DATA CO-ORDINATION OF SEMANTIC SUPERVISING IN P2P DATABASE SYSTEM

[1]GANESAN VEERAPPAN, [2]SURESH GNANA DHAS
[1]Asst. Professor, Dept. of MCA, Sri Venkateswara College of Engg. & Tech., Thiruvallur Dt.
[2]Professor, Holycross Engineering College, Srivaikuntam City, Thoothukudi Dt.
goldicon.ganesan@gmail.com, sureshc.me@gmail.com

ABSTRACT

The peer-to-peer database management System (P2PDBMS) runs to the high level of the standard database management system, to connect its database with other peer databases on the network. The P2PDBMS are completely independent in which no central server (or) any other centralization are not present. Peers are mostly independent from other peers and they interoperate in a local point-point links. Peer-to-peer is a decentralized model where each peer has equivalent abilities providing the data (or) services to other peers. Each peer manages its own data. In this paper P2P system, we will have some new idea to contribute the databases through a number of techniques for semantic based data modeling, Query management, and update processing and data coordination mechanism. We allow peer to specify Coordination formula to explain how the data which is in one peer should relate to data in an acquaintances. In effective P2P databases, the data placement and data Coordination plays an important role. The data placement is to distribute data among peers so that queries can be executed. The data Coordination will have the ability of peers to manage at runtime semantic interdependence in a decentralized and distributed manner among databases. The main focus of the proposed work is the adaptive techniques for development of a databases management system compatible with the P2P paradigm.

KEYWORDS

Peer-to-peer database management System (P2PDBMS), decentralized model, Query management, adaptive techniques

1. INTRODUCTION

Nowadays, Peer-to-Peer System becomes popularity in the research area and their use in future application. Peers are largely autonomous in particular in what data they store; in which nodes they establish semantic data dependency links with a coordinate their data etc. The Peer-to-Peer systems are generally used for object sharing but they can provide an infra structure for any other application. The existing p2p system allows end-users to share their fields with a good response time. Peer-to-Peer system is an open-ended network of distributed peers, where each peer can exchange data &services with a set of other peers called acquaintances [6]. In the present DBMS, it was not able to analyze data management problem as such as structured data objects, contents update, and data semantics between data. Due to the data management problem the p2p application still lacks scalability. In the distributed databases all the peers are homogenous it controls all query processing and centralized model, replicated in all databases. There is no central controlling site in the p2p system & all the metadata is distributed among the peers. DBMS provides all functionality like query processing, query optimization, view & integrity constraints with the relationship between data. The database architecture is different compare to the p2p system technology.

Implementing DBMS in p2p network system is harness and having various difficulties. Each peer contains a database and their peer can be able to share to its database, to send queries to the network and to route queries coming from the network [1]. Local Relational Model is a data model designed especially for p2p network with a local database each with a set of acquaintances, define the p2p network topology. In each acquaintance link, domain relation helps to translate rule between data items & coordination formulas define semantic dependencies between two database coordination rules are responsible for data coordination with acquaintances & nodes [7].

2. P2P NETWORK Vs DISTRIBUTED DATA MANAGEMENT

In p2p system database resemble heterogeneous distributed databases called multi-database system. The Multi Database Management System (MDBMS) enables data sharing among heterogeneous local databases. A MDBMS allows user to access and share data without requiring physical database integration [11]. Distributed database technology is the collection of participating sites & communication topology is known a priori. It is not the case of highly dynamic p2p network. In p2p network each peer maintains its own databases & exchange data &services with other peers. Due to decentralization features of p2p databases can reside and managed locally at peers and it decides on their own how to develop their databases. There is no central controlling site & all metadata is distributed among the peer. Peers are fully independent whereas flexibility cannot be found in database system. The p2p database management architecture distinguishes from distributed databases though they share the similarity that they both are in a distributed environment [3].

3. PEER DATABASE REFERENCE ARCHITECTURE

In the peer reference architecture is the separation of the functionality into an interface used for submitting the queries.

