; Wan, C.; Song, Y.; Zhang, Y.; Wang, X. Optimal Cloud Computing Resource Allocation for Demand Side Management in Smart Grid. *IEEE Trans. Smart Grid* **2017**, *8*, 1943–1955. [CrossRef]

213. Amir, M.; Prajapati, A.K.; Refaat, S.S. Dynamic Performance Evaluation of Grid-Connected Hybrid Renewable Energy-Based Power Generation for Stability and Power Quality Enhancement in Smart Grid. *Front. Energy Res.* **2022**, *10*, 1–16. [CrossRef]

214. Colak, I.; Sagiroglu, S.; Fulli, G.; Yesilbudak, M.; Covrig, C.-F.F. A survey on the critical issues in smart grid technologies. *Renew. Sustain. Energy Rev.* **2016**, *54*, 396–405. [CrossRef]

215. Singh, H.; Bawa, S. A MapReduce-based scalable discovery and indexing of structured big data. *Futur. Gener. Comput. Syst.* **2017**, *73*, 32–43. [CrossRef]

216. Qureshi, N.M.F.; Shin, D.R.; Siddiqui, I.F.; Chowdhry, B.S. Storage-Tag-Aware Scheduler for Hadoop Cluster. *IEEE Access* **2017**, *5*, 13742–13755. [CrossRef]

217. Zhu, Z.; Zeng, F.; Qi, G.; Li, Y.; Jie, H.; Mazur, N. Power system structure optimization based on reinforcement learning and sparse constraints under DoS attacks in cloud environments. *Simul. Model. Pract. Theory* **2021**, *110*, 102272. [CrossRef]

218. Zhang, S.; Luo, X.; Litvinov, E. Serverless computing for cloud-based power grid emergency generation dispatch. *Int. J. Electr. Power Energy Syst.* **2021**, *124*, 106366. [CrossRef]

219. Latif, S.N.A.; Chiong, M.S.; Rajoo, S.; Takada, A.; Chun, Y.Y.; Tahara, K.; Ikegami, Y. The trend and status of energy resources and greenhouse gas emissions in the Malaysia power generation mix. *Energies* **2021**, *14*, 2200. [CrossRef]

220. Aceto, G.; Persico, V.; Pescapé, A. Industry 4.0 and Health: Internet of Things, Big Data, and Cloud Computing for Healthcare 4.0. *J. Ind. Inf. Integr.* **2020**, *18*, 100129. [CrossRef]

221. Attaran, M.; Woods, J. Cloud computing technology: Improving small business performance using the Internet. *J. Small Bus. Entrep.* **2019**, *31*, 495–519. [CrossRef]

222. Zhang, G.; Zhang, H.; Zheng, Y.; Liu, Y.; Li, T.; Wang, M.; Wang, X. Research on power grid operation and maintenance cost based on analysis of internal and external influence factors. *E3S Web Conf.* **2021**, *329*, 01068. [CrossRef]

223. Real-Time Monitoring of a Solar Charge Controller for Stand-Alone Photovoltaic Systems. *Int. J. Recent Technol. Eng.* **2019**, *8*, 6295–6300. [CrossRef]

224. Schmidt, L.; Talwar, K.; Santurkar, S.; Tsipras, D.; Madry, A. Adversarially robust generalization requires more data. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 5014–5026.

225. Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A.J. Machine learning algorithm validation with a limited sample size. *PLoS ONE* **2019**, *14*, e0224365. [CrossRef]

226. Wang, Z.Q.; Bai, W.J.; Dong, A.Q.; Zhang, F.H.; Li, B. A Global Data Model for Electric Power Data Centers. *J. Signal Process. Syst.* **2021**, *93*, 201–208. [CrossRef]

227. Huang, L.; Zhuang, W.; Sun, M.; Zhang, H. Research and application of microservice in power grid dispatching control system. In Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 12–14 June 2020; pp. 1895–1899. [CrossRef]

228. Lv, Z.; Li, X.; Wang, W.; Zhang, B.; Hu, J.; Feng, S. Government affairs service platform for smart city. *Futur. Gener. Comput. Syst.* **2018**, *81*, 443–451. [CrossRef]

229. Rahman, A.; Srikumar, V.; Smith, A.D. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. *Appl. Energy* **2018**, *212*, 372–385. [CrossRef]