

## OGSA-DAI—Uses and Applications

Richa Batra<sup>1</sup> & Ramandeep Kaur<sup>2</sup>

<sup>1</sup>richabatra@yahoo.com <sup>2</sup>ramanz2112@yahoo.co.in

### ABSTRACT

Services of OGSA-DAI enable databases to be accessible through grid service interface. Databases are to be incorporated within OGSA framework. The use of OGSA-DAI helps the access of heterogeneous disparate resources in uniform way. The OGSA-DAI middleware product, available for more than 2 years allows data resources, such as relational or XML databases, to be accessed, integrated and federated via web services. The web services of OGSA-DAI are brought in action through Grid Environment. In this paper we have discussed the various applications and the results of the use of OGSA-DAI on various databases.

### 1. INTRODUCTION

The OGSA-DAI (The Open Grid Services Architecture - Data Access and Integration) project started in February 2002 and since then it has undergone 3 main phases. OGSA-DAI is a middleware product that allows data resources, such as relational or XML databases, to be accessed, integrated and federated via web services. OGSA-DAI services can be used to provide web services that offer data integration services to clients. The main purpose of OGSA-DAI is to provide the required data resource sharing not only to support the access to disparate resources but also to transform, integrate and deliver the data. To give proper access of resources the OGSA-DAI is a framework that executes workflows. OGSA-DAI is emboldened so as to allow different resources to be exposed on Grids and also to provide web services that can be combined to provide higher-level web services that support data federation and distributed query processing. One of the important features of OGSA-DAI is a workflow engine, which helps to define retrieval, update and transformation operations to the grid users.

### 2. PRESENT WORK

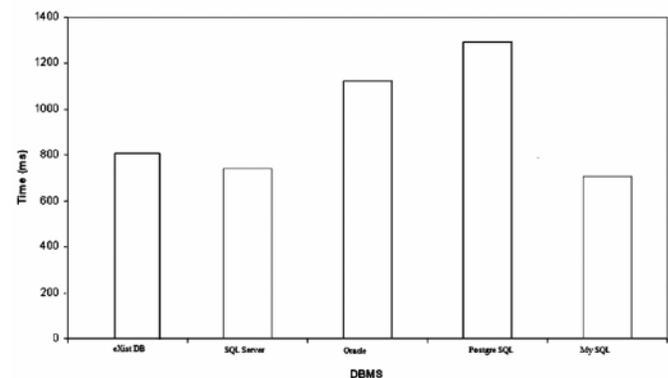
Multiple data can be accessed in a scope of single workflow. The workflow engine supports both synchronous and asynchronous execution. In synchronous mode results are declared after completion of entire workflow. Hence to have immediate results or to have instant output the asynchronous mode is preferred as the results are returned after the first request submission to the OGSA-DAI server. So by the use of asynchronous mode of OGSA-DAI there is support for registration and discovery of databases and also there is interaction with those databases. The structure of the results returned, the method and also the location of their delivery can be set by client. OGSA-DAI provides transformation activities

that support processing such as transforming SQL rows to XML-based WebRowSets.

OGSA-DAI is used by research and business users from all around the world, from various sectors including geographical information systems, meteorology, transport, computer-aided design, engineering, astronomy and medical research.

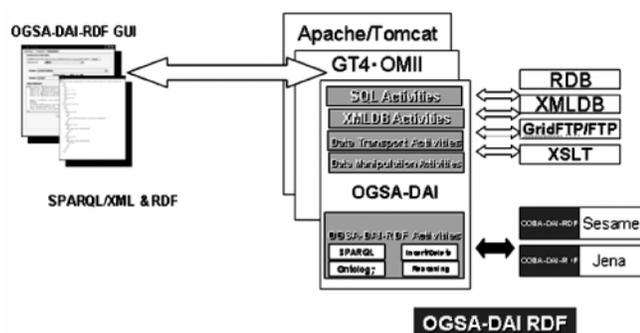
The OGSA-DAI middleware has been tested with drivers for eXist, IBM DB2, Microsoft SQL Server, MySQL, Oracle and PostgreSQL databases using Globus Toolkit 4.0.5. On the basis of data already being tested by OMII-UK more analysis were done taking more array size and more number of rows. The results are being shown in following graph.

Mean Time for 15000 rows by DBMS



In order to extend a new data resource type called RDF (Resource Description Framework) has been developed. OGSA-DAI RDF middleware extends OGSA-DAI access to RDF database systems, e.g., in Sesame and Jena 2. There has also been implementation of several OGSA-DAI activities for handling RDF data. The SPARQL query language is being used for query processing.