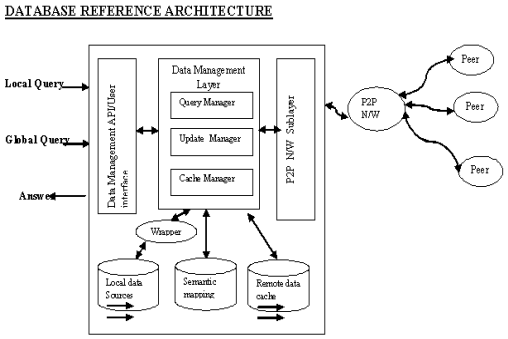

Fig: 1 Peer Database Query Reference Architecture

The data management layer handles query processing & metadata information. P2P network sub layer manages that p2p network connectivity provides communication to the data management layer. The fig.1 show that Queries are submitted using a user interface (or) data management API & handled by the data management layer. Queries can refer to data stored locally (or) globally in the system. The query manager process all the query request and retrieves semantic mapping from a repository when the system integrates heterogeneous data sources. The semantic mapping repository contains Meta information that allows the query manager to identify peers in the system which store relevant data to the query and reformulate the original query to understand by other peers. The execution of the query may be varying according to the different p2p implementations [8].

Data is sent to the peer where the query was initiated and combined at this peer. In specialized peer it is provide for query execution & coordination, As a result data returned by the peers involved in the execution of query may be cached locally to speed up future execution of similar queries and is responsible for executing the local portion of a global query when data is requested by a remote peer [10]. A wrapper can hide data, query language between local data source and the data management layer. The update manager is updating the data which coordinates the execution of update between the peers which the replica of data is updated. The p2p network sub layer provides communication

services to the data management layer. Peers are joined the system which was assigned specific location in the network & are required to share all network responsibility.

4. LRM IN P2P SYSTEM

In similar to the multi-database system, the database coordination presents a new solution to the p2p system which allows independent database system to communicate freely in decentralized manner.Fig.2 show the Local Relational Model (LRM) having some assumption with the set of all a data in a p2p network consists of local databases, with a set of acquaintances. Peers are fully independent in choosing their acquaintance they can join (or) leave the network at anytime. The data model is to allow for inconsistent data bases & to support semantic interoperability in the absence of global schema [9, 12].

The LRM is assumed that all peer nodes have identical architectures consisting of an LRM layer running on a local DBMS. The LRM layer has 4 modules User interface (UI), Query manager (QM), Update manager (UM) & Wrapper.

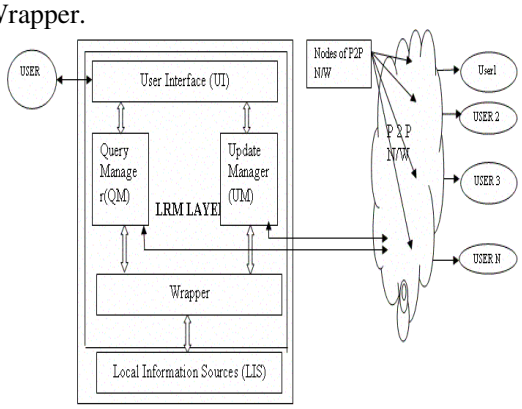

Fig.2 Local relational model in Peer to peer system

The UI allows the user to define queries, receive messages & results from other nodes & control other modules in the local node [3]. QM & UM is responsible for query & update propagation and database coordination function are implemented within QM &UM. QM process both user queries& queries coming from other nodes. Peers are communicating through QM &UM by XML msg. The communication language between the wrapper & LDB is LDB dependent (eg. sql).Each data source has a wrapper layer that maps a sub-queries into its native query language. The LRM model manages domain relation, coordination formulas coordination rules, acquaintances & interest group.

5. DATA COORDINATION IN PEER TO PEER

In data Coordination model the nodes consists of a source databases (SDB) and Source schema (SS) it describes a shared part of SDB. The fig.3 exposes the Peer-to-Peer data management that consists of user interface (UI), Database Manager (DBM), JXTA layer and Wrapper. JXTA Layer is responsible for all nodes activities on the network, by discovering the new nodes and interest groups, joining and leaving groups, sending and receiving queries and query results. Wrapper manages connection to SDB and responsible for extraction and maintenance of the source schema. Different database may require different databases drivers, their module is adjustable depending on the underlying databases All the layers represents procedure calls, In the JXTA Layer the advertisement repository stores al discovered and locally created JXTA advertisement.

