Results, Conclusion and Future Extension of Research Work

Content

6.1 Major Contributions of the Research Work ....................................................... 167
6.2 Features and Benefits of the Research Work .................................................. 169
6.3 Future Extension of the Research Work ....................................................... 170
Chapter 6

Results, Conclusion and Future Extension of Research Work

This chapter discusses the results, conclusion and future extension of the research work. It discusses the major contributions drawn out of the research work. It also covers the features and benefits of the research work in detail.

6.1 Major Contributions of the Research Work

The work presented in this thesis provides an insight into the world of building a dynamic and scalable architecture for a multi-agent knowledge-based system accessing distributed database grid which is composed of both data grid technology and agent oriented computing.

Data grid and agent technologies are one of the best combinations to build dynamic and scalable systems. Data grid computing is a specialized case of a distributed system. It is characterized by geographically diverse & heterogeneous data resources. These resources are under the control of different administrative domains with different policies and levels of security. Data grid computing is highly dynamic in nature. Whereas, agent oriented computing has many features and characteristics that encourage us to adopt them in the development and set up of the distributed database grid. Some of the characteristics are autonomy, adaptability, proactivity, goal orientation, communication, collaboration and cooperation. The generic framework developed here exploits features of both data grid and agent technologies to create manageable, extensible and flexible architecture.

One of the major contributions of this research is to provide an alternate and novel approach to existing ETL (Extraction, Transformation and Loading) approach for data access & integration. Modern research applications analyze a large data which reside in data repositories. These data repositories are often owned by different organizations. Apart from this, data within business organizations generally exist in multiple and heterogeneous databases which are owned by different administrative domains. Therefore, both research and business organizations require these heterogeneous and geographically separated data on a real time basis for analysis and decision making. Volume of data, diversity of data sources
and data freshness requirements makes data access & integration a complex task. Data integration challenges have also been addressed by the ETL approach earlier. ETL approach comprises with three steps: extract, transform and loading. An ETL approach first extracts data from outside sources. It, then, transforms it to fit operational needs and finally, it loads it into the end target (can be a database, operational data store, data mart or data warehouse). However, the ETL approach fails to integrate the real time data due to the latency involved in cleaning, transforming and moving the data. Therefore, ETL approach is having its own limitations as they fail to integrate the data in real time for which organizations generally strive. These limitations of ETL are overcome by the developed framework as it accesses and integrates the data on a real time basis from geographically distributed database grid.

Another contribution of this research is to provide a generic framework which is a based on SOA (Service-oriented Architecture) platform for data integration within an organization in knowledge-based fashion. A service-oriented architecture is a set of principles and methodologies for designing and developing software in the form of interoperable services. The grid is said to be a distributed architecture developed in a service-oriented perspective. On the other hand, agents and multi-agent systems are said to be autonomous, intelligent and interactive entities that may use and provide services. Services offered by generic framework are wrapped up into agents and different collaborative task agents are working together to provide the core, user & knowledge level services to the client applications and users. To offer knowledge level services, knowledge-based component is implemented by incorporating fuzzy interface. Fuzzy logic representations are more intuitively satisfying than classical Boolean (bivalent) logic. Furthermore, it is more precise and compact compared to classical rule based representations. As, we have developed the generic framework and multi-agent system provides scalability and extensibility, new agents will be added and existing agents will be customized as per the requirements of the client applications and users.

Also, this research leads to the development of the Intelligent EIS (Enterprise Information System). Corporate sector, academic sector, R & D centers and industries such as manufacturing, banking, medicine, entertainment and service sectors generate large amounts of data, which are available in the structured form. In this scenario, such implicit and automatic information and knowledge management of the available databases incorporated in the given application has strong potentials. The whole solution leads us towards the Intelligent EIS and can be used for analysis and decision making on real time data. Full
automation of the generic framework can be carried out for commercialization and a claim can be made for the patent for that.

Apart from this, as a part of the research, an experimental system has been implemented to evaluate the benefits of integrated framework. The framework leads towards the EII (Enterprise Information Integration) which helps to shorten the time required to carry out data integration and reduces the data management costs and maintenance cost over time. Software agents have been implemented to be the best for handling distributed information issues like information retrieval, integrity and navigation assistance in a knowledge-based fashion.

