

Review of research on big data technology in the field of petroleum exploration

Ke Hou^{1,*}, Xinhao Li², and Mingcheng Guo²

¹ Xi'an Shiyou University, School of Economics and Management, 710065 Xi'an Shannxi, China

² Xi'an Shiyou University, School of Economics and Management, 710065 Xi'an Shannxi, China

Abstract. In the era of big data, the rich data resources of the petroleum industry not only provide auxiliary decision-making support for the production and scientific research of petroleum enterprises, but also play a vital role in the digital transformation of enterprises. From the perspective of petroleum exploration data management technology, this paper divides its past development process into database technology stage, data warehouse technology stage, data lake technology stage, data integration platform stage, and special application of big data technology by combing the relevant literature. At present, it is in the initial stage of the construction of the data middle office. The core problem that needs to be solved is to match the calculation type with the most suitable computing framework, optimize the allocation of computing resources, and give full play to the processing capabilities of the data middle office.

1 Introduction

With the increasing amount of data in the petroleum industry and intelligent scenarios, traditional petroleum data platforms can no longer meet the needs of current data services. The field of petroleum exploration is the front-end link of the entire industrial chain of petroleum industry. Exploration data is not only used to analyze the stratigraphic structure before petroleum extraction, but also used to optimize petroleum extraction methods to increase petroleum extraction and other situations. From the perspective of business needs, the field of petroleum exploration is more suitable as a pilot field for building data middle office. For the petroleum industry, the key technical problem is to match the calculation types and computing frameworks of the data middle office in the petroleum business system. POSC provides a set of unified international standards for the design and implementation of data management and application systems for petroleum exploration and development. Therefore, on the basis of POSC standard, this paper discusses how to match calculation types and computing frameworks in the construction of data middle office.

* Corresponding author: kehoul@188.com

2 POSC standard and business type

In 1990, the five major petroleum companies in the United States jointly initiated and established the POSC organization. The petroleum industry data standard defined and published by POSC have become industry standards. Among them, the software integration standard mainly includes: EPICENTRE data model, data access and exchange (DAE), PEF data exchange format, and basic computing standard (BCS) which provides computing system technical standards for petroleum exploration and development, and supports the interoperability and data exchange of application programs. PetroChina's exploration and development business involves the main links such as petroleum exploration, reservoir evaluation, development and production, petroleum and gas processing, storage, transportation and sales. The specific sub-business has a clear matching relationship with the type of POSC business [1].

3 The development of petroleum exploration data management technology

At present, the research and development of data management technology in the petroleum industry has gone through five stages (database technology stage, data warehouse technology stage, data lake technology stage, data integration platform stage, special application of big data technology), and it is moving towards the data middle office stage.

3.1 Database technology stage

The early construction of digital petroleum fields helped the petroleum industry to accumulate large-scale archive data. With the continuous increasing in the scale and types of real-time data, the original data storage technology has been unable to meet the demand. Zhang Jiachen combined with data compression technology to save data storage space; at the same time, taking the cache optimization of the database compression system in different storage environments as the entry point, corresponding measures such as cache optimization were given [2].

3.2 Data warehouse technology stage

The research of the data warehouse technology stage is mainly focused on using various professional databases to establish a petroleum industry data warehouse which is oriented to subject application, and the construction of a comprehensive petroleum data platform with data set sharing and data mining capabilities. On the basis of data warehouse dimensional modelling theory, distributed message queue and distributed streaming computing, Liu Yanjun proposed a distributed real-time data warehouse scheme for the multi-data scenario, rapid iteration, high scalability and data reliability [3].

3.3 Data lake technology stage

As the types of unstructured data that can be obtained and analyzed in the petroleum industry are increasing, the value of data utilization in petroleum production and scientific research is rapidly emerging. The processing capabilities of traditional relational databases can no longer meet the needs of high-concurrency and time-efficient applications. Scholars combine current advanced theory and technology to study and use data lake technology from the perspective of petroleum data storage, management, and application. In order to

realize the sharing of various professional databases, improve the efficiency and accuracy of data utilization and promote the construction of an integrated data fusion platform, Li Guoxin proposed a plan to build a data lake that supports the full life cycle [4]. Aiming at the time-validity of data update in centralized data lakes, Tan Jingxin proposed a data lake design idea with distributed architecture [5].

