

## Big Data and Cloud Computing Platform for Energy Internet Application

Zewu Peng<sup>1</sup>, Bojian Wen<sup>1</sup>, Jianwen Huang<sup>1</sup>, Hanyang Xie<sup>1</sup>, Lihao Wei<sup>1</sup>, Zhanhui Xiao<sup>1</sup>, Jiang Jiang<sup>1</sup>, Rui Fu<sup>2</sup>

1. Information Center, Guangdong Power Grid Co., Ltd. Guangzhou, Guangdong, 510080, China

2.State Key Lab of Power Systems, Department of Electrical Engineering, Tsinghua University Beijing, 100084, China

**Keywords:** MapReduce, Spark, Lambda Architecture, BaaS, Microservice Architecture, Complementary Operation of Multi-form Energy System

**Abstract:** With the development of power grid, energy internet has become a popular and promising trend in electrical power industry. According to the deep merge of operational technology and internet information technology, the implementation of the next-generation cyber-physical systems has become the core message, shaping the future of energy industry. The growth of energy internet depends on the availability of high performance computing and analytics technology to process a massive amount of data set. In this paper, we have discussed two core techniques, which aims at supporting the growth of energy internet, big data and clouding computing. The focus of the techniques is the platform architecture that is particularly adaptive to handle a variety of characteristics of energy industry data. As a summarization, this paper proposes a solution for energy big data and IoT cloud computing platform. The application scenarios focus on the energy consumption monitoring and complementary operation of multi-form energy system that support all aspects within a data lifetime cycle, such as acquisition, storage, analytics, and visualization.

### 1. Introduction

In recent years, big data quickly becomes a “hot topic” in industry and academia. 《Nature》, 《Science》 and other magazines have published special issues to explore the challenges and opportunities of big data [1]-[2]. “Data have penetrated into every industry and business area today and have become an important factor in production. The discovery and use of big data signal a new wave of productivity growth and consumer surplus,” said McKinsey & Company, a leading management consulting firm.

Big data is a term for large and complex datasets that traditional data processing approaches are not suitable to deal with them. Big data usually comes from the Internet, enterprise systems, Internet of Things, and other information systems. Data collection and preparation, storage management, data processing, data analysis, and knowledge presentation [3] would generate new insights to support decision-making and business intelligent operation.

Energy Internet is the direction and trend of the energy industry evolution. Energy Internet is a new generation of physical information systems which integrate the advanced communication information technology and physical energy systems. The essence is to provide comprehensive energy services based on multi-energy complementarity. The core is the energy production, transportation, distribution, and consumption based on distributed and renewable energy. At the same time, it can integrate heat, electric power, gas, and even water together [4]. Energy Internet business model includes two aspects: integrated energy, that is, electric power, gas, and other complementary multi-exchange; integrated services, such as energy security monitoring, energy efficiency management, operation and maintenance services, and green energy deployment services [4]. Energy Internet is based on integrated, high-speed communication network, and makes full use of advanced sensing, measurement technology, optimization control technology, and decision support technology to realize the reliable, safe, economical, efficient, and environment-friendly goal. The intelligent operation of the Energy Internet is based on a high degree of “observability” and

"controllability." The observability and controllability are based on the need to obtain real-time panoramic data that can reflect the operating status of the system [5]. Thus, the high-performance data analysis technology will become a significant support for the development of Energy Internet. This paper discusses the development of the two core technologies of supporting the development of Energy Internet: big data and cloud computing. We mainly focus on the big data platform and cloud computing platform's architecture and then summarize the use cases based on the proposed framework.

## 2. Big Data Platform

In this section, we study and compare the architecture of big data platform plus related applications, and combine the demand of Energy Internet, to put forward a set of large data platform architecture which has excellent computing ability and can adapt to various business requirements.

The system is based on the Apache Hadoop ecosystem, the Hadoop Distributed File System (HDFS) is used as the underlying file system. YARN, as resource management and task scheduling tool, uniformly distributes Hadoop MapReduce and Spark to support multiple business scenarios and business requirements. The specific architecture is shown in Fig. 1.

Based on this architecture, we performed a detailed investigation on storage, computing, security, management, and other aspects.

