Building Data Management Components for Business Applications

Craig THOMSON, Kostas KAVOUSSANAKIS, Arthur TREW
EPCC, The University Of Edinburgh, Edinburgh, EH9 3JZ, Scotland
Tel: +44 131 6505030, Fax: +44 131 6505555, Email: c.thomson@epcc.ed.ac.uk
kavousan@epcc.ed.ac.uk, a.trew@epcc.ed.ac.uk

Abstract: With the growing use of information technology-based tools across a broad spectrum of industries, it has become increasingly important to manage data effectively. These tools have also begun to offer the possibility of greater collaboration between organisations. This has led to new requirements to better share and integrate data. The BEinGRID[1] (Business Experiments in Grid) project has been set up to investigate the real requirements of industry as they use Grid software to develop their business. In the context of data management, the goal has been to enhance Grid middleware and bring its capabilities closer to those required by business. This paper describes some of the findings of this investigation and documents work already done to extend the OGSA-DAI middleware to meet the challenges identified in the business experiments.

Keywords: Grid, OGSA-DAI, BEinGRID, Data Management

1.   Introduction

Data management is important in a wide variety of business situations. High tech companies are processing increasingly large and complex datasets for diverse tasks like ship design and geological modelling. Companies are also increasing their reliance on electronic storage and processing of information in the day-to-day running of their businesses. Even small organisations will often use electronic order and invoicing software.

The BEinGRID project was set up to explore the technology gap between the current uses of European funded Grid middleware and the businesses that could be using it. The goal is to increase the uptake of Grid software. One of the main ways of achieving this aim is by enhancing existing Grid middleware to make it better suited to the needs of business.

This paper examines business use cases for data management and the requirements that came from them. It then details some of the enhancements undertaken as part of the BEinGRID project to extend the OGSA-DAI[2] middleware. The results of the project will be illustrated by a use case, which describes how the OGSA-DAI middleware and extensions will be deployed in a real business scenario.

2.   Objectives

The purpose of this paper is to highlight some of the innovative work done by the BEinGRID project in the area of data management. This involves both component-based extensions to existing Grid middleware as well as descriptions of underlying patterns which would be of use to businesses starting to look at Grid-based enhancements to their current business models. The overall objectives of the data management part of the project are:

- Analyse a number of business pilots to identify the gap between their needs and the capabilities of current Grid middleware
- Develop recommendations on the use of Grid middleware by business

- Extend and Enhance Grid middleware based on real problems

This work will help to make Grid middleware more relevant and accessible to a wider audience of business users. It aims to bridge the gap between the early adopters of Grid technology and the wider market of potential users. The results of the project will provide benefits for companies with data management problems in a broad range of business sectors and allows for greater exploitation of the research results developed within the EU.

3.   Methodology

An initial group of Business Experiments have been analysed from the point of view of data management. The project took the approach of selecting a number of business experiments to form the basis of the requirements capture. Though it was not possible to know at the time of selection how representative these experiments would be of the business landscape, they were chosen to cover a broad range of industrial sectors and areas of interest.

The experiments produced requirements, documentation and designs, which outlined the work they intended to do to achieve their business objectives. These documents were then analysed to extract the requirements that appeared more than once.

These common technical requirements were then refined and patterns were developed to describe useful generic behaviours that could address the requirements. The intention of these patterns was to make them generic enough to be applicable to more than one business sector, but specific enough to relate directly to Grid techniques.

A number of components have been developed to implement these design patterns. The components have been developed to extend the OGSA-DAI middleware. The choice of middleware to extend was based on in house expertise as well as on the analysis of the middleware being used by the business experiments.

Finally a second wave of experiments is being used to validate the utility of the components. The experiments will use the components produced during the project and provide feedback to motivate further development that is continuing in parallel.

