Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 5, No. 3, 2023

Research on the Application of Big Data Analysis Platform in the Enterprise Management Optimization

Ying Feng [1], Chunyan Luo [2]

[1] The First Affiliated Hospital of Guizhou University of Traditional Chinese Medicine, Gui Yang City, Guizhou, China
[2] School of Information Engineering, Guizhou University of Traditional Chinese Medicine, Gui Yang City, Guizhou, China

Abstract: As companies attempt to make analysis an important data support component of daily decision-making, big data analysis technology is rapidly expanding across all industries. Although there are many software tools and libraries available to assist analysts and software engineers in developing solutions, enterprises are looking for reliable analysis platforms that can meet their specific goals and requirements. In order to minimize costs, such platforms also need to coexist with existing IT infrastructure and reuse the knowledge and resources already accumulated within the organization. To meet these requirements, this article proposes the Data Analysis Solution Engineering (DASE) framework - a knowledge driven approach supported by semantic web technology, for the design and development of requirements engineering and new data analysis platforms. It includes capturing data analysis platform requirements through a knowledge base, and enterprises learning how to use this data analysis platform to analyze all daily production data involved in the engineering data analysis platform. This article analyzes the DASE framework through knowledge modeling, requirement modeling, data architecture modeling, and platform design modeling, and demonstrates how it promotes knowledge and requirement driven data analysis platform engineering. The resulting data analysis platform is considered user-friendly, easy to maintain, and flexible in handling changes in requirements. This work contributes to the knowledge system of knowledge driven requirements engineering and data analysis platform engineering by providing customized models and reference architectures for different analytical application fields.

Keywords: Data Analysis Solution Engineering (DASE); Data Analytics Platform; Semantic Web.

1. Introduction

The emergence of big data technology represents the development of the economic and social era and the changes of market models. It integrates information from various sources, and transform it into data that we can utilize [1]. This data has become an important asset for various enterprises. Using technology of big data for thematic analysis and mining not only allows for macro monitoring of data, but also support enterprises to reduce costs, improve efficiency, develop new products, and make more scientific and accurate business decisions. Nowadays, information technologies such as big data technology and AI technology have entered an era of rapid development. Data analysis platforms have become necessary analysis systems for various industries to explore and analyze data, assist daily business operations, and develop new business markets. Data analysis platform technology is also constantly being updated, making enterprise operations more intelligent, precise, and diversified. Especially when new demands arise over time. Even minor changes in organizational analysis requirements, such as adding new algorithms or integrating new data sources, require a thorough redesign of existing data analysis platforms and processes [3].

2. Background

The traditional big data analysis systems is to address communication barriers caused by inconsistent data indicator calibers in various industries, incomplete indicator systems, difficulty in tracing indicator issues, and inaccurate analysis results, thereby affecting decision-making in business operations and management, product improvement, and other aspects[4].The system sorts out users' data analysis needs based on a combination of methods such as word frequency statistics, Carnot model, Better Worse influence system, and sensitivity curve. Then, establish an analysis framework from multiple dimensions such as business breadth, depth, and span, and propose optimization strategies and suggestions based on the analysis results. Then, establish an analysis framework from multiple dimensions such as business breadth, depth, and span, and propose optimization strategies and suggestions based on the analysis results. At present, the big data analysis platform has optimized the logic of data collection, data processing, data analysis, and data life-cycle management on the original architecture, and combined machine learning algorithms, fuzzy C-means clustering algorithms, cloud technology, and other integration into the big data analysis platform to build a more intelligent platform based on the Data Analysis Solution Engineering (DASE) framework. Effectively utilizing data to improve internal management has become a key issue faced by enterprises. The development of big data analysis platforms and new data analysis tools can accurately and quickly analyze large amounts of data in different fields. The deployment of these advanced tools and technologies enables people to collect, store, process, and analyze large amounts of data in real-time. This enable to enhance business processes, operational efficiency, and provide decision-making support and management optimization for enterprises.