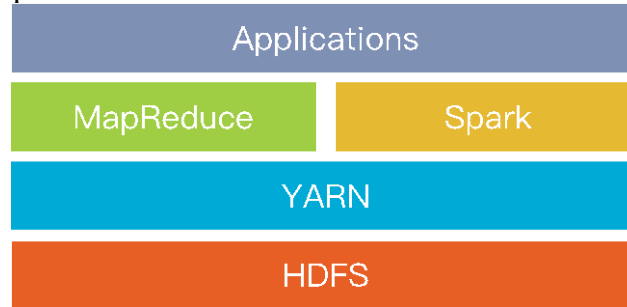


Fig. 1 Big data platform architecture.

### A. Lambda Architecture

The new generation of big data platform features in support of semi-structured, unstructured data processing, support data visualization, also with higher processing performance and easier operation and maintenance features. Lambda architecture is the reference model for the next generation of big data platform, which first proposed by Twitter. As shown in Fig. 2, lambda architecture contains three layers, batch layer, speed layer and serving layer [6].

Ideally, any data access procedure can start with a single argument input and a function result output, but if the data reaches a significant level (such as PB) and needs to support real-time queries, it can be very expensive. Lambda Architecture introduces the concept of batch view, which pre-evaluates part of the query results so that when the query needs to be executed, the results can be read from the batch view. A pre-computed view can be indexed, thus enable fast random read.

The batch view is generated from the batch layer, the use of batch layer is to make the large data small, to use computing resources to improve real-time query performance effectively.

Speed layer uses the same data processing logic as the batch layer, but the main difference is that the speed layer deals with real-time streaming data and batch layer deals with bulk off-line data. Another difference is that to meet the minimum computing delay, and the speed layer does not read all new data at the same time, on the contrary, it will receive one piece of new data and update the speed view. Speed layer is an incremental calculation, but not re-operation.

### B. Storage

Energy data are of multi-source heterogeneous features [7], for structure data storage, open source database MySQL and PostgreSQL are the better choices. For unstructured data storage, NoSQL database such as HBase, MongoDB and Cassandra are the most appropriate.

The raw data imported through the energy internet cloud computing platform is stored in MongoDB (cluster mode), and the data schema of the cloud computing platform is shared. When Spark /

MapReduce calculation engine is ready, the data will be automatically imported into HDFS through Kafka. The results will be exported back to MongoDB via Kafka, as well as to other data sources already deployed in the system, such as PostgreSQL or HDFS.

### C. Security

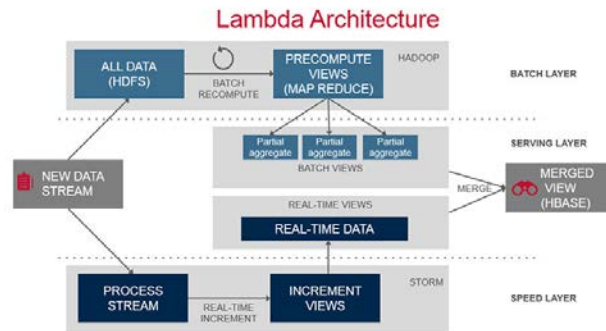


Fig. 2 Lambda architecture.

Hadoop's security certification is based on the Kerberos. Kerberos is a network authentication protocol. Users only need to enter authentication information to verify that they can access multiple Kerberos-based services by obtaining Kerberos ticket. Single sign-in of the machine can also be done based on this protocol. Hadoop itself does not create user accounts but uses the Kerberos protocol for user authentication.

## 3. Cloud Computing Platform

### A. System Architecture

#### 1) Monolithic architecture

Monolithic architecture is a very common form of computer software architecture, which is often layered by functions. Common layers are the presentation layer, business logic layer, and data layer [10]. The business logic layer can be modularized to components by different business responsibilities and functions. The “monolithic architecture” is a “single block” at the physical deployment architecture level. It is usually compiled, packaged, deployed, and maintained as a single application.

#### 2) Micro-service Architecture

The micro-service architecture is a new concept, which divides the application into a series of small services. Each service focuses on a single function, runs in a separate process, and has a clear boundary between services. Services can communicate with each other with light protocols (such as HTTP / RESTful), to achieve a complete application that meets business and user demands [11].

### B. Service Type

Cloud computing services currently have a variety of ways, which are recognized by IaaS, PaaS, SaaS, BaaS, etc. [8].

#### 1) IaaS

IaaS is short for “Infrastructure-as-a-Service”, usually as a hardware server rental service. IaaS service providers provide off-site server, virtualization, storage, and network hardware for users to rent. It can save maintenance costs and office space for users. Some of the successful IaaS products include Amazon's AWS cloud service, Microsoft's Azure, Alibaba's Aliyun, Tencent's Qcloud.

