Research on Smart Grid Load Forecasting Cloud Platform Based on Hadoop Architecture

Lin Qian\textsuperscript{a}, Mingjie Xu\textsuperscript{b}, Jun Yu\textsuperscript{c}, Guangxin Zhu\textsuperscript{d}, Lin Wang\textsuperscript{e}, Haiyang Chen\textsuperscript{f} and Zhu Mei\textsuperscript{g}

State Grid Electric Power Research Institute (SGEPRI), Nanjing, China

\textsuperscript{a}qianlin@sgepri.sgcc.com.cn, \textsuperscript{b}xumingjie@sgepri.sgcc.com.cn, \textsuperscript{c}yujun@sgepri.sgcc.com.cn, \textsuperscript{d}zhuguangxin@sgepri.sgcc.com.cn, \\
\textsuperscript{e}wanglin18@sgepri.sgcc.com.cn, \textsuperscript{f}chenhaiyang@sgepri.sgcc.com.cn, \textsuperscript{g}meizhu2016@aliyun.com

Abstract. In view of the shortcomings and shortcomings of traditional smart grid load forecasting system, this paper uses Hadoop architecture to build cloud computing platform, optimizes the layout and architecture of the entire smart grid load forecasting cloud platform, and forms the main framework. The basic coordination model is used to carry out multi-level coordination and optimization service, and its practicability is proved by an example. Through the research on the framework of the smart grid load forecasting platform based on cloud computing, improving the level of load forecasting can deepen the application of cloud computing in power system, lay a solid foundation for smart grid dispatching and load balancing, and provide a strong guarantee for the stability analysis of online upgrade of the system.

1. Introduction

With the gradual development of smart grid, a large number of smart terminal devices have been put into use in the industry, and the environment, means and objectives of load forecasting will undergo tremendous changes. Traditional load forecasting system has outstanding heterogeneity, and it is difficult to grasp the diversity, complexity and intermittence of load. The forecasting process often needs to weigh the rationality of mathematical model and real-time calculation, and the low utilization of servers. The demand of smart grid for load forecasting is getting higher and higher. It will enter a new stage, including the massive data management of load forecasting, the emergence of volatile loads, the increasing demand for load forecasting density and accuracy, which will have a great impact on improving the level of load forecasting.

With the emergence of cloud computing, power cloud came into being. Cloud computing can achieve massive data processing and rational allocation of resources. Relevant technologies have been widely used in medicine, education, e-commerce and other fields, and have been successfully applied in the field of power system. Cloud computing will provide services for load forecasting of smart grid.

In view of the shortcomings and shortcomings of traditional smart grid load forecasting system, this paper uses Hadoop architecture to build cloud computing platform, optimizes the layout and architecture of the entire smart grid load forecasting cloud platform, and forms the main framework.
The basic coordination model is used to carry out multi-level coordination and optimization service, and its practicability is proved by an example.

2. Research status and main problems of load forecasting at home and abroad

Load forecasting is a traditional research problem in the field of power system, which plays an important role in the safe operation of power system. It is the basic work of energy-saving dispatching, power generation planning, power grid planning, and stability calculation and so on. Many examples and data show that accurate load forecasting is very important for rational design of power supply structure, formulation of fuel plan, implementation of energy saving and emission reduction measures, and acceleration of regional economic development.

Power system load forecasting is based on the historical data of power load, economy, society, meteorology and other factors. It explores the influence of historical data of power load on future load, and seeks the internal relationship between power load and various related factors, so as to make a scientific forecasting of future power load. Power load forecasting is a branch of forecasting, which can draw on the experience and methods of forecasting in other fields. For example, foreign magazines such as International Journal of Forecasting and Journal of Forecasting focus on the latest forecasting methods, forecasting development and practical application. Forecast, a domestic magazine, is more concerned with economic forecasting, but many of its methods can also be used in power system load forecasting.

