Analyses on functional capabilities of BizTalk Server, Oracle BPEL Process Manger and WebSphere Process Server for applications in Grid middleware

R. Goranova
University of Sofia "St. Kliment Ohridski", Faculty of Mathematics and Informatics, Sofia, Bulgaria

Abstract. The problems which are considered in this article are related to the examination and the choice of appropriate service oriented application, which can be used for process orchestration in conditions of the Grid middleware g-Lite. Our vision for requirements to these applications is presented. On the base of shown characteristics and analyses on the functional capabilities we give some reasons for possible solutions of the described problems.

Introduction

The Grid technology is one new way of using computers, network and devices. It concept is based on Resources sharing in large scale, use of standard open protocols and delivering of qualities of services. Grid systems and applications aim to integrate and manage these resources and services within distributed, heterogeneous and dynamic Virtual Organisations. The virtual organizations (VO) [1] are sets of separate people, institutions and organizations, which share data, work on the same field (field of study or industrial field) and share common interests. Also the virtual organizations define the requirements, the conditions and the ways for access to the Grid.

For scientific tasks or computation of financial, bio-medical or physical tasks, the Grid is the most suitable technology, because it provides better effectiveness decreasing the time for computation. In practice it provides unlimited access to computational power and storage resource with only one user's login (single sign on) to the system.

Due to the heterogeneous nature of Grid environments and Grid applications the Grid architecture needs to cope with a large set of requirements. To standard this requirements, Global Grid Forum (GGF) – a community of users, developers, and vendors for grid computing standardization, developed OGSA.

The Open Grid Services Architecture (OGSA)[2] describes an architecture for a service-oriented grid computing environment for business and scientific use. OGSA is based on Web service technologies, WSDL and SOAP. OGSA assure interoperability on heterogeneous systems so that different types of resources can communicate and share information.

In their nature Web services are platform independent, language independent and provide reusable but usually simple functionality. The end user often needs more complex functionality, which can be achieved as composition of Web services. Often such kind of composition is called workflow.

For management and orchestration of the Web services workflow (process), the different Web services providers as IBM, BEA and Microsoft had different approaches. For example IBM used WSFL language for business process description and Microsoft used XLANG. In 2002, OASIS defined a specification called BPEL4WS, which combined graphical process orientation from WSFL and the structure process construction from XLANG, in a new BPEL standard.

The role of BPEL4WS standard is to define new Web services on the base of existing ones. It is a language for implementation of process orchestration, which is supported from the main Web services vendors as IBM, BEA, Oracle and Microsoft.

The Grid, as technology, is very close to the ideas of Service Oriented Architecture. Furthermore, the upper layers of the most Grid middleware architectures are service-oriented. The Grid has its futures as effective use of distributed resources and open protocols and standards.

Combined with the features of SOA, and possibilities for workflow (process) orchestration, the Grid will increase its usability.

The problem we are interested of is grid process orchestration in the grid middleware g-Lite. In this article we consider the choice of appropriate service oriented application, which can be used for process orchestration in conditions of the Grid middleware g-Lite. We also present our vision for requirements, which this application has to address.

Grid Projects and the Grid middleware g-Lite

The European DataGrid Project (EDG) [3] is three years running project, which started in 2001 and successfully ended in the end of March 2004. The aim of the project was development of Grid middleware for sharing data and computational resources for the needs of scientific European society. This project played basic role in development and deployment of new computational Grid middleware, called LCG for the biggest physical experiment built ever – the Large Hadron Collider (LHC) [4].

Discovering new fundamental particles and analysing their properties with the LHC accelerator is possible only through statistical analysis of the massive amounts of data gathered by the LHC detectors ATLAS, CMS, ALICE and LHCb, and detailed comparison with compute-intensive theoretical simulations.

LHC Computational Grid Project (LCG Project) [5] is a project, specially created to build and maintain data storage and analysis infrastructure for the entire high energy physics community that will use the LHC accelerator. Based on EDG software components, Virtual Data Toolkit (VDT) and Globus Toolkit 2.4 components, the new Grid middleware LCG was created.

The Enabling Grids for E-SciencE (EGEE) is two years running project, started in the beginning of April 2004 and successfully ended in the end of March 2006. The aim of this project is to develop and deploy new Grid infrastructure for scientific research. In this project participated 12 federations and over 70 institutions, including Bulgaria and University of Sofia”St. Kliment Ohridski”. The project has two priority scientific directions, serving the needs of Bio-medical experiments and needs of LHC. The EGEE project uses for its computational and data storage environment, the Grid middleware g-Lite [6].

