AGENT-BASED ARCHITECTURE FOR ENTERPRISE-WIDE DATA INTEGRATION

Muhammad Ahmed, Muhammad Khalid Khan, Muhammad Taha Jilani, Maaz Bin Ahmed
College of Computing and Information Sciences, PAF - Karachi Institute of Economics and Technology, Karachi, Pakistan
Corresponding Author: mkhalidkhan@gmail.com

ABSTRACT: *Organizations are experiencing a rapid increase in their size all across the globe. This trend has made the departments to act as an autonomous entity. However, the data has to be shared to achieve organizational goals. This is Data Integration. An addition to this concept comes when the merger of different organizations takes place. Now the sharing of data comes under the heading of Enterprise Wide Data Integration. This is a more complex scenario, as the state and the sources of data are different which give rise to the problems such as inconsistency, incompatibility, etc. The toughest challenge that IT faces today is to form a single virtual representation of multiple data sources and the transformation of enterprise data into accurate, comprehensible, and actionable information for achieving competitive advantage of enterprise business. In this paper, we have proposed a Multi Agent-based Architecture for the Enterprise-wide Data Integration. We believe that use of agent technology provides not only robustness and efficiency, but also the intelligence that is essential for such a complex task.*

KEYWORDS: Agent based Architecture, Data Integration

INTRODUCTION

The recent advancement in information systems, architectures increases the demand of forming a unified view of enterprise data as they are spread over various heterogeneous data sources like relational databases, mainframes, different operating systems, hierarchical repositories, and websites. Most of the applications utilize data in some shared resources. The data span from legacy systems to the emerging web systems.

Data is what you run your business on and it is the ability to capture, dispense and utilize it at the right time and in the right ways to help your business excel. Data about customers, employees, products and more is the heart of the business. To integrate the data from heterogeneous data sources and to give it a unified single representation to the users; this is known as Data Integration. These data sources might be database systems, but more often are unconventional data sources such as structured files, legacy systems, or websites.

Let us discuss the benefits that can be achieved; the benefits can be tremendous and surface at multiple levels.

- Accelerated transformation to e-business by integrating data, information and technology assets.
- Creates a single view of a customer or other business entities.
- Better customer service with better informed staff, partners and employees through improved efficiency of information delivery and utilization.
- Provides an infrastructure for Enterprise Data Management.
- Takes full advantage of web and application server functionality needed for next generation systems, including interfaces, transactions and security.
- Interoperable through standards of past, present, and future, including COM, CORBA, EJB as well as ODBC, JDBC, ADO/OLEDB and XML.
- Hides complexity of data and information for both end-user and development, while maintaining security and control over data access.

As a result of changes in business circumstances, corporate Information Systems that are running independently are required to share data or information that flows across the business processes. The organizations that are formed due to mergers or business reorganizations [8], Due to this mergers or reorganizations the resultant will be an 'Enterprise'; and the data which flows across the enterprise is known as to be 'Enterprise Data'.

Organizations are usually divided into groups such as departments, divisions or cost centers that have considerable autonomy and different level of decision-making power. It is the main strength of an 'Enterprise' that these groups still work autonomously but with cohesive bound and overall single goal direction. There are many application architectures that are designed for Enterprise-wide Data Integration for coping with problems of inconsistency, heterogeneity among the diverse data sources that are to be integrated, and incompatibility among these data sources. Such as Federated Architecture, InfoSleuth and Knowledge Sifter.

In Federated Architecture for data integration [1] the whole enterprise is separated out into corporate domains, inter domain communications are carried out through using Federal Highway and this highway is implemented by using Message-Oriented-Middleware. There are no inter domain Informational Dependencies, but in each domain there may exist informational and processing dependencies. The Interaction between domain and federal highway is the responsibility of 'Interface Agent' (IA). IA decides which is the relevant message to pick-up from the highway and when to release message to the highway. IA keeps track of message ID and description of the source domain. Federated Architecture makes possible accessibility to data and preserves the necessary autonomy of the systems.

