The Transformation of Enterprise Data and Business Under the Data Center Structure-Taking Kuaishou Group as an Example

Xiaoling Sun[1, †], Zihao Xiong[2, *, †], Haocheng Yang[3, †]

[1]School of Management, Wuhan Donghu University, Wuhan, China

[2]Institute of International Education, Hubei University of Economics, Wuhan, China

[3]School of International Economics and Trade, Lanzhou University of Finance and Economics, Lanzhou, China

*Corresponding author: xiongzihao@email.hbue.edu.cn

[†]These authors contributed equally

Abstract. The construction of data center plays a huge role in enterprise management and decision-making. The purpose of this paper is to study "the operation mode of the enterprise under the data middle-office architecture and the impact it serves on various businesses". Firstly, based on the discussion of the data, the concept of the data center is defined. Secondly, it discusses the structure of the data middle platform in the enterprise, and puts forward a theoretical model of its operation mode based on the discussion, and discusses the role of the data middle platform in the business. Finally, in the process of summarizing the defects and challenges faced by the middle and Taiwan, put forward development suggestions, which will help enterprises to improve the efficiency when building and using the middle platform. This research will help enterprises to further develop the value of middle and Taiwan, which is of great significance for the digital transformation of their own business and the maximization of efficiency.

Keywords: Data middle office; Data and business; Change.

1. Introduction

Recently, the "data middle platform" has rapidly become popular in the market, and both the large and small companies are rushing to accelerate the process of developing data middle platform.[1] Even though many people in the industry have a heated discussion on the management of "data middle platform" in enterprise data and business, most people still feel confused.

The concept of "middle desk" comes from Alibaba's strategy of "small front desk, big middle platform" [2]. The function of "middle platform" is to integrate resources and help "front desk" carry out various businesses [3]. Centering on the concept of "middle desk", there are some concepts such as "data middle platform", "organization middle platform" and "business middle platform", among which "data middle platform" has the greatest impact on the enterprise [4]. Not only can the data middle platform help the enterprise, but it can also be used in other areas. For example, in order to solve the problems of data dispersion, nonstandard and inconsistent in the data governance of some colleges and universities, the whole life cycle of data asset collection, cleaning, integration and sharing can be effectively managed through the data middle platform [5].

In order to better analyze the impact of data middle platform on the enterprise, we collected data from various aspects and found that the research on "data middle platform" has some problems: the definition of data middle platform is unclear. [6]"How do we now define the data middle platform" become a hot topic, and about the problems like “what is the operation mode of the middle platform”, “how to play a role in the enterprise”, and “what is the impact on the business”, as if every company has its own understanding, such as Alibaba’s strategy of "small front desk, big middle platform", and these cannot attract everyone's attention, because everyone wants to define their own data middle platform. As a result, the interpretation of the data platform becomes various, and the definition is not clear; The other problem is that the architectural model of data middle platform is not clear. According to the research and development data from various companies in the market, the architecture is very different, and the enterprise will encounter different problems in operation or public display. In

practical application, we encounter far more difficulties than these, so the scientific response to the problems brought by the "data middle platform" should be also learned by the enterprise.

The data middle platform enables us to better understand more informatization theories, but the existing literature on "the impact of the data middle platform on enterprise data and business" cannot solve the problems well. Therefore, we will take Kuaishou Group as an example to analyze the application and influence of the data middle platform on the enterprise's data and business from various aspects, understand the shortcomings in the operation, and put forward some development suggestions.

