Journal of Theoretical and Applied Information Technology
31st December 2012. Vol. 46 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DESIGN AND IMPLEMENTATION OF METADATA MANAGEMENT IN THE NATIONAL EARTHQUAKE NETWORK

[1] JIANWEI ZHANG, [2]YUNTIAN TENG, [1]HAI LIU, [1]SHILONG MA

[1] State Key Laboratory of Software Development Environment, BeiHang University, Beijing, China

[2] Institute of Geophysics, China Earthquake Administration, Beijing, China

ABSTRACT

Due to the large numbers of units and organizations in earthquake precursor industry, there exist some problems as follows: a) how to dynamically manage the relationships among different units; b) how to synchronize the metadata from lower unit to upper unit in a quick way; c) how to manage the metadata restriction in distributed system. Based on the current situation, the demand for earthquake precursor industry, and the analysis of existing metadata management programs, we first proposed a novel metadata management model –DAC (Dynamic Automatic Constraints) Model, which considerate all the property of metadata in this system, i.e. diversity, distribution and restriction. Then we design and implement the metadata management system for earthquake precursor industry by DAC model. Experiments show that our system meets the demand and is appropriate for applications in the practical environment.

Keywords: *Metadata, Distributed Database System, MVC, Restriction, DAR Model*

1. INTRODUCTION

National Earthquake Precursor Management System is developed to satisfy the demand for the Earthquake Precursor Industry, whose instruments and data resources should be interconnected, shared and coordinated [1-2]. However, for the past decades, it has deployed large amounts of heterogeneous, distributed and autonomous instruments, which leads several problems as follows: a) how to dynamically manage the relationships among different units; b) how to synchronize the metadata from lower unit to upper unit in a quick way; c) how to manage the metadata restriction in distributed system. As a result, an efficient metadata management mechanism is needed urgently [3-5].

In this paper, we first mine, extract and classify all the metadata in system, and then propose the needs for the metadata management according to the characteristics of metadata in National Earthquake Precursor Management System. After that, we put forward our metadata management model –DAC Model (Dynamic Automatic Constraints Model), and design and implement the metadata management architecture using DAC model at last.

The remainder of this paper is organized as follows: Section II introduces the National Earthquake Precursor Management System; Section III analyzes the metadata in real applications; Section IV describes our DAC model; Section V designs the architecture of metadata management system based on DAC model; Section VI gives the experimental results; and Section VII concludes the paper.

2. NATIONAL EARTHQUAKE PRECURSOR MANAGEMENT SYSTEM

2.1 Physical Topology

National Earthquake Precursor Management System is a large-scale distributed hierarchical system deployed in China. Figure 1 shows the physical topology of this system. From top to bottom, the system is divided into 5 layers, i.e. 5 academic centers, 1 national center, 32 regional centers, 360 stations, and many instruments. Obviously, there are a lot of metadata in such a large-scale distributed system.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

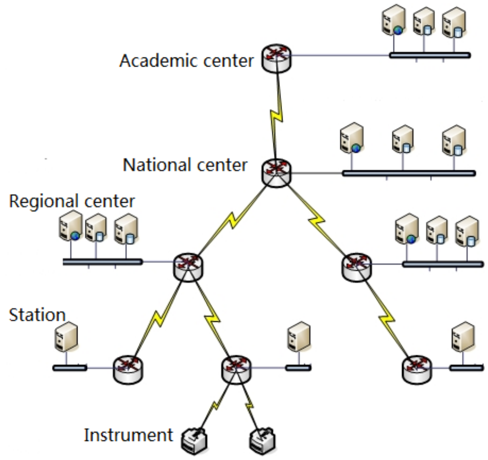

Figure 1: Physical Topology Of National Earthquake Precursor Management System

2.2 Logical Topology

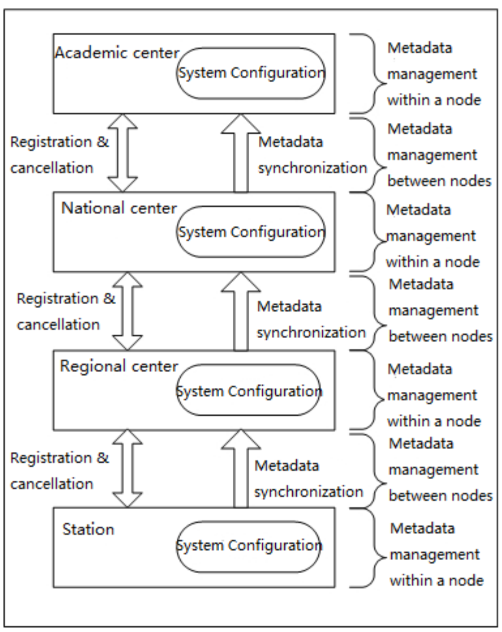

Figure 2: Logical Topology Of National Earthquake Precursor Management System

Metadata management is an important part of National Earthquake Precursor Management System. It provides state description and associated description of nodes, instruments, resources and et al for whole system, which guarantees the stable running in system.

