SciVerse ScienceDirect

Physics Procedia 24 (2012) 2180 – 2186

Physics Procedia

2012 International Conference on Applied Physics and Industrial Engineering

The Application of Semantics Web in Digital Library Knowledge Management

Du Liangxian,Qi Junxia,Guo Pengfei

Handan College

Abstract

The semantic grid can bring different community together to solve the problem, so large human interest is increasing, and based on the grid technology suppliers, to share information of urgent need. TIn this article, we will discuss the details of the research status and technical support HBUTiGrid project, elaborate design scheme of software architecture, HBUTiGrid infrastructure. We also exist vital experience HBUTiGrid project construction, introduces the software tools, support it.

Keywords: Semantics Web; Digital Library; Knowledge Management

1. Introduction

In recent years, the network is composed of a massive data and information. This growth brings many problems for the user. But more and more difficult to find, visit and safeguarding the information, isolation and lack of interoperability application software seriously hamper the communication between different systems. Since resources and the application, make use of resources platform, multiphase is ideal. Therefore, the mesh. The goal is to make the grid human resources utilization of electric power facilities. Once the computer infrastructure to support resource sharing and cooperation with the grid, the user can use their grid, do not need to pay attention to those resources are directly from, and how they deal with the underlying computer infrastructure in these facilities.

With the further development of the grid, quicken the grid technology; there will be a result, as users reduce the opacity to huge resources and services are not easily be human user. Using the semantic web technology interprets the resources and services will greatly reduce the human efforts and use resources and services more efficient. All the resources and services of the grid should fully describe as machine - process able. This is the semantic meaning and shall clarify the claim for the power grid service facilities,

doi:10.1016/j.phpro.2012.02.320

the grid applications. Make the computer and work in cooperation to support a flexible and calculation of worldwide.

HBUTiGrid project's main purpose is to solve the problem in the resource sharing and interoperability, increase scalability multiple digital, the development and utilization of digital information environment, and provide efficient information service, by using grid technology and semantic areas. This explains the software architecture of the infrastructure and HBUTiGrid design details.

2. Current status of semantic grids

Many different grid technology standards have been presented, and a huge amount of middleware systems and tools have been developed to support applications in grid middleware services, including the creation, deployment and remote service call, safety, resources monitoring, scheduling, high-speed data transfer, data management, composition and application of workflow management [1] complain these standards and tools in the following chapters.

Tim berners-lee and his partner, a new idea of semantic web, a highly interconnected network data can easily access and understand any desktop or hand-held machine. This data is a common language, can understand the various software agent, Ontology into common from various database information, And the rules for some information allows software agents to describe those terms. This data format, ontology and reasoning software would operate a great application, in the home page, analysis and processing of all original data stored in the database, and all information and data of text, images, video, audio, communication network includes [7].

Big companies have major projects, and greatly improve the efficiency of the processing and analysis data, using the interaction between semantic web to improve data structure, establishing a hidden. For example, Cincinnati children's hospital medical center by using semantic web tools to find potential of the causes of heart disease. The goal of this project is to find out which genes affect biology is most likely. The relevant information from different database, have different sources and in incompatible formats. The team has holes through four or five database for everyone, this is true, time - task. Later research team translated into RDF and storage formats, and then use the information in the semantic web of knowledge integration and to assess. The semantic web technology was also used in health. Sapphire, by using semantic technology integration a wide range of data from local health care providers, hospitals, environmental protection and scientific literature into a single health can better current detection and analysis of emerging and the public health problem, health officials for information through different network platform [7]. This system developing an ontology of disease, now will explain the flu cases, flulike symptoms such as automatic report them to the centers for disease control and preention, this is a key sapphire.

There are many manufacturers support the semantic web, some of the large famous companies such as IBM (Snobase ontology management system), hewlett-packard (supp - orts framework, reading, writing, query, and reasoning), the oracle (now add new features, and inquires support RDF storage).