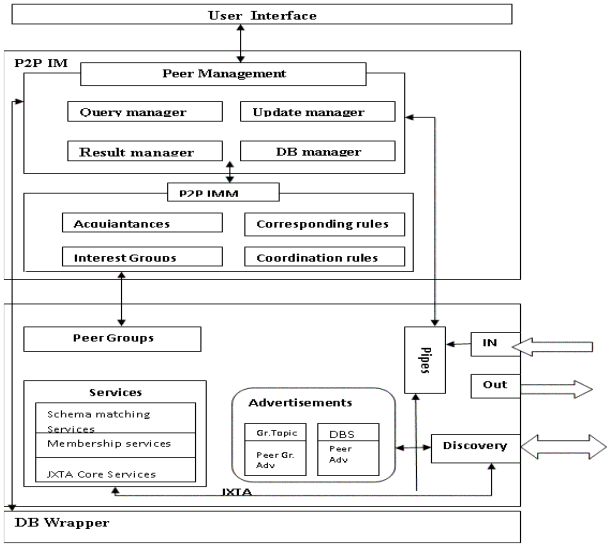

Fig. 3 Data flows in various service level layers.

The peer advertisement includes the source schema information. The Services module implements the core JXTA services and DB-related services .The P2P Management module allows users to control other modules and repositories from both the DBM and JXTA Layer. It makes it possible to create a new communication link, to make a new acquaintance or to modify a coordination rule. The control lines are shown as thick arrows from P2PManagement to other components. Query Planner processes all input queries. It uses acquaintance queries, acquaintances and interest groups information in order to detect groups and nodes for propagation [13].

The Query Propagation (QP) module takes this information as input and uses correspondence rules for query rewriting. It uses pipes to send translated queries to acquaintances. When necessary, QP submits queries to the source database. Results Handler receives results coming from acquaintances and translates them using Correspondence Rules. These results for a user query, and then it reports them to UI. Otherwise, it sends them backward to the node which sent respective network query. Results Handler gets results coming from Wrapper, and sends them either to UI or to the network. Update Handler provides all functionality necessary for updates processing. The peer can read acquaintance queries for *all* peers from a file and broadcast this file to all peers on the network. Once received this file, each peer looks for relevant for that peer acquaintance queries, reads them, and creates necessary IPE connections [2, 4]. During the lifetime of a network, each node accumulates this information. The current version of the P2PDBMS implements Acquaintances and Coordination Rules, and partially implements Interest Groups and Correspondence Rules. Discovering nodes and publishing node's resources on the network; remotely monitor other nodes and send queries to acquaintances, receive and reconcile incoming query results discover network topology defined by paths of interdependent acquaintance queries; execute global update procedure on the network.

6. VISION OF DATA COORDINATION IN P2PDBMS

During the development process P2PDBMS identifying relevant sources of information in P2P network is resource discovery. In P2P architecture, the resource discovery poses additional performance problems since there is no central metadata repository and thus large number of global metadata needs to be processed. The resource discovery still requires robust propagation techniques for both peer and resources in P2P. Data coordination metadata influencing the interaction among peers is decided at run time [5]. Feature, which allows us to deal strong dynamics of P2P networks. Due to dynamics in p2p network, In order to achieve good quality services from the database network a significant effort towards metadata management should be done. Query propagation and answering depends on the level of metadata development and management. Performances of P2PDBMS of query propagation within P2P system have recursive and direct method. Recursive method when a query

reaches a node, the query is reissued from that node and waits for a response from all the nodes that it queried. With the direct method when a node receives the query it will immediately answer the query and then pass the query onto all the other adjacent nodes. Here, the problem of a correctness and completeness of the query results arises. The results are correct with respect to their inter-database schema mappings and complete with respect to the database storing data relevant to the query [14]. P2P can use the classical approaches to query processing, the coordination is effectively constructed. Peer database management system is the efficient management of updates to the data [15,16].

