ISSN: 2664-9640

Big data governance system based on Data Middle Platform technology

Shu Xu Lang[1], Xiao Lin Zhang[1], Hou Chun Quan[2], Mao Ye[2], Lan Bai[2]

[1] University of Science and Technology Liaoning, Liaoning Anshan, China;

[2]Ansteel Group Mining Corporation Limited, Liaoning Anshan, China.

Abstract

With the development of 5G, broadband business and grid operation, the amount of data involved is also rising, and data is becoming a new engine for economic transformation and development, as well as an effective tool for social governance.The whole big data governance system is divided into three parts, metadata collection, digital blood relationship, data value and heat.We will classify and integrate the collected metadata, and import it into the system to form a data governance directory, yes The data governance directory, it is convenient for the data maintenance and management personnel to maintain the data center desk, form a database, set the audit standard in advance, convenient for the data import is easy to manage the data.When the increase of data volume will certainly produce database redundancy, we determine the data heat table through the in-process governance ability and the analysis of data blood relationship.Through the heat tables, we can know which data tables are extremely important tables, and which are the tables that can be processed offline to reduce redundancy.

Keywords

Data governance system, production and operation,data center desk.

1. Introduction

With 5G, broadband services, grid operation Development, the amount of data involved is also rising. With the increase of data volume, the data platform and data are increasingly complex, thus causing data quality problems, data use problems, data security and other problems; the complex data environment brings new challenges to data operation and maintenance, data development, and thus brings many fault problems.

2. Problems with traditional data

1. Data source problems: the data platform and data are increasingly complex, so they bring data quality problems, data use problems, data security and other problems; the complex data environment for data operation and maintenance, also bring many fault problems; 2. Data processing problems: data platform and number According to the rapid growth of storage resource consumption, data development investment, developers need to consult through continuous cross-department coordination to understand the related business time long, communication costs increase, demand development cycle longer; 3. Data value: main data only quantity, but low data value density, redundant data; data is not convenient, data unclear; data operation problems: long, data processing links, data correlation, data problem positioning; data related platform more and miscellaneous, but also to the data operation and maintenance difficulty.

ISSN: 2664-9640

3. Data center

In 2015, accompanied by the " Data Middle Taiwan "concept, Alibaba Group launched the" big middle Taiwan and small front desk "."The strategy was first inspired by Finnish company Supercei Games Finland. With only 300 employees, the company launched many hit games in a short time, and became the most profitable game company in the world. Judging from the Gartner hype cycle, the word "big data" is nearing the peak of speculation, as shown in Figure 1

