

Introduction to the Management and Development of Data Resource Sharing and Exchange Platform in the New Era

Shiping Liu^{1,a}, Wen Liao^{1,b} and Shan Li^{1,c}

¹Chengdu Neusoft University, Chengdu, Sichuan, 611844

^aliushiping@nsu.edu.cn, ^bliaowen@nsu.edu.cn, ^clishan@nsu.edu.cn

Keywords: Data exchange; Exchange platform; Management and monitoring; Exchange process

Abstract. In the process of designing data exchange platform's management and development system, we have been faced with many contradictions, and the most important thing is the tradeoff between usability and rich functional support. This article starts with the needs of the whole, from the perspective of users, expounds the process of its management and development.

Introduction

As a supporting part of the E-government system, the data sharing and exchanging platform has been playing a central role in the construction of e-government for a decade. If the function is very rich, but directly exposed to the customer, and let customers combine and apply in the business system on their own, then the ease of use will not perform well. On the contrary, if it looks very concise, who needs the data and the relationship definition is clear, it will give customers a thin impression of function. However, the balance of powerful functions and usability is always our common goal. If we can find a better solution, there will be no doubt that platforms and systems will be highly competitive.

A Link between Directory and Data Source

It's normally through the directory to see what is available, what can be provided, and what you need. However, it is worth noting that the directory represents the business perspective of the data structure, which is divided according to different dimensions. Behind each registration, the corresponding data source should be associated, and the data source also needs to be registered and managed by the exchange platform.

Therefore, we can track the corresponding data sources, even the data source access status, related exchange tasks and processes, exchange volume and other information through the directory in the monitoring page.

Definition of Exchange Process and Task

The processes and tasks defined on the exchange platform must also be associated with different data sources, and the exchange task itself has a certain business conceptual category. So, different data sources are registered in directory system, in other words, which are related directory entries, what data subjects are registered, or associated multiple topic categories. In this way, the definition of exchange tasks and processes, as well as the statistics of exchange volume, are also very easy to provide data monitoring and statistical information around topics. However, Data themes and topics are the concepts of the directory category.

By the way, the definitions of data exchange process and task in the exchange platform is best to provide debugging functions, and transport the running console output to the main IDE, which is more suitable for engineers' use habits, and it is also easier to track data during debugging.

Close to the Management Monitoring in Business Scenario

So, in fact, data resource sharing and exchanging will provide a portal and front page, and also provide a overall display architecture through the correlation among several main modules, such as directory, data source, exchange task and process, and exchange volume statistics. Each part can be extended to monitor the state of its internal affair , based on statistics from any perspective, and access to security authority management from any angle, etc.

The display of state monitoring, statistics and other information must be showed in the way of flatness, visualization and graphical representation, if use too many words and lists, it will be difficult to express complicated relations, and can not make it intuitive and vivid.

It need to support mouse-over window tracking, such as the mouse over a directory, then the associated data source can be listed; stay on a data source, relevant exchange tasks and processes can be listed; Stay in an exchange process where the associated data source and directory definitions can be listed.

Thus, perhaps we need a complete repository design to support these functions. Any component, process, and task on the exchange platform requires a controllable instantiation ability, so that make It is possible to instantiate a process and execute on a specified node as needed, and support can be achieved through a container. This also provides a prerequisite for the support of the subsequent cloud environment.

Encapsulates the Necessary API

Based on the above functions, encapsulateing the appropriate API, provide frameworks of redevelopment , can support the ability to dock with other management systems and applications systems.

Data Exchange and Big Data

For big data, our understanding is still relatively shallow. In the eyes of many people, big data is actually a high-speed parallel processing system, which takes advantage of the nodes of relatively cheap distributed server network to share data processing power. But this pattern seems to be more appropriate for the cloud center to use virtual machines. With Hadoop as an example, its design pattern and architecture are widely distributed parallel, automatic sharding, automatic addressing, automatic allocation task and automatic result set summary. In this process, there is no relationship with the data exchange platform.

So what are the essential differences between the data exchange platform and the big data platform?

Usage scenarios and application premises are different:

The data exchange platform is based on the distributed existence of disparate data sources and the need for interaction based on business needs. The data must first be decentralized and distributed across the network. The formation of this kind of boundary is actually formed by the fact that the government, large enterprises and other business systems are unable to integrate with each other, or artificial non-fusion.

Big data, such as Hadoop, has a completely different application premise. Hadoop will first focus on the huge amount of data source as the data input, and then the data is divided by oneself, its recruited a number of can do the same work (Map subdivision data processing, Reduce data acquisition, storage and backup Node Node, etc.) of the members (Job). So, the exchange platform is typically distributed, and Hadoop is a typical large-scale cluster and concurrency. Therefore, Hadoop is more suitable for processing data processing scenarios with large data volumes and less complex process correlation and transaction correlation (such as scanning a class of information).

Data structures and expressions are different:

Data exchange platform, more traditional data structure, and structure definition relatively complex model. The processing process emphasizes the business process itself, such as the rules of

data capture, exchange process, data routing, field control, and so on. Perhaps the exchange process will be very complex. Can be temporarily defined as: business model.

Big data for this, although its program structure and code can be done more complex. However, the big data highlights one thing: the atomicity of data processing and the internode independence! Failing that, you can't easily break up a business into so many small pieces of processing that you can process in parallel.

As mentioned above, how we define the connection between big data and data resource sharing exchange platform, we don't actually see the better successful cases and application patterns. However, there is nothing to explore in this direction.

For data embedded in business systems, and data that is still remote, it is virtually impossible to incorporate a highly parallel system, and this step is still dependent on the sharing of data resources. After the data is collected and formed, the big data provides the upper layer fusion algorithm and the engine, and extracts the deep information. From this point of view, the data exchange platform is the preliminary extension and preparation of big data processing. This is the data processing and use stage to correlate data exchange platform and big data system.

In addition, you can try to introduce Hadoop's platform and processing methods to a node environment on the exchange platform. A data exchange node environment to encapsulate calls to large data platforms such as Hadoop. In a higher level of data management, invocation, and service, it is necessary to take Hadoop as the platform for data processing and give the necessary results. This result will be the input of the next step in the exchange process. Thus, the processing of Hadoop becomes a processing activity of the data exchange process. Through local parallelization, large data can be used to enrich the processing power of the exchange process. In the future, the exchange platform may grow into a higher stage of data ecological environment and full life cycle management services.

Conclusion

To sum up, the common thinking point of the data resource sharing and exchange platform is: what is offered? what do you need? How to exchange and obtain it? How is the result of the implementation? is it working well? and how much did it exchange? The main task of our management and development is to find a balance from these issues.

References

- [1] Liu Yanhui, Dong Bi, Zhang Danfeng. Research on Distributed Application of Data Exchange Platform [J]. *Computer Engineering and Design*,2009(16)
- [2] Zhang Yijun. Uniform standards break through barriers to build industry data sharing system [J]. *Publishing Research*. 2016(02)
- [3] Yang Qing. Study on the Application of Data Exchange Platform in Environmental Information [J]. *Value Engineering*. 2013(15)