In the past twenty or thirty years, many experts and scholars at home and abroad have done a lot of research on the theory and method of load forecasting, and have achieved a lot of results. From a methodological point of view, forecasting methods have roughly experienced four stages of development: (1) traditional historical data statistical forecasting methods represented by linear regression method and moving average method; (2) time series forecasting methods (including AR, MA and ARMA) proposed by Box-Jenkins; (3) grey model and combination forecasting method; (4) neural network and support vector machine as the stages of development. The main learning algorithm and the intelligent forecasting method stage of the optimization algorithm based on particle swarm optimization and genetic algorithm. The core content of load forecasting is to improve the accuracy of load forecasting, so what forecasting method to use has always been the focus of the development of load forecasting.

As the demand of load forecasting accuracy and density is getting higher and higher in smart grid, the difficulty of load forecasting is increasing. Traditional load forecasting tends to be forecasting technology, while smart grid load forecasting pushes forecasting to a refined service management level. Through the analysis of relevant literature, the following problems are found in load forecasting: (1) In terms of data storage and processing, with the application of a large number of terminal devices (smart meters, embedded smart appliances, etc.) and the advancement of the process of "informatization, digitalization, automation and interaction", data information acquisition, processing, storage and calculation will encounter bottlenecks, and new ones need to be adopted. Method. In traditional load forecasting, data processing speed is slow and response time is long, and data migration and disaster preparedness of various information data platforms can not meet the current needs. In addition, the traditional load clustering analysis method will fall into the local optimal solution or even non-convergence, which will seriously affect the development of load forecasting. (2) From the perspective of forecasting mathematical model, with the increasing demand for accuracy and density of load forecasting in smart grid, it is necessary to study the real-time and self-adaptation. Mathematical model of sex. However, the influence factors considered by traditional mathematical models are not comprehensive or specific, and the identification degree of the influence factors is not high, which results in the weak prediction ability at some extreme points or inflection points, and makes the load model under multi-factors less versatile. The load analysis model in the early stage, the mid-term prediction algorithm model and the revised load prediction model in the later stage are not refined enough; (3) From the perspective of special forecasting objects, the influx of stochastic and intermittent loads leads to the instability of load forecasting accuracy at the present stage. This part of
the load often refers to the typical regional large user load, high-speed electrified railway load, distributed wind farm power generation load, solar photovoltaic power generation load, and electric vehicle charging load and so on. With the development of smart grid, the impact of these loads forecasting accuracy on regional load forecasting will gradually increase. These loads often use load-dedicated substations to connect or disconnect from the network, and the forecasting results will directly affect the work of power flow calculation and security checking. (4) From the aspect of load forecasting system, most of the existing load forecasting systems are based on multi-layer system Bro. Wser/Server (B/S) structure, using. NET or J2EE technology. For different prediction objects (systems, buses, large users, etc.), the system is prone to form heterogeneous phenomena, which results in a bad situation of unreasonable allocation of resources. For example, many provincial power companies often adopt a system for system load forecasting, a system for bus load forecasting, or a system for large regional power users. However, due to the different time between the system and the system, it is difficult to achieve the same function and configuration of the system. For data sharing, it is basically only at the level of predictive results sharing. (5) In terms of server operation, CPU utilization of load forecasting network, system and database server is low, and most of the time is idle or unsaturated, which wastes resources seriously. Researchers found in the actual system research and development, operation and maintenance, the general provincial system load forecasting server deployment using distributed mode, that is, provincial and local end deployment of separate servers; bus load forecasting often uses centralized deployment mode, that is, provincial and local end shared servers. In fact, data processing, storage, load analysis, forecasting model calculation, correction calculation of all kinds of servers do not need to run for a long time, and CPU utilization is generally less than 10%. Generally speaking, the utilization rate of load forecasting server is relatively low.