#### 2) PaaS

PaaS is short for “Platform-as-a-Service”. PaaS provide a complete application runtime environment, middleware, database, and so on. Developers only need to upload and deploy the code; the application can run up. PaaS not only reduces the cost of IT operation and maintenance, but also eliminates a lot of development workload. PaaS platform include Google's App Engine, Baidu's App Engine (BAE), Sina's App Engine (SAE).

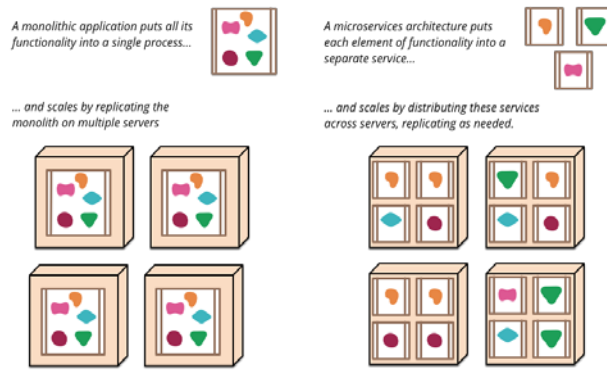


Fig. 3 Monolithic architecture vs Micro-service architecture.

### 3) SaaS

SaaS is short for “Software-as-a-Service”. Usually, it provides an entirely web-based application, the user does not need to consider any software development, deployment, and do not have to worry about server hardware, bandwidth, just directly purchase and the software is ready to go.

### 4) BaaS

BaaS is short for “Backend-as-a-Service”. It is a new PaaS-based cloud service. It is designed to provide backend cloud services for mobile and web applications, including cloud data, storage, account management, messaging, social media integration, big data analysis interfaces and other services [9]. The BaaS platform provides users with a variety of core components and micro services (middleware) to complete specific functions and meet the individual demands.

## C. Internet of Things

The Internet of Things is to let devices sense, communicate, interact, and collaborate within a network. From the perspective of monolithic architecture, the traditional IoT cloud computing platform consists of three layers [12], as shown in Fig. 4.

The protocol layer acts as a gateway and accesses the IoT devices through MQTT, XMPP, or other proprietary protocols, and provides APIs and corresponding SDKs for other Internet devices and applications [13-14]. IoT devices can also directly connect to the gateway through WiFi, 3G, 4G, and other network communication module, and complete the data transfer. To ensure the stability of data access and push services in the high-concurrency environment, load balancing mechanism are usually implemented, and the number of gateways is extended according to the traffic [13,15].

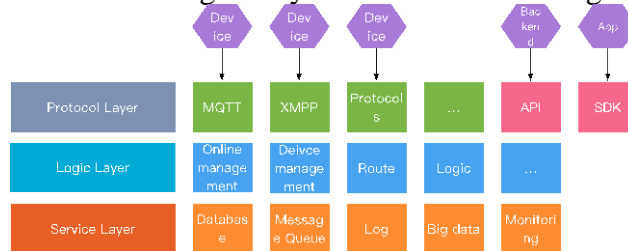


Fig. 4 An IoT cloud computing architecture.

The business logic layer is mainly based on the user's request parameters, matching the target hardware, the business data transfer distribution, the completion of the business system data subscription and distribution, to achieve intelligent device control and monitoring [14-15].

Core services layer provides all the essential services, including data storage database, message queue, log management, big data platform docking, and monitoring warning services [16].

## D. Cloud Computing Platform for Energy Internet

The energy internet cloud computing platform abstracts and unifies backend system's requirements in most scenarios, and implements a common, powerful, customizable BaaS (Backend as a Service) service through cloud engine:

Cloud engine is the hosting service launched by the energy internet cloud computing platform. The engine is based on Node.js runtime environment and able to run backend applications. The engine support multi-instance service load balancing and scaling.

#### 1) Cloud functions, hooks, and background tasks

Energy Internet cloud computing platform provides cloud engine open function, function hook, and background tasks. Through the cloud function, users can easily customize the business logic based on data interaction and can be exposed through the routing interface. Function hook allows the user to define a function that will be called when target event happened, such as new object saved, or object modified. Background tasks enable you to set non-blocking tasks and perform tasks such as big data processing.