Based on the hierarchical Earthquake Precursor System, our metadata management system is also logically divided into layers, i.e. academic center, national center, regional centers and station. Besides, the application server, database server and backup server are managed separately in the first three layers. Figure 2 shows the logical topology of this system.

1) *Station*

Station is a node which owns independent servers, and can self-manage its resources. It is located in the bottom of logical topology of system. The application server and database server in station layer share one server, and there is no backup server. Since this is the bottom layer and there is no aggregation of metadata from the lower layer, the metadata management of station is the most simplest in the whole four layers.

2) Regional center

Regional center is located between station and national center in the system. On one hand, it approves the station register and gathers metadata information from all attaching stations; on the other hand, it registers to the national center, and provides metadata information from both itself and its attaching stations to national center. Besides, the application server, database server and backup server are managed separately in this layer. Thus, regional center is a typical node in metadata management system, since it involves almost all the contents of metadata management in the Earthquake Precursor System.

3) National center

National center is a critical layer of whole four layers. On one hand, it gathers metadata information from all attaching stations and regional centers, and has a complete metadata structure of entire system; on the other hand, it provides corresponding metadata for different academic centers. In short, this layer not only gathers the information of all nodes and instruments, but also owns a complete physical topology of entire system and associates with the academic centers. All above makes national center become a layer whose amount of metadata is largest.

4) Academic center

Academic center is a node which does specialized academic research in earthquake field. Now there are 5 academic centers, i.e. Geomagnetism, Electric, Strain, Fluid and Gravity. Each academic center can only obtain the corresponding metadata from national center, and one can’t read the original data from others in order to ensure the independence.

We analyzed the characteristics and the relationship of this four layer system, and found there are two categories among layers. The first category includes national center, regional center and station, since these three layers form a complete structure, where higher layer obtains the metadata information from lower layer by synchronizing, and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the number of nodes in lower layer is uncertain. The second category is academic center, since it only contacts with the national center, and the number of nodes in this layer is determined to be 5.

3. METADATA ANALYSIS

In this section, we extract the metadata in the Earthquake Precursor Management System, and classify them according to their relevance.

3.1 Metadata Contents

Metadata is a structure description of data. It is used in describing information of content, coverage, quality, management, owner, delivery mode and et al of data, and is a bridge between the data and users. It can provide not only standard and general description and search tools for various information resources, but also integration tools for distributed information system formed by a variety of resources, such as Earthquake Precursor Management System.

Thus, system is imperfect if there is no metadata. Considering the scale and requirements of various devices and data resources in our system, and the actual deployment environment, we extract metadata as follows:

1) Instrument metadata

Including code, name, model number, record mode of instrument, name, address and contact phone of manufacturers, performance indicator description, et al.

2) Station metadata

Including code, name, latitude, longitude, address and duty phone of station, et al

3) Regional center metadata

Including IP of application server and database server in node, register information of attaching stations, whether it has registered to the national center, whether there is a backup repository, et al.

4) National center metadata

Including IP of application server and database server in node, register information of attaching regional centers and academic centers, information of backup repository, et al.

5) Academic center metadata

Including IP of application server and database server in node, whether it has registered to the national center, whether there is a backup repository, and et al.

6) Backup database metadata

Including whether it has registered to the main repository, IP of main repository and itself, et al

7) Earthquake Precursor metadata

3.2 Metadata Classification

Since the metadata extracted in section A is fragmented, we classify them according to their type, correlation degree and actual demands for different function modules of system as follows in order to provide better metadata management service.

1) Node management metadata

Including metadata of academic centers, national center, regional centers and stations

2) System configuration metadata

Including metadata of instrument, station configuration and backup repository

3) Earthquake Precursor metadata

Above categories obviously shows that the Earthquake Precursor Metadata Management is mainly composed of the first two modules.

