ISSN: 2664-9640

Big Data Governance Approach for Mines Based on Data Central

Wen Kai Sun[1], Xiao Lin Zhang [1], Xiao Cheng Han[2]

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

[2]Ansteel Tendering Company Limited, Liaoning Anshan, China;

Abstract

The whole big data governance system is divided into four parts, metadata collection, data governance capability beforehand, data governance capability during the event, digital bloodline, data value and heat. We will collect the metadata for classification and integration, import the system, form the data governance directory, with the data governance directory, it is convenient for data maintenance and management personnel to maintain the mine data center, in the mine data to form a database, we set the audit criteria in advance, to facilitate the import of data is easy to manage the data. Mine data as a representative of the old data, the database redundancy is relatively high, we determine the data hotness table through the ability of governance in the matter and analysis of data lineage. Through the heat table, we can then know which data tables are extremely important and which ones can be processed offline to reduce redundancy.

Keywords

Big data, data governance, mine data governance.

1. Background

As a core production factor of the digital economy, data is becoming a new engine for economic transformation and development, as well as an effective tool for social governance. With the development of 5G, broadband services, and grid-based operations, the amount of data involved is also on the rise. With the increase of data volume, the problems and adjustments faced by enterprises to increase with it.

When we face a large amount of data, data quality issues, data usage issues and data security issues are three issues that deserve attention and are often encountered. Along with the increasing complexity of data platforms and data, there are also problems such as low data value density with large development investment, increased redundant data, and difficulties in operation and maintenance.

2. Introduction

When we face a large amount of data, data quality problems, data usage problems and data security problems are three problems that deserve attention and are often encountered. Along with the increasing complexity of data platforms and data, data also has problems such as low data value density with large development investment, increased redundant data, and difficulties in operation and maintenance.

3. Big Data Governance

The data center requires the whole enterprise to share a data technology platform, build a data system and share data service capability. In fact, due to the uneven development of each business line in an enterprise, each has its own independent data processing architecture, which makes it very difficult to share data, so to build a data middle platform is not only a

ISSN: 2664-9640

change to the technical architecture, but also a change to the business operation mode of the whole enterprise, which requires the support of the enterprise in terms of organizational structure and resources. It requires understanding the business situation of the whole enterprise and sorting out the business, as well as technical support and organizational support, otherwise it is difficult to implement.

3.1. Data governance solutions

Throughout the data governance process, we believe that data governance is integrated with data assets, and each governance solution has a separate set of governance system. Our expected data governance model is integrated with the data generation process, and should not be an isolated system. The best way to achieve data governance and production is to have governance beforehand and governance during the process. It is the same as data production is throughout, and try to avoid remedial governance afterwards [1]. 2.2 Metadata collection and data asset release services

First of all, the enterprise metadata is collected. In data collection to facilitate data management, a model of classification is done for data in different formats, and the model corresponds to various table structures in the database. A variety of ways can be used for data collection. It can be imported through the database. In the use of database import, manually select the database, select the table, you can also draw up your own rules for filtering. In addition, we want the system to automatically import, we can write a timed task to reverse engineer the data import at a specified point in time, while detecting whether the data table has been changed when the data changes are detected, we will also reflect the data version to the database. We can also manage the imported metadata at a deeper level by marking the data for desensitization, choosing whether to encrypt it or not, cleaning and backing it up regularly, etc.

The collected metadata is formed into an asset catalog.

Different catalog types can be formed according to different patterns of different companies. Easy management and viewing of data in the administrator interface.

Import the asset information according to the asset catalog to further improve the asset business information. We put the final formed data into the information center, where we can view a variety of data, and also select the corresponding library for direct or indirect data retrieval, and contains a variety of search methods to provide users to retrieve.

When the maintenance is completed, a knowledge base of asset data will be formed, and the formation of the knowledge base will facilitate user access and user asset application. In the formation of the database in the macro level will form a data map. In the process of user data application will also design the data security aspects, the data desensitization process, the ability to add data watermark to do a through. [2]

For data security aspects, we define different levels for different data and label them with security levels. Certain security storage recommendations are also given in the data protection process. Access and export recommendations are also tagged to ensure data security. Relative to data with low data sensitivity, some data need to be encrypted in plaintext or ciphertext processing, etc. [3]

In addition, we also define rules for sensitive information data, such as telephone number ID card and so on rules, a high degree of freedom can also customize the rules handwritten regular expressions to form the final filtering rules.