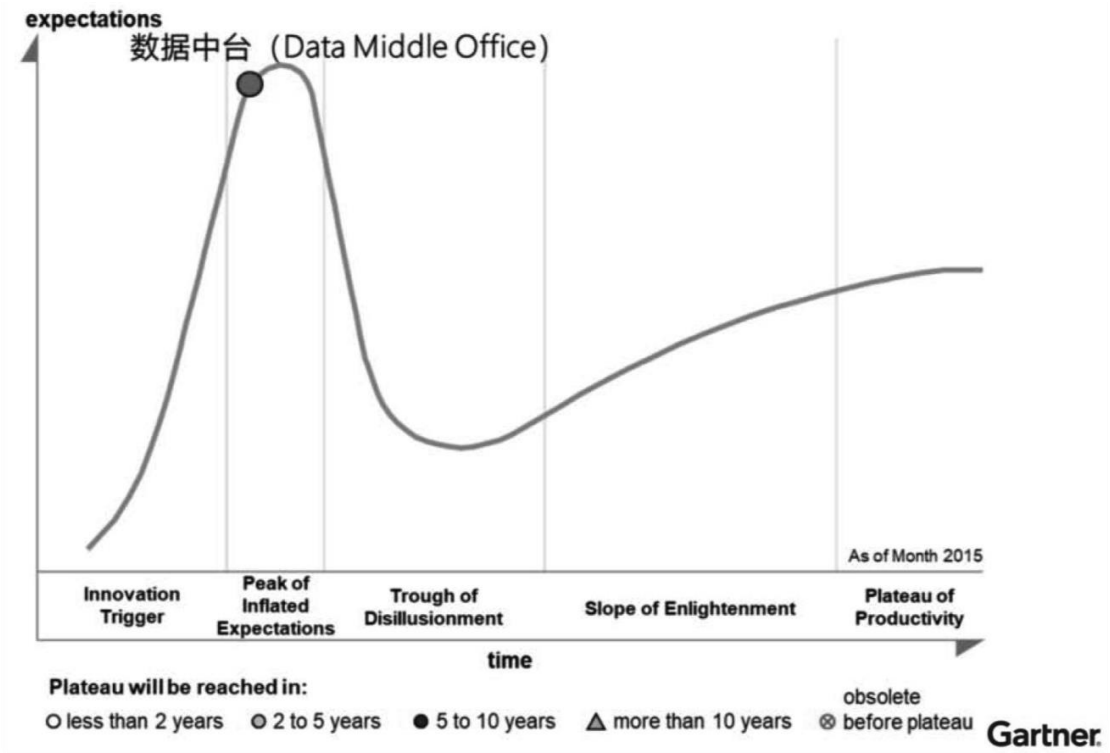

Figure 1

In the development of the traditional platform architecture "front ground and background", the front ground is composed of systems that directly use or interact with different end users, and the background is managed by a class of core resources (computing) Data) of the background system composition. However, as the company's business develops, a large number of companies are directly connected to the receiving systems in response to the continuous needs of users. Front desk systems continue to grow, and vertical business systems fill chimneys, causing reduced business flexibility and responsiveness. Now the company's application system is divided into front desk, middle platform and background, each using a different strategy. The middle of data aims to make up for the response speed of background data development and front ground business development cannot keep up with the problem [1].Unlike data warehouses, data fairs, and data lakes, a data center is not a traditional data platform. It means By collecting, storing, managing, calculating and processing massive data through data technology, it finally forms a unified standard data asset layer, and provides efficient business-level services through a unified data API interface. A reusable and shared intelligent data platform that supports services and improves the efficiency of implementing data values. Data offices are created from the data to serve the business and reflect value. The value of data is mainly reflected in four aspects: cost reduction, and efficiency improvement, business growth and organizational change. Among them, cost reduction and efficiency improvement are the value that can be achieved through traditional data systems, while business growth and organizational optimization are data intermediate Unique advantages of

ISSN: 2664-9640

office practice. Although data centers are different from traditional data platforms, the two are not separate. The structure of the data center is based on the data warehouse and the data center, which inherits the design of the data warehouse model. The data center level includes data model and data resource management, data service opening, data applications and top tag management, and ultimately maximize the value of data resources.

4. Innovation point

In the whole process of data governance, we believe that data governance is integrated with data assets, and each governance plan has a separate governance system. Our expected data governance model is both and data generation Cheng is integrated, it should not be an isolated system. Data management is with production, it is with data production is through, avoid remedial management as far as possible.

4.1. Highlight the data value

Data management spans multiple parts and is a "protracted battle", not overnight. Although the intelligent data collection and user image construction can be realized by relying on the digital technology and SAAS data center platform, it can also achieve group operation and management. However, in each link of data management, professional staff need to consult, support, cooperate and coordinate, operate effectively, maintain the interaction with the data to improve the stickiness, so that the data can reflect the truth Positive value.

4.2. Metadata collection and data asset release services

Taking enterprises as an example, the enterprise metadata is collected first. In the data collection is convenient for data management, for different formats of data classification model, the model corresponds to various table structures in the database. Collect the data in various ways. You can import it through the database. In using the database import, manually select the database, select the table, you can also customize the rules for filtering. In addition, we hope that the system can automatically import, we can write the timing task, reverse engineering the data import at the specified time point, and check at the same time Test whether the data table has been changed? When the data change is detected, we will also reflect the data version on the database. At the same time, we can also carry out a deeper management of the imported metadata, we can desensitize the data tag, whether to choose encryption, whether to regularly clean and back up and so on [2].

Form the collected metadata into an asset catalog. According to different models, form different directory types. Easy management and viewing of data in the management interface.

Import the asset information according to the asset catalogue to further improve the asset business information. We put the last formed data uniformly into the information center, in the information In the center, we can view a variety of data, but also can choose the corresponding library for direct or indirect data retrieval, and contains a variety of retrieval ways, to provide users to search.

The completion of maintenance will form an asset data knowledge base, and the knowledge base is convenient for users to consult and users to apply for assets. In forming a database will form a data map.In the process of user data application, data security problems will also design data security, desensitize the data, and make the ability to add data watermark throughout.