InfoSleuth is an agent-based system of data integration in a distributed environment with heterogeneous data sources [2]. In InfoSleuth system, there are groups of agents that collaborate with each other for achieving a complex task of forming a uniform view on data from different sources. The framework consists of Network Agent, Resource Agent, Matchmaker Agent and Control Agent. In InfoSleuth; agents communicate with each other by Knowledge Query and Manipulation Language (KQML). The users only need to know one uniform language to access the data. The system works well in a dynamic and web-based environment.

Knowledge Sifter is another agent-based system that enables users to access heterogeneous data sources [3]. The Knowledge Sifter Agent has layered architecture; the layers are user layer, knowledge management layer and Data layer. Specialized agents like Integration Agent, Web Service Agent, and Control Agent that resides at various layers and perform specific functions. Knowledge Sifter allows users to search information, both organic and open source and access data sources like web, open source repositories, XML databases and the emerging semantic web. Other agent systems, specializing in data integration and analysis include CoBase, Garlic, HERMES, InfoMaster, RETSINA, SIMS, and TSIMMIS.

Now-a-days most of the organizational Information Systems are in dire need of the systems that are self-driven and capable of making decisions by analyzing the requirement specifications. That is, there is no human intervention so that the systems don't require the continuous supervision. In the area of Data Engineering, the solution to almost all the problems lies at the intersection of database systems and Artificial Intelligence [4]. The Agent Technology comes with the solutions to various technological areas and it is useful to know what properties Agents may have to understand why they will change our life [5]. Following are the properties:

- Autonomy
- Intelligence
- Communication
- Collaboration/ Cooperation
- Mobility
- Proactiveness
- Reactive
- Security

In the next section we will try to summarize the data integration on enterprise level. Section 3 will introduce the working benefits of Multi-agent Systems. In section 4 and 5 we will propose our architecture and it's benefits respectively. Section 6 is the conclusion.

I. ENTERPRISE-WIDE DATA INTEGRATION

Today the organizations are more likely to move towards the information sharing at inter and intra organizational level. To accomplish this, is not an easy job. There are several issues that are to be addressed appropriately like heterogeneity among the data sources, level of security in data sharing, and consistent mechanism of information flow among the business entities with in the enterprise. But still there are many reasons for doing Enterprise-wide data integration and are listed as follows:

- Accessibility of corporate information
- Integration of information systems architectures for better data sharing over inter or intra organizational level
- Merger of organizations
- Acting e-business imperatives
- Organizations want to integrate their.
- re IS's with the IS's of other organizations

i. Enterprise Architecture

Enterprise Architecture is a comprehensive model for an enterprise. It is the master plan guiding for integration. Organizations have instead moved towards centralized architecture, adopting resource and data sharing in decentralized architecture and demand more flexible and cohesive environment for their information system architecture [6].

ii. Enterprise-wide Integration

When we are talking about Enterprise-wide integration we mean to say the integration of the resources that shares some common data. Enterprise model is the blueprint of Integration plan for the enterprise. Enterprise integration includes the integration of various data sources like XML data for Business-to-business interchange, packaged data from ERP systems. Legacy systems, or relational databases (structured and semi structured data). The interoperability of the dispersed data sources is carried out by utilizing simple communication protocols and these data sources are classified into business domains.

The main advantage of the Enterprise-wide data integration is that business users believe that they are interacting with one large homogeneous data source, but the fact is that this homogeneous view of data is the virtual representation of the data that is integrated from various data sources. Users will never know the underlying complexities and platforms of the data sources. Here comes the concept of virtual representation or a virtual data warehouse, virtual data warehouse is only the abstraction to the users to provide them a homogeneous environment.

Many architectures have been developed for enterprise-wide integration; for example, Federated databases architectures, mediator based architecture, or agent based architectures that are utilizing data integration services. By enterprise data integration many organizations are formed into enterprises to gain competitive advantage over their competitors. Enterprise Data Integration gives benefits like transformation of data into useful knowledge, data analysis capabilities, and enterprise data management [7].