3.3 Metadata Characteristics

After the extraction and classification in above section, we analyze the characteristics of metadata as follows:

1) Diversity

National Earthquake Precursor Management System consists of 5 academic centers, 1 national center, 32 regional centers, 360 stations, and more than 90 kinds of instruments. And each node consists of application server, database server and backup server. Therefore, our system contains a wide variety of metadata.

2) Distribution

National Earthquake Precursor Management System is a nationwide deployed system. Thus, metadata is scattered all of the China, i.e. it is geographical distribution.

3) Automatic

In system, according to real needs, each node can configure station and instrument, register node and backup, and has a right to self-manage their own metadata. Since operations within a node will not affect other nodes, metadata has a characteristic of automatic.

4) Constraints

The characteristic of constraints includes validity constraints, consistency constrains and unique constrains. The first one plays a role when metadata comes into the system, it ensures that the above metadata can meet the given specific format or requirements. The second one means that if metadata has replicas in distributed system, all replicas should be modified synchronically. And the last one means that metadata is global unique in whole system, it doesn't allow two different

Journal of Theoretical and Applied Information Technology
31st December 2012. Vol. 46 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

metadata have a same identity.

4. METADATA MANAGEMENT MODEL: DAC MODEL

Combined with the previous analysis of metadata characteristics, this section proposes a metadata management model— DAC model (Dynamic Automatic Constraints model). This model is used to dynamically manage relationship among different layers in system, provide automatic synchronization mechanism to ensure that the metadata from lower-layer can be synchronized to upper-layer in time and in constraint. There are three characteristics of DAC model:

1) Dynamic

This model allows that the lower node can dynamically register to upper node, as well as dynamically cancel. But note that although lower node can select which upper node to be registered, but it can't register to more than one node simultaneously.

2) Automatic

When two nodes establish an upper-lower relationship, the automatic synchronization mechanism of metadata will be established at the same time. This mechanism is used to ensure that the metadata from lower-layer can be automatically synchronized to upper-layer without human intervention.

3) Constraints

It means that the metadata in system should guarantee their constraints, including validity constraints, consistency constrains and unique constrains. The constraints detection will do when metadata come into the system and during their synchronization process.

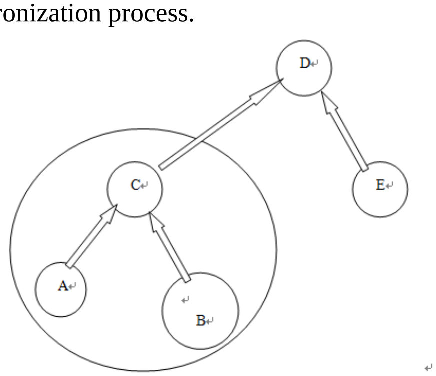

Figure 3: Example Of DAC Model

Figure 3 shows our DAC model. Assume that node A, C, D and E are single node, i.e. they don't have sub-node, while node B is a composite node, i.e. it has sub-node. Node A can choose whether to register to node C or not. When node A registers successfully to node c, a upper-lower relationship will be established as well as the automatic synchronization mechanism of metadata. Then the metadata of node A will automatically synchronized to node C without human intervention.

Besides, both single node A and composite node B can register to node C equally. And all nodes registered to node C can be regarded as a new composite node, which can register again to another node like singe node. Thus, DAC model is recursive, and appropriate for multi-layer nodes.

There are three rules set for DAC model.

1) The upper-lower relationship is built up by registration or cancellation. It is not the original attribute of node.
2) There will be a tree structure (no ring) after the complete registration.
3) Each node can't register to more than one node simultaneously.

Rule 1 is used for expanding the model flexibility, i.e. any two nodes can establish an upper-lower relationship according to the demand. Rule 2 is used for limiting the applications of DAC model, since DAC model is designed for solving metadata management problems in distributed system with tree structure. Excluding network structure makes DAC model be more targeted. Rule 3 is used for ensuring that all metadata in a register tree will be converged to the root node.