3. Big Data Analysis Platform based on DASE Architecture

DASE architecture, also known as Data Analytics Solution Engineering framework, is a knowledge-based requirements engineering and platform design method that uses semantic data models to improve the design and development of data analysis platforms. The construction premise of the DASE

framework is that requires the professional knowledge of software engineers, data scientists, and domain experts. The use of semantic web technology provides opportunities for building rich information models that can integrate knowledge from different fields and support flexible software design. The data analysis platform is an interactive application that implements a universal analysis workflow for common, frequently used, and recurring business problems. Examples of this recurring problem include using existing prediction services or using model building services to train machine learning models using new datasets. The data analysis platform will have an interactive GUI supported by a knowledge base and a set of services for data analysts in the organization. The data analysis platform developed through the DASE framework (as shown in Figure 1) can be customized according to specific organizational requirements. They are flexible enough to be quickly modified, and when used to analyze data in new fields or adopt new methods, they can reduce the cognitive burden on data analysts, thereby reducing learning time.

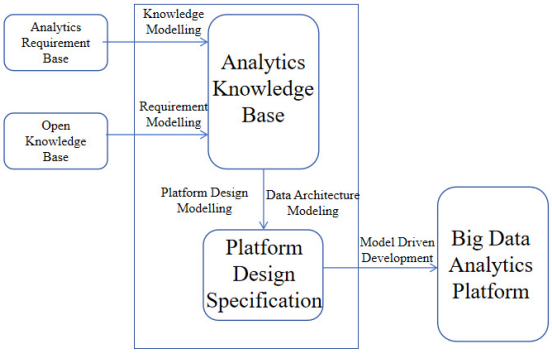

Fig 1. Data Analysis Solution Engineering Architecture Diagram

Figure 1 shows the construction of the DASE framework and its main modules. In the DASE architecture, open knowledge is defined as the collection of all the requirement points that support the analysis requirements of relevant data, which are represented in the semantic web as linked data through different units. The Analytic Requirements Base is the core theoretical point of the requirements module in this structure. It is a meta-model that can represent the theory of data analysis platform requirements in enterprises. Analytics Requirement and Open Knowledge models are mainly used for Knowledge Modeling and Requirement Modeling of enterprise big data analysis platforms to create an enterprise analysis knowledge base: an integrated information base that reflects enterprise analysis activities, resources, and professional knowledge. Data Architecture Modeling requires interfaces between various layers of the big data analysis platform, and the data architecture is related to whether the entire platform can operate according to expected requirements. Platform Design Modeling is the principle of creating Platform Design Specification using enterprise analysis knowledge base and enterprise IT infrastructure. Due to the DASE framework's design principles based on service-oriented and workflow architecture, enterprises is able to freely utilize IT components (databases, middleware, and machine learning models) from existing infrastructure and expose them as services used by data analysis platforms.

3.1. Knowledge Modeling

In the Knowledge Modeling module, the platform develops an analysis knowledge base by collecting and integrating all knowledge and resources related to analysis, such as professional domain knowledge, detailed information on experiments and results of the past analysis, and information on the resources and services contained in the organizational IT infrastructure. Enterprises is able to query data in open knowledge repositories repeatedly, and Establish universal classification algorithms, clustering algorithms, and linked data. Open knowledge can also be extended and enterprise specific data can be modeled based on semantic web and linked data standards.

3.2. Requirement Modeling

Requirement Modeling is a model established for the analysis of enterprise requirements, mainly used for the identification and classification of company specific data related analysis requirements. collect detailed requirements for analysis, and combine them with open knowledge to conduct in-depth analysis of a series of enterprise behaviors such as daily processes and data management, in order to identify problems and provide data support for optimizing enterprise management. The model of analysis includes: basic analysis, advanced analysis, data application analysis, data management analysis, etc. The specific model is shown in Table 2, Requirements Modeling.

Table 1 is an analysis model established for the data analysis content of different enterprises. In Fundamental Analysis modeling, it is a universal analysis model for each enterprise and can handle most analysis scenarios.