3. Smart grid load forecasting cloud platform based on Hadoop architecture

At present, the work of load forecasting mainly focuses on platform development, forecasting technology research and the establishment of forecasting management center. However, the demand of power system internal forecasting is not single, and the heterogeneity of system platform is prominent (diversification of infrastructure such as servers, diversification of development platforms, diversification of data standards, diversification of forecasting technology and diversification of objects). The establishment of load forecasting system based on cloud computing can solve the problems of data sharing and slow processing speed, and adapt to the development of smart grid. Figure 1 shows the cloud architecture of load forecasting for smart grid.

The whole platform is mainly divided into "cloud", "end" and the support and development tools of corresponding management business. The development tools include Hadoop cloud computing system suite (Linux operating system, Hadoop+HDFS+HBase cloud computing system and Java development program), virtualization platform (Virtual Box, VMware, Xen and KVM, etc.), Map Reduce program development tool Eclipse+IBM Map Reduce Tools Eclipse Plugins, etc. "End" refers to the user end. Users of different specialties of dispatching and control centers at all levels (control specialty, mode specialty, communication specialty, automation specialty, comprehensive specialty and relay protection specialty) pass through the terminal equipment (with "four screens" - computer, TV, mobile phone, tablet computer as the main part) and through the Internet (with the current brand-new Internet technology, so that users can pass through it.
Figure 1. Cloud Computing Application Architecture.

It is possible to provide practical remote cloud services over the Internet, building a broad bridge between users and the cloud, sending requests to load forecasting clouds and obtaining services. And "cloud" mainly includes two parts: service and management. Services include IaaS, PaaS and SaaS. Management refers to Cloud Management. Cloud Management mainly includes billing management,
service management, user management and security management. It is mainly used to manage user
service requirements. User management includes user account management, single sign-on and
configuration management. Service management mainly refers to that cloud services can improve the
quality of service operation on the basis of abiding by SOA design specifications. Accounting
management is the statistics of users’ use of network resources and services. Security management
guarantees users to access and obtain services reasonably and legally.

The adoption of SOA is conducive to the unification of technical standards, system reconfiguration
and business flexibility. And its combined efforts with Saa S will promote the development of cloud
computing in power load forecasting. SOA truly reconstructs and reorganizes the above systems to
form a comprehensive and orderly system. The above parts form the "cloud" of the architecture, and
the "end" of the user is mainly presented by virtualization technology such as desktop virtualization.

4. Multi-dimensional and multi-level coordinated load forecasting optimization service of cloud
platform
Load forecasting in smart grid often results in unbalanced forecasting results. Coordination mechanisms
need to be established, such as between upper and lower power grids, between industries and regions,
between buses and systems, etc. Smart grid forecasting is divided into point, line, surface and volume
multi-dimensional and multi-level load forecasting. At the same time, it brings coordination among all
levels.

"Point" refers to typical users, high-speed rail and intermittent distributed load forecasting. "line"
refers to bus load forecasting, "surface" refers to traditional regional system load forecasting, "volume"
refers to provincial, network and whole network load forecasting.

There is coordination between "point" and "line", which refers to the coordination between typical
users and bus load forecasting. On the one hand, it is helpful to analyze the load components of buses,
on the other hand, it is helpful to correct the buses load forecasting results by comparing the
forecasting results.

There is coordination between "line" and "surface". It refers to the coordination between bus load
forecasting and regional system load forecasting. This process is often involved in safety check. Bus is
the node of safety check. Several buses constitute a safety check section. The bus load is coordinated
by the bus result, and the bus load is trimmed by the bus result.

There is coordination between "face" and "body". It refers to the coordination between provincial
power grids and superiors. The coordination process takes into account the load, network loss and
simultaneous rate of external links between provinces. As formula (1).

\[ P^0 = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_n P_n + \Delta P_l + \Delta P_s \]  

In formula (1), \( P^0 \) is the forecasted load of the superior power grid; \( P_i \) is the forecasted load of the
provincial power grid; and \( \lambda_i \) is the reliability coefficient of load of each provincial power grid
considering simultaneous rate and accuracy (i=1,2,..., n, n is the number of provincial power grids);
\( \Delta P_l \) is the comprehensive load of tie-line; \( \Delta P_s \) is the loss load.