2. Research on the architecture of Kuaishou Data Center

2.1 The structure of the data center

As a mainstay of large modern Internet enterprises, Zhongtai includes Kuaishou, Baidu and Tencent. It plays the role of connecting and organizing the front combat department and the back technical resource department.[7] The middle office is responsible for the integration and processing of background resources, such as the comparison of data quarters and the optimization of the expected model. For the front office, according to the completion of regional tasks by the person in charge of each business and the analysis of the resources provided by the background, it adjusts the front office business and puts forward corresponding measures for some problems. For example, in order to solve the problem that wind farms cannot effectively utilize the remote monitoring system of wind turbines, China's wind power industry has built a set of technology architecture based on Hadoop&Spark. Capable of storing all historical data, real-time data, relationship data and video image data and other various enterprise business system data of new energy enterprise intelligence center [8]. Now the middle stage needs computer technical support to operate. These technologies mainly include springboot, mybatisplus, mybatis reverse engineering, Kafka, redis, es, logback-Kafka, Apollo: Unified Configuration Center, Multiple Data Sources AbstractRoutingDataSource, JWT authentication, groovy script (Java dynamic compilation code GroovyClassLoader), webSocket full-duplex communication (long connection), okhttp3 (springcloud feign), mongodb, lombok, Page Helper plug-in, swagger interface document generation, guava toolkit, springcloud zuul gateway, unified exception handling RestControllerAdvice, MySQL row lock select for update, Shared lock: select lock in share model, optimistic lock: add version number field through table, thread pool ThreadPoolExecutor, and task scheduling Schedule/distributed task scheduling springxxl, etc.

2.2 The operation mode of the data center

For Kuaishou, a company that mainly focuses on short video products, Kuaishou's data platform is built around products, production, consumption and so on. Let's take the overseas Kuaishou short video Kwai as an example of how the data center is built and operated. For a short video product, the data on the user side is very important. The first is the data access part. Kafka is mostly used in Kuaishou Group. The back-end technical department will set up buried points in different interfaces and windows of Kwai product. When users access these interfaces and windows, these buried points will be triggered, and then the user's original data will be recorded and stored in a temporary database. Back-end technicians will extract the field information from the original data in the database, and then import it to the ODS layer to use the JSON format. Then the second part is data distribution, which is generally executed in the form of Hiva table to add new data to the middle station.

