The Research and Design of the Service Platform on Comprehensively Processing Finance-oriented Mass Network Data

Guizhen Yang

ABSTRACT

The service platform on comprehensively treating online finance-oriented mass data supports flexibly to form data warehouse, data fair, data files to process, transform the huge data, and realize atomization and parameterization of batch jobs, and simplify operation, control flow, and provide the program interface which is flexible and customized, and the platform has a good extensibility. The service platform is based on SOA, and has the structure of cloud computing, includes many techniques on ETL and the tools of ETL, and has features of being united, high effective, and extendable. This research breaks the monopoly situation of foreign manufacturers, and promotes the enterprises which have the traditional advantages to go on the information-oriented road, and accomplishes fully “resource sharing, low investment, low costs, low emission and high efficiency”, which deserves to be developed and promoted.

Research Significance

IDC (International Data Company) has predicted that the expenditure used on cloud computing will increase three times in five years, which is 25% of the whole cost on IP. The various data of the current enterprises’ business system are going up in the mass growth of the level of GB, TB and PB, and the corresponding storage method has turned to network from the single machine. The traditional means on dealing with information, such as the database technique, can only complete the input, search and statistics of data,
which are the functions with low level, but can’t use and refresh fully the mass data in time, and do comprehensive research hardly, it is so on Chinese finance industry as well. So far the Chinese five biggest state-owned commercial banks have gathered huge business data, and they are collecting data still. By the research to data from different sources, and historical stages, the banks can distinguish the groups of customers with potential value and find out the development trend of future financial market, develop the financing products with competitiveness according to the features of target customers and the requirement of the financial market. So the demand of banks to the analysis of mass data is urgent. In addition, today financial industry is facing increasingly fierce competition under the condition of high speed development on information technology, it is the first to think about high sharing of information and security of data in the system construction. Along with aggravation of the competition between domestic banks, the five biggest state-owned commercial banks continuously deepen the concept with customer-centric and the core of excellent business, which made the higher request to perfect the system of banks. The technology of “cloud computing”, however, will be the main choice of banks to enhance the data security and accelerate to share information, improve the service quality, reduce costs, and get the advantage to win the competition.

The service platform on comprehensively treating online finance-oriented mass data combines distributed technology of cloud computing, SOA, ETL, and the job scheduling, and is based on SOA, and uses the system structure of cloud computing, and integrates multiple technologies of ETL, different tools of ETL, which for financial industry establishes the service platform that is united, highly efficient, and extensible finance-oriented to treat mass data. The service platform on comprehensively treating online finance-oriented mass data supports flexibly to form data warehouse, data fair, data files to process, transform the huge data, and realize atomization and parameterization of batch jobs, and simplify operation, control flow, and provide the program interface which is flexible and customized, and the platform has a good extensibility. The service platform is based on SOA. This system integrates customers, contracts, trade, account, products and other business data for financial institutions, provides customer view, user relationship management, marketing management, account analysis, quality monitoring, risk alarm, business flow, and other function modules. The system supports to build operation-oriented service platform, and charge the users according to functions and amounts of data, the user provides data in terms of formula, then click the corresponding function module to get the target results.

Along with increasing requirements on integrating data from various walks of life all over the world, the demands to “the service platform on comprehensively treating online finance-oriented mass data” will be more and more. This system provides decision support for creation and transformation of fields of bank, security, insurance, and other financial services, and can promote the combination with and the infiltration to the traditional financial fields. The service platform belongs to the fields that our country develops firstly and supports priorly. From the general view, this service platform can
promote the development of Chinese financial informatization, adjust industrial structure, meantime strengthen the competitive power on enterprise brand, enhance the enterprise popularity in the field of the international information technology, break the monopoly situation of foreign manufacturers on financial software, boost the enterprises with traditional advantages to take the new information–based road, perfect socialistic market economic system, conform to the features of the development of “the Twelfth Five-Year Plan”, and the service platform is significant to build moderately prosperous society, and drive the cause of socialism with Chinese characteristics.

Research Contents

Today the vocations of bank, security, insurance and others have gathered huge data of business data and unstructured data, and established data warehouse, data fair, and image storage platforms based on the mass data. Some data of financial institutions have gone up to the level of TB, even PB, so that the processing quality of mass data, the treatment of complexly heterogeneous environment, the changing flow of process, and increasing costs, maybe have become many thorny problems. To solve above problems efficiently, it is the goal for the service platform to tap the potentials to the mass data of financial institutions, and provide a proper platform to treat huge data for financial institutions.

The service platform combines distributed technology of cloud computing, SOA, ETL, and the job scheduling, and is based on SOA, and uses the system structure of cloud computing, and integrates multiple technologies of ETL, different tools of ETL, and establishes for financial industry the service platform that is united, highly efficient, and extensible finance-oriented to treat mass data.

The service platform on comprehensively treating finance-oriented mass network data supports flexibly to form data warehouse, data fair, data files to process, transform the huge data, and realize atomization and parameterization of batch jobs, and simplify operation, control flow, and provide the program interface which is flexible and customized, and the platform has a good extensibility. The service platform is based on SOA. This system integrates customers, contracts, trade, account, products and other business data for financial institutions, provides customer view, user relationship management, marketing management, account analysis, quality monitoring, risk alarm, business flow, and other function modules. The system supports to build operation-oriented service platform, and charge the users according to functions and amounts of data, the user provides data in terms of formula, then click the corresponding function module to get the target results.
KEY QUESTIONS PLANNED TO SETTLE

Integrating Multiple Advanced Technologies

The system combines distributed technology of cloud computing, SOA, ETL, and the job scheduling, image processing methods, and workflow technology. This system is based on the architecture computing model of hadoop, breaks the speed limit of the traditional database to the process of mass data, and improves work efficiency highly to deal with data by the concurrent access and treatment of high data. Hadoop has the irreplaceable advantages on extendibility, robustness, calculate performance, and costs, and lays the firm solid foundation to build the finance-oriented to process mass data comprehensively.