4.3. Build an electronic standard system and strengthen data management

These standards include basic general standards, and data Coding standards, data element standards, metadata standards, indicator adjustment standards, data exchange and sharing standards, and data management standards.

ISSN: 2664-9640

4.4. Data value and heat

In the process of data governance, both tables and models have declaration cycles. By identifying the use of data, we create a set of [3] system of data heat analysis.

Data heat information collection, mainly collects the information from the perspective of the database execution log, the database platform obtains the database operation log, and obtains the use of the table status through the log. If it is not enough to rely on the data tables alone, the numbers can also be used According to the popularity of library browsing, browsing weight, digital origin dependence for evaluation. According to this true to the data heat score, evaluate the data heat data we need to do is to guarantee. On the contrary, whether the low heat of the data is considered to do the data offline processing. In heat processing, we also consider the relationship of blood of data. Some data may be less used by users, but it plays a huge role in maintaining the whole data relationship, so we weighted the table according to this. In this way, the data processing will have a more customer official governance system and ability, convenient for operation and maintenance personnel reference and guiding significance.

We get through the heat Source analysis, you can find the data heat distribution map, and find the database of the main branch of the data heat. Find topology divisions and analyze data trends. The way of the data heat details can be exported.

4.5. Search for target data and accurately locate the enterprise data asset knowledge base

Build enterprise data asset knowledge base, provide data asset retrieval capability, classify query and retrieval capability, support the comprehensive display of models, indicators, dimensions and task assets; facilitate data operation and maintenance management personnel and data consumption personnel to explore data assets.1. Database asset retrieval: provide the keyword, Asset classification for data asset retrieval;2. Asset knowledge base: precipitate data assets such as model, index, dimension, standard documents, provide access for data operation and maintenance managers and data consumers; 3. Data asset classification display: provide different asset types for display, covering business, management, technology and different scene requirements; 4. data sharing capability: support and data service sharing platform, data discovery to the closed loop of data application; 5. data knowledge collaboration: provide communication comments, problem reporting collaboration, drive Users can improve the data asset knowledge base.

4.6. Sensitive data protection and management

Provide sensitive data standard rule management, sensitive data scanning discovery, sensitive data marking, sensitive data protection strategy and desensitization processing full closed-loop management ability; provide data sensitive data protection base support ability; sensitive data standard rule management, manage sensitive data classification, sensitive data characteristics and sensitive data protection principles; sensitive data scanning, provide cycle scheduling capability, regular sensitive data scanning and discovery, support sensitive data scanning files, database for database, support traditional relationship Type database and large database scanning, support full and sampling scanning mechanism; data marking indicates sensitive type and level of data, and support the system to automatically mark and manually marking; sensitive data protection strategy, provide sensitive data and protection strategy according to sensitive data standards and protection principles; provide sensitive data summary statistical analysis capability of data distribution; provide data desensitization capability, support data static desensitization and dynamic desensitization capability, provide encryption and fuzzy data fuzzy strategy [4].

ISSN: 2664-9640

5. Conclusion

In today's society, the coronavirus disease continues, and the international situation is unstable. excess production capacity The lack of personalized products, the inability to effectively allocate production resources, and the increasing saturation of the large equipment market have greatly affected the company's supply chain. Reducing costs and improving efficiency have become solutions to corporate profitability. In the supply chain, the intermediate stage of building a smart supply chain aims to reduce costs and improve the operational efficiency of companies. At present, the data center-based data governance system is still in the early development stage, and there is still a lot of room for development.

References

[1]Liu Xiaofeng. Theoretical Research and Path Exploration of Patent Navigation Service Based on Data Center Platform Technology [J]. Intelligence Exploration, 2022 (05): 5 9-64.

[2] Yang Lin, Wan Jun, Qiu Qingyuan. Research on Data Management Engineering and Application Technology of Intelligent Oil Field [J]. Information System Engineering, 2022 (05): 149-152.

[3] Feng Dongyu, Wang Pengda, Huang Feilong, Yue Shuang. A Cache Strategy Based on Data Heat [J]. Information and Computer (Theoretical edition), 2021,33 (12): 196-199.

[4] Zhang Zhenghao, Li Yong, Zhang Zhenjiang. Controlled and accountable sensitive data sharing scheme [J / OL]. Computer Research and Development: 1-12 [2022-06-10].