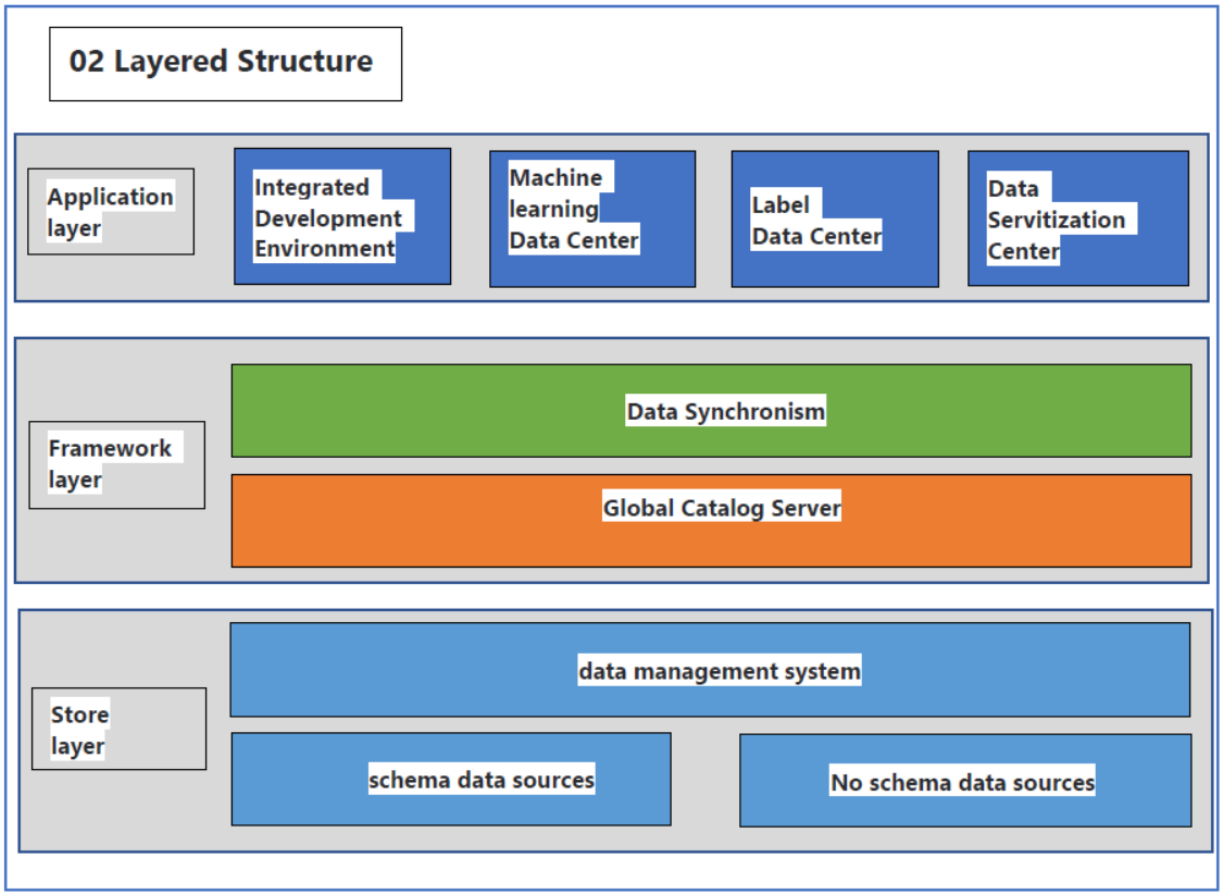

Figure 1. Kuaishou big data | Data middle platform technology exchange meeting

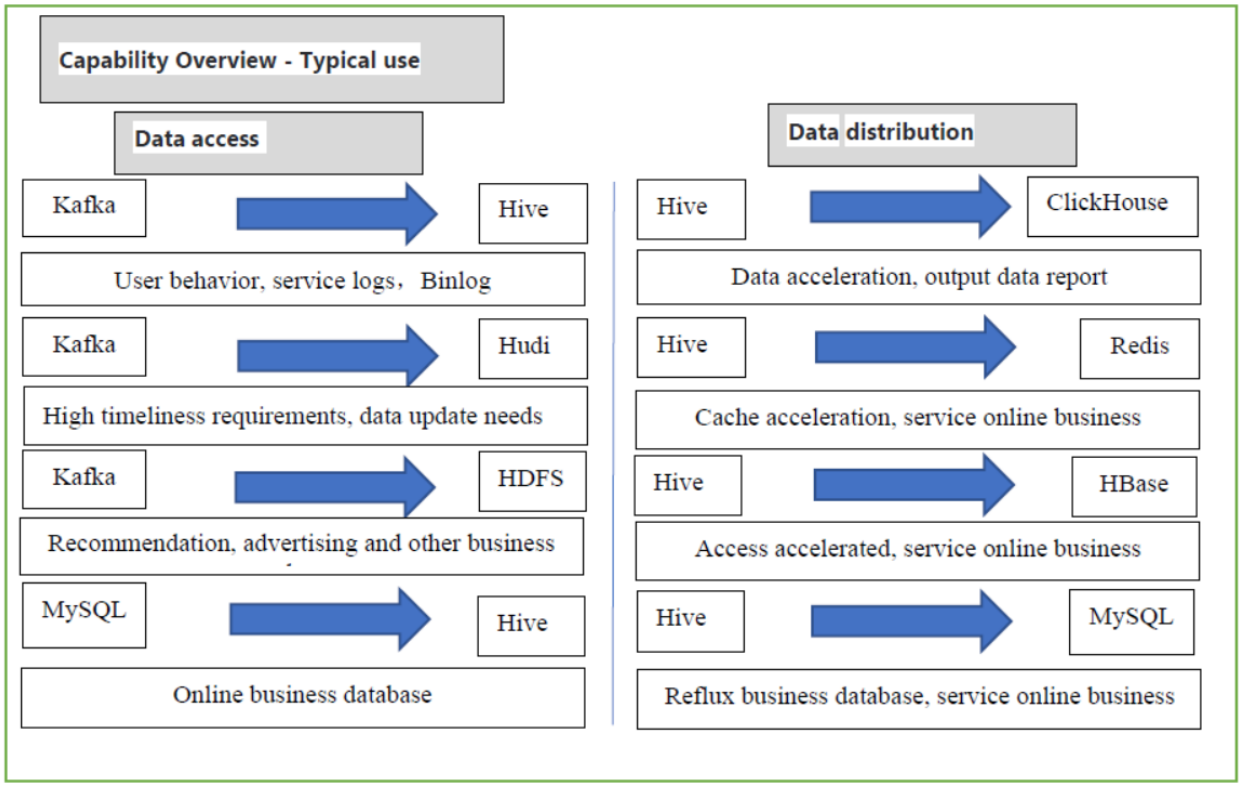

Figure 2. Kuai Shou Data access and Data distribution

Table 1. Data platform construction link

Data exchange and sharing	Intelligent gateway, shared exchange platform, API gateway, service bus
Data analysis	Kafka, Flink, Hbase, Hive, maxcompute,
Data fusion	OLTP, OLAP, RDS, CMDB, HDFS,
Data governance	ETL, Spark, Storm, MapReduce, Slipstream, Workflow
Data collection	api, agent, snmp, syslog, Sql Loader, ftp