1) Capital Structure Analysis is the analysis of the value composition and proportion of various types of capital in a company. It largely determines the company's debt repayment and refinancing ability, determines the company's future profitability, and is an important indicator of the company's financial situation.

2) Asset Management, as an important component of modern enterprise management system, also it represents the efficiency of internal management and the ability to prevent and resolve various business risks. The entire process of asset management generally involves a series of uncontrollable factors such as credit risk, liquidity risk, market risk, and legal risk. By establishing an internal control system to control each stages of the company's business activities, and control and prevent various risks related to the acquisition, management, and disposal of company assets, thereby improving the company's risk response ability, accelerating risk response speed, and reducing unnecessary losses.

3) User Analysis refers to the process of comprehensive analysis and description of a user or groups. This process is able to help enterprises better understand user needs, habits, and behaviors, thereby better carrying out product design, marketing, and service provision work.

In Advanced Analysis modeling, The collection, integration, analysis, and research of various types of data in the daily operation of enterprises play a very important role in the development and decision-making of enterprises:

1) The purpose of strategic decision analysis is to analyze the internal and external environment of the enterprise, evaluate and sort out important factors that affect the formation of the enterprise strategy, in order to facilitate the next step of strategic selection and formulation.

2) Enterprise investment is a very complex economic activity. The purpose of investment decision analysis is to analyze and evaluate the economic and social benefits of

investment projects in order to make corresponding conclusions on the necessity, technical feasibility, and economic profitability of investment projects, as the basis for investment decisions.

Table 1. Requirements Modeling

Fundamental Analysis	Capital Analysis	Shareholders' Equity Ratio
		Capital Liability Ratio
		Long Term Debt Rratio
		Interest Bearing Debt Ratio
	Asset Analysis	Control Environment Analysis
		Risk Assessment Analysis
		Control Activity Analysis
		Information and Communication Analysis
		Supervision activities Analysis
	User Analysis	Attribute Analysis
		User Analysis
		User Tags Analysis
		User Behavior Analysis
	Product Analysis	Plan Analysis
		Feedback Analysis
		Summary Analysis
	Customize Analysis	Content Based on Customer Needs
Advanced Analysis	Strategic Decision Analysis	External Environment Analysis
		Internal Environment Analysis
		Strategic Development Analysis
	Investment Decision Analysis	Investment Necessity Analysis
		Analysis of Construction Conditions
		Core Technology Analysis
		Investment Benefit Analysis
		Investment Risk Analysis
	Post investment evaluation analysis	Objective Evaluation Analysis
		Benefit Evaluation Analysis
		Impact Evaluation Analysis
		Continuous Evaluation Analysis
		Management Evaluation Analysis
	Cost Analysis	Overall Level Analysis
		Cost Component Analysis
		Analysis of Influencing Factors
		Change Analysis
	Financing Plan Analysis	Financing Channel Analysis
		Analysis of Financing Structure
		Financing Cost Analysis
		Financing Risk Analysis
	Daily Business Analysis	Business Decision Analysis
		Business Budget Analysis
		Financial Analysis
		Human Resource Analysis
		Other Analysis

3) Post investment evaluation refers to the objective analysis of commercial activities including purpose, execution process, benefits, effects, and impacts of completed projects. Its aim is to identify successful factors and lessons learned through the inspection and summary of investment activities, and provide suggestions for future project decision-making and improving investment decision-making management level. At the same time, it also proposes improvement suggestions for the problems that arise in the implementation and operation of the evaluated project, in order to achieve the goal of improving investment efficiency.

4) In order to effectively control the total investment of the project, it is necessary to implement project cost analysis work. Its purpose is to seek the possibility of reducing project costs and identify measures that should be taken through analysis. Secondly, accumulate project cost accounting information. To provide scientific and effective historical reference basis for predicting project cost management decisions, thereby making the cost accounting of new projects more scientific, economical, and reasonable.