6.1 IMPLEMENTATION DATA COORDINATION

The target of update propagation management is to allow high performance in presenting up-to-date answers their queries, when they are plenty of resources. In p2p the emphasis, is on coordinating databases, than integration. Due to coordination mechanism peers effectively manage at runtime semantic inter among databases in distributed & collaborative manner. P2PDBMS implements a fully decentralized data coordination model. This explains by having 4 notions. Interest group, Acquaintances, corresponding nodes & coordination rules. The interest groups will allow us to gather peers according to relevant queries; will increase the relevance of query answer. An acquaintance will provide with a set of paths for propagation from one node to other nodes in interest group. Corresponding rules will ensure the proper information flow along these paths, and coordination's rules will define the query propagation policies along these paths.

The Data Coordination manages queries between acquaintances, to give appropriate query answering & update propagation. It enables peer to evaluate queries against their local databases, receive & reconcile query results, locate to other peers. With the concept of data coordination we take an example of healthcare domain. The patient was recently relocated from one place (Bangalore) to another place (Chennai) how she is going to keep control of her health condition she should undertake regular blood examines. The doctor Mr. Ram who is in Chennai needs all the medical records of the patient Ms.Swetha from Bangalore Hospital and also details from the personal doctor Sakthi. Assume that the involved databases are heterogeneous and they use different relation and attributes names to represent similarity, different formats for patients ids, different format for data.
Let we take, the following peer location and databases schema

PEER1:Bangalore:Database:BangaloreHospitalDB(ID#, Fname, lname, Date, Disease, Examines)

PEER2:Bangalore:Database:PreviousDoctorDB (PID, Fname, lname, Date, Disease, Prescription, allergies);

PEER 3: Chennai: Database: Current Doctor DB (ID, Name, Address, illness, Treatment, Examines)
To search all the relevant information for the patient in P2P, specific query should be followed.

SELECT name, disease, examines
FROM 'Bangalore hospital' , 'Dr.Sakthi' WHERE name="Swetha" and disease='Diabetes"

By the decentralized features of P2P, the autonomous databases can be managed locally at peers through database Coordination. Interest groups will be able to answer a given query, for instance the Bangalore hospital and Dr.Sakthi from Bangalore (PEER 1:Bangalore Hospital DB,PEER 2:Previous Doctor DB).Next we have to create links from one node to another. Acquaintances are the nodes that have data and can be used to answer specific query with respect to the databases of the acquainted nodes .Node is an acquaintances we have to know to compute and how to propagate a query and to reconcile it with the results coming from other acquaintances. Eg: PEER is acquaintances of PEER 2 with the specific query $Q^{Peer1->peer2}$ (FN, LNBD, D) which describes First name, last name, Birth date, Disease.

Each acquaintance is associated with one or more corresponding rules and a set of Coordination rules. Corresponding rules will take care of the heterogeneity problem. How to translate queries that are to be sent to a particular acquaintances (Previous Doctor DB: fname, lname→ Current Doctor DB:Name). Coordination rules will specify under what condition, when and where to propagate a query to specific acquaintances. One possible implementation of the coordination rules is Event

Condition Action (ECA) rules. The Event could be any query coming from the user (or) from another peer. Condition refers to the query properties; Action could be a translation and propagation of a give query to a particular acquaintance.

Eg. ECA rule constructed for the Query 'Q'

Event : Q

Condition : if name==”Swetha” Then Execute Q;

Endif

Action : propagate to peer3.

