Research on Heterogeneous Data Access Based on OGSA-DAI

Zheng CuiFang [1, a], Zhang Qing [2,b] and Cheng Zheng[3,c]

[1] Institute Computer Application, Chinese Academy of Engineering Physics，Mianyang Sichuan P.R. China.

[2] Institute Computer Application, Chinese Academy of Engineering Physics，Mianyang Sichuan P.R. China.

[a]zcfkb@sina.com, [b]zhangq@caep.ac.cn, [c]chengz@caep.ac.cn

Keywords: Grid; OGSA-DAI; data service; database grid; heterogeneous data.

Abstract. The existence of a large number of distributed heterogeneous data sources, to access betweent application systems brings much inconvenience.OGSA-DAI middleware simplifies the date access and integration in the grid environment. This paper focusing on the innermechanism of OGSA-DAI, describes its architecture, main modules and the relationship among modules, and on this basis, the paper conducts a preliminary study on the heterogeneous data access working mechanism of OGSA-DAI.

Introduction

Along with the rapid development of Internet and grid technology, the network has become an important means of information exchange and communication, abundant data resources have appeared on the Internet, these data resources distribution in different server site, and data storage and expression are also different. How to effectively access and integrate these distributed and heterogeneous data has become a hot topic in grid research. So the Open Grid Forum Organization design of a grid middleware OGSA-DAI (open grid service architecture data access interface), it appears to solve the problem of heterogeneous data access, promote the development of grid.

OGSA-DAI is grid middleware for data access and integration in grid environment [1], based on OGSA standard, in GT (Globus Toolkit, it is a set of software toolkit for writing application based on grid) is developed, grid provides a unified interface for data access. It is widely used in a variety of data integration systems; support a variety of platforms, a variety of database systems. Currently OGSA-DAI support database types are: relational database (such as: SQL, Server Oracle, MySQL, etc.), XML database (such as: Xindice) and documents (such as: BinX file). It can realize the access to these heterogeneous data sources, at the same time, it can also realize the function of data conversion, and the use of the language is independent of the client application

At present, AstroGrid, Biogrid, BioSimGrid and other projects are used in foreign countries about OGSA-DAI, and the research based on OGSA-DAI is still in the initial stage in our country, and the application is almost blank. Therefore, the author studies the OGSA-DAI architecture, analyzed the interaction process of the data access. At last, the paper introduces the process of heterogeneous data access, but this is only a theoretical guidance, in practice, the access of application system will involve more and more complex issues, and we need to study.

OGSA-DAI Overview

OGSA-DAI Architecture

OGSA-DAI provide a basic framework for integrating existing data resources into a grid environment [2]. The architecture of OGSA-DAI is shown in Fig.1 [3]:

1. Data layer: The layer package for each heterogeneous data source, so that it becomes the GDR (grid data resources) set of OGSA-DAI, and for access to a layer of business logic layer.
2. Interface of data layer and business logic layer: Information communication in data layer and business logic layer by using data resource visitor, each type of data service resources have

their specific data visitor to control the access of data resources. Can call JDBC drives, XMLDB drives, etc.

3. Business logic layer: This layer on the core functionality of the OGSA-DAI package, it can accept, responds to the client's request. The layer package OGSA-DAI all the services to the grid display, including: GDS (grid data service), DAISGR (data access and integration service group registry) and GDSF (grid data service factory) [4]. Among :

   1) DAISGR，Used to find data resources is a permanent service. This service is used to find the needed GDS, and it can also be found by it to create the desired GDSF。

   2) GDSF，Used to declare the data resources is a permanent service. This service is used to create an instance of GDS to access specific data resources.。

   3) GDS，Used to access data resources is created by the GDSF with a limited life cycle of services. Through this service, you can access a data resource。

This layer contains one or more GDSR. It is mainly used to execute Perform documents, generate a response document, perform various operations on data resources, to realize data transmission, but also to support session management and property management.

4. Interface of presentation layer and business logic layer interface: it is mainly used for the communication between the two layers, the presentation layer transmission service resource name, and attributes name and data service resource allocation information to business logic layer data. In the business logic layer, said layer transmission response and results of the document data and data service resources to support the event information.