4.   Requirements and Use Cases

The first step in producing useful Grid middleware extensions was to identify the important data management requirements and use cases. Eighteen Business Experiments were analysed and they produced a number of requirements. These included:
- Ease of use of middleware
- Allowing middleware to interface with existing software
- The ability to react to changes in a database

One of the experiments which provides a motivation for why a number of these requirements is useful is Business Experiment 24 – GRID2(B2B). This experiment is part of the second wave and aims to develop an extension to existing Business-to-Business (B2B) software.

The Experiment will allow B2B platforms to significantly evolve from the current state-of-the art. Currently data and process synchronisation between the participants of the B2B network requires a human operator, logging in on a portal or generating and processing files that represent supply-chain activities. What is missing is an affordable B2B platform extension to automate this synchronisation. While bigger companies can adopt new software, SMEs can only afford synchronisation if they can retain their original (legacy) infrastructure.

The technical partners of the project are EPCC, CINECA and Joinet. Joinet is the developer of the MaNeM[3] B2B platform which is being used in this experiment. In addition three of Joinet's customers are part of the consortium. Ducati, who make sports motorcycles, and PM and Bentivogli, two of Ducati's suppliers.

MaNeM is used to manage the flow of information between partners in a B2B network. Different legacy software exists to perform supply-chain operations inside each company. The B2B platform provides workflows that manage the interaction between the companies but data exists in parallel in the legacy systems and the same information has to be inputted twice. This is achieved either through the use of a custom script, which is run manually or by data entry by an employee at the company.

The goal of the experiment is to produce a stand alone extension to B2B platforms (not just MaNeM, but potentially others). This extension will allow information changes in one system to automatically update the B2B platform and other legacy systems.

To achieve the goal, a new application is being developed. To succeed it needs to:
- Be flexible
  - The intention is to market the solution to multiple B2B platform vendors
- Interface with existing legacy applications
  - One of the key benefits to the SMEs in B2B networks is that they keep their legacy systems
- React to changes in the legacy system data
  - The information from the legacy system should be automatically propagated to the B2B platform
- Be easy to set up
  - The IT providers of the SMEs involved in a B2B should not have to spend a lot of time deploying the software

These requirements are critical to the success of the experiment and match well with requirements that came from the first phase of business experiments.

For example one first wave experiment was developing an extension to a point of sale application for pizza shops. One of their requirements was that the extension would be easy to install by the technicians who maintained the existing point of sale infrastructure.

In another case there was a requirement to be able to react to updates to data. The particular example here was a travel agent and a tour operator. The travel agent might book a tour with a customer. They would then have to manually enter customer information once for their own records and again to add the customer to the tour operators system. The desire was to have an automatic method of keeping the customer information synchronised.

5. Design Patterns

After examining the use cases and requirements of the business experiments a number of design patterns were identified. They describe generic architectures, which can be used to solve some of the data management problems that emerged from the analysis of the business experiments.

The first two design patterns implemented as components are the Data Source Publisher and the Primary-Secondary Replicator.

5.1 Data Source Publisher

This pattern describes a mechanism for allowing data to be made available for access at another location. It also provides a layer that can be used to translate to or abstract the data type. The goal is to allow an existing system to be grid enabled so that it can be accessed via other grid middleware components.

The pattern works by adding a component, which communicates with the existing data source. This component provides another interface, which allows the information to be accessed remotely. The intention is that this pattern allows any existing applications to use their existing procedures to access data.

5.2   *Primary-Secondary Replicator*

This pattern allows a backup of a data source to be prepared and made available if the primary data source fails. This allows a more robust system: if one machine goes down for some reason, the secondary replicas can continue to provide a service.

The underlying idea is identifying a change in one database and reacting to it. This very generic pattern describes an event-based reaction to a change in data. Like the data source publisher, this pattern also allows for an interaction with an existing system. Actions which affect the data source can be monitored and actions taken which communicate with other remote systems.

These design patterns provide further motivation to develop these components. They describe a broader setting to help ensure that the enhancements made to existing middleware through the development of new components are done in a generic way.

6.   Component Development

Once the design patterns had been established the next step was to produce components that provide concrete implementations of the patterns. The first two components have been developed. They provide enhancements to the OGSA-DAI middleware. Another requirement for these components was that they be general purpose and flexible.