The EGEE middleware (g-Lite) architecture is defined as Service Oriented Architecture (SOA). The main reason is that SOA facilitates interoperability among Grid services and also allows easier integration with new incoming standards as OGSA.

The g-Lite environment was developed as result of experience of the previous projects EDG Project and LCG Project, and combination of the service oriented concepts with already developed software components as LCG grid middleware and EDG middleware.

The g-Lite middleware consists of software elements and services for resource brokering, monitoring, data storage and computing. According to EGEE architecture, g-Lite services are grouped into logical groups. The most basic of them are Security Services, Information and Monitoring Services, Data Services and Job Management Services

Security Services comprises the Authentication, Authorization and Auditing services, which are responsible for identification process and access to resources.

Information and Monitoring Services provide a mechanism for publishing and consuming information used also and for monitoring purposes.

Job Management Services includes the main services related to job management as computing element services, workload management, package manager services and accounting.

Data Services provide a mechanism for data and file access based on storage element service, catalog service and data movement services.

The databases which are used in g-Lite environment, for storing data and accounting information, are MySQL and Oracle. At the moment the environment is supported for Scientific Linux 3 (CERN), Scientific Linux 4 (CERN) and some distributions of Debian Linux operating systems.

Grid process and Grid processes orchestration

Before to define, what we understand behind the terms Grid processes and Grid process orchestration, we will give a short description of the terms process and process orchestration and WS-BPEL standard, in the way IBM, Microsoft and other software providers use them.

Web service business process execution language (WS-BPEL) is an XML based programming language for description of high level business processes. WS-BPEL provides methods for processes description and web services interaction. BPEL is the industry standard for orchestration, supported from IBM, Microsoft, BEA and Oracle.

Service Oriented Architecture (SOA) is component based architecture. It defines methods for integration of application and processes, as connected services. The services are self-contained and do not depend on the context or state of the other service. The key concept of SOA is a loose coupling between interacting services, which communicate with each other through well defined interfaces and protocols.

Business Process [7] in the terms of service oriented architecture is any system or procedure that an organization uses to achieve a large business goal. It is composed from in individual tasks, which are executed in a specific order. The business process is a sequence of actions, which together meet the needs of the business.

Business process choreography [7] is the capability to bring together individual services to accomplish a larger piece of work. There are two aspects to business process choreography: the model and the runtime. The model is a description of the steps and logic required to complete the business process. The runtime (engines), are responsible for tracking the state of the process instances, handling human intervention, dealing with exceptions, and routing messages to the appropriate instance of the process. The business process orchestration is arrangement of services in end-to-end business process.

As we consider, the definitions above and the nature of the Grid, we can define Grid process as any system or procedure that a virtual organization uses to achieve a specific goal. The Grid process is a strong specified sequence of Grid services. Grid process choreography is the process of modeling, engineering and executing of Grid processes or their instances.

The Grid service is a web service that conforms to a set of conventions (interfaces and behaviors) that define how a client interacts with it. The Open Grid Services Infrastructure (OGSI), created by Global Grid Forum (GGF) is a standard that, defines mechanisms for creation, management, and exchange of information among Grid services.

As we mentioned above OGSA is service oriented architecture, which defined requirements to grid services for grid middleware. Based on OGSA, OGSI defines WS-Resource specifications for grid services. In general they are: WS-Resource Lifetime, WS-Resource Properties, WS-Resource References, Service Group and Base Faults. Web services, which implements these interfaces, complete the OGSA requirements for grid services.

Let consider grid service FTS, form g-Lite grid middleware, for example. The File Transfer Service (FTS) [8] is the lowest-level data movement service defined in the gLite architecture. It is responsible for moving sets of files from one site (cluster) to another. It is designed for point to point movement of physical files.

The FTS service has interface *FileTransfer* and provides function for: job submission (*submit*), job cancellation (*cancel*), gets job status (*getTransferJobStatus*), gets file status (*getFileStatus*) and etc.