3. Current research and technology

3.1 Grid Technology

The grid based on programmable controller hardware and software infrastructure to ensure a cheap, consistent and reliable access to computing resources anytime [4]. In the interest of the grid, especially the industrial interest, improve the development of several different standards and information exchange agreement, creation and management:

1) OGSA

This is an important open standard, establish the grid. Ogsa-dai designated grid service framework and grid service standards. The grid services, it is compatible with the framework of network service, not only can define technical interoperability and a wide range of light, the service also supports the new mechanism of service, service and management of the notification service or resources. To provide a safe, ogsa-dai is the grid environment and distinctive resources to multiple suppliers shield, the development of specific safety service, tools and application, will they seamless connection.

2) WSRF

WSRF ogsa-dai inherited merits. This is a set of test, based on the network service standard, defines the rendering of the specific method of Resource - WS to exchange information and related to the definition, including the XML declaration and executive connections between network services and one or more resources and related mechanism of grouping and Resource - about WS addressing, describes how to define a picturesque national resources, and web services description, and how to make state through network Resource - WS service interface [3].

3) Globus Toolkits

It's a open-source toolkit, and has emerged as a standard infrastructure for establishing Grids by 1998. With GT, the Grids architects can establish Grid system using standard building blocks.

4) Grid Services

Web services is an integral whole, its customers to provide some ability, through the exchange messages defined sequence, service network on some operating [1]. Therefore, it is the implementation of the flexible and service. The creator of the sun, JAXR studio, JAX - WS 2.0 is based on a Java development kit use network services and provide integrated environment to support development life cycle. Based on grid services, network service, with a standard XML [11] information system and mature communication protocol to support interoperable machine - to - online interactive machine. Grid structure and interface definitions in WSDL, customers can be exposed to call the service method and interact clubs provide remote.

3.2 Semantic Grid

The semantic information grid extend existing services and meaning of the grid work is clear for human and machine understanding, improve interoperability. This is a set of service entity is according to the definition of the universal, make its function can be recognized and accept any application. Ontology is a conceptual model of formal concept and concept of relation, make a consensus of the specific meaning of words. Web ontology language (owl [5]) and network service (11 modeling ontology) is generally [6], is used to define the standards body so that they can be compatible, can understand the RDF.

4. The software architecture stack in hbutigrid infrastructure

The main design scheme, HBUTiGrid is called the semantic web services in the operation of the resources of the digital library database. First, the service provider of the semantic web service advertising, then UDDI Globus platform based on it and registration information found by ontology. When a request service, service dispatching center submitted semantical scheduling, originally concepts in the ontology to find relevant clauses of the user needs extracting. If there are several service concept, and related resources scheduling, namely next launch dispatching center services, according to their priority of hardware environment resources scheduling algorithm, and expounds [8],at last, the grid center elects the most suitable service and return its URI to consumer—The paper aims at how to realize this proposal and will present a detailed software stack used for constructing the HBUTiGrid infrastructure. The software architecture stacks under the HBUTiGrid project is showed in "Fig .1".