The components meet the requirements identified by analyzing the first phase of business experiments and contribute to the solution being developed for the GRID2(B2B) experiment.

6.1   *Technology Background*

The base on which the components have been developed is the OGSA-DAI middleware. The Open Grid Services Architecture - Data Access and Integration (OGSA-DAI) project, currently funded as part of the Open Middleware Infrastructure Institute UK (OMII-UK), aims to provide the e-Science community with a middleware solution to provide access to and integration of data for applications working across administrative domains. OGSA-DAI offers services that add data access and integration capabilities to the core functionality of service-oriented Grids. Structured data resources, whether these are databases, files, or other types of data, can be made available to Grid applications.

OGSA-DAI provides access to a variety of different database types and allows data to be published via a web service interface. It also contains a variety of activities, which allow data access, transformation and delivery. These activities are also an extension point of the middleware so that it can be customised to meet the particular requirements of a given project.

6.2   *Data Source Publisher Component*

A key requirement for the uptake of a new technology is the ease of adoption. Many companies whose core business is not information technology are turning to it to help improve their competitiveness and efficiency. One way to help them do this is to make it easy to use the new technology. The goal of the Data Source Publisher is not to extend the existing OGSA-DAI functionality, which already implements the underlying Data Source Publisher design pattern. Instead it automates the deployment procedure. This addresses the ease of use requirement of the GRID2(B2B) experiment and will be used to reduce the complexity of installation.

The Data Source Publisher provides a simple, GUI based installer, which deploys OGSA-DAI and publishes a data source via web services. The real benefit over following the instructions in the user guide is not a technical one. It is much more convenient and requires much less effort on the part of the person installing the middleware if everything

they need is bundled together and can be installed in a few simple steps. In order to install OGSA-DAI and deploy it on a computer you will need to download and install correct versions of its pre-requisites as well as database drivers.

By using the Data Source Publisher you reduce these requirements and simplify the process to install OGSA-DAI. You no longer need to use a number of command line tools, everything is handled inside a GUI installer. You have one download which contains the correct versions of all the software required to set up OGSA-DAI. All you have to do is configure the component for your application.

6.3 OGSA-DAI Triggers

The Primary-Secondary replicator pattern was defined in the analysis of business experiments. Replication is already handled natively inside many relational databases. There are limitations however when trying to move information between databases developed by different vendors. The more general idea of replication is reacting to a change in a database and performing an action that affects something else (another database for example). The OGSA-DAI Trigger component seeks to enhance OGSA-DAI by providing a mechanism for an OGSA-DAI workflow to be executed when a database is modified.

By providing a general mechanism for reacting to a change in a database, the OGSA-DAI Trigger component allows all the database access, transformation, and data delivery activities of OGSA-DAI to be used in response to a database change. An OGSA-DAI workflow can be executed automatically whenever a relational database changes.

