

# Overview of Construction and Application of Big Data Analysis Platform

Difei Zhang

Zhongke Xingtu Co., Ltd, Beijing 101399, China

**Abstract:** Big data technology is widely spread around the world, and is constantly developing and applying. In order to enhance the application value of big data analysis platform, it is necessary to constantly improve the data analysis and processing capacity of big data platform, so as to build a complete data analysis platform, realize resource sharing and real-time data collection. As a key point of contemporary information development, big data analysis platform is of great significance to promote social data exchange. Based on this, this paper focuses on two aspects: first, it describes the construction process and content of big data platform; second, it summarizes the relevant application and development of big data platform for reference.

**Key words:** Big data; Platform construction; Application

**Publication date:** November, 2020

**Publication online:** 30 November, 2020

\***Corresponding author:** Difei Zhang, 3051309461@qq.com

In the era of big data, the improvement of big data analysis and processing ability provides technical support for the analysis and mining of massive information. In the construction of big data analysis platform, the system architecture is particularly important, which is directly related to the creation of platform functions. At present, there are four types of big data analysis architectures, among which Hadoop is the most widely used one to build big data analysis platform<sup>[1-2]</sup>. On this basis, according to the characteristics of big data analysis platform and system architecture requirements, the path of building big data platform is introduced in detail.

## 1 Construction of big data analysis platform

In big data analysis applications, building a platform is particularly important because it is directly related to the effectiveness of the analysis application. The construction of big data analysis platform involves various fields and high technical requirements, and the effective implementation of platform functions is emphasized in the design of technical architecture<sup>[3]</sup>. For example, Hadoop is the most common application in big data analysis platform, and it can guarantee good big data analysis performance in platform function and analysis application. Therefore, in the construction of big data analysis platform, it is necessary to realize platform system functions based on mainstream platform architecture and apply cloud computing technology to improve the platform construction efficiency.

### 1.1 Commonness analysis

There are four major data analysis platforms in the mainstream architecture. The implementation of platform architecture needs the technical support of cloud computing and big data technology, which is also the common function of the four mainstream architectures. The big data platform architecture includes five main parts: Cloud computing engine, high-performance data mining algorithm library, workflow engine and cloud storage are directly related to the efficiency of building big data analysis platform. In the process of building the big data analysis platform, it can be seen from Figure 1 show that the application value of the big data analysis platform is guaranteed by building a "five big" functional system, and a large number of information processing and mining are realized. At the same time,

in the process of building high-performance large-scale data mining algorithm library, we have realized the analysis and mining application of complex large-scale data. Therefore, in terms of generality, the architecture of the four big data analysis platforms focuses on using big data analysis and mining algorithms to effectively process big data.

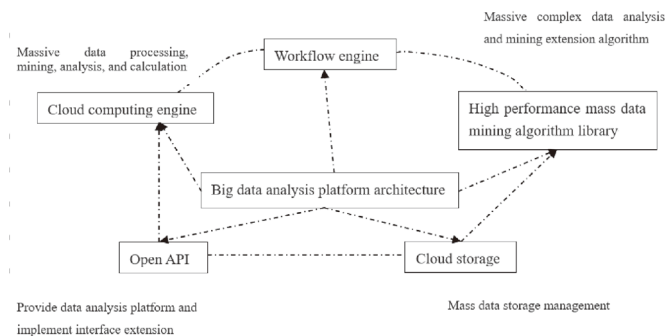


Figure 1. Architecture of big data analysis platform

## 1.2 Difference analysis

Due to the composition and architecture characteristics of data analysis platform, there are some differences in data analysis platform. This paper mainly refers to four mainstream data architecture platforms, and each platform architecture has its own differences. Because the extension content of each platform architecture is different, the main differences are reflected in four aspects. The following describes the differences of each platform.

### 1.2.1 Hadoop

The scope of this project is different from other projects, covering programming model, i.e. design framework and algorithm, and HDFS (distributed system file). This project mainly deals with the content of system files and Hadoop common analysis. Therefore, it is different from "spark, storm, samza".