5) Financing has become one of the main method for most enterprises to expand their business abilities. The purpose of financing plan analysis is to enable enterprise shareholders, project review committees, financing institutions, and other relevant parties to understand the financing channels, financing structure, financing costs, financing risks, and financing plan arrangements of projects through project financing plan analysis, Provide basis and assistance for project investment decision-making, as well as financing during subsequent construction and operation stages.

6) After the project is completed, it will enter the commercial operation phase. The purpose of daily business analysis is to guide the development of production, improve operations, strengthen management, seek reasonable and effective ways to utilize human, material, and financial resources, tap into the internal potential of the enterprise, enhance core competitiveness, and improve the economic efficiency of the enterprise under high-quality and low consumption conditions. Daily business analysis of enterprises mainly includes business decision analysis, business budget analysis, financial analysis, human resource analysis, etc.

3.3. Data Architecture Modeling

The big data analysis platform requires interfaces between various layers to connect, and the data architecture determines

whether the entire platform can operate according to expected requirements. The overall modeling of the data architecture is shown in Figure 2. The external system configures the data storage path or IP address to generate a configuration file. The data collection interface accesses the data by reading the configuration file, converts and loads the data through data collection, and provides services for data processing. Data can be divided into three types: small amount of data, large amount of data, and real-time data. Different types of data processing strategies are different. For small amounts of data, it can be stored on a single machine node and calculated using local memory; Offline batch data requires distributed storage and parallel computing using clusters; Online real-time data can be used for streaming computing using clusters. The data processing process needs to be recorded in the historical version library, while loading the model library and processed files for model construction. The final result file is transmitted to external systems such as web pages, clients, or interactive terminals such as API requests.