5. Presentation layer: this layer in the way of Web Service to package the functions to be realized of DSR and to provide service interface. OGSA-DAI provide two ways to realize, one is applicable to the WSRF specification, and the other is applicable to the WSI specification.

6. Client layer: the client can interact with the DSR through the corresponding DS, the client can choose the WSRF standard or WSI standard.
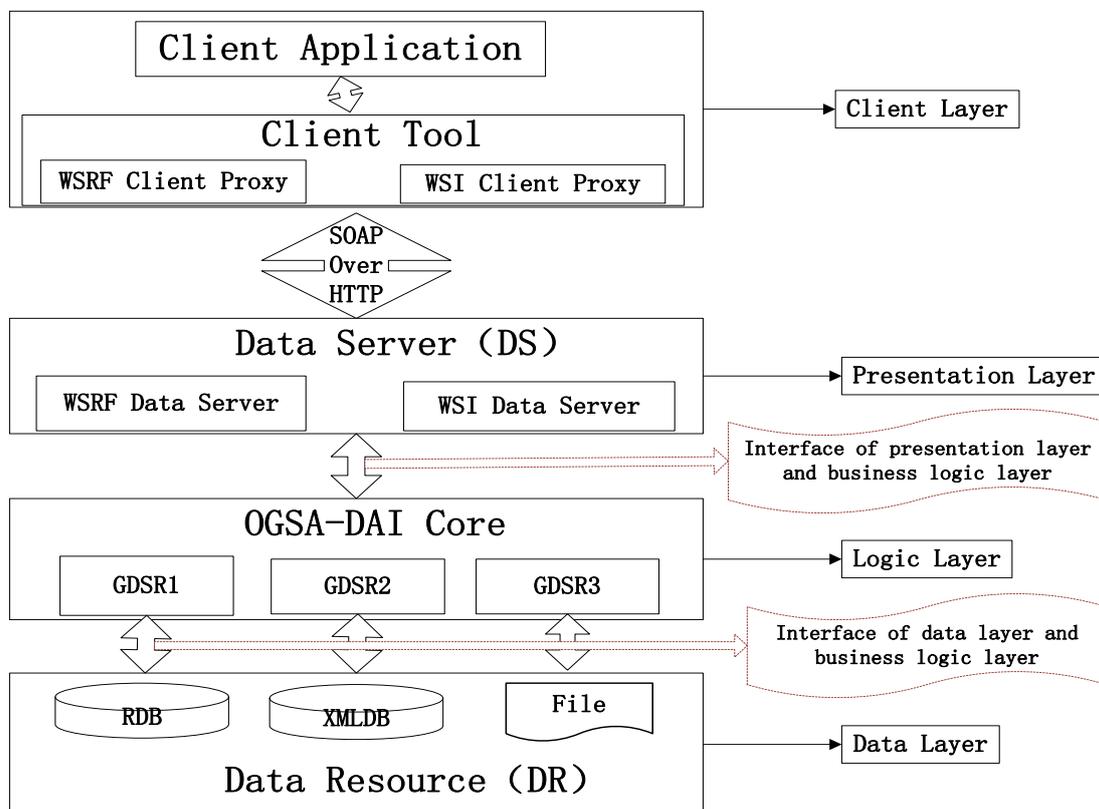

Fig. 1 OGSA-DAI Architecture

Interaction process of OGSA-DAI

OGSA-DAI uses SOA architecture, complete the interaction of DSR with the way of the Web Services interface of the document. In the process of internal interaction, allocate a session for each request; make conversation to do specific work. Several important components used in the OGSA-DAI interaction model [6]:

1. Perform Docment: The client uses the XML language to express the activities they want DSR to perform.
2. Response Document: Returns the result of perform document to the client which is processed by GDSR. The response document describes the execution status of perform document and the session to which the identity request is added.
3. Activity: The operations which is Executed by DSR, including data resources query and update processing, data conversion and data transfer.
4. Session: Object that stores the state of the DSR when DSR process multiple requests.