3. Research on the Architecture of Kuaishou Data Platform

3.1 How does the data serve the business under the middle platform structure?

Under the application of the middle platform architecture, the data serves the business and makes an analysis of the business. For this part, Kuaishou has its own middle stage. For example, the product platform is a kind of data platform, which contains data in many directions: product side, consumer side, production side and so on. First of all, how to select the data we need from the massive monitoring data in the data center. We need to understand its business scope, data types and analysis methods. Kuaishou's overseas business is mainly short video products, and the main products are Kwai and snack video. Its overseas user scale reached 180 million users by the end of June 2021. Its business is mainly distributed in Brazil, Indonesia, Spanish, Middle East, South Asia and Turkey. Among them, the key areas of concern in the Spanish-speaking region are Colombia, Mexico, and other Spanish-speaking countries. The focus areas in the Middle East are Saudi Arabia, Egypt and others in the Middle East. Brazil and Indonesia are the best fast players in these regions. There are 45 million users in Brazil alone. Its global users are more than 1 billion. There are a lot of data in the Kanban in the middle of the product, such as DAU, MAU, active times, active 7 times, active 30 times, the usage time of the device app, VV, DNU, DRU, CAC, and the time proportion of different vertical categories.

For data analysis, in the monthly report and mid-year report, we generally calculate the data gap between this month and last month and the gap with the target, that is, the month-on-month growth and task completion rate, to test the completion of the business to the task target. When predicting the target, we will generally establish a model, use the historical data of DNU and CAC to establish a regression equation, make a judgment on the fitting degree, and then continuously optimize it to roughly predict the target that the business may achieve or want to achieve in the next quarter, at the end of the year or in different periods. In enhancing user stickiness, we will adopt the ab experimental method. (The ab test actually comes from hypothesis testing. We now have two random and uniform sample groups A and B, and make some changes to one of the groups A. After the experiment, we analyze the behavior data of the two groups of users, and determine whether the change has a significant impact on the core indicators we are concerned about through the significance test.) Kuaishou used this method when studying the impact of different hashtags on different users.

Based on an understanding of the business, middle-office data types, and analysis methods. Kuaishou internationalization strategy BP department will adjust the business objectives and strategic policies of different regions, so as to further optimize the direction and content of the business.Here are some data center related metrics:

Table 2. The data middle station data index of Kwai

DAU	DAU(Daily Active User)
MAU	MAU(Monthly Active User)
Active second stay	New Day 1 retention (i.e., "secondary retention") that varies per day
Device app usage time	Device app usage time
VV	“Video View”
CAC	Customer Acquisition Cost

3.2 The impact of some problems and challenges arising from the data defects in the middle station on the business

Of course, there are some problems and challenges in the Kuaishou data center. First of all, the data scale is large: ① The number of messages per second has reached the scale of trillion.[9] ② Single message structure is complex ③There are many data processing links. Secondly, there are many kinds of data sources:①Complex/different scenario②Schemaless & a data source with a schema. Schema data sources generally use MySQL, Hive, Druid, TiDB and other forms to create database table structures that can only be used; schemaless data sources generally use Kafka, Redis, HDFS, HBase and other technologies that do not constrain a series of serialized data sources. ③ Ensure the stability of data source. Here, how to reduce the cost of maintaining data and how to establish the data transmission between schema data sources and schema-free data sources are important issues of concern. Finally, the guarantee requirements are high:①Very low data latency② Zero tolerance for data loss. For data leakage and loss, employees and immediate leaders will be criticized throughout the company, and those who are serious will be dismissed directly. ③To keep the cost low, the cost of Kuaishou Group's products is mainly divided into three aspects: user-related costs, content-related costs, and operation and maintenance costs. For user-related costs: mainly fission costs (for new users) and maintenance costs (old users); followed by content costs: divided into content production, production costs and delivery costs; finally, operation and maintenance costs: mainly for server maintenance and labor costs. The most common problem in data analysis is that when screening data, the data caliber is prone to errors, resulting in a large deviation in the accuracy of business analysis. These are the challenges and problems faced by Kuaishou when building a data platform.