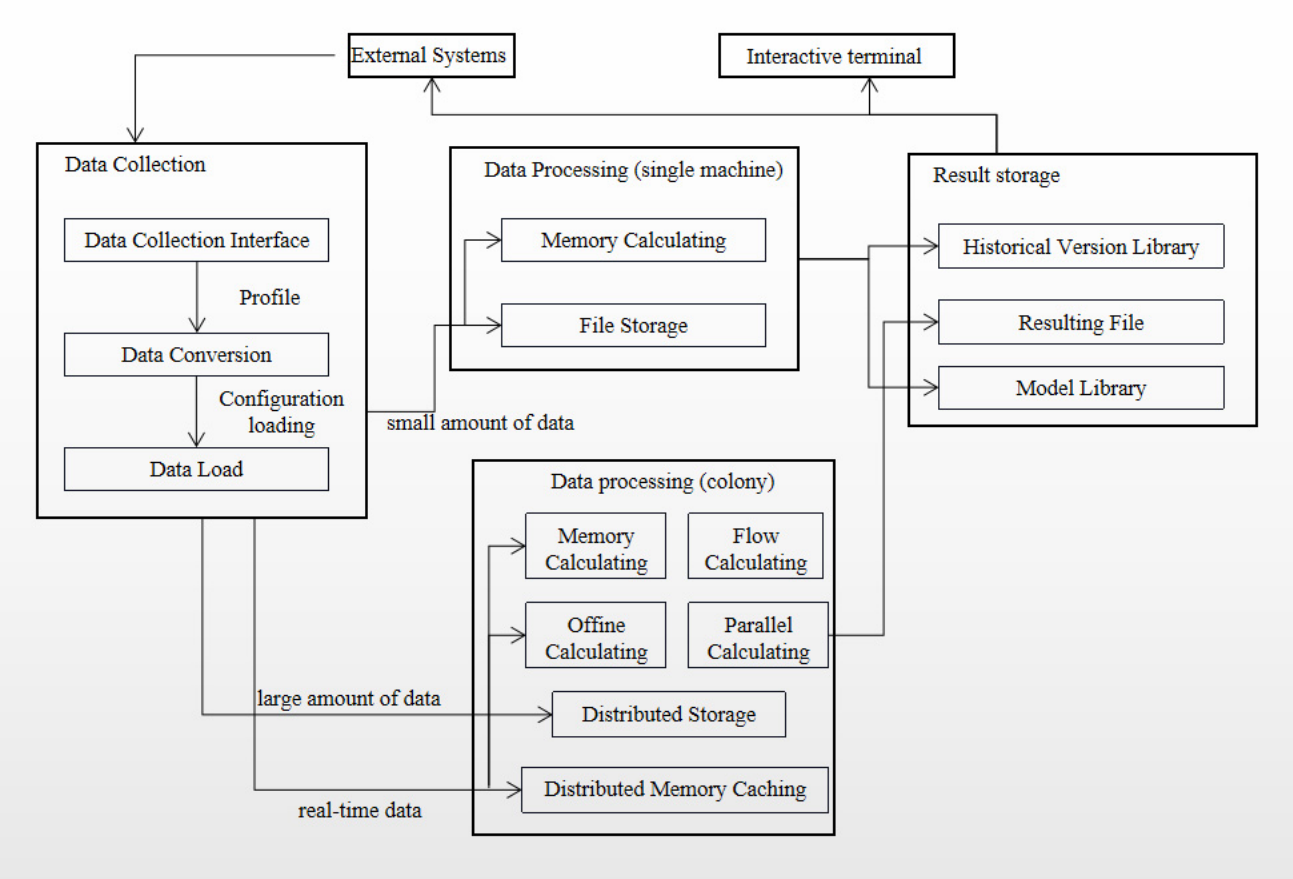

Fig 2. Data Architecture Modeling

3.4. Platform Design Modeling

Once Knowledge Modeling and Requirement Modeling are completed in the Analysis Knowledge Base, the components of the platform can be instantiated as platform design specifications linked to organizational IT infrastructure and organizational analysis knowledge base. The general big data analysis platform modeling (see Figure 3) mainly includes Subject Management Module, Data Exploration Module, Data Preprocessing Module, and Algorithm Module.

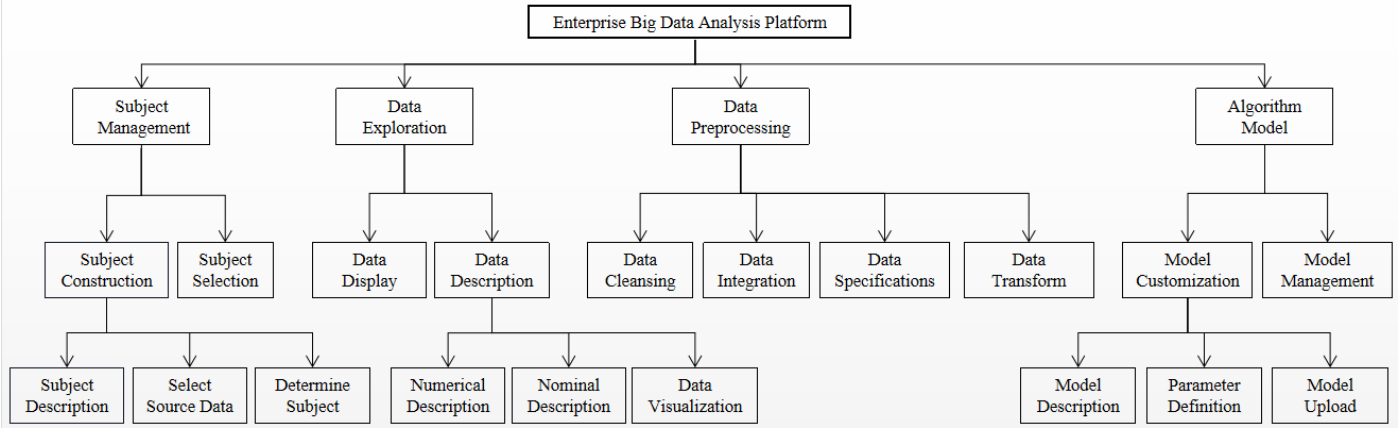

Fig 3. The architecture of enterprise big data analysis platform

As shown in the figure, the Subject Management Module mainly addresses data security, data analysts provide detailed descriptions of subjects based on their business, select corresponding data sources, and ultimately determine the