In next section, this model will be applied in Earthquake Precursor Industry to achieve a metadata management system.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5. ARCHITECTURE OF METADATA MANAGEMENT SYSTEM

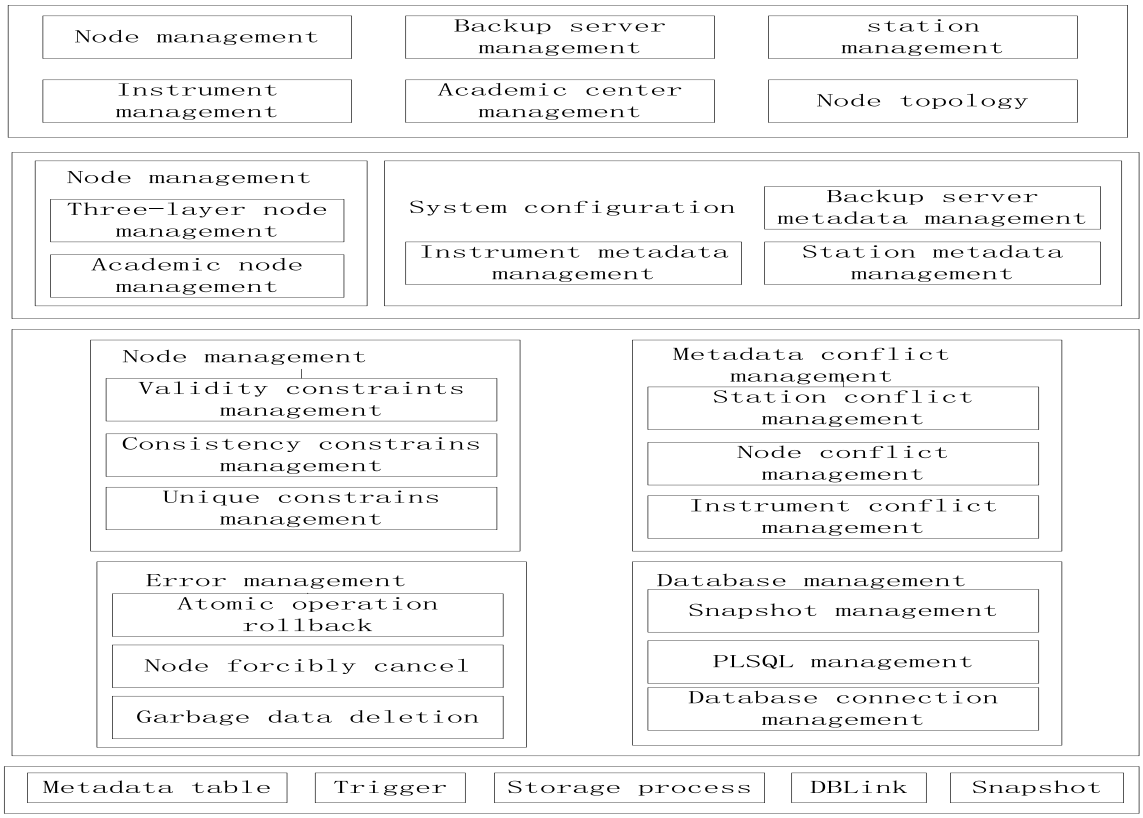

Figure 4: Architecture Of Metadata Management System

Metadata management system consists of four layers, i.e. resource layer, logic layer, business layer and presentation layer (from bottom to top). Following gives a brief introduction.

1) Resource layer

Resource layer locates in the bottom of the system, supporting logical operations for upper-layer. It belongs to the model layer of MVC architecture, and contains metadata tables, triggers, storage process, DBLink, snapshot, JOB, and et al.

2) Logic layer

Logic layer uses resources from resource layer to provide logical management (such as management of metadata constraints, metadata conflict, error, database, and et al) for business layer to complete its corresponding function.

3) Business layer

Presentation layer displays the operations or their sets to users, while business layer integrates the operations with similar object according to the characteristics of each operation. Besides, it links the user operation with system, and constitutes the control layer of MVC architecture with the logical layer.

4) Presentation layer

Presentation layer provides users a website to use this metadata management system. It includes management of node, backup repository and academic center, configuration of instrument and station, node topology, and et al.

6. EXPERIMENTS

In this section, we first deployed a test environment, and then tested the main function of our metadata management system. Besides, we did constraints test and disaster recovery test for some critical business and analyzed the results. At last, we evaluated the function and performance of this system.

6.1 Test Environment

With the support of Institute of Geophysics, China Earthquake Administration, we deployed a test environment using 19 Dell blade servers in two pilot provinces. Following is the details.