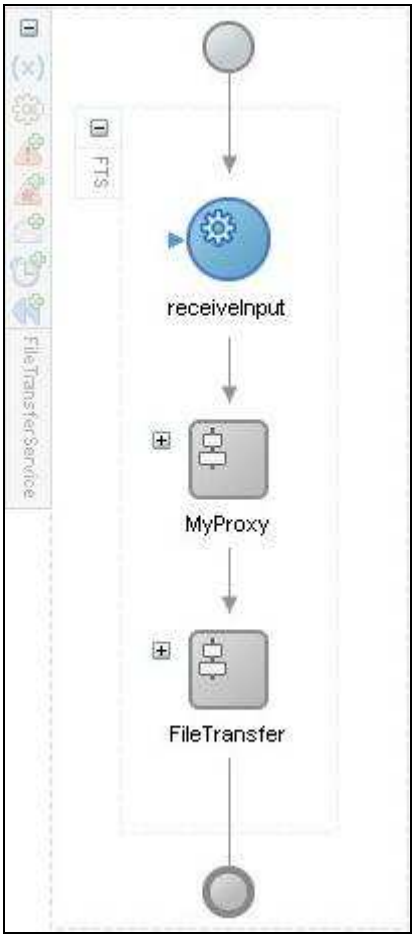
In the terms of FTS, a job is transfer job and it consists of a set of files to be transferred. Basically a job contains of a list of Files and a MyProxy password. Before submitting a job, the client is expected to upload an appropriate credential (user grid certificate) into the same MyProxy server used by the FTS.

If we describe this grid process in terms of BPEL4WS it will look like this:

We can consider this grid process as synchronous composition of the following sub processes *MyProxy* and *FileTransfer*.

In sub process MyProxy, grid user proxy is created, invoking *MyProxy.create()* service.

In sub process FileTransfer, job transfer is submitted, by invoking services *setTransferJobElements*, *setCredential* and *FileTransfer.submit*;



Requirements

On this stage of our research we can not say for sure if every application which possesses capabilities for processes orchestration will be appropriate one and for Grid processes orchestration, but we could determine, the requirements to these applications.

We believe that, first of all, the application, we search has to meet advantages of *service oriented architecture* [9]. Service oriented architecture provides us the ability to develop and modify integration applications dynamically. Service-oriented applications have advantages as reusable functionality, platform-, language- independency, and operating interoperability. Also service-oriented applications are more flexible and promote loose coupling between their components so that they can be reused.

In second place, the application has to meet requirements of *Grid application* [10]. This requirement comes natural from that, the application will be placed in Grid environment. Basically, a Grid application is a collection of work items or jobs that carry out a complex computing task by using Grid resources. A Grid-enabled application runs in a dynamic, sometimes loosely defined, and heavily networked environment. So a candidate application has to possess ability to be adapted for Grid infrastructure and to possess possibilities to handle with services for Grid security infrastructure, Grid job flow, Grid resource management and Grid data transfer.

As we consider, requirements above, we could say that we are looking for application, that is service-oriented, that provides functionality for process orchestration that is simple enough to be integrated in Grid environment that is platform independent, that requires less hardware resources and that can be used for free or under production license.

Instruments for process orchestration

During examination of appropriate product for process orchestration we focused our efforts on the following process servers: WebSphere Process Server, BizTalk Server and Oracle BPEL Process Manager. We chose exactly these products, because they are one of the leading tools for business process orchestration at the moment. They are built in different technologies, but used common standard WS-BPEL.

IBM Web Sphere Process Server

IBM Web Sphere Process Server [11] is a deployment and runtime environment for business integration applications. It enables the deployment of applications built in a service-oriented architecture structure. It provides capabilities for services and processes orchestration.

WebSphere Process Server has multi layered and complex architecture. It is built on a J2EE runtime provided by WebSphere Application Server. On top of it, WebSphere Process Server implements a layer called the SOA core. The next two layers implement a number of components and services that can be used in an integration solution. The product, provide many functionalities as business process management, business object maps and human tasks. With installation of process server is provided also powerful graphical tool called WebSphere Integration Developer for process creation, developing and deploy. It supports the WS-BPEL standard, import from BPEL file and export to BPEL file. WebSphere Process Server is supported for the most known distribution of Windows, Linux, Solaris and AIX operating systems. It requires minimal 2 GB RAM to function normally and around 6 GB HDD. Full version of the product can be downloaded from the site of IBM, under special conditions of an academic program, otherwise for business development or production purpose, the product have to be purchased.

We have to mention that, our tries to install the product on Scientific Linux 3 and Scientific Linux 4 operation system was unsuccessful. The reason is that this Linux distribution is not supported.

BizTalk Server

BizTalk Server is Microsoft product for business process orchestration. It is supported for all Windows platform. The main requirements for installation of the server are software as Microsoft Visual Studio 2005, .Net and Microsoft SQL Server 2005.

BizTalk Server engine is composed of components for messaging and orchestration. The messaging component gives opportunities for communication with outer software. The orchestration is responsible for creation and management of graphical defined processes. With installation of the product is provided also a tool for development and deployment of business processes called Enterprise Application Integration and tools for XML creation (BizTalk Editor), XML mapping (BizTalk Mapper) and process specification (Orchestra Designer).

BizTalk Server can be downloaded from Microsoft site, as a trial version, and can be used for two months. It requires minimal 1 GB RAM and 6 GB HDD space.