subject. Each subject corresponds to a dataset, data processing process, and algorithm model specific to their own business. In the Data Exploration module, data analysts need to determine how to preprocess the data, choose features, and build a suitable model based on the distribution of data features. After loading the data source, each data analyst has different analysis standard of data. At this point, the module is able provide basic data description fields based on the type of data column, which mainly includes discrete and numerical types. Different types of data columns could display indicators that data analysts are concerned about. In the Data Preprocessing Module which enable data analysts to have a clearer understanding of the data and prepare for data preprocessing and model construction. The data preprocessing module could perform semi-automatic preprocessing on the data to be analyzed. The reason why it is called semi-automatic preprocessing is mainly because data processing is closely integrated with business, and the platform can construct different business subject. Due to the inability to learn the corresponding business background of each subject, the system is able to provide corresponding data processing suggestions through data distribution. In the Algorithm Module, data analysts could use this module to define the algorithm model and upload the implemented algorithm model file. Other data analysts can input specified parameters to execute the model, and the final platform will output the results according to the output type.

For example:

Fields:	Agreement Description

Algorithm Model Name: The name of the algorithm model needs to be the same as the name of the algorithm model file to be uploaded, and the model needs to have the same method name as the model, such as the Random Forest algorithm model file.

Algorithm Model Upload: Currently supported algorithm model file.

Model type: It is divided into single machine type and cluster type. If it is selected as a standalone, the data source can choose a local text file type; If it is a colony model, the data source can select files in the colony.

Public or Private: The algorithm could be chosen between public or private. If private is selected, the algorithm model can only be used by the subject created by the current data analyst; If you choose to make it public, the algorithm model can be used by everyone.

Parameter Type: The parameter type is the input and output basis when the model is called.

Input Parameter:

(1) Dataset type: this type is directly related to the model type. If the type is a standalone model, the locally preprocessed dataset will be loaded; If the type is cluster mode, load the dataset on the cluster and then call for model training.

(2) List type: specify the features and targets in the training data, and corresponding features and targets can be selected based on the input dataset.

(3) Model type: specify the features and targets in the training data, and corresponding features and targets can be selected based on the input dataset.

(4) Numerical type: The parameters input for model training.

Onput Parameter:

(1) List type: predicted or evaluated results.

(2) Numerical type: information such as the training accuracy or recall rate of the model.

(3) Model type: the system saves the corresponding model to a specific location according to the model name.

4. Optimization of Big Data Analysis Platform

The emergence of big data shows the development of the information age and the transformation of market models. It integrates information from various sources and formats, and transforms it into data that we can utilize, which is able to improve business efficiency. Enterprises establish a data processing department and data centers to analyses technology research and development, supply chain, sales, production, and human resources. They quickly integrate information resources for classification, analysis, storage, and processing, in order to fully utilizing the characteristics of fast processing speed, large information volume, and remote management of data centers. Learning the latest data analysis techniques and methods can quickly integrate resources and improve analysis efficiency, It is significant to improve enterprise management and business analysis. Although the big data analysis platform has entered the stage of intelligent analysis and has relatively complete functions and services, there are still many features that need to be optimized, such as insufficient standardization of providers and users, and insufficient data resources. For example, the understanding of the data platform is not accurate enough, there is a lack of product service processes, incremental requirements are difficult to meet, and the deep analysis of the data analysis platform is limited. In response to the shortcomings of the current big data analysis platform, optimization can be carried out in terms of data, service processes, business function expansion, and system improvement.