#### 2) Realtime data streaming

HTTP protocol is well known as the "request/response" model, but this model does have a lot of limitations. In many scenarios, we want the server to actively send notifications to the browser, such as notifying the browser to refresh the data display in real time when received new meter data. HTML5's WebSocket protocol allows the server to send packets to the client directly, so it can make real-time data streaming possible.

To ensure the communication connectivity and service scalability of WebSocket in the multi-instance cloud engine environment, we take the necessary measures to prevent WebSocket from becoming the bottleneck of the whole service. If the user's service has high throughput, the single instance of the cloud engine may cause the packets blocking. To avoid such situation, we use Redis as a global message queue to handle all messages' distribution and subscription. This will not only ensure the reliability and efficiency, and let the whole cloud engine scalable.

### 4. Complementary Operation of Multi-form Energy System

One major goal of developing Energy Internet is to promote complementary operation of multi-form energy system that will achieves energy cascading utilization and boosts the energy efficiency. Distributed combined cooling heating and power (CCHP) system develops as a novel energy supply way in recent years, which is regarded as environment-friendly, economical, and reliable option for the future [17]. CCHP is one of the key components for successful operation of multi-form energy system.

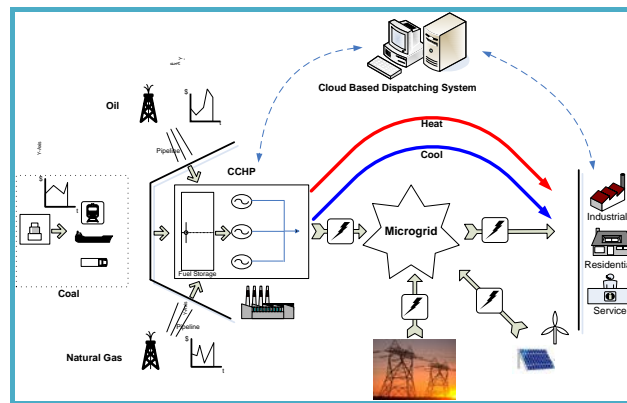


Fig. 5 CCHP based multi-energy form system.

The typical primary movers for CCHP includes steam-turbines, gas-turbines, reciprocating internal combustion engines, and micro-turbines etc., which are integrated with thermally-activated cooling technologies. Due to its unique characteristics and complexities, the new research needs to be done on system design, control, operation, and planning. Consequently, optimization and economic evaluation techniques based on thermo-dynamics, thermo-economic principles are critical to the successful deployment of CCHP system. A data-driven method for better operating multi-form energy system, with a consideration of uncertainty on price and demand, would be beneficial for directing practical system operation.

The multi-form energy system that employs more than one thermos-dynamic cycle. For example, a CCHP refers to the combination of gas turbine generators with exhaust waste heat boilers and thermally-activated chilling machines for the production of electric power, heating, and cooling energy together [18]. The overall efficiency of CCHP can reach around 80%~90% due to energy cascading utilization [19]. The CCHP is particularly fit for multi-form energy system where

distributed solar, wind, energy storage, and load are systematically integrated. The CCHP can provide most of heating and cooling energy, and partially fulfill electricity demand. The outstanding electricity demand is compensated by power grid. (Fig. 5) The overall reliability of energy supply is improved by a combination of distributed and centralized models. Our platform implements a general methodology to optimize CCHP performance and introduces a dynamic model to evaluate CCHP operating cost and combine power network analysis with heating/cooling energy dispatch. [20] The data-driven method is demonstrated on a simulated data set. The reduction of total cost is about 20% compared to the old design of electricity, heating, and cooling energy being produced separately.

## 5. Demo system

The demonstration system uses javascript es2016, node.js, react.js, bootstrap, python, webpack, leaflet, plotly.js and docker to develop the web app based on our cloud computing platform.

The demonstration system is connected to a variety of power consumption and electricity generation device located on the campus of Tsinghua University, including teaching buildings, dormitories, and other functional buildings.

### A. Data Emulation

The device data are updated every minute. The updated data includes current electricity consumption/generation and costs. We use Python to implement a simulator as data upload tool, the simulator upload data from cloud engine's open API.

Once the data are stored in the database, the cloud engine will start a background task to let the big data platform perform data calculation and analysis. The results will be automatically stored in the corresponding database and ready for the frontend page.