OGSA-DAI uses the delivery documents form to convey information. The interaction process is shown in fig. 2.
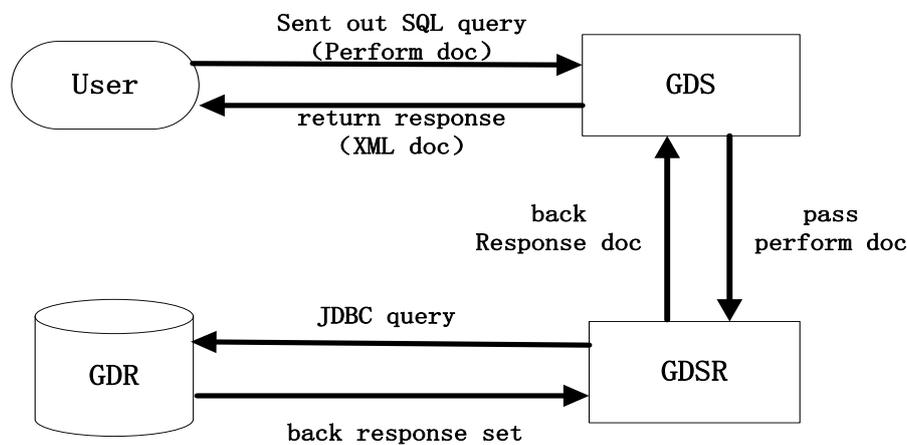
Fig. 2 OGSA-DAI interaction process

OGSA-DAI interactive steps are as follows:
1. Users send Perform Document to GDS; this document contains all kinds of information needed by OGSA-DAI;
2. GDS sends Perform Document to the GDSR that represents the actual GDR;
3. GDSR interpretation verification Perform Document and to form a specific operation behavior，these behaviors interact with the underlying database;
4. The database will perform the results back to the GDSR;
5. GDSR nest processing result in Response Document and back to GDS;
6. GDS Returns a response document to the user.

Implementation of heterogeneous data access

OGSA-DAI Supports interaction with various data sources, using XML document interface to achieve the various operations of the data，Communicate between client and server through SOAP message. In this paper, we use access relational database and XML database as an example to analyze the process of accessing heterogeneous data by using OGSA-DAI.

Environment building
1. Environmental requirement

1) Operating system: Windows、Linux etc;

2) Server software：MyEclipse，database (such as Oracle、MySQL、SQL Server etc)，OGSA-DAI，GT，Java Web Service Core，Tomcat，J2sdk and Apache Ant etc;

3) Client software: OGSA-DAI client application interface，Globus Toolkit client.

2. Installation Steps

Here outlined the installation steps are as follows, the specific reference Guide install [7] of Globus official website。

1) Support environment build

Install supporting tools, such as database、Eclipse、Java JDK、Apache Ant、Tomcat and Java Web Service Core ect;

2) Install and deploy of OGSA-DAI

Download the GT4 package and install it on Eclipse development tools, and then deploy the GT4 to Tomcat; install OGSA-DAI, deploy Tomcat to OGSA-DAI;

3) Configure OGSA-DAI access database

Deploy a variety of database resources into OGSA-DAI services. Remote database resources are registered to the OGSA-DAI service through the GUI of OGSA-DAI, used to receive the operation of OGSA-DAI.

Environment building is a complex process, need to slowly explore, involves a variety of software configuration, one step is not the right way to lose.

Access procedure.

In OGSA－DAI, client sent a Perform Document to GDS，GDS Ready to receive Perform Document，run the database query, transfer the results of the query, transfer data, the specific process as shown in Fig.3:
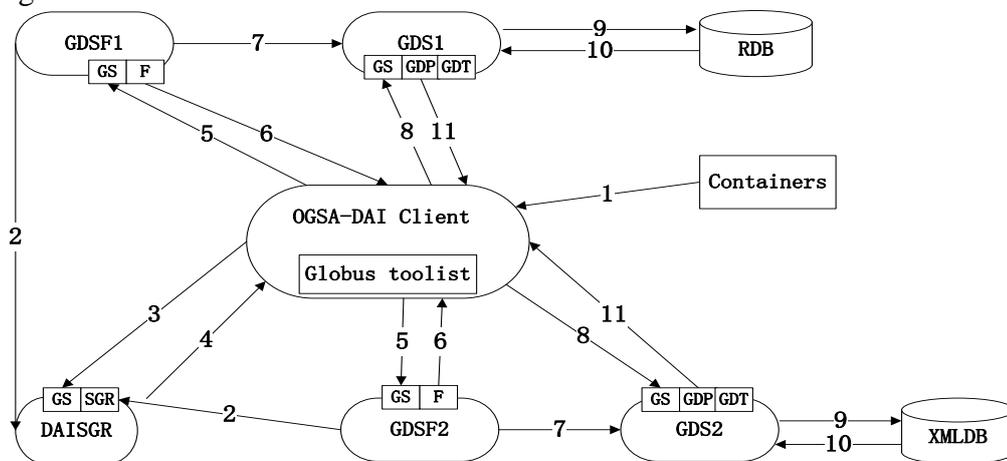
Fig. 3 OGSA－DAI work flow

