

A Virtual Distributed Database Model for Creating a Database Federation

Jaspreet Singh Pahwa
J.S.Pahwa@cs.cardiff.ac.uk

Abstract. A typical AEC (Architecture / Engineering / Construction) industry project involves many individuals and companies forming a consortium for the duration of a project. This paper describes a virtual distributed database model called the MDSS (Multiple Database Search Service) Federation with Grid enabled database search for meeting the data needs of consortium members. The Product Class Database (PCD) System, the Supplier Database (SD) System and the Grid enabled MDSS System [4] are important components of the MDSS Federation. The MDSS System is based on the Open Grid Services Architecture (OGSA) [1] model and is being built using Globus Toolkit 3 (core).

1 Introduction

The MDSS Federation is based on the concept of information sharing. Commercial organisations share information about the products and services they supply using various channels of media such as television, radio, yellow pages etc. to reach a wider user base. The MDSS Federation aims to enable sharing of such product information using Product Class data structure. The MDSS System is essentially a search paradigm that supports database search operations based on predefined criteria. It also supports data access mechanism in a virtual distributed database where participating component SD Systems are part of the MDSS Federation. Product suppliers provide information about products in their SD Systems using product Class templates and advertise them to Virtual Organisations (VOs) by participating in the MDSS Federation where they are searched by the MDSS System for consortia members based on the criteria such as product specification, delivery time, cost and availability.

2 The MDSS Federation

The Federated Database System (FDBS) is a collection of cooperating but autonomous component Database Systems (DBSs) [3]. The FDBS provides an integrated and transparent mechanism to access information from a number of component DBSs for different class of federation users. However, traditional FDBS architectures lack in functionality and scope to address the needs of emerging business models for collaborative working. The MDSS Federation is an effort in this direction. The MDSS Federation is a channel for product suppliers to advertise products and services to the members of a VO. The approach we aim to adopt in the MDSS Federation for accessing information from a large number

of component DBS is different from the approach of traditional FDBS Architectures. Schema integration is an important aspect of FDBS Systems. The Grid enabled MDSS System is presently being designed to access information from a large number of component SD Systems without the need of forming external and federated schemas. Participating component DBSs in the MDSS Federation provide a list of operations which can be invoked by the VOs via XML based Web Service technology [2]. In this way, the local autonomy of SD System is maintained whilst participating in the federation.

3 Semantic Heterogeneity

Whilst creating an FDBS, detecting semantic heterogeneity is a difficult problem as the schemas of participating Component Data Models (CDMs) do not provide enough semantics to interpret data consistently [3]. Differences in data models lead to semantic heterogeneity. The SD schema conforming to the PCD Schema provides a consistent data model across all product suppliers. The PCD System aims to address the issue of semantic heterogeneity by explicitly defining specifications for each product attribute. For example in the PCD system, a "Window" Product Class can have specifications (properties) such as width, height, weight, wood type, panel shape etc [4]. Specifications are created whilst creating a Product Class. Since a Product Class is a consistent entity, it serves as a mechanism for creating complex products with a large number of specifications conforming to different specification types whilst maintaining data semantics. A Product class is subscribed to by large number of product suppliers for creating products in their SD Systems. Therefore consistent semantic interpretation of product data is possible with the aid of Product Classes in the MDSS Federation. We also aim to provide versioning support to Product Class for creating existing products with enhanced features and functionality. For building the MDSS Federation, prototype PCD and MDSS Systems have been created which are presently undergoing a testing phase.

References

1. I. Foster, C. Kesselman, J. Nick, S. Tuecke. *The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration*. Open Grid Service Infrastructure WG, Global Grid Forum, June 22, 2002, <http://www.globus.org/research/papers.html>
2. Graham, S., Simeonov, S., Boubez, T., Davis, D., Daniels, G., Nakamura, Y., Neyama, R., *Building Web Services with Java: Making sense of XML, SOAP, WSDL, and UDD*. Sams Publishing 2002.
3. Sheth, A.P., Larson, J.A., Federated database systems for managing distributed, heterogeneous, and autonomous databases, September 1990 ACM Computing Surveys (CSUR), Volume 22 Issue 3, Special issue on heterogeneous databases, Pages: 183 - 236, Year of Publication: 1990 ,ISSN:0360-0300
4. Pahwa J.S., Burnap P., Joita L., Gray W.A., Rana O.F., Miles J., Creating a Virtual Distributed Database-Data Definition and Search Model for Collaborative Virtual Teams in the Construction Industry. Paper accepted in the Application Track stream of 21st Annual British National Conference on Databases. <http://www.macs.hw.ac.uk/BNCOD21/>