### B. Data Visualization

We use plotly.js, leaflet, and other web components for data visualization. Fig. 6 shows the number of devices and the total power, total consumption/generation, and the total cost. Fig. 7 uses a heatmap graph to show the energy consumption/generation of different devices. Fig. 8 shows the geographical coordinates of the buildings and user can view the selected building's device list and readings. Fig. 9 and Fig. 10 show the electricity consumption and generation data of the specified building. Fig. 11 shows the details of the selected building's devices.

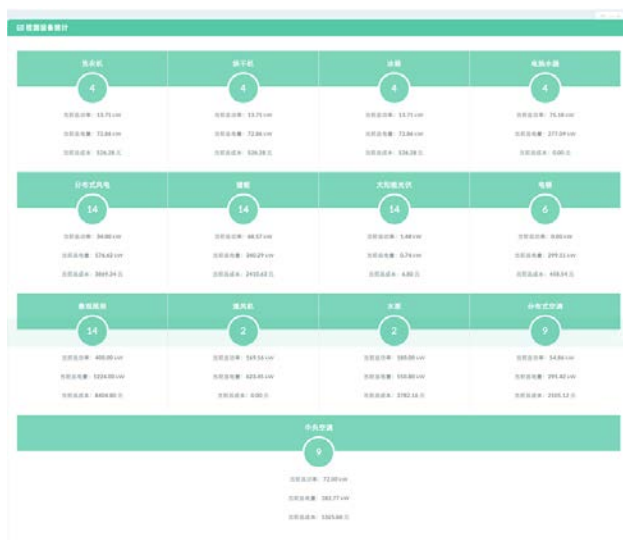


Fig. 6 Device list

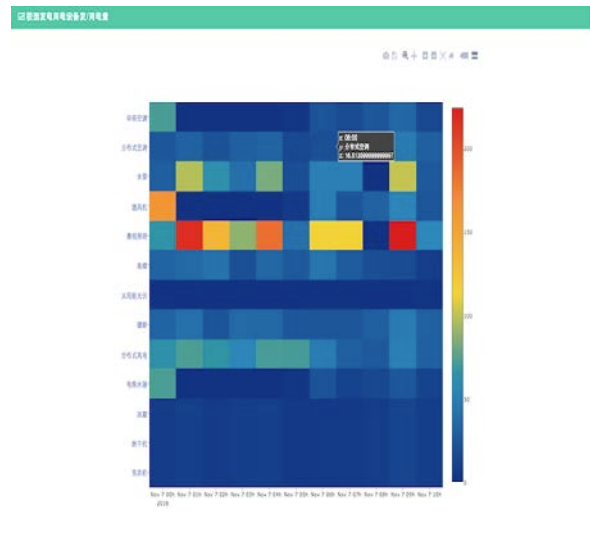


Fig. 7 Consumption heatmap of different devices.





Fig. 8 Buildings and devices on the map.

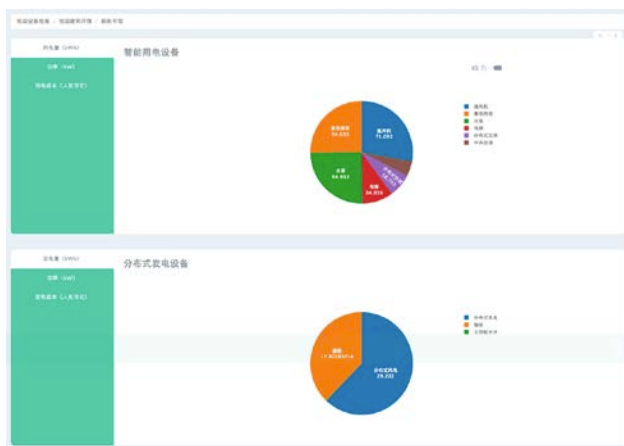


Fig. 9 Device statistics.



Fig. 10 Detailed device statistics.

### C. Realtime Data Streaming

Based on the WebSocket protocol, user's browser will open a long connection when a user is loading the web page. At the same time, data will send through WebSocket in real-time. Also, with react.js's support, the web page can update without a refresh.

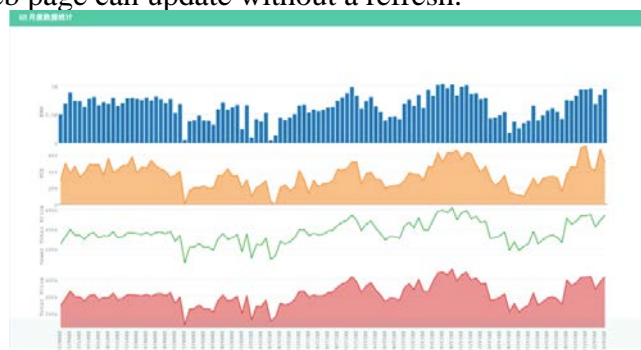


Fig. 11 Monthly data statistics.