5. Coordination optimization model and example analysis
Basic coordination method, correlation coordination model and load index coordination method can be
used in multi-level load forecasting coordination. In this paper, the basic coordination model is used.
In one-dimensional two-level coordination, the total sum of load forecasting for lower power grids
equals the result of load forecasting for higher power grids, that is, \( r_0 = \sum_{i=1}^{n} r_i \), n is the number of
lower power grids. The goal of coordination of multi-stage load forecasting is to obtain the revised
load forecasting values \( x_i, i = 0, 1,..., n \). Considering the reliability \( \omega_i \) of load forecasting for all levels
of power grids, the weighted sum of squares of relative corrections is minimized.
\[
\begin{align*}
\min f &= \sum_{i=0}^{n} \omega_i \left( \frac{r_i - x_i}{r_i} \right)^2 \\
\text{s.t.} \ x_0 &= \sum_{i=1}^{n} x_i
\end{align*}
\]

In formula (2), weight \(i\) is expressed as the reliability of prediction accuracy. The higher the prediction accuracy is, the greater the \(\omega_i\) is. Formula (2) is expressed as a general quadratic programming problem, which can be solved by Lagrange multiplier method. The Lagrange functions constructed are as follows:

\[
L(x_0, x_i, \lambda) = \sum_{i=0}^{n} \omega_i \left( \frac{r_i - x_i}{r_i} \right)^2 + \lambda (x_0 - \sum_{i=1}^{n} x_i)
\]

The partial derivatives of \(x_0\), \(x_i\) and lambda are obtained by the above formulas.

\[
\begin{align*}
\frac{\partial L}{\partial x_0} &= \frac{\omega_0}{r_0^2} (-2r_0^2 + 2x_0) + \lambda = 0 \\
\frac{\partial L}{\partial x_i} &= \frac{\omega_i}{r_i^2} (-2r_i^2 + 2x_i) - \lambda = 0 \\
\frac{\partial L}{\partial \lambda} &= x_0 - \sum_{i=1}^{n} x_i = 0
\end{align*}
\]

So there is

\[
x_0 = r_0 + \frac{r_0^2}{\omega_0} \left( \sum_{i=0}^{n} \frac{r_i - r_0}{r_i} \right) \\
x_i = r_i - \frac{r_i^2}{\omega_i} \left( \sum_{i=0}^{n} \frac{r_i - r_0}{r_i} \right)
\]

It is not difficult to find that the coordination results of all levels of forecasting are obtained by adding the unbalanced quantity \((\sum_{i=0}^{n} r_i - r_0)\) according to the weight \(\frac{r_i}{\omega_i}\) on the basis of the original forecasting results.

Through the above analysis, this paper uses the database sample data of a typical large user load analysis and management system of a certain area to refine the bad data, refines the forecasting process to get the original forecasting results, respectively, through the general revision of the system and the optimization and coordination of multi-level load forecasting, to revise and coordinate the original forecasting results, as shown in Table 1, the average of a certain time in a month. Data comparison and error analysis. Among them, user 1 ~ 3 is the typical user load. In the table, the general amendment refers to the amendment of single lower (upper) level load to upper (lower) level load. Both of them reduce the error of original load forecasting, but the latter is more effective and improves the overall forecasting accuracy. The prediction level of the upper and lower power grids has been significantly improved.
Table 1. Comparing the original forecast with the average data of general correction and optimization coordination at a certain time of a month.