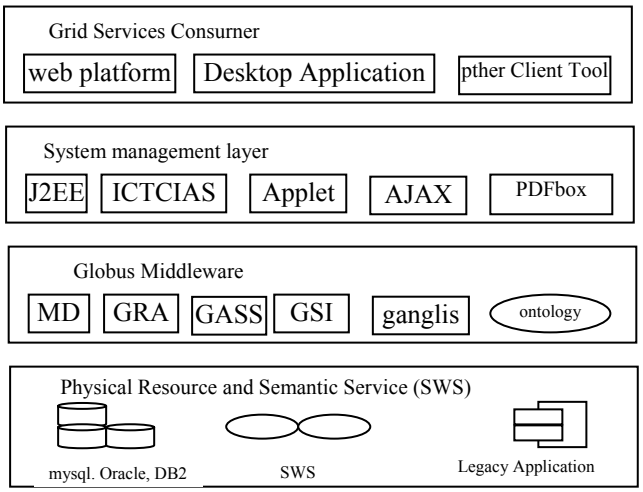

Figure 1. Software Architecture for the HBUTiGrid project

4.1 The Fabric Layer

The lowest layer is the fabric layer, on which are physical data and resources, and the semantic web service whose methods operate on those resources, as well as a number of legacy applications that will be reused in system integration. For this layer, the project adopts windows or Linux as the operating system. The semantic web services of digital Library is on this layer. Every Service provider deploys their services in their server respectively. Thus a server for deploying Web Services is needed. We select Jboss which is an excellent application server supports EJB as the server.

4.2 The Middleware layer

Middleware layer fabric layer. The goal of this project is to find a free one of the grid computing resources. Network services and good hardware environment is called the client. Depending on factors, which is part of the CPU idle, the percentage of storage and free network load? In this layer, we use cluster monitoring tools, data obtained ganglion layers. In the server's gmetad and responsible for storing data from a list of clustering user's chosen, so we can be calculated based on the data value. And it provides wonderful GUI for watching.The“Fig.2” is a monitoring interface the ganglia GUI provide for user.

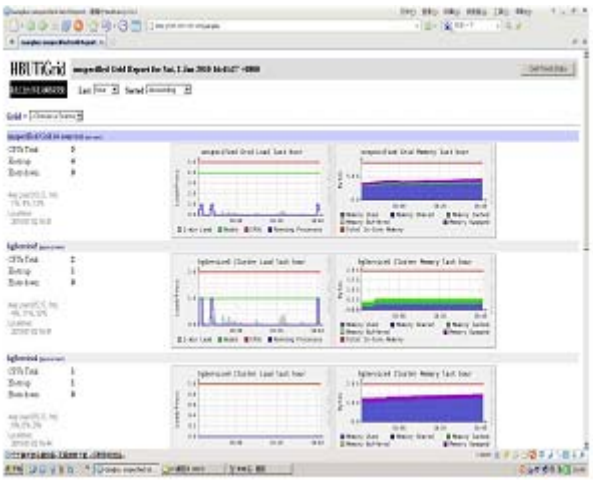

Figure 2. The Grid Cluster Monitoring Interface

But the shortage of this proposal is that the ganglia' server must install on a Linux Operating System. Therefore, the globus middleware is installed on Linux OS.

Introduce[2], as a Grid development tool, aims to lightening the grid developer's load, encapsulates the complex orders used in GT and provide a GUI supports service creation, service deployment and service deletion using the underlying GT and WSRF standards. With introduce, grid developers only concentrate on the application development and deploy it as a Grid Service in globus container, which is easier and time-saving than before. Based on the globus platform, we develop two kinds of Grid Service, one is Grid Monitoring and Registration Service (GMRS), monitoring factors mentioned above of web services in the fabric layer and register information of web services advertised; the other is Grid Dispatching Service (GDS), orchestrates and control access to the Web Service providers in term of priority.

We use semantic matching calculations game. The semantic matching is invoked the mapping relationship of service, write in an ontological visitors, services, written in the ontology.

Describe a service function, the input and output, and other programmers and machine can understand, we use a noumenal signification to repair service provision applies to describe by adding semantic tagging of WSDL document. This is the semantic web services. In application-layer (will introduce in part), searching speed 450.finding performance requirements of the matching relation, and inquires is due to the service efficiency of consumers. Therefore, we calculated degree in related, each semester registration service, the ontological web services, through the grid monitoring and registration service. When the consumer demand, puts forward the extract, terms GDS in first, then the search term in the ontology and choose network services related content, the network service GDS election has high value evaluation of hardware and network services to the consumer.