II. INTELLIGENT AGENTS

Intelligent Agents are software elements that work without the assistance of users by making choices. Choices are based on rules that software developers have identified and built into the software. It is expected that Intelligent Agents will be a strong focus for AI development in years to come, to automate simple tasks.

An agent is merely a software program but with the difference that they can react, assist, take decisions and benefit the users. As the time passes they get more complex then ever. The efforts of researchers from several different fields have eventually come at common agenda; the development of software agents. On the one hand, researchers from the fields of human-computer interaction, intelligent and adaptive interfaces, knowledge acquisition, and end-user programming, have concerned themselves with the implications of the agent metaphor and its concrete representations, learning, explanation, agent authoring, and other aspects of interaction between humans and software agents. On the other hand, researchers in the fields of distributed artificial intelligence, robotics, artificial life, and distributed object computing have contributed expertise in the areas of negotiation and planning, situated action, agent-to-agent protocols, concurrency, and component-based frameworks.

We feel that the use of agent technology will make an scalable and interoperable architecture. To prove our point of

view that the use of agent in the system will definitely be efficient, we are providing here some agent properties.

i. Autonomy and Deliberation:

The ability to act and make decisions without being controlled by anyone else [Oxford advanced learner's dictionary]. An autonomous agent is a system situated within and a part of an environment that senses the environment and acts on it over time in pursuit of its own agenda so as to affect what it senses in the next. For Enterprise Data Integration this is a definite need because the data sources are residing on different locations and are in different standards. To make the query breakups and formalizations of sub queries require such behavior.

M. Wooldridge and N. Jennings define a deliberative agent as "one that contains an explicitly represented symbolic model of the world, in which decisions (...) are made via logical (or at least pseudo-logical) reasoning, based on pattern matching and symbolic manipulation". In our architecture, the ontology-based searching is used. This requires deliberation because the decision is taken on runtime according to the environment. Actually the autonomous behavior and deliberation go hand in hand.

ii. Communication:

One of the key properties of agents is the ability to speak with human (Interface Agents), or devices. Communication between agents are carried out through different languages. Commonly used languages are: KQML: Knowledge Query and Manipulation Language is a language and protocol for exchanging information and knowledge using "performatives". As our architecture is a multi-agent one, we need to transfer knowledge from one agent to the other.

iii. Collaboration and Delegation:

Agents are collaborative when they are able to work together. Also the agent is able to communicate and negotiate with others; it is deliberative and may coordinate its actions with others. Collaborative agents are useful among other things when a task involves several systems on the network. The negotiation is the main issue of Delegation means an agent can ask the other agent to do some work. When an interface agent receives data from the user, it needs to send it to query agent and this collaboration will result in the fetching of data.

iv. Mobility:

Since Java appeared, we may find a lot of mobile agents that can move in a network seeking out information or completing some tasks. Mobile agents can move from one system to another system autonomously. One of the main issues of the mobility is the potential security weakness of mobile agents, because it can also be used to run malicious codes in a network. This is very important as our architecture is a distributed one and collect the information, the agent has to go from one repository to other.

v. Learning:

Agents can learn not only from user inputs, but from other agents in a network to increase their knowledge-base. They can modify their behavior by utilizing its new knowledge. Despite the fact that learning is an important factor of intelligence, there are only a few agents able to learn; most of the time they have fixed (pre-compiled) rules and knowledge base. Agent has to make an intelligent decision. This can be achieved only if the learning capability is there. In our architecture the path of mobile agent is set according to the knowledge base result. Every time the data is retrieved from any repository, its keywords are stored in the knowledge base and as the time passes the retrieval of information will take less time because of informed decisions.