| Prediction item | Actual load (MW) | Original prediction | General correction | Optimal coordination of multi-stage load forecasting |
|----------------|------------------|---------------------|--------------------|-----------------------------------------------------|
|                |                  | predicted value (MW) | Prediction error (%) | Correction value (MW) | Correction error (%) | Optimizing coordination value (MW) | Optimizing Coordination Error (%) |
| user 1         | 20.23            | 21.27               | 5.14               | 21.05                  | 4.05                 | 20.97                           | 3.66                          |
| user 2         | 32.58            | 30.97               | -4.94              | 31.46                  | -3.44                | 31.58                           | -3.07                         |
| user 3         | 23.97            | 25.17               | 5.01               | 24.93                  | 4.01                 | 24.83                           | 3.59                          |

6. Conclusion and prospect
In view of the shortcomings of traditional load forecasting technology and system, and combined with the new demand of smart grid load forecasting, this paper puts forward the basic method of building smart grid load forecasting platform under cloud computing, and builds the framework of smart grid load forecasting platform based on cloud computing.

But the combination of cloud computing and power load forecasting has just sprouted. Although the basic framework and some load forecasting services are proposed in this paper, the practical problems of cloud computing in the development of power system load forecasting platform are not considered thoroughly. With the development of smart grid, it should be expanded accordingly.

Acknowledgments
This work was financially supported by the science and technology projects of headquarters of State Grid Corporation (Research and application of non-stop state Adaptive upgrade technology for information system).

References
[1] C.Svingos, T.Mailis, H.Kllapi, et al. Real time processing of streaming and static information. In: 2016 IEEE International Conference on Big Data (Big Data). Washington, DC: 2016, 410-415.
[2] BAI Hong-wei, MA Zhi-wei, SONG Ya-qi. Processing of smart grid monitoring data based on cloud computing [J]. East China Electric Power, 2011, 39 (9): 1485-1487.
[3] FAN Yin-jiu, MI Juan. Cloud computing in the smart distribution automation applied research [J]. Computing Knowledge and Technology, 2011, 39 (1): 1485-1487.
[4] HE Mai, LIU Jun-yong. Experimental system design of power network loss calculation and analysis based on cloud computing platform [J]. Experimental Technology and Management, 2011, 28 (7): 112-116.
[5] A.Ringlstetter, S.Scherzinger and T.F.Bissyande. Data Model Evolution Using Object-NoSQL Mappers: Folklore or State of the Art? In: 2016 IEEE/ACM 2nd International Workshop on Big Data Software Engineering (BIGDSE). Austin, TX: 2016, 33-36.
[6] QIN Hai, SHI Wei-feng, ZHANG Dan. Service oriented dynamic architecture description language SO-DADL [J]. Computer Engineering and Applications, 2011, 47 (6): 68-74.
[7] Ye Q, Qian F, Gu Q, et al. Research and application of big data technology in China southern grid [C]. IEEE International Conference on Power System Technology. IEEE, 2016.1-6.
[8] Chen L, Chao Y, Ma Y. Risk Warning System Based on Big Data Applied in the Power Informatization of State Grid [C]. In: International Conference on Information Science and Control Engineering. IEEE Computer Society, 2016: 578-582.
[9] S.S.Refaat, H.AbuRub and A.Mohamed. Big data, better energy management and control decisions for distribution system in smart grid. In 2016 IEEE International Conference on Big Data (Big Data), Washington, DC: 2016, pp.3115-3120.
[10] Anish Gaikwad, Penn Markham, Pouyan Pourbeik. Implementation of the WECC Composite Load Model for utilities using the component-based modeling approach [A]. 2016 IEEE/PES Transmission and Distribution Conference and Exposition, 2016.

[11] S.Zhang, D.Zhang, Y.Zhang, et al. The research on smart power consumption technology based on big data. In: 2016 International Conference on Smart Grid and Clean Energy Technologies (ICSGCE), Chengdu, China: 2016, pp.12-18.

[12] Wesin Alves, Daniel Martins, Ubiratan Bezerra, et al. A Hybrid Approach for Big Data Outlier Detection from Electric Power SCADA System [J]. IEEE Latin America Transactions, 2017, Vol.15 (No.1): 57-64.