4.3 System management Layer

Middleware is the core layer in the infrastructure, such as body, system management level is the brain. Based on this layer is constructed based on J2EE technology. It use struts technology to control transfer. This process is as follows: first, ICTCLAS30 (Chinese lexical analyzer) extract from the user's terms, the application submitted to query the consumer demand, then the result was sent to the middleware, grid middleware layer is automatic invocated, Secondly, URI elected network service system is back, when

the customer management level, resources, and operational methods, through the network service in the fabric layer.

Figure 3. The Visualization Interface for User

The information that we use small grid realize graphical visualization cluster. The result shows content-based image retrieval efficiency of a clear web, we adopt the AJAX, JavaScript and Dom technology and implement all interactive function through by ordinary JavaScript browsers support. We also use tools, Pdf box analysis in the implementation of a PDF document full text retrieval. "To", is common user interface. The results show that in every node on the detailed information of the grid and their relations are grid servers node of the center.

4.4 Application Layer

This application in this layer, namely the network platform, desktop applications and other customer tools, is real consumer grid service provider, is usually based on an Internet access through the portal. The grid services, authorized management is the most suitable for the election of service customers. Finally, consumer demand, latent semantic web services and access to remote in the grid.

5. Conclusions

Future work after the project are concentrated in solving the problem in the grid services, the semantic assert middleware has stronger robustness of cooperation and service, now we have complete description of the natural composition of network services.

References

[1]Hastings, S., Langella, S., Oster, S., Saltz, J.: Distributed datamanagement and integration: the Mobius project. Proceedings of the Global Grid Forum 11 (GGF11) Semantic Grid Applications Workshop, Honolulu, HI, pp. 20–38, 2004.

[2]Shannon Hastings,Scott Oster ,Stephen Langella. Introduce: An Open Source Toolkit for Rapid Development of Strongly Typed Grid Services.Grid Copmuting,PP.407-427, May,2007.

[3]Karl Czajkowski ,Donald F Ferguson,Ian Foster,The WS-Resource Framework.http://www.globus.org/wsrf/specs/ws-wsrf.pdf ,March 2004.

[4]M. Bubak, M. Malawski, and K. Zajac, Architecture of the Grid for Interactive Applications, Proc. Int. Conf. Computational Science LNCS 2657, Springer, pp. 207-213,2003.

[5]David Martin, Mark Burstein, Jerry Hobbs,OWL-S: Semantic Markup for Web Services.http://www.w3.org/Submission/OWL-S/,Nov 2004.

[6]http://www.wsmo.org/ accessed March3, 2010.

[7]Feigenbaum, Lee, Ivan Herman, Tonya Hongsermeier, Eric Neumann, and Susie Stephens. “The Semantic Web in Action.” Scientific American, vol. 297, pp. 90-97, Dec. 2007.

[8]Saltz, J., Oster, S., Hastings, S., Kurc, T., Sanchez, W.,Kher, M., Manisundaram, A., Shanbhag, K., Covitz, P.:caGrid: design and implementation of the core architecture of the cancer biomedical informatics Grid.Bioinformatics,vol.2, pp.1910–1916, 2006.

[9]Chervenak, A., Deelman, E., Kesselman, C., Allcock, B.,Foster, I., Nefedova, V., Lee, J., Sim, A., Shoshahi, A.,Drach, B., Williams, D., Middleton, D.: High-performanceremote access to climate simulation data: a challenge problem for data Grid technologies. Parallel Comput,vol.29,pp.1335–1356, 2003.

[10]Dan Brickley, R.V. Guha, RDF Vocabulary Description Language 1.0: RDF Schema.http://www.w3.org/TR/rdf-schema/, Feb 2004.

[11]Henry S. Thompson, David Beech, Murray Maloney,XML Schema Part 1: Structures Second Edition.http://www.w3.org/XML/Schema, Oct 2004.

[12]Deborah L. McGuinness, Frank van Harmelen ,OWL Web Ontology Language Overview ,http://www.w3.org/TR/owl-features/, Feb 2004.