III. PROPOSED ARCHITECTURE

The conceptual architecture is depicted in figure 1. The architecture is divided into three levels: Data Level to manage the data sources, Management Level to perform integration of the data sources and to manage and control the resources. Finally the User Level to give users a graphical interface to obtain a single virtual representation. Specialized agents to perform predefined functionalities reside at different levels of the collaborative agents. If coordination may occur without collaboration, it needs negotiation.

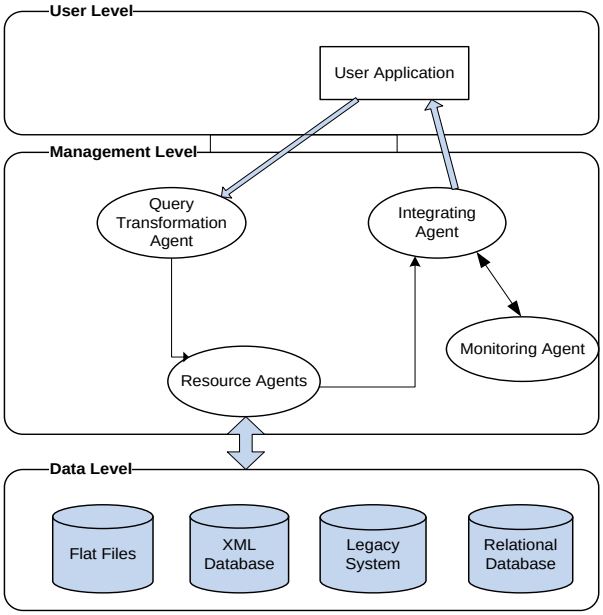

Figure 1

Ontology driven techniques provides lots of benefits in searching and has become an emerging trend in the recent past. Ontology can be defined as the collection of concepts pertaining to any specific domain, for example product ontology is containing the concepts that represent the categories of products. Similarly the sub-queries formulated on the basis of queries specified by users are represented in the form of entities; these entities may belong to specific ontology or domain of ontology. It is quite obvious, that now these sub-queries should be forwarded to the corresponding data repository so that data can be fetched.

The Resource Discovery Agent provides this information by the help of a knowledge base. This knowledge base stores the rules that will be used in finding the relevant data repository against the sub-query. The ontology based technique is also used to formulate these rules and that's how the transformation made by Query Transformation Agent can be understood by the Resource Discovery Agent. Handling of heterogeneous representations is another responsibility of the Resource Discovery Agent

Finally the results from the sub-queries are provided back to the Query Transformation Agent that formulate the result and

sends it back to the Interface Agent. Inter-agents communication is carried out by using Knowledge Query and Manipulation Language (KQML). Following is the list of KQML performatives used in our architecture:

(ask-one
: Sender Interface Agent
: Content GET_SALES RECORD
: Receiver Query Transformation Agent
: Language AGENT SCRIPT
: Ontology INTEGRATED_SEARCH)
(tell
: Sender Query Transformation Agent
: Content WAIT
: Receiver Interface Agent
: Language AGENT SCRIPT
: Ontology INTEGRATED_SEARCH)
(ask-one
: Sender Query Transformation Agent
: Content PRODUCT INFO
: Receiver Resource Discovery Agent
: Language AGENT SCRIPT
: Ontology INTEGRATED_SEARCH)
(ask-one
: Sender Query Transformation Agent
: Content TRANSACTION DETAILS
: Receiver Resource Discovery Agent
: Language AGENT SCRIPT
: Ontology INTEGRATED_SEARCH)
(tell
: Sender Resource Discovery Agent
: Content PRODUCT INFO
: Receiver Query Transformation Agent
: Language AGENT SCRIPT
: Ontology INTEGRATED_SEARCH)
(tell
: Sender Resource Discovery Agent
: Content TRANSACTION DETAILS
: Receiver Query Transformation Agent
: Language AGENT SCRIPT
: Ontology INTEGRATED_SEARCH)

The message flow goes on until …

(tell
: Sender Query Transformation Agent
: Content QUERY_COMPLETE
: Receiver Interface Agent
: Language AGENT SCRIPT
: Ontology INTEGRATED_SEARCH)