This component will be used to help interface with different B2B platforms in GRID2(B2B). It provides access to a flexible grid middleware for data management and a mechanism, which allows the B2B extension to react to changes in the legacy system.
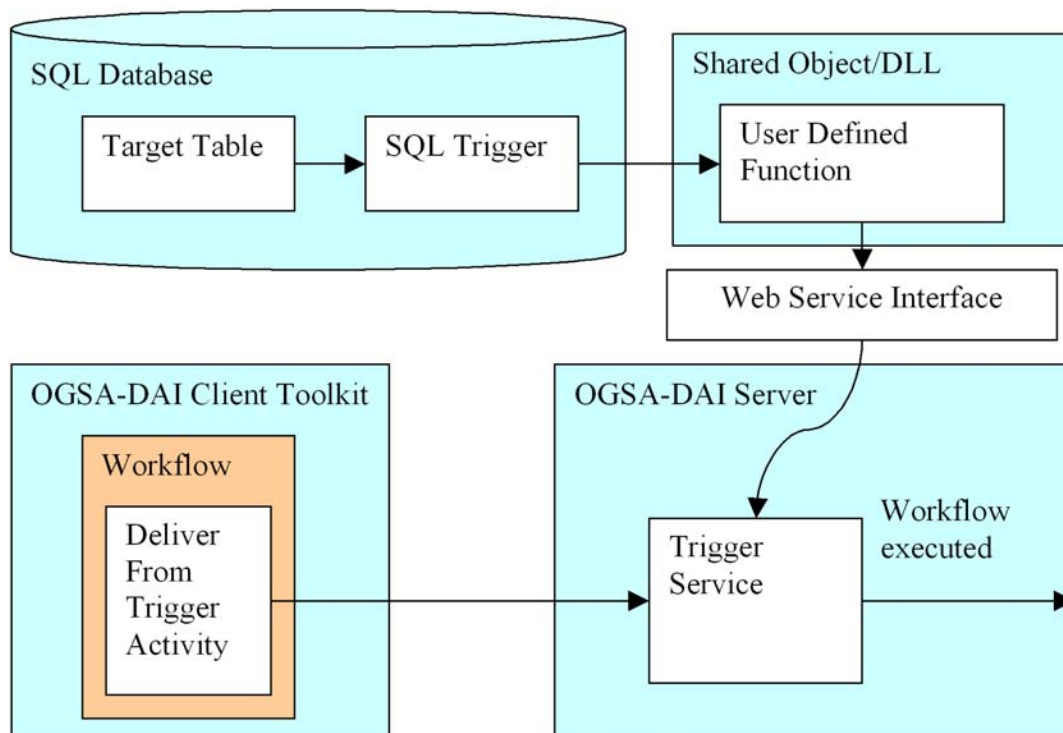
The design of the component is made up of a number of parts:
- An SQL Trigger
- A User Defined Function
- A Web Service Trigger Event Interface
- An OGSA-DAI Service and associated Activities

They provide a mechanism for notifying OGSA-DAI that a database has changed, and allow OGSA-DAI workflows to be stored and executed when the changes occur.

6.3.1   SQL Trigger

The first step in notifying OGSA-DAI that a database event has occurred is to be able to execute an operation whenever the database changes. SQL offers this in the form of triggers. Triggers operate on tables and allow SQL statements to be run when items are added to, updated in or deleted from a table in a database.

In order to minimise the chance of errors and to make it as easy as possible to deploy, the component builds a script that contains the SQL commands required to set up the trigger.

6.3.2   User Defined Function

A conventional SQL Trigger will allow you to run standard SQL commands. No built in command will let us notify OGSA-DAI of the database changes. SQL offers an extension mechanism for its database through user defined functions. This user defined function can then be added to a table and used like the standard set of functions (like SUM() for example). A function has been written to notify a trigger web service that something has changed in the database. It also communicates the values of the changed row. This user defined function is called from the custom trigger.

6.3.3   OGSA-DAI Trigger Web Service

In order to tie in the mechanism with OGSA-DAI we need a target for the trigger information and a place to retrieve it from. In order to do this we need a new Web Service. The UDF can call it directly to notify OGSA-DAI when a table row has changed.

This web service interface also allows standard OGSA-DAI workflows to be submitted by the component user. They are stored along with an identifier for the trigger they should be associated with. When a trigger fires and a notification is sent, any matching workflows are automatically executed.

6.3.4   OGSA-DAI Trigger Activities

The event trigger workflow requires a method of accessing the row information that has been passed into the web service. In order to do this a DeliverFromTrigger activity has been written which presents the row information in an OGSA-DAI format. This activity may form part of the workflow submitted to the web service interface.

7.   Results

The outcomes of the initial phase of development are:
- Identification of common requirements from initial Business Experiments
- Description of Design Patterns based on these requirements
- Development of 2 new components to help meet the requirements
- Validation of these components by a new Business Experiment

They provide an analysis of the gaps in existing Grid middleware and their application to a variety of business sectors. They also provide the first steps in closing these gaps and making data management Grid middleware more relevant to business users.