Journal of Theoretical and Applied Information Technology
31st December 2012. Vol. 46 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1 : Description Of Test Environment

Node's Type	External Type	academic center, national center, regional center, stations
	Internal Type	Database server, Backup server, Application server, GIS server
Hardware	Database server	2CPU, 2G memory
	Backup server	2CPU, 2G memory
	Application server	2CPU, 2G memory
	GIS server	2CPU, 2G memory
Software	Operation system	Suse Linux Enterprise 10
	Database system	Oracle10g for Suse Linux 3.Enterprise 10
	Web server	Apache Tomcat Version 6.0.10
	Compiler	jdk-1_5-linux-i586
	GIS	ARC GIS9.1

Table 1 describes the test environment. Figure 5 shows its topology. According to the actual demands, our environment includes Geomagnetism academic center, national center, 2 regional centers (Shanxi, Shandong) and 3 stations (Datong, Yantai, Tai'an).

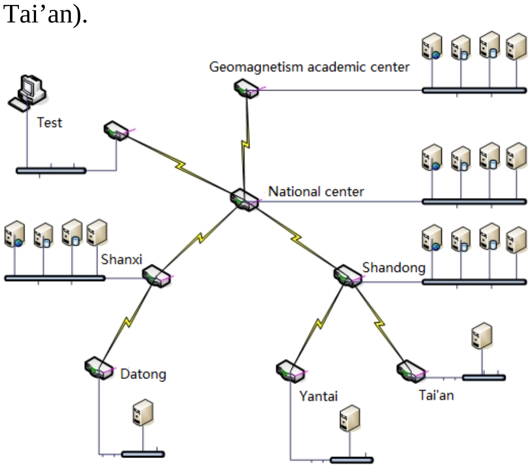

Figure 5: Topology Of Test Environment

6.2 Function Test

This section tests the main function of Precursor metadata management. Following is the results.

1) The main functions (including: registration of node and academic center, configuration of instrument and station, metadata synchronization, et al) have been realized and are able to work properly.
2) Station configuration and synchronization from regional center to national center costs less than 3 minutes, which meets the time requirements of our system.

6.3 Constraints Test

Function test exams whether the function of system is correct or not when all the inputs are correct, while constraints test exams whether system is stable when some inputs are error.

Results show that our system has the validity constraints of metadata, since it can give a prompt response to the error input. Besides, when users configures a station existed already to regional center, this configuration will not be succeeded. Thus, our system has the ability of conflict management, which ensures the unique constraints of metadata.

6.4 Disaster Recovery Test

Disaster recovery test exams whether system can recover from the irresistible catastrophic accident or not. When the network between Shanxi and Datong meshed, there was no information of new configured instrument in Shanxi regional center. However, after 1minutes when network recovered, this new configured instrument appeared in Shanxi regional center correctly.

7. CONCLUSION

In order to meet the demands of National Earthquake Precursor Management System, this paper researches deeply on the metadata information of system. We first analyze, extract and classify the metadata, and then proposed a novel metadata management model — DAC model (Dynamic Automatic Constraints model). After that, we design and implement the architecture of our metadata management system according to MVC pattern and our DAC model. We deploy a four-layer test environment and test out management system at last. The experiment results show that our management system works well, and has good fault tolerance and disaster recovery.

ACKNOWLEDGEMENTS

The work was supported by the National Natural Science Foundation of China (No. 61100066) and Special Fund Research in the Public Interest (No.201008002).

REFERENCES:

[1] I. Foster, C. Kesselman, S. Tuecke, "The Anatomy of the Grid: Enabling Scalable Virtual Organizations", *International Supercomputer Applications*, Vol. 15, No. 3, pp. 200-222, 2001.

Journal of Theoretical and Applied Information Technology
31st December 2012. Vol. 46 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[2] L. Zeng, B. Benatallah, A. Ngu, et al., "QoS-aware Middleware for Web Services Composition", *IEEE Transactions on Software Engineering*, Vol. 30, No.5, pp. 311-327, 2004.

[3] E. M. Maximilien, M. P. Singh, "A Framework and Ontology for Dynamic Web Services Selection", *Internet Computing*, Vol. 8, No. 5, pp.84-93, 2004.

[4] I. B. Arpinar, R. Zhang, B. Aleman-Meza, et al., "Ontology-driven Web Services Composition Platform", *Information Systems and E-Business Management*, vol. 3, n. 2, pp.175-199, 2005.

[5] V. Tosic, K. Patel, B. Pagurek, "WSOL—Web Service Offerings Language", *Lecture Notes in Computer Science*, Vol. 2512, pp.57-67, 2002.