Integrating data from heterogeneous data sources is a frustrating job; the users must know the various languages to access data from various data sources and then compile the queries in order to view desired result data. Architecture defined in this paper, users need to know only one formal language to query from the system. Agent-based architecture undertakes this complex task and formulates the results by extracting relevant data from relevant data sources by utilizing simple protocols. Then system gives abstraction to the users that the data is coming from a single source. This abstraction is achieved by forming a single virtual repository of data. System analyzes the queries and identifies the relevant data sources then sends those queries to the identified sources through agents. After retrieving partial results from the data sources, it integrates those results and transforms them into single uniform view that is to be represented to the users.

Query to the system is more like a SQL phrase; the GUI provided by the interface agent gives convenience to the users by restricting them to single uniform language. Display the results as in single uniform representation.

i. Benefits of Proposed Architecture

- Inexpensive in terms of development and implementation as there is no need of implementing expensive middle-ware softwares.
- Abstraction through Monitoring Agent. Providing security features at different operating levels of the system.
- Ontology based searching along the domains.
- Enterprise-wide flow of data and searching through mobile agent.
- Forms a virtual data warehouse instead of moving huge amounts of data from different data sources to the single repository. This will reduce the time of ETL (Extract, Transform, and Load) process as it will overcome this problem and take advantage over data warehouse.
- Provides unified data manipulation language to users for querying the system. Query is divided into sub-queries by query transformation agents and they are responsible for communicating with data sources of different types, different operating systems and having different manipulating languages.
- Creates abstraction through resource agents, the users feel that the data is coming from a single homogeneous data source, but in fact it is coming from heterogeneous distributed data sources.
- Scalability; dynamically searches the data sources for relevant data. The system can extend its search across the enterprise level to web based data sources for integrating the data sources that are required for compiling results.
- Flexibility.

V. CONCLUSION

Enterprise data integration depends on sophisticated technology and complex architectures. In this paper we have proposed an agent-based architecture for data integration having three fundamental levels. There is a network of agents that reside at these different levels for accomplishing the tasks like integration, data extraction, monitoring and resource management. In proposed architecture we have developed a single uniform language for manipulating the heterogeneous data sources that will provide ease in development and deployment in a complex environment carrying various data sources. This has also made our architecture flexible as well.

REFERENCES

[1] Inji Wijegunaratne, George Fernandez, John Valtoudis, *A Federated Architecture for Enterprise Data Integration*: *2000 Australian Software Engineering Conference*, Australia, 2000.

[2] Tomasz Ksiezyk, Gale Martin, Qin Jia: *InfoSleuth - Agent-Based System for Data Integration and Analysis*". *25th Annual International Computer Software and Applications Conference*, USA, 2001.

[3] L. Kerschberg, M. Chowdhury, A.Damiano, H. Jeong, S. Mitchel, J.Si, and S.Smith: *Knowledge Sifter Ontology-Driven Search over Heterogeneous Databases". 16th International Conference on Scientific and Statistical Database Management*, Greece, 2004.

[4] Marti A. Hearst: *Trends and Controversies - Information integration*, IEEE Intelligent Systems, Volume 13, pp. 6-13, Jan./Feb. 1998.

[5] Muhammad Khalid Khan: *Multi-Agent Architecture for Distributed and Dynamic Network Management, 2nd International Conference Frontier of Information Technology*, Pakistan, 2004.

[6] H. Michael Chung, Graham McLeod: *Enterprise Architecture, Implementation, and Infrastructure Management, 35th Annual Hawaii International Conference on System Sciences,* USA, 2002.

[7] Charles J. Petrie: *Agent-Based Engineering, the Web, and Intelligence,* IEEE Intelligent Systems, Volume 11, pp. 24 -29, Dec. 1996.

[8] Manfred A. Jeusfeld, Aldo de Moor: *Concept Integration Precedes Enterprise Integration, 34th Annual Hawaii International Conference on System Sciences,* USA, 2001.