The following key, business-focused requirements have been identified:
- Ease of use
- Integration with legacy systems
- Reacting to changes in an existing database

The Data Source Publisher aims to address ease of use of Grid middleware by simplifying the installation process. The OGSA-DAI Trigger component assists in the integration of existing legacy software with middleware and provides a mechanism to react to changes in an existing database.

The requirements have already been validated by the selection of GRID2(B2B) which is a compelling example of the generality and utility of the components.

8.   Future Work

In addition to the components already produced as part of the experiment, further development is planned. More components will be produced which will further extend the capabilities of the OGSA-DAI middleware.

8.1   Query Translator

The Query Translator helps integrate the databases from different companies together. It does this by allowing the users to produce different views onto the same data. The underlying databases can remain different, but a common schema can be presented on top of the actual database. This allows a company to present a different view of its data to different customers without changing the application they use internally.

This component builds on top of OGSA-DAI. It will allow OGSA-DAI to present an SQL View on to a database that it exposes. This will allow two similar, but not identical, tables to present a common view of their data.

8.2   JDBC Driver Interface to OGSA-DAI

As we have seen, the requirement to interface with existing systems is important. One way of allowing this is to implement an interface, which the legacy application already supports. With the addition of a JDBC driver interface in front of OGSA-DAI it would be possible to very quickly provide data integration to an application, which already supports JDBC with few changes to that application.

The application could continue to use the same interface as before, but could now support integrated data sources through OGSA-DAI.

9.   Business Benefits

A new technology is only of benefit if there is a market for the product, and that product meets the needs of its intended users. In order to better explore the potential benefits of some of the data management components developed as part of BEinGRID, we will again look at the GRID2(B2B) use case.

The GRID2(B2B) experiment will allow the B2B platform provider to enhance its market share by providing a more advanced service. It will allow the manufacturer and suppliers to improve the communication between their organisations. Finally it presents an opportunity to diversify into selling the technology to other B2B providers. It will do this by extensive use of the components developed as part of the BEinGRID project.

The OGSA-DAI trigger component provides functionality that the current B2B software does not provide. It allows for the automatic update of information across organisational boundaries. This allows the B2B provider to give a more advanced service.

The trigger component also increases the speed and frequency with which information is exchanged. It automates the process and minimises the risk of human error. This allows for a more up to date picture of the state of the B2B collaboration. This has benefits in efficiency. In a manufacturing scenario, such as the one in GRID2(B2B) it is important to have up to date delivery information to ensure that the production line is optimally filled.

In order for the B2B platform extensions to be successful, the ICT suppliers as well as the B2B platform developer have to use it. In order to make adoption of the extensions easy, it is essential that the non-Grid software companies can easily deploy and adapt the software. The use of the Data Source Publisher will simplify the installation process.

10. Conclusions

This paper demonstrates the improvements that have been made to Grid-based middleware for data management as part of BEinGRID. It also shows the impact the resulting components can make in a real business scenario. The GRID2(B2B) experiment gives a business-focused use case to highlight the potential benefits of using Grid technology in a business setting. The integration of specific components, Data Source Publisher and OGSA-DAI Trigger, with the OGSA-DAI middleware demonstrates how innovative work can be translated into a real product.

It is intended that the results of BEinGRID be of use to businesses looking at Grid solutions for the first time. To aid knowledge transfer and improve the take-up of Grid Technology, this information will be made available via the Gridipedia website [4].

Acknowledgements

References

[1] BEinGRID, (2008). BEinGRID project information [online]. Available: http://www.beingrid.eu [accessed 11 July 2008]
[2] OGSA-DAI, (2008). OGSA-DAI information [online]. Available: http://www.ogsadai.org.uk [accessed 11 July 2008]
[3] MaNeM, (2008), MaNeM [online]. Available: http://www.joinetspa.com/supply-chain/prodottosoluzioni.cfm?wid_cat=14&wid_pro=8 [accessed 11 July 2008]
[4] Gridipedia: The European Grid Marketplace, (2008). Gridipedia repository [online]. Available: http://www.gridipedia.eu [accessed 11 July 2008]