4.1. Refinement of Data

To optimize the enterprise big data analysis platform, it is necessary to have data governance, which is a standard established on data access, use, storage, and protection. After a series of continuous evaluation, supervision, and guidance, it ensures the effective utilization of data and reflects its value [5]. A data analysis platform is a platform with a large amount of data. It establishes a comprehensive data management system, clarifies the primary and secondary relationships of various entities in the data circulation process, and builds a reasonable data circulation mode. Maximize the value of data resources, promote corporate governance capabilities and improve systems. Enterprises should pay attention to data asset management, data quality management, data life-cycle management, data security, permission management, the application of data exploration and analysis throughout the entire process. By implementing employees' understanding and use of the data analysis platform, it is important for enterprises to build special group for data analysis, strengthen employee training on the use of functional modules and standardize work methods, and strengthen publicity for users to enhance their awareness of the value of the data analysis platform. Provide product usage training and collect feedback on issues, solve problems encountered by employees in real time, enhance their problem-solving abilities, improve their efficiency in using data, and enhance their recognition. Optimize the framework of the reflection system, clarify data analysis indicators, and enable employees to enhance their

knowledge of data analysis.

4.2. Optimize the Process of Big Data Analysis Platform

In the design phase of the big data analysis platform, effective communication should be conducted with users to clarify their needs, and different attributes should be designed for different users to make their use more convenient. Provide systematic user training to users to understand the functionality and value of platform products, optimize delivery processes, and make it easier for users to use. Build a bridge for data communication and exchange between enterprises and providers, and equip different professional analysts at different analysis stages, enabling analysts to quickly solve users' different problems. During the usage phase, the update iteration of the platform should meet the incremental needs of the industry based on the market. After user feedback, the product update iteration should be done well, so that users can experience the value and efficiency of the enterprise analysis platform during use. Periodic research can provide a more accurate understanding of user needs and ensure that the product meets the majority of user needs as much as possible. The platform has strong analytical capabilities and security mechanisms, and users should also enhance their knowledge of data security. A comprehensive security mechanism system should be established at each level to prevent data leakage risks.

4.3. Platform Function Extension

It is necessary to continuously expand the functions on the original basis to adapt to user needs. User needs are constantly changing, and it is important to follow up on the needs of users for using the enterprise data analysis platform, continuously update iterative products, and improve the functionality of the data analysis platform. Suppliers should actively explore boundaries on the existing functions of the platform to make the product functions of the data analysis platform more powerful, in order to meet incremental needs. Always keep up with the development of the data analysis industry, learn other product functions, integrate chart drawing and other functions into the data analysis platform, improve users' user experience and work efficiency. By setting permissions, users can display the analysis results during use, ensuring data security and making it more convenient to search for data to avoid file loss and other situations. Enhance the viewing experience by combining data analysis with visualization, making the presentation of data analysis more intuitive. Provide access to different functional component products according to different needs, ensuring the functionality required by users to prevent redundancy of product functions, reducing the threshold for data analysis, and enabling non data analysts to use the data analysis platform.

4.4. Big Data Platform System Optimization

Provide fast query services to achieve fast response to data. Implement intelligent upgrade of the system, using multiple events and other customization to analyze the subject; realize revenue analysis function through customized indicator calculation; design partition storage for data; support real-time import and query of large amounts of data, and provide a real-time query engine to make search more efficient; Conduct multiple types of data collection to achieve the strategy of upgrading the data analysis platform to full end data collection. The system should have deeper and more powerful secondary development capabilities, and there will be more scenarios where the system's performance is more powerful and can be flexibly applied. Developing towards a native direction of data cloud to achieve containerized deployment that is more convenient for enterprises to conduct testing and development. Reasonably allocate storage resources, reduce operational and maintenance costs, unify resource system management and scheduling, and improve operational efficiency. Utilize machine learning and artificial intelligence technologies to automate data processing and perform data clustering and other analysis.