Although it seems that BizTalk Server is not an appropriate application because it is a Microsoft product and basically is not compatible with Linux, it is an object of our consideration because of the concept for web services interoperability and the BPEL4WS common standard. In theory, every BPEL process can be deployed on every process server engine, which supports the BPEL standard.

The BizTalk Server was successfully installed on Windows 2003 Server, but not tested yet with g-Lite grid processes.

Oracle BPEL Process Manager

The Oracle BPEL Process Manager enables enterprises to model, deploy and manage BPEL processes. It comprises an easy-to-use BPEL modeler, a scalable BPEL engine, an extensible WSDL binding framework, a monitoring console and a set of built-in integration services (transformation, user task, java embedding).

The Oracle BPEL Process Manager differs from other process integration products, in that it provides native and comprehensive BPEL support. It is simpler and has cross-platform support (can be used on the top of Oracle Application Server but also WebLogic, WebSphere and JBoss application servers). BPEL Process Manager runs on any J2EE servers. It requires minimal 512 MB RAM and 300 MB HDD space for installation. The product can be downloaded from Oracle site, after registration and accepted license agreement and can not be used for production or business purposes. With installation of the product is provided only Server and Console tools. For process design and development, separate products have to be installed, as JDeveloper, Eclipse and Application Server.

We have to mention that the product was successfully installed on Windows XP and Scientific Linux 3 operation systems. JDeveloper tool was used for grid process definition. During installation of Scientific Linux operation system we found that, basic requirements for installation of the product on Linux is availability of graphical environment. Also additional database have to be installed.

Analyses on functional capabilities

The table bellow, (Table 1) shows a brief summery for presented process servers. We have to mention that, the collected information is for versions of WebSphere Process Server 6.0, BizTalk Server 2006 and Oracle BPEL Process Manager 10g.

Table 1. Process servers' characteristics

Process Servers	Platform	License (Trial, Free)	Software requirements	Hardware requirements	Provided functionality
Websphere Process Server	AIX, Solaris, Windows, Linux	Under Academic Initiative(not for production and business purpose)	JDK	RAM: 2GB HDD: 6 GB	Integration Developer, Application Server
BizTalk Server	Windows	Trial (2 months)	.Net, MS VS 2005, MS SQL Server 2005	RAM: 1 GB HDD: 6 GB	BizTalk Editor, Mapper, Designer
Oracle BPEL Process Manager	Linux, Windows, Solaris	License agreement (not for production and business purpose)	JDK, JDeveloper, Eclipse	RAM: 512 MB HDD: 300 MB	Server Console

During our examination we have been faced to two main problems: platform independency and license agreements. All of the reviewed products provide distribution for specific operating systems, mainly for Enterprise Red Hat Linux, Suse Linux, Window and some distributions of AIX and Solaris. All of them are accessible and can be used under specific license agreements or evaluation copies. We can say that, all of them answer the requirements of the service-oriented architecture and functionality for process orchestration. But only Oracle BPEL Process Manager is simple enough in the sense of component system and requires less hardware resources.

Conclusion

In this paper we made a brief overview of some of the most reliable process servers. As a result of above characteristics and presented functionality we can conclude that Oracle BPEL Process Manager, answer to the most requirements we placed. In the result of defined problems in this article,

object of further research will be examination how in practice this process server can be integrated into Scientific Linux operating system, and also what database can be used with it. If integration is successful, we can continue with searching solution for license problem. It is possible old version of the product (as oracle BPEL Process Manager 2.0) to be downloaded for free and can be used in production environment.

Reference

[1] I. Foster, C. Kesselman, S. Tuecke. The Anatomy of the Grid: Enabling Scalable Virtual Organizations, 2001;
[2] I. Foster, C. Kesselman, S. Tuecke. The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration, 2002;
[3] EDG [http://eu-dataGrid.web.cern.ch/eu-dataGrid/]
[4] The Large Hadron Collider [http://lhc.web.cern.ch/lhc]
[5] LCG [http://lcg.web.cern.ch/LCG/overview.html]
[6] g-Lite middleware [http://cern.ch/glite/]
[7] IBM, Information Integration for BPEL on WebSphere Process Server, 2005;
[8] Gavin McCance. EGEE gLite User's Guide: g-Lite File Transfer Service - JAVA, 2005
[9] R. Kodali. What is service-oriented architecture? An introduction to SOA, 2005;
[10] B. Jacob. IBM - Design an application for Grid, 2003;
[11] A. Goering. WebSphere Process Server made easy, Part 1: Architecture. Application Server lays the foundation, IBM, 2006;