As shown in the following figure, DAI-RDF provides internal product-independent access layer, which wraps product dependency. We provided several RDF repository drivers for OGSA-DAI-RDF. This interface also supports the http server interfaces, which many existing RDF repository provides. To support RDF Dataset, DAI-RDF introduces new resource type "RDF" to the original OGSA-DAI resource types such as "XML" and "Relational". Thus, several heterogeneous data resources including RDF data resources can be handled with a single OGSA-DAI data service. An RDF Dataset is a set of RDF Graphs. An RDF Graph is a set of RDF triples, which has Subject, Predicate and Object. Each graph is uniquely identified with URI. Thus, an RDF data resource holds a set of Graphs with URI.



In DAI-RDF, an RDF data resource in the context of OGSA-DAI is equal to a RDF Dataset of W3C SPARQL. An RDF Dataset is a set of RDF Graphs. An RDF Graph is a set of RDF triples, which has Subject, Predicate and Object. Each graph is uniquely identified with URI. Thus, an RDF data resource holds a set of Graphs with URI.

To support RDF Dataset, DAI-RDF introduces new resource type "RDF" to the original OGSA-DAI resource types such as "XML" and "Relational". Thus, several heterogeneous data resources including RDF data resources can be handled with a single OGSA-DAI data service.

Basically, all DAI-RDF activities support following input/output types.

- URI reference: The identifier to the data to be processed.
- Data Stream: Piped data stream itself.

### 3. ACTIVITY FRAMEWORK WITH OTHER OGSA-DAI PROVIDED ACTIVITIES

OGSA-DAI activity framework provides to construct a complex data processing by combining several data processing activities. Combining OGSA-DAI data conversion and transfer activities can combine OGSA-DAI-RDF activities with other OGSA-DAI activities so that it is easily possible to construct a distributed RDF data

processing. The perform document shown below is a simple example which sends the result of SPARQL query to another site.

```
<perform"
<documentation>
SPARQL query example
</documentation>
<sparqlQueryName="MyActivity">
<queryRequest rdb="/var/ogsadai/repository/
db.dat">
PREFIX db:http://db.x-try.de/rdf#
Select ?code where{
?employee db:code?code
?employee db:yoj:year"2007"
</queryRequest>
</queryResponseName="Jean2Output">
</sparqlQueryName>
<deliverToFile name="Shoe Result">
<from Localfrom="Jean2Output">
<toFile> /abc/results.txt </toFile>
</deliverToFile>
</perform>
```

### 4. SUMMARY AND FUTURE WORK

In this paper we have explained the use of OGSA-DAI and also its applications, which clearly show that OGSA-DAI can handle the heterogeneous databases and uses triple handling formats, which can be of great help in handling databases. Hence more of its framework and modification of RDF can help us in easy transformation and deliver of databases from one site to other using SPARQL as the language. The future work is on the modification of RDF with improved dependability and security integration as RDF will be the most important requirement of OGSA-DAI.

### REFERENCES

- [1] S. Tuecke, K. Czajkowski, I. Foster, J. Frey, S. Graham, C. Kesselman, D. Snelling, P. Vanderpilt, Open Grid Services Infrastructure, Version 1.0, <http://www.gridforum.org/ogsi-wg>, March 13, 2003.
- [2] Presentations and notes from the first OGSA-DAI users' meeting are available from: <http://www.ogsadai.org.uk/docs/docs.php#ug1>.
- [3] OGSA-DAI Open Grid Services Architecture Data-Access and Integration.

- [4] <http://jena.sourceforge.net/>
- [5] I.Foster, C. Kesselman, J. Nick, S. Tuecke. Grid Services for Distributed System Integration. *Computer*, 35(6), 2002.
- [6] W3C Document Object Model. See <http://www.w3.org/TR/DOM-Level-2-Core/core.html/>.
- [7] <http://www.openrdf.org/>
- [8] GGF DAIS WG, Grid Data Service Specification, July 2003
- [9] Global Grid Forum. <http://gridforum.org> 2003
- [10] Deelman E., Singh, G., Atkinson, M.P.,Chervenek, A., Chue Hong, N.P., Kesselman, C., Patil, S., Pearlman, L., Sui, M-H., Grid-Based Metadata Services. Submitted to 16th International Conference on Scientific andStatistical Database Management, June 2004.