5. Conclusion

A big data analysis platform based on the DASE framework, which is a relatively new architecture for big data analysis platforms, can achieve knowledge driven requirement engineering and platform development for data analysis platforms. By utilizing a semantically rich knowledge base to represent requirements and integrating them with other metadata (domain knowledge, analytical knowledge, and IT infrastructure) in the analysis platform, the DASE framework supports integration between requirements and software architecture models, as well as model driven platform design and development processes. The resulting data analysis platform is knowledge driven, user-friendly, and easy to maintain. They help reduce the technical knowledge and cognitive burden required by data analysts.

By implementing and evaluating the DASE framework, we have identified multiple future research directions. This article does not explore the use of model driven development paradigms to generate data analysis platforms. The workflow logic definition in the framework can serve as the foundation for future such work, and combine with existing research on exploring workflow automation using ontology [6]. The resulting workflow definition can be used in conjunction with appropriate model driven development engines to fully automate the generation of data analysis platform software.

In addition, the DASE framework needs to be improved so that it can fully utilize the power of semantic web technology, such as enhancing semantic reasoning capabilities. In this way, the generated data analysis platform can provide intelligent recommendations to analysts, further reducing the cognitive burden related to analysis.

Once the DASE framework is implemented within an organization, the organization will accumulate an analytical knowledge base over time, as well as metadata, such as how analysts utilize this knowledge. For example, organizations can generate statistical data on the performance of the same analysis model on different datasets, as well as what is the best or most popular analysis algorithm performing. This information can be used to build value-added applications, such as meta learning and scalable artificial intelligence systems specifically designed for specific organizational environments. The data analysis platform aims to learn and adapt to the experience gained by utilizing knowledge extracted from previous analysis tasks or from different fields or problems [7]. AI is committed to creating a set of analytical techniques that can generate more interpretive high-performance machine learning models and enable humans to understand, trust, and manage artificial intelligence systems [8]. By emphasizing the importance of having a knowledge base and service-oriented modules As a component of the analysis and artificial intelligence platform, the DASE

framework provides a practical solution for managing, learning, and related analytical knowledge and semantics.

The value of big data is being seen and utilized by people now and in the future. The intelligent processing of large-scale data collections will be applied to big data technology, and data analysis will become the core of big data technology. With the rapid development of technology, people's pace of life is also becoming faster and higher, and the requirements for data processing efficiency are also increasing. Real time data processing will be widely adopted. The development of cloud computing technology is becoming faster and wider, and its application scope is also becoming more and more extensive. With the continuous development and improvement of data analysis platforms, the ability to process data will be significantly improved. Big data technology has formed an industrial scale, and open-source software has been developed, which will become a new driving force for the development of big data. In the future, big data will accelerate its integration with the real economy, promoting the integration and development of big data analysis related products and services in fields such as education, industry, electricity, transportation, healthcare, government, internet, finance, telecommunications, etc. The big data analysis market in the future will continue to develop and grow.

References

[1] Zhixia L .(2020).Cost Management and Optimization in the Era of Big Data. Finance and Accounting Learning, (19),pp. 113-114.

[2] Xue L. (2022). Research on optimization of enterprise data analysis platform based on user demand. Master. Central China Normal University.

[3] Sculley D, Holt G, Golovin D, Davydov E, Phillips T, Ebner D, Chaudhary V, Young M (2014) Machine learning: The high interest credit card of technical debt. In: SE4ML: Software Engineering for Machine Learning (NIPS 2014 Workshop).

[4] Tingting Z. (2022) Research on optimization Measures for the Construction of Chengdu Statistical Big Data Platform. Master. Lanzhou University.

[5] Gil Y, Ratnakar V, Deelman E, Mehta G, Kim J (2007) Wings for pegasus: Creating large-scale scientifc applications using semantic representations of computational workfows. In: Proceedings of the 19th National Conference on Innovative Applications of Artifcial Intelligence - Volume 2, AAAI Press, pp 1767–1774.

[6] Lemke C, Budka M, Gabrys B (2015) Metalearning: a survey of trends and technologies. Artif Intell Rev 44:117–130.

[7] Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, García S, Gil-López S, Molina D, Benjamins R et al. (2020). Explainable artifcial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fus 58:82–115.