3.4 Data integration platform

In order to facilitate data management and make full use of existing data, the research at this stage focused on the overall strategy of building an integrated petroleum data platform. Gong Faming used Noe4j's storage model and two-tier index structure retrieval algorithm to build a petroleum data information storage system, which saved more than 10% of storage space and increased the information retrieval speed by 30 times [6]. Taking the business data and business systems of petroleum companies as the starting point, Dong Xisong used Hadoop to build a data sharing platform for data extraction, conversion, cleaning and loading [7].

3.5 Special application of big data technology stage

In the process of integrating the data management business of the petroleum industry with an integrated platform, it was found that the processing capacity of the original business system could no longer meet the demand of massive data processing. At this stage, big data technology has not only been widely used in seismic exploration, production and development, but also showed great value in oil and gas transportation, refining and chemical industry and refined oil sales. The research directions at this stage are mainly divided into the following aspects: the application research of big data technology in the field of oil and gas exploration, research on the application of petroleum big data technology in the field of production and development and the application research of big data technology in the field of oil and gas transportation.

3.6 Data middle office stage

The data middle office is a set of sustainable mechanism, which supports the service-oriented comprehensive platform of business activities by collecting, calculating, storing, processing and analyzing massive data [8]. The data middle office can meet the high availability requirements of multiple data processing types for different data application situation in the petroleum industry, and can meet the high availability requirements of multiple data processing types for different data application scenarios in the petroleum industry, and make up for the weak adaptability of special big data technology applications to the environment. The data middle office can provide higher data processing efficiency for the upper business system on the premise of ensuring the universality of the system.

The data middle office divides the calculation into four forms: batch calculation, stream computing, online query and Ad-hoc analysis, and different scenarios are realized with different storage and calculation frameworks. The outstanding advantage of the data middle office is making full use of internal and external data responding quickly in the face of complicated and scattered massive data, breaking the status of petroleum data islands in both storage and processing, and creating data assets with continuous value-added lowering the threshold of data service [9].

4 Analysis of calculation types of petroleum exploration data processing Job

The various applications supported by the data development layer of the data middle office for petroleum industry will contain inevitably many different types of business processing requirements. In most cases, these pending businesses are complex jobs composed of various types of operations. Therefore, it is necessary to analyze and study the calculation types corresponding to the simple jobs that constitute the above complex jobs. At present, the typical calculation types involved in the petroleum industry business system are mainly offline calculation and cyclic iterative calculation.

4.1 Offline computing

While the sorting of Common Middle Point (CMP) gathers [10] in seismic data processing stage, it is necessary to match each trace header data in seismic data according to data attribute characteristics, which belongs to batch computing mode (offline computing mode) of basic computing types. The data processing of logging process also belongs to batch calculation type, but it has high time-validity requirement, which usually requires submitting logging interpretation result report within two hours after receiving logging data, and also includes calculation process for various data, such as calculation of shale content, porosity, water saturation and formation permeability.

4.2 Iterative calculation

The static correction link needs to complete the tasks of velocity analysis, dynamic correction, cutting and superposition, time difference correction with zero dip angle (NMO) and time migration before falling through multiple iterations [11], until the residual static correction is small enough, which belongs to the typical iterative calculation type. Liu Yiwen proposed an improved static correction method for complex seismic data imaging in mountainous areas. For medium and long wave static correction, the model and static correction are used for quantitative calibration. In order to ensure that the static correction results meet the requirements in the quality control process, the cycle operation is adopted in the process of model evaluation and static correction quantitative evaluation [12].