Powerful Function to Treat Mass Data of Financial Industry

Our company has plentiful experiences for many years to process mass data of financial profession, and joins most up-to-date technology to treat mass data, including the technology of concurrent processing, and the job scheduling, Hadoop architecture, and so on, provides mighty functions to deal with the mass data for the financial industry.

Loosely-coupled, Flexible Data Treatment Mode

It is by using SOA to treat real-timely, run, and monitor loose couplings, make the system pliable, suit dynamic business requirement and environment.

Support to Build the Marketing-oriented Service Platform

This system integrates customers, contracts, trade, account, products and other business data for financial institutions, provides customer view, user relationship management, marketing management, account analysis, quality monitoring, risk alarm, business flow, and other function modules. The system supports to build operation-oriented service platform, and charges the users according to functions and amounts of data, the user provides data in terms of formula, then click the corresponding function module to get the target results.

THE GENERAL FRAMEWORK OF FUNCTIONS OF THE PLATFORM

Using the advanced technology in the corresponding field to build the whole architecture, Figure 1. The system physical architecture is as follows:
KEY TECHNOLOGY AND CREATIVITY

Cloud Computing:

This service platform uses distributed cloud computing with high performance, to complete the storage of mass files and data, and uniform programming, running environment of mass data. Cloud computing is mainly based on the two cores of virtualization and distribution technologies, the virtualization platform assigns the server to be multiple virtual machines that can be configured, monitors and manages all virtual machines among the system of whole groups, and distributes and allocates resources according to the actual usages. Virtualization can not only delete the difference of large scale heterogeneous server, but also its computing pool has the super computing power. Distributed parallel architecture is another core technology, which is used on integrating plenty of machines with low configuration to be a one with high performance to provide the ability to store mass data and processing service.

SOA:

The service-oriented architecture is a component module, which links different function units of applications by the favorable interface between services. The interface is independent to the hardware platform to realize
services, the operating system and the programming language. It is by using SOA to treat real-timely, run, and monitor loose couplings, make the system pliable, suit dynamic business requirements and environment.

ETL:

Three letters of ETL represent Extract, Transform, and Load respectively. ETL is responsible to add the data of distributed and heterogeneous sources, such as, relationship data, graphic data file into the middle tier, then clean, transform, integrate, and finally load them into data warehouse or data fair as the basis of analysis and deepening online. The self-developed tools of ETL or integrated main tools of ETL provide a uniform, clean data view for users, and provide a high-quality data source.

Job Scheduling:

It is the usual problem that there is a great number of batch jobs of data warehouse, data fair, and the process flow and the dependencies are complicated and various with low performance. The system makes use of job scheduling technology to define the scheduling policies and rules of batch jobs by the uniform operating platform and visual interfaces, so that control, execute and monitor batch jobs of the cross platform and cross system. The system uses jobs scheduling to simplify the management, and enhance the overall performance according to the features of huge data, numerous source data files, and complex logic.

Image Processing:

Processing and uploading images use our image scanning technology with independent intellectual property rights.

Scanning Image:

This software can call every scanning instruments to scan images on IE platform as a plug-in, and save the image automatically according to the predefined template, and compress the image data to meet the requirement to the file size for transmission.

Processing Image:

This software can preview, zoom the scanned images. And it can delete the black spots, cut the black box, and rectify deflection automatically, rotate and complete other functions.
Classifying Images:

The software can provide classified folders to the scanned images, so that make it easy for users to sort images according to real business requirements.

File Operation:

The software supports the default operations, such as, copy, delete, new folder, etc. and supports to drag and drop thumbnails.

Uploading Image:

The software can upload images in the special folder and formats, so that the server can store and refresh the image files. It can keep file accuracy during transmission.

TECHNOLOGY DISSEMINATION AND SOCIAL EFFICIENCY

It is five or six years for the large concentration of the data of financial industry from the time when the concept was proposed to today. Chinese banks have centralized data on a large scale, some of them have already finished the task, and obtained a lot of achievements. But along with large concentration of data, the problems of troubles of real operations and future trends after the large concentration of data become the hot spots of society. Today it has only finished the construction of physical devices needed for largely centralizing data, the real integration and centralized application of data resources, however, have not started. It’s short of communication and integration between every application systems, so that it can’t complete the integrated management to data resources, and deepen and analyze data resources deeply.

The requirements of the internationally financial market to analyze data warehouse and data analysis will grow continuously in the next five years, the technology of the data warehouse developed by this service platform will be improved further, which meantime will make further efforts to promote R & D investment of domestic financial field to the corresponding technology of data warehouse, including development and progress of cloud computing, business intelligence analysis, tools of ETL, super parallel database system, mass data mining, SOA technology, jobs scheduling, and greatly promote the widespread use of the application of domestic commercial intelligence in the financial industry, and shorten the gap between the IT of domestic banks and the financial information technology of foreign banks, and strengthen the competitive power on enterprise brand, enhance the enterprise popularity in the field of the international information technology, break the monopoly situation of foreign manufacturers on financial software, and has a certain role to promote the development of domestic financial information industry,
accelerate to transform the economic development pattern, and keep the stable, faster economic development.

ACKNOWLEDGEMENTS

The kind of service platform: the research project of Guangzhou municipal university (project number: 2012A148); Project name: The research and development of the visual service platform based on mass network data; The introduction of the writer: Yangguizhen (1967/11), female, lecturer, master, the main direction of research: information service and vocational education.

REFERENCES