3.2. Ex ante governance capability

The development of a data standard system is divided into two main links in governance, one is the layer volume and the other is the incremental volume. The way of layer volume is a passive way of governance maintenance. Incremental is a way to strongly control the data, and

ISSN: 2664-9640

what we want to do is to manage the incremental volume and dispose of the layer volume in a specific time.

For the standard system modeling, the current standard is mainly divided into the following kinds of assurance. There are mainly field standards, word root standards, layered domain standards, prefix standards, terminology standards, indicator standards, and dimensional standards. These standards all have different roles for data governance. For example, when creating table data, we use these standards to name the database, rather than modifying the naming format according to our own set of standards. There are several advantages to using this approach for data integration. Using the same standards when collecting identity information facilitates data integration, data management and correlation. It is also easier to understand the meaning of the data by using the same set of standards. The standards allow everyone to form a set of the same data language, which facilitates managers and developers to communicate for data integration and management, all of which can circumvent the barriers of communication difficulties and reduce data ambiguity. At the same time, our managers can also plan data according to a set of data standards system they specify.

Data standard audit, a check mechanism before the release of data, before the task goes online, to provide a standardized check mechanism. Divided into the following parts to score the data, table naming specification, field naming specification, in the task release whether there is a large table scan and so on non-compliance, temporary table there is no clean-up check, in advance a strong control of the check in the data governance before the check action. [4]

3.3. Governance capabilities in the matter

Data audit rules configuration, more from the data quality for security, the system provides many provide rules, such as whether the data is associated, data tables are consistent, between tables and tables, between files and files whether there is consistency. Normative audit, check whether the data is non-empty, whether the audit is over-length. Fluctuation audit, mainly to check the fluctuation of the trend, accuracy audit, etc. In certain audits, you can use the prescribed audit script template to audit the data, and another way is to use self-written script snippets to audit the data, to stitch together the audit system.

Data quality auditing, we can check the data quality, we can create, we can add a specific table, we can select the target schema, and finally select the rules to specify, and finally form the logic of the check, which will be executed as needed. This can be done using a periodic approach or a one-time approach. More auditing tools need to be combined with ETL tools throughout, and the ETL approach is to use the auditing process and the data scheduling process for integration.

Data quality issues are alerted, we list the results of the audits, and send the list of audited data with problems to the person responsible for the audited data to form a closed loop of data processing.

3.4. Data Value and Heat

In the process of data governance, we have statement cycles for tables and models, and we have created a system for data heat analysis by identifying the data usage.

Data heat information collection, mainly from the perspective of database execution logs to collect information, database platform to obtain database operation logs, through the logs to obtain table usage. If the data table alone is not enough, in addition can also use the database browsing heat, browsing weight, digital origin dependence for evaluation. Then according to this really score the data hotness, the evaluation of the data hotness is high data we need to do is to safeguard. On the contrary, the data with low hotness is considered to do the processing of data offline. In the heat processing also consider the relationship of data lineage, part of the data may be in the user use accounted for less, but in the maintenance of the entire data relationship plays a huge role, we will be based on this table for weighting processing. This will

ISSN: 2664-9640

have a more guest governance system and ability in data processing, to facilitate the reference and guidance meaning of operations and maintenance personnel. [5]

We can find the data heat distribution map through the heat source analysis, and find the database of the main divisions of data heat. Finding the topological divisions also has the analysis of data trends. The data hotness details can be exported in a way.

