

Grid-enable COBrA-CT

Yin Chen

Feb 2006

In the past few years, the grid has attracted enormous attention and gained popularity by supporting distributed resources sharing and aggregation across large scale, multiple administrative virtual organizations in science and business. Compare to the web, the grid offers upgraded performance in terms of security, fault tolerance, scalability, dynamic, heterogeneous, etc. By enabling COBrA-CT operation accessible through Grids, the software capabilities are expected to be enhanced greatly.

One of simplest ways to Grid-enable COBrA-CT is to assemble OGSA-DAI with CORrA-CT. Aimed to assist with access and integration of distributed data resources via the grid, OGSA-DAI provided various interfaces supporting data operation, transforming and delivering with many popular (relational or XML) databases, such as Oracle, DB2, SQL Server, MySQL, Xindice, eXist etc., and file systems, such as CSV, BinX, EMBL, OMIM etc. Wrapped by OGSA-DAI, a data resource can be deployed within a Grid environment, thereby provides a means for user to expose their web applications onto Grids. Projects that employ OGSA-DAI middleware including, myGrid, Biogrid, OGSA-WebDB, GEON, etc. appeared to be successful.

We propose to use OGSA-DAI middleware to glue XML database and Protégé client GO data viewer in COBrA-CT system. The reasons are as follows,

1. It is a simple method to add features of the grid to COBrA-CT, to improve the software capabilities in terms of scalability, security, high performance, etc.
2. OGSA-DAI provides interface to interact with XML database (Apache Xindice and eXist), data operations are straightforward.
3. It may reduce our development task greatly;
4. Reusing fully tested software could reduce test task of the new software and enhance code quality.

System Architecture

As shown in the Fig 1.1, the client API will be implemented which can be plug-in to Protégé. It responds for triggering OGSA-DAI *activities*, uploading OG data, and receiving return data sets.

OGSA-DAI *Activities* are the operations that a data service resource can perform, including data resource manipulation, data transformation and data delivery operations. The interactions of OGSA-DAI activities is illustrated by Fig 1.2 On receiving request from client, activities invoke Data Resource Accessor's (DRA) methods to connect with specific data resources, and return response message in a form of XML document.

Two type of XML data resources are supported by current OGSA-DAI release, Xindice and eXist. OGSA-DAI provide DRA to interact with these databases.

