Big Data Technology Progress And Development Trend

Wei Jin
Northwestern Polytechnical University Mingde College 710124

Abstract: In the era of Internet+, big data technology has been widely used in all walks of life, and has achieved good application results, including university education industry. From the perspective of the overall structure, big data technology can not only maintain campus network security, but also improve the university education platform, assist students to obtain valuable information, and promote effective communication between teachers and students. This article will layer on the progress and development of big data technology, and put forward some personal opinions.

1. Introduction
Big data is an emerging product in the network era. It can effectively solve the problem of information integration processing, optimize data computing mode, improve data utilization efficiency, and maintain network security. This article will briefly introduce the concept of big data technology, analyze the progress of big data, and discuss the development trend of big data technology with the university network as an example.

2. Concept of Big Data Technology
Big data technology is an emerging product of the Internet+ era, which uses new thinking patterns to process and process data information. In the age of the Internet, there is a wide variety of data and information, and the amount of information is extremely large. If traditional processing methods are used, it is inevitable to deal with huge data information. In this context, big data technology has emerged, which not only helps innovative data processing. The method can also form a complete technical system, correctly process and process data information, correct data errors in time, and comprehensively improve data application effects.

In addition, big data technology has four V features, which can use different access methods to complete data access and operation. The so-called four-V features refer to Volume, Variety, and Velocity. And Value (low value density), in which Volume (large mass) big data technology has ultra-large-scale unstructured data, which can achieve rapid data growth, the total number of which accounts for 80% to 90%, and the growth rate is high. 10 to 50 times the structured data. Variety is mainly embodied in four aspects – first, the heterogeneous diversification of big data. Second, the data forms are diversified (data forms mainly include images, texts, documents, video and machine data, etc.); third, patterns and modeless are not very obvious; fourth, grammar and sentence qi are diversified. Velocity (rapidization) mainly means that big data technology can quickly perform real-time analysis work, achieve efficient data input and processing, and discard, with immediate results. Value (low value density) means that big data technology usually collects a huge amount of information, including many irrelevant information, can predictive analysis of future models and trends, and do in-depth analysis on artificial intelligence, machine learning and business intelligence.
3. The Progress of Big Data

3.1 Data Collection

Data collection is the basic working method of big data technology, and the most important way. From the micro perspective, data collection methods are divided into two types: centralized collection and distributed collection. Both of these acquisition methods integrate and analyze the collected data information, and extract valuable information from the massive data to support the development of the application industry. Secondly, data exists in different ways. Big data technology can be used for structured and unstructured collection of different data, and information systems are used to identify data information. Third, big data technology can clean and quality control data information, and integrate data with a variety of technologies to form a data set, which in turn makes the data collection and finishing method more perfect. In addition, the real-time data collection methods are divided into the following three types:

First, the front-end log collection access. This kind of acquisition method has the advantages of strong effectiveness, high reliability and good usability. In the process of technology selection, the most valuable information can be analyzed through the comparison of open source data collection tools flume, scribe and chukwa tests.

Second, the back-end log collection access. FileCollect, considering that many online environment variables can not be changed; in order to reduce intrusion, most of them are currently using Go language to achieve file collection, and are also working on refactoring this module.

Third, business data access. This method mainly uses Canal to realize incremental data of synchronous services, and its access method can be subdivided into the following four types:

① Uniform data access: For the processing specification of the data stream link, all data access data centers must be uniformly reported to the Kafka cluster through the data collection gateway to avoid the processing problems of multiple access methods at the back end.

② Data Real-Time Cleaning (ETL): In order to alleviate the resource pressure of the storage computing cluster, improve the reusability of data, decompress, decrypt and confuse data at high speed, complete the defective data information, and do the abnormal data processing work. Data processing needs to be cleaned in real time using big data technology.

③ Data cache reuse: In order to avoid the instability of large data streams (400+ billions/day) being written to HDFS and HDFS clients, and improving the real-time performance of data, it is necessary to rewrite the data after real-time cleaning to Kafka and do a good job. Save the work and then move the KG-Camus to HDFS by offline calculation (batch) (configure the corresponding job plan through the job scheduling system). For now, the basis for real-time computing is to use Storm/JStorm to consume directly from Kafka to design the most complete storm-kafka component.

④ Offline calculation (batch processing): This method is mainly implemented by spark and spark SQL. Its overall performance is 5 to 10 times higher than hive. Moreover, hive script has been converted to Spark or Spark SQL, and some complicated jobs are still used. Hive/Spark way to achieve.

3.2 Perfection of Data System Architecture

From the perspective of micro-structure analysis, the big data technology system architecture is mainly composed of a data preparation system, a data storage management system, a calculation processing system, a data analysis system, and a data result output system. Each of these five systems has its own characteristics and plays an important role in improving data analysis and processing capabilities and practical application capabilities. In the case of a data storage management system, the system can overcome the shortcomings of the traditional data storage mode and store massive amounts of data.