### 1.2.2 Spark

The application scope of the framework is relatively wide, because when the framework is designed, it mainly relies on the basis of rod, and the main focus is to process data and manage the effectiveness of data generation.

### 1.2.3 Storm

The performance of this framework is mainly due to the influence of two components, one is the stateless nimbus component, the other is the stateless supervisor component. Because of the stateless

component, the focus of framework performance has changed.

### 1.2.4 Samza

The framework mainly focuses on data processing, analysis, and command execution. This framework can greatly meet the needs of data analysis, and can also effectively analyze massive data.

## 2 Overview of large number platform application

### 2.1 Necessity and challenge of smart grid big data application

The main requirements and challenges of big data analysis project for power consumption mode of power users are as follows. The first is the preprocessing of data noise. When collecting data, due to the influence of acquisition equipment and conditions, the collected data noise contains more outliers and missing values<sup>[4]</sup>. Therefore, it is necessary to find and process all kinds of noise data timely and accurately to ensure that the analysis results are efficient in the process of modeling. The second is the merging of multi-source heterogeneous data. In the application of the target project, it is necessary to process and analyze the user's information, geographic location and power equipment to provide consistent data for subsequent modeling and analysis. The third is power mode analysis model. According to the management requirements of current market operation, power grid companies must analyze specific objectives in different scenarios and design analysis models for different power consumption modes. The fourth is a platform for enhancing big data. The purpose of building application platform is to realize automatic analysis and large-scale application of power big data. The whole process of big data is collected, integrated, preprocessed and further analyzed.

### 2.2 Availability of big data

Now, the emergence of big data is confusing. This will cause problems when analyzing big data, and data quality is an important factor in the process of big data analysis. There are five aspects of data quality availability integrity, data accuracy, data consistency, data timeliness and importance.

### 2.3 Lack of professional tools for big data

Communication and network technology is developing continuously. Use information technology to record data resources. However, data analysis technology is now limited, can not make better use of human ideals, this is the problem to be solved. Therefore, it is possible to further optimize the design of objects by implementing computer programs, and the efficiency of software management, distribution and utilization has been improved through computer technology. But in today's complex and highly specialized environment, these old tools can no longer process all human data. With the rapid growth of telecommunication service data and medical health big data, there is no mature data analysis tool to process and process these data.

### 2.4 Countermeasures for big data analysis

Therefore, the availability and timeliness of big data analysis must be improved. Secondly, the introduction of cloud storage technology. The storage mode of big data not only affects the analysis and processing efficiency of big data, but also develops a method of storing data in the cloud, which can store data more effectively and cheaply.

## 3 Summary

In China, rich data resources, diversified application scenarios, potential consumer market and the economic and innovative value of big data are

attracting the attention of all walks of life. Therefore, the infrastructure needed in the era of big data is to build a big data analysis and application platform. It can bring a lot of value to the development and deployment of big data. For example, the first aspect is that it can break through the technological limit, provide good services and infrastructure for big data development, and accelerate the pace of innovation. The second is to use the fruits of today's rich big data technology innovation to make the next generation technology perform well in application services. The third part refers to the technical products and industry solutions that can solve the problem of big data, so as to accelerate the development of big data industry in China.

## References

- [1] Wang X. China mobile digital intelligence integration Hubble analysis platform helps supply chain digitalization [J]. Bidding and procurement management, 2020(9): 20-22.
- [2] Cui YB. Research on the construction of big data analysis platform for local literature in digital humanistic era [J]. Henan Library Journal, 2020, 40(9): 112-113.
- [3] Wu HY. Research on the construction of agricultural big data analysis platform [J]. Market weekly, 2020, 33(9): 5-6 + 40.
- [4] Liang B, Zhen XJ, Li JL, et al. Research on the design of big data analysis and application platform for forestry and fruit -- a case study of Xinjiang production and Construction Corps [J]. Journal of Central South University of forestry science and technology, 2020, 40(9): 173-182.