## 6. Conclusion

Energy Internet is a new generation of physical information systems that is deeply integrated with existing energy systems and internet technologies. Its essence is to use the concept of the Internet to transform the traditional energy system, and apply the internet technology into the traditional energy system [21]. The development of Energy Internet counts on big data analysis and data mining. In this paper, we propose a framework containing both big data and cloud computing platform, which are suitable for characteristics of energy data. Next step we will integrate solar, battery, electric vehicles, and other green distributed energy devices into our platform, to push the combination of distributed energy, multi-form energy, and green energy technology development.

## References

- [1] DAVID G. Big Data [J]. Nature, 2008, 455(7209): 1-136.
- [2] WOUTER L, JOHN W. Dealing with Big Data [J]. Science, 2011, 331(6018): 639-806.
- [3] Han, Jiawei, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.
- [4] Dong, Zhao, Wen. “cong zhi neng dian wang dao neng yuan hu lian wang : ji ben gai nian yu yan jiu kuang jia”. Dian Li Xi Tong Zi Dong Hua, 2014, 38(15): 1-11.
- [5] Song, Zhou, Zhu.” zhi neng dian wang da shu ju chu li ji shu xian zhuang yu tiao zhan”. Dian Wang Ji Shu, 2013, 37(4): 927-935.
- [6] Nathan Marz. Big Data Lambda Architecture [EB/OL]. (2016-08-03)[2012-09-05]. <http://www.databasetube.com/database/big-data-lambda-architecture/>
- [7] Li, Geng, Zheng. “neng yuan hu lian wang xing shi xia de dian li da shu ju fa zhan qu shi”. Xian Dai Dian Li, 2015, 32(5): 10-14.
- [8] LUO Jun-Zhou, et al. Cloud computing: architecture and key technologies [J]. Journal of China Institute of Communications, 2011, 32(7): 3-21.
- [9] SAREEN Pankaj. Cloud computing: types, architecture, applications, concerns, virtualization and role of it governance in cloud [J]. International Journal of Advanced Research in Computer Science and Software Engineering, 2013, 3(3).
- [10] MACHADO Rodolfo, EL-KHOURY Rodolph. Monolithic architecture [M]. Prestel Publishing, 1995.
- [11] NEWMAN Sam. Building Microservices [M]. O'Reilly Media, Inc., 2015.
- [12] Zhao. “gou jian ji yu yun ji suan de wu lian wang yun ying ping tai”. Dian Xin Ke Xue, 2010, 6: 48-52.
- [13] Liu, Cui, Chen. wu lian wang guan jian ji shu yu ying yong. Ji Suan Ji Ke Xue, 2010, 37(6): 1-10.
- [14] Huang, Deng. “wu lian wang wang guan ji shu yu ying yong”. Dian Xin Ke Xue, 2010, 4: 20-24.
- [15] Qian, Wang. “wu lian wang ji shu yu ying yong yan jiu”. Dian Zi Xue Bao, 2012, 40(5): 1023-1029.
- [16] Sun, Liu, Li. “wu lian wang : gai nian, jia gou yu guan jian ji shu yan jiu zong shu”. Bei Jing You Dian Da Xue Xue Bao, 2010, 33(3): 1-9.
- [17] D. Wu, "Multiple Objective Thermodynamic Optimization and Application Study of Distributed Combined Cooling Heating and Power System", PhD Dissertation, Shanghai Jiao tong University, 2008.
- [18] Workshop, "Combined Cycle Resources in ERCOT Nodal Market", ERCOT, 2010.



- [19] Y. Liu, "Optimal Design of Distributed Combined Cooling, Heating, and Power System", MS Thesis, North China Electric University, 2012.
- [20] F. Gao, "Integration of Wind Generation with Storage Techniques", MS Thesis, Department of Economics, Iowa State University, 2008.
- [21] Cao, Yang, Zhang." neng yuan hu lian wang : xin xi yu neng yuan de ji chu she shi yi ti hua". Nan Fang Dian Wang Ji Shu, 2014, 8(4): 1-10.