1. Detailed implementation steps are as follows:

1) Run WebService Container，at this time ,GDSF represents the database;
2) GDSF registers in the DAISGR;
3) The user selects to meet the GDSF of user need through the DAISGR;
4) GDSF return the created GDS instance of GSH to client;
5) Users choose direct query GDSF to understand the database;
6) Return GSH of Meeting the needs of GDSF to the OGSA-DAI client；
7) User request to create a GDS;
8) Client call a specific GDS according to the GSH which is just returned by GDSF.and sent requset perform docment to GDS instance by the form of XML;
9) GDS will sent received documents and some of the context information to the implementation engine, then the document is processed by the execution engine (Specific

include: parse and validate the execution documents, identify the activities specified in the execution document, to create an instance of these activities and the examples of these activities, and instances of the pipeline between these activities, create and start the instance of the active processor, etc.

10) The database returns the results to GDS;

11) GDS sends the results to the client in the form of response document.

2. The interface type is as follows:

1) GS (Grid Service)：OGSI grid server;

2) SGR（Service Group Registration）：OGSI Service group registration;

3) F（Factory)：OGSI factory;

4) GDT (Grid Data Transport)：OGSA-DAI grid data transmission;

5) GDP (Grid Data Perform)：OGSA-DAI grid data execution.

Conclusions

Overall, using OGSA-DAI to achieve heterogeneous data access have cross platform, cross language, scalability, loose coupling and other advantages, can solve the problem of heterogeneous data access process. However, OGSA-DAI is still in the process of research, there are still some problems [8]. Such as: OGSA-DAI is directly connected to each heterogeneous data source by calling the data service, The response time of this database connection mode is long, affect the system performance; in addition, OGSA-DAI is only independent package the single database with different types, it is not a good solution to the problem of database semantic conflict. How to solve these problems, so that OGSA-DAI can finish heterogeneous data access processing with high quality and high efficiency will be the focus of the next.

Acknowledgements

References

[1] Dj.M. Maric, P.F. Meier and S.K. Estreicher: Mater. Sci. Forum Vol. 83-87 (1992), p. 119

[1] OGSA-DAI WSRF2.2 http://www.ogsadai.org.uk/documentation/ogsadai-wsrf-2.2/doc/.

[2] LIU Jian-Xin, YAN Bao-Ping .Research on the architecture and key technology of OGSA-DAI [J].Computer Applications，2005，24（11）：81-87.

[3] OGSA-DAI Architecture http://www.ogsadai.org.uk/documentation/ogsadai-wsrf-2.2/doc/

background/architecture.html.

[4] Neil P Chue Hong,OGSA-DAI Overview.

[5] [27] Interacting with Data Service Resources http://www.ogsadai.org.uk/documentation/ogsadai-wsrf-2.2/doc/interaction/index.html.

[6] LUO Qing-di,JIANG Xiu-feng.Research on Isomerous Data Integration Based on OGSA-DAI in Grid Environment[J].Computer Technology and Development,2009,19(3):144-147.

[7] GT4 Admin Guide，http://www.globus.org/toolkit/docs/4.0/admin/docbook/

[8] YANG Xue-Seng. Integrate Heterogeneous Database System Using OGSA-DAI Middleware [J].

Information Technology, 2011, 35(1):121-123.