3.3 Data Calculation Mode Management

At present, the computing model commonly used by big data is cloud computing, an emerging computing model that integrates distributed computing, Utility Computing, Grid Computing, and
Parallel Computing, virtualization, network storage (Network Storage Technologies) and load balancing (Load Balance) advantages [3]. Secondly, cloud computing technology has diversified features, can achieve the organic combination of multiple computing modes, and complete the calculation and processing of massive data in the first time, thus effectively improving the computing efficiency. In addition, to ensure the accuracy of data calculation, the data analysis, processing and integration capabilities of the cloud computing model must be continuously enhanced.

4. The Development Trend of Big Data Technology

4.1 Data Network Security Management

In the internal network construction of colleges and universities, big data technology is widely used in network security management, which makes the campus network security protection system more perfect, and the network problem recognition ability has been effectively improved. Secondly, big data technology can realize the aggregation, sharing, analysis, integration and effective allocation of network security information, and effectively enhance the emergency response capability of network system security risks. Third, big data technology can perform network security monitoring and early warning work at any time to prevent data loss, stealing and disclosure, and timely eliminate network viruses. In addition, in the data collection, storage and transmission work, big data can ensure the security of these workflows and build a complete network security system. Therefore, under the support of big data technology, the campus network security protection system will be more robust.

4.2 Data Platform Construction

To promote the good development of big data technology in college education and improve the quality of education, it is necessary to build a good data platform. In the process of building the platform, teachers should pay attention to the main role of students and form a data platform with students. The platform provides students with free data information and a wealth of learning resources. In addition, teachers should patiently teach students to use the spare time to log in to the data platform, download the learning content they need, and view valuable data information. On the other hand, teachers can use the data platform to strengthen their interaction with students, keep abreast of the students' learning situation, and actively ask students whether they have difficulties in daily learning and homework, and help students answer questions. At the same time, teachers can let students share their own learning insights on the data platform and upload valuable learning information, thus effectively improving the sharing of the data platform [4].

4.3 Big Data Platform Data Storage Model is Maturing

Looking at the development trend of big data technology, the data storage model of big data platform is becoming more mature and perfect. At present, the big data platform data storage model consists of DCL (English full name is Data Cache Layer, ie data buffer layer), DDL (full name is Data Detail Layer, ie data detail layer), Common (common data layer), DSL (full name is Data Summary Layer (data summary layer), DAL (English name is Data Application Layer, Chinese name is data application layer), Analysis (data analysis layer), Temp (temporary mention layer). Among them, DCL (English data Cache Layer, the data buffer layer) is mainly reported by the client or the storage service system, and is cleaned, decoded and converted to prepare the data filtering work. DDL (full name is Data Detail Layer) is to further detail the data by filtering the data of the storage interface buffer layer. Common (Common Data Layer) is mainly used for external business system data and storage dimension table data. DSL (full name is Data Summary Layer) has three functions: First, storage according to business themes, lightweight summary data, user behavior wide table data and detailed data, while doing a good job with public data layer data management The user behavior theme data storage work; second, provide the basic data required for the data application layer statistics work; third, do the data cluster protection layer data cluster protection work. DAL (English name is Data Application Layer, Chinese name is data application layer) is not only responsible for Operations
Analysis (Storage Operations Analysis), Metrics System (indicator system), Online Service (online service), User Analysis (user analysis), etc. Work, and can achieve effective storage of external output data and some hot data external services, and DSL layer loading query for data within a certain period. Analysis (data analysis layer) is mainly to store the common data layer, the data detail layer, the data summary layer, and then to count them, and provide data resources that meet the intermediate results for data mining needs work such as list, advertisement, recommendation, etc. Temp (temporary mention layer) is used to store temporary data output by temporary mention and data quality check work.

5. Conclusion

In summary, big data technology is an emerging product of the Internet+ era, which uses new thinking patterns to process data information. Compared with traditional technologies, big data technology not only helps innovative data processing methods, but also can form a complete technical system, correctly process data information, correct data errors in time, and comprehensively improve data application effects. For the moment, data collection is the basic working method of big data technology. It can integrate and analyze the collected data information, extract valuable information from massive data, and ensure the accuracy of data calculation to support the development of application industry. In addition, under the support of big data technology, the campus network security protection system will be more robust, and the data platform construction and database management model will become more mature. On the other hand, in the process of building a big data platform, teachers should pay attention to the role of students, and form a data platform with students to provide students with free data information and rich learning resources.

References:
[1] Wang Zishu, Hu Di. New Progress in Environmental Management in China and Prospects of Environmental Big Data Technology Application [J]. China Environmental Management, 2018 (10:1).
[2] Duan Qingling, Liu Yiran, Zhang Lu et al. Research progress and development trend of big data technology in aquaculture [J]. Science and Technology Innovation and Application, 2018 (06:2).
[3] Li Yan, Chen Fenglei. Analysis of Big Data Technology for Network Education [J]. China High-tech Security, 2016 (07:3).
[4] Jiang Lizhen, Qi Shengkui, Lu Zhengxia, et al. Application analysis of big data technology in power operation [J]. Science and Technology Communication, 2015 (04:4).
[5] Ye Lijia. Application Research of Big Data Technology in Communication Field [J]. China New Communication, 2018 (04:5).