The research goal of the experimental system is to provide a uniform query and browse mechanism for multiple distributed university databases, shield their differences in data schema or location, provide a multi-agent system for management of the existing databases in a knowledge-based fashion, knowledge-based access of the domain information to enhance the utilization of scarce resources, and provide a new solution for the research on the integration of the domain databases.

6.2 Features and Benefits of the Research Work

In this section, the features and benefits of the research work have been summarized. This research work:

- aims at integrating agent oriented approach and mechanisms that meet the requirements of data grid applications.
- offers a generic framework to develop multi-agent knowledge-based system accessing distributed database grid.
- provides a single point and transparent access through a standardized uniform interface to heterogeneous geographically distributed data resources like relational, XML and file system resources to client applications and users that hides underlying heterogeneity.
- provides the data grid service through multiple agents.
- provides an agent oriented interface to client applications and users to access the data grid services in a platform and language independent way by integrating agent and multi-agent system with the data grid.
- provides an open, scalable and extensible framework in which new functionalities can be added by adding specific services provided by means of agents.
• offers an interface to manage the data grid nodes to privileged users and automates the process of configuration and deployment of OGSA-DAI (Open Grid Services Architecture – Data Access and Integration) resources.

• provides on demand real time access, integration and delivery of heterogeneous data resources and eliminates the need of building data mart and data warehouse at a certain level.

• provides a single and a virtualized view of the enterprise data assets (digital objects) and allows data providers of enterprises to retain control of their data.

• enhances business intelligence and quickening decision making process by providing knowledge-based component through fuzzy set theory and fuzzy logic that leads to a novel model for scientific and business applications called Intelligent EIS (Enterprise Information System).

• provides a generic framework that is acceptable and supportable by existing database providers.

• provides a generic framework as a base for developing future applications like distributed data mining, web mining and more.

• offers flexibility to integrate any other artificial intelligent technique through the realization of an agent for knowledge-based access like neural networks, neuro-fuzzy approach, genetic algorithms and many more.

The generic framework is developed as an outcome of research which uses OGSA-DAI as a data grid middleware which is 100% JAVA and runs under Windows, UNIX, and Linux platform. The agents of the framework use web services and so client applications can be written in any language and on any platform that supports web services.

6.3 Future Extension of the Research Work

During the research work, as a case study, the university domain was taken and as an experimental system, students’ performance evaluation was implemented. With bases of that, the generic framework can be applied to implement different applications for a variety of domains like Digital Libraries, GIS (Geographical Information Systems), Medical Research & Development, Banking Systems and many more. Such applications generally deal with a large number of users which requires these applications to be large-scale, heterogeneous, distributed and dynamic. Grid and agent technologies play a major role in these types of
applications as they provide a significant enhancement to these applications to make them more scalable, dynamic and reliable.

Also, currently the focus of this research work is to implement data grid environment accessed by multi-agent system. But, this implementation can be extended by adding the capabilities of computational grid. Several grid middleware exists like Globus Toolkit, Unicore etc. We can use any of these middleware with OGSA-DAI to have the power of both types of grid: data grid and computational grid. Apart from this, the developed generic framework can be extended in several research fields like distributed data mining, knowledge grid and so forth.

The grid infrastructure is growing up very quickly and is going to be more complex and robust with advancement of technology. As mentioned earlier, the basic grid architecture offers core services and high level services would be built upon these core services. Therefore, the components for distributed data analysis and knowledge discovery process would be integrated as a part of higher level services which lead the current grid infrastructure towards the knowledge grid.

With the advent of Internet, nowadays we are dealing with an overwhelming amount of data as many scientific and commercial applications generate enormous amount of data and data mining is the process of discovering useful patterns in these data. The manual process of finding patterns in data would be excessively costly and may prone errors. Therefore, this process must be automatic, or at least, semi-automatic. Grid computing has emerged as a new technology which enables the integration of heterogeneous computing and data resources in order to provide a global computing space. Therefore, we may extend the generic framework by integrating various data mining strategies like association rule learning, classification, clustering, sequential pattern mining etc. which lead towards the novel architecture for distributed data mining in grid.

I hope that the research work carried out will helpful to the society and will lead to new discoveries in the field of computer science & applications.