4. Development space

In addition to the problems and challenges faced by Kuaishou's data platform, there are also some difficulties and challenges in the development of the data platform itself. First of all, the construction cost of the data center itself is huge, and now this technology is mainly used in the construction of large Internet companies and some cities. How to use and empower the data platform to more industries is a difficulty that needs to be overcome. Secondly, the goal of the data center is to centralize and share the scattered data capabilities, and realize the ideal of "one point release, all sharing". However, before the continuity problem of the data center cannot be completely solved, the centralized data center also brings the risk of centralization. For example, once the centralized data is deleted, the impact on enterprise applications is comprehensive. The better the data center does, the higher the sharing ability, the higher the risk, which is the "sword of Damocles" hanging on the continuity of the data center, that is, "a little failure, a comprehensive impact"[10].

5. Conclusion

As an emerging technology, data center is becoming a platform support for industrial digital transformation and industrial economic transformation. At present, the concept and application of data middle ground are still imperfect. Taking Kwai group as an example, based on the analysis of its data center, we understand that the data center is supported by a large number of computer technologies, adopts the cloud computing architecture mode, fully cloud the data, unifies and standardizes the management, builds a hierarchical structure, and enables all departments of the enterprise to share resources. Through comparative analysis, we put forward effective suggestions for business. When the data middle office is running, compare the data in different periods, establish a model, use the regression equation to calculate the completion rate, and then predict the possible goals of the business in different periods according to the model, so as to help the enterprise develop. In the research of data middle office, we found that middle office has great potential in enterprise

management and decision-making, helping enterprises improve efficiency, and putting forward suggestions on development prospects.

When creating and using the middle office, grasp the depth of the middle office in the business, cooperate with each other among organizations, and reduce the cost of cooperation; Strengthen the nonfunctional attributes of the middle platform, improve the stability of the middle platform, and maximize the use of the middle platform to achieve goals.

In the future, with the continuous development and upgrading of technology, the data middle office will continue to improve, and more managers will use the data middle office to assist decision-making.[11] The digital middle office should take customer needs as the development direction, improve the business and data integration ability, and achieve channel interoperability, so as to meet the challenges brought by enterprise transformation and upgrading, and promote more coordinated and long-term development of enterprises.

References

[1] How to build a data middle platform based on DataWorks? [OL]. https://www.sohu.com/a/443224380_120054107.

[2] Alibaba announced the dismantling of the middle platform, firstly is to optimize the data middle platform's architecture? [OL]. https://blog.csdn.net/alex_yangchuansheng/article/details/112211108.

[3] Data middle platform, business middle platform, technology middle platform, AI middle platform.....they are all the same [OL]. https://www.jianshu.com/p/b98d94e66d05.

[4] Great middle platform: the base of digital government [OL]. https://xueqiu.com/5894228201/135185023

[5] Li Wenyan, Shi Yu, Chen Zhao, Pan Bo, Yang Chenglong, Peng Shenfeng. Smart middle platform construction of new energy enterprises based on big data platform[J]. The Solar Energy, 2022, (08): 92 - 96.

[6] Li Guangqian. What is data center? [J] China information circle, 2019 (06): 72 - 75.

[7] The analysis of enterprise digital transformation based on data middle platform perspective DOI: 10.19840/j.cnki.FA.2021.01.007.

[8] Han Hongguang. Analysis of Campus Information Prediction Model Based on Data Middle Platform Model [J]. Electronic Technology, 222, 51 (08): 67 - 69.

[9] Kuaishou big data | Data middle platform technology exchange meeting.

[10] Analysis and solution of common problems in the front-end middle platform system [OL]. https://www.jianshu.com/p/2252a8a20a1f.

[11] Zhang Youguo. Digital Zhongtai Helps Accelerate the Application and Innovation of Intelligent Buildings [J]. Intelligent Buildings and Smart Cities, 2022 (05): 138 - 140. DOI: 10.13655/j.cnki.ibci.2022.05.041.