5 Conclusion

The informatization construction of petroleum industry has experienced many years, from the past database, data warehouse, data lake and integrated platform construction stage to the special application and development stage of big data technology, which reflects the transformation process from data accumulation to data assetization. The data middle office has stronger processing capacity and wider application scenarios than the previous stages, which can meet the new requirements of petroleum exploration data management well. Through combing the relevant literature of petroleum exploration data management technology and analyzing the data processing methods of exploration operations, we find that the research on petroleum exploration data management mainly focuses on the storage methods, while the research on data processing methods emphasizes the realization of efficient processing according to specific business requirements. However, faced with the heterogeneity and complexity of massive petroleum exploration data and business types, how to achieve the best matching between business and calculation restricts the processing capacity of the whole system. According to the calculation characteristics of different

business data processing, it is necessary to select the most suitable calculation framework to perform the corresponding calculation tasks, so that the system has high generality and better system performance. Therefore, matching calculation types and computing frameworks involved in massive data processing operations will be the key research topic in the future development of massive petroleum exploration data management technology.

This research was supported by the Key Scientific Research Plan of Shaanxi Provincial Department of Education (Project Name: Study on the Construction of Supervision System for Large Energy Enterprises in Shaanxi Province. Project Number: 20JZ076) and the Teaching Project for School of Economics and Management of Xi'an Shiyou University (Project Name: Research on Comprehensive Teaching Mode of "Database Course Group" for Economic and Management Majors)

References

1. Wang, H.P., Ge,J., & Wang, J. Oilfield business analysis and modeling method based on POSC[J].Computer System Applications,19(03):100-102 (2010)
2. Zhang, J.C., Liu, X.G., & Wang, G. Cache optimization of compressed databases in multiple storage environments [J]. Computer Applications, 38(05): 1404-1409+1435 (2018)
3. Liu Y.J., Feng, Y., Wu, Q.H., & Huang, B.Q. Distributed real-time data warehouse construction technology for large-scale equipment status analysis[J]. Computer Integrated Manufacturing System, 23(10): 2324-2333 (2017)
4. Tan, J.X., Liu, Y.L., & Li, H.J. Research on the construction method of distributed data lake driven by virtualization model[J]. Computer Science and Exploration, 13(09): 1493-1503 (2019)
5. Li, G.X., Wang, F., Pi, X.J., & Liu, H. Thoughts and suggestions on the optimization and application of unconventional oil and gas reservoir geological engineering integration data[J]. China Petroleum Exploration, 24(02): 147-152 (2019)
6. Gong, F.M., & Li, Y.R. Research on Ontology Data Storage in Mass Petroleum Field Based on Neo4j[J]. Computer Science, 45(S1): 549-554 (2018)
7. Dong, X.S. Research on the construction of big data integration and sharing exchange platform for oilfield enterprises [C]. Chinese Society of Automation. Proceedings of China Automation Conference 2018 (CAC2018). Chinese Society of Automation: Chinese Society of Automation: 145-150 (2018)
8. Tan, H. Detailed explanation of Alibaba Cloud Data Center [N]. China Information Weekly, 2019-10-28 (014) (2019)
9. Zhong, H. Alibaba's strategic thinking and structure in China and Taiwan [M]. Beijing: Machinery Industry Press (2017)
10. Liu, L.F., Wei, X.C., Huang, Z.Y., Wang, L., & Li, L.H. Parallel gather extraction technology based on database[J]. Petroleum Geophysical Prospecting, 2015, 50(03): 565-568+8 (2015)
11. Yang, C.Z., Jiang, H.Z., Feng, Y.Q., & Du, C.J. Comparative analysis of the application effects of human-computer interaction and automatic tomographic static correction technology[J]. Geophysical and Geochemical Exploration, 42(04): 745-752 (2018)
12. Liu, Y.W., Luo, Y., Yin, L.L., & Guo, Q. Datum static correction methods and quality control strategies for complex mountainous exploration areas in southern Jungfrau[J]. Geophysical and Geochemical Exploration, 42(06):1209-1214 (2018)