4. Mine Data Governance Solution

In solving mine data, according to the characteristics of mine data sources, mine data will face the disadvantages of difficult governance and high redundancy in the process of governance, and a new insight model table will be generated when importing the system to assist mine data governance. [6]

With the imported data, a virtual directory of assets is created to facilitate mine system managers to maintain information assets. After importing the mine data, a data knowledge base is formed. An process is performed to import and govern the data. [7]

After importing the flash data we can then develop a specific data standard management system based on the mine data. Establishing a relative data model, the established data model can then be used for the virtual reconfiguration of the old data platform in the platform for the mines. The table structure is redefined and the table names and landing standards are redefined. This allows the database to be operated again to enable auditing, allowing data that meets the criteria of the auditing rules and filtering dangerous data that does not. [8] When there is a danger of audit warning, the problem is submitted to the mine data manager, so that the mine data can be reconstructed for processing, which makes it easier to maintain the mine data later. When the mine data is online in the data governance console, we can then tag the data according to the bloodline analysis tags and evaluate the data hotness. A comprehensive data heat map is formed, and we can refer to this heat map for data processing, and we can weight the hot data for maintenance, while we can recommend the cold data for offline processing. In this way, the hot data can be maintained, and the hot data can reduce the redundancy of the database and optimize the database structure.

5. Summary

In understanding the big data governance middleware system, we learned about the big data governance platform management method. Data governance has been a hot word in recent years, and his emergence has enabled the management and maintenance of old data to become transportable.

In the big data platform stage, the user demand for data information continues to rise, the scope of the user from the data information department to expand to the whole enterprise, data governance can no longer just for the data information department, need to become an office environment for the whole enterprise users, need to use the whole enterprise users as the center, from the perspective of providing services to users, control the data information while providing users with the ability to self-service access to big data. Help enterprises achieve digital transformation.

In the old data model, including mining data, there are difficulties in governance, maintenance and so on. We gradually structure the old data into a new data governance center system, so that the old data with high redundancy can become easy to maintain, operable, reduce redundancy, reduce the burden on the server, and enhance the difficulty of data management for managers.

In the past, we relied on the primary and foreign key relationships to determine the relationship network between each table, but with the data governance platform, we can use this platform

ISSN: 2664-9640

to further understand and understand the data divisions and linkages of mine data tables through the interface visualization.

At this stage, various fields have started to build big data platforms, expecting to use the capabilities of big data to achieve digital transformation. Big data platform is actually the construction of data information, the traditional data platform encountered all the difficulties big data platform may encounter, in view of the change in the volume of data information, big data platform will certainly appear new problems.

In the era of big data, enterprises urgently need to establish user-centered self-service big data governance. Information sorting, data control, connecting users, and intelligence are the four main stages of realizing self-service big data governance, and mastering a series of key technologies and technical principles is an important foundation for realizing self-service big data governance.

References

[1]Marelli Luca,Testa Giuseppe,Van Hoyweghen Ine. Big Tech platforms in health research: Re-purposing big data governance in light of the General Data Protection Regulation's research exemption[J]. Big Data & Society,2021,8(1).

[2]Basukie Jessica,Wang Yichuan,Li Shuyang. Corrigendum to "Big Data Governance and Algorithmic Management in Sharing Economy Platforms: A Case of Ridesharing in Emerging Markets" Technological Forecasting & Social Change 161 (2020) 120310[J]. Technological Forecasting and Social Change,2020(prepublish).

[3]Longzhi Yang,Jie Li,Noe Elisa,Tom Prickett,Fei Chao. Towards Big data Governance in Cybersecurity [J]. Data-Enabled Discovery and Applications,2019,3(3).

[4]Maria Kovacova,Tomas Kliestik,Aurel Pera,Iulia Grecu,Gheorghe Grecu. Big Data Governance of Automated Algorithmic Decision-Making Processes[J]. Review of Contemporary Philosophy, 2019,18.

[5]Won gu Jang, 이경호. Big Data Governance Model for Effective Operation in Cyberspace[J]. The Korea Journal of BigData,2019,4(1).

[7]. Science - Earth Science; Reports from University of Florence Highlight Recent Findings in Earth Science (Big data managing in a landslide early warning system: experience from a ground-based interferometric radar application)[J]. Science Letter,2017.

[8]Mari Luca,Petri Dario. The metrological culture in the context of big data: managing data-driven decision confidence[J]. IEEE Instrumentation & Measurement Magazine,2017,20(5).