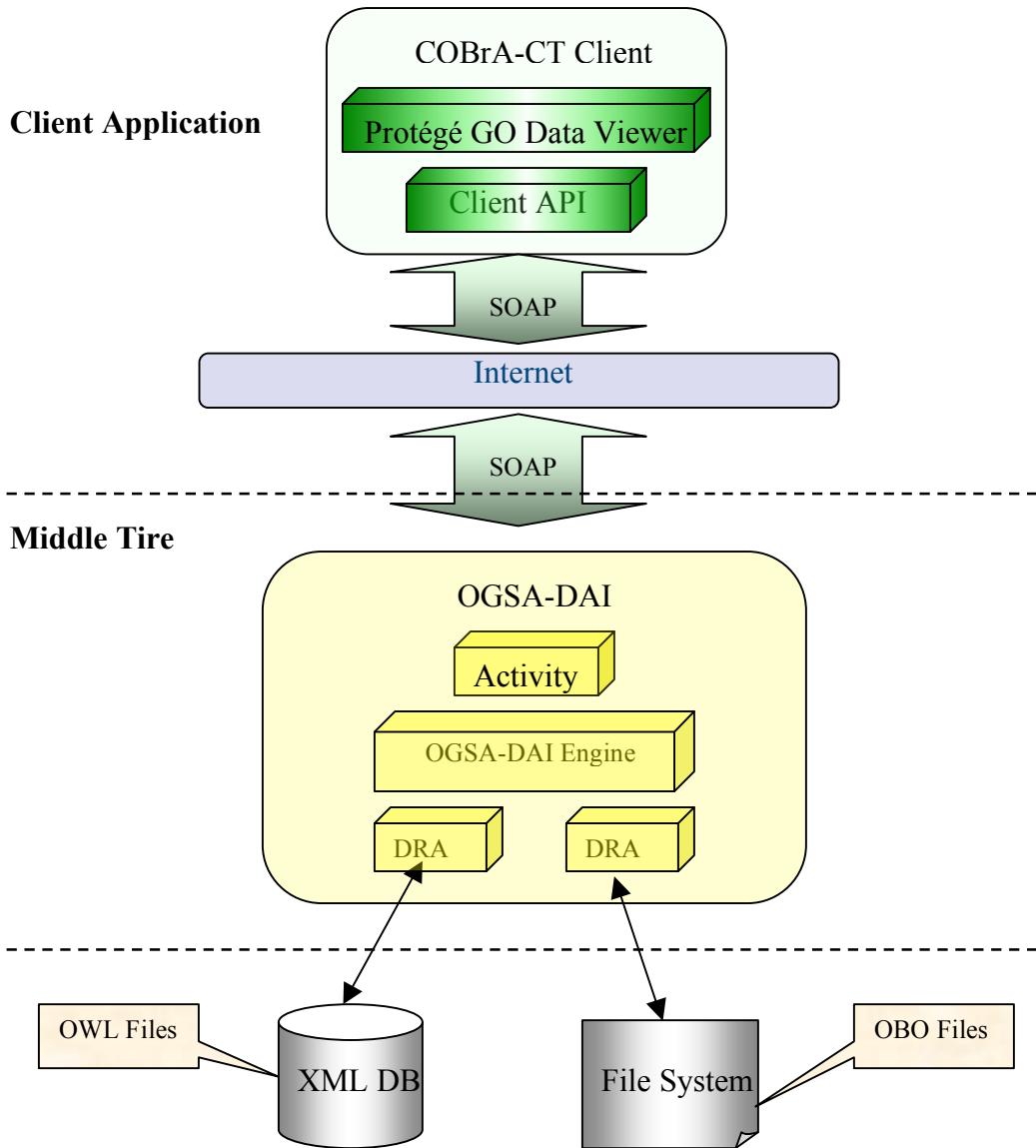


Fig 1.1 System Architecture

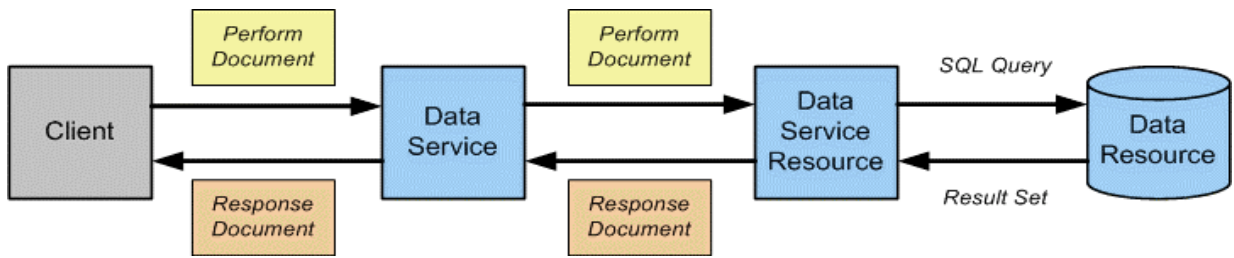


Fig 1.2 OGSA-DAI Activities Interactions [1]

Unfortunately, current version of OGSA-DAI doesn't support Berkeley XML Database. Following table shows parts of features of three XML databases products.

	Berkeley XML DB	Apache Xindice	eXist
Supporting Query Language	XQuery , XPath	XPath, XUpdate	XUpdate, XPath XQuery
Documents size	Large size	Small to medium sized	<i>Not mention</i>
Update	Document level	<ul style="list-style-type: none"> • Document-level • Node-level 	<ul style="list-style-type: none"> • Document-level • Node-level
Support File Types	Both XML and non-XML documents	Only XML documents	
Deployment	<ul style="list-style-type: none"> • Stand-alone DB • Java Library 	<ul style="list-style-type: none"> • Stand-alone DB • Java Library • Servlet 	<ul style="list-style-type: none"> • Stand-alone DB • Java Library • Servlet

Fig 1.3 Comparison of Berkeley XML DB, Apache Xindice and eXist

Initial investigation shows Xindice and eXist probably be able to serve our basic requirements, storing and loading OG data. However, we concern about capabilities of both products for handling large size of XML documents. To take the advantage of existing OGSA-DAI technologies, we suggest, at the first development stage, to employ Xindice or eXist XML Database. OGSA-DAI may have plans to improve XQuery capabilities in their release 9 in October 2006. We will push their action to support Berkeley XML DB interaction. If it happens, we expect the later tasks of switching to Berkeley XML DB would not be difficult.