7. CONCLUSIONS

In this survey we enlightened the concept of Peer-to-Peer database management system. In particular, we have identified an peer database architecture for the system. In our architecture a peer contains a query interface, a Data Management layer and P2P network sub layer. We have identified Resource discovery, metadata maintenance, Query propagation and answering, Query processing and update propagation in P2PDBMS. Within each component we have classified the problem and solution addressed by current work. The Local Relational Model architecture identified how the database coordination has a new idea in Peer System. All the four notion of database Coordination can be explained with an example of Healthcare domain. Similar to MDBMS, Data Coordination is implemented in the P2PDBMS, the P2P technology to exploit the power of available distributed database management. The technique achieved in P2PDBMS will be completely autonomous without any central server should be achieved. As P2P technologies become more popular in different domains they are required to deal with the inherent differences in the way peer represent their data.P2P DBMS are a hot area of research and will continue to be until these challenges are adequately addressed.

Acknowledgments

We are grateful to my supervisor, Co-author Professor, Dr. Suresh Gnana Dhas, Principal Holycross Engineering and Technology. His knowledge, guidance, support, motivation, and faith in my abilities were instrumental throughout the course of my research. My parents and my sister Mrs. N. V. Revathi, Encourage and gave me great support for the fruitful discussion, as well as some sporting enjoyment.

REFERENCES

[1] **Kementsietsidid,A.,Arenas,M.,Miller,R.J., “**Data sharing through query translation in peer-to-peer system”, Submitted for publication (2004).

[2] **M.Arenas,V.Kantere,A.Kementsietsidis,I.Kiringa,R.Miller, and J.Mylopoulos.,** “The Hyperion project” : From data integration to data Coordination,2003.

[3] **Wee Siong Ng, Beng Chin Ooi, Kian-Lee Tan, Aoying Zhou**, ”Peerdb:A P2P-based system for distributed data sharing,” ICDE,2003.

[4] **A.Kementsietsidis, M.Arenas and R.Miller**, ”Data mapping in peer-to-peer system”, ICDE,2003.

[5] **F. Giunchiglia and I. Zaihrayeu**, “Making peer databases interact – a vision for an architecture supporting data coordination,” 6th International Workshop on Cooperative Information Agents (CIA-2002), Madrid, Spain, September 18 -20, 2002.

[6] **Anastasios Kementsietsidis**, “Data Sharing and Querying for Peer-to-Peer Data Management Systems”, University of Toronto, 2003.

[7] **W. S. Ng, B. C. Ooi, and K. L. Tan**. “Bestpeer: A selfconfigurable peer-to-peer system”. In Proceedings of the 18th International Conference on Data Engineering, page 272, San Jose, CA, April 2002 (Poster Paper).

[8] **ZHUGE, H.** 2003. Active e-document framework ADF: Model and tool. Information and Management, **41**(1):87-97.

[9] **Bernstein. P.A., Giunchiglia, F., Kementsietsidis, A., Mylopoulos, J., Serafini, L., and Zaihrayeu, I.** "Data Management for Peer-to-Peer Computing: A Vision", WebDB02 – Fifth International Workshop on the Web and Databases, 2002.

[10] **Halevy, A**. "Answering queries using views: a survey". VLDB Journal 2001.

[11] **Lenzerini, M**. "Data Integration: A Theoretical Perspective". PODS 2002: 233-246, 2002.

[12] **Serafini, L., Giunchiglia, F., Mylopoulos, J., and Bernstein, P.A**. "The Local Relational model: Model and Proof Theory". IRST Technical Report 0112-23, Istituto Trentino dicultura, December 2001.

[13] **P. Bernstein, F. Giunchiglia, A. Kementsietsidis, J. Mylopoulos, L. Serafini and I. Zaiharayeu**, " Data Management for Peer-to-Peer Computing: A Vision", In WebDB, 2002.

[14] **E. Franconi, G. Kuper, A. Lopatenko and I. Zaihrayeu**. "Queries and Updates in the coDB Peer to Peer Database System", In VLDB, 2004.

[15] **D. Calvanese, G. De Giacomo, M. Lenzerini, R. Rosati and G. Vetere**, "Hyper: A Framework for Peer-to-Peer Data Integration on Grids", In ICSNW, 2004.

[16] **C. Yu and L. Popa, "**Constraint-Based XML Query Rewriting for Data Integration", In SIGMOD, 2004.

Biography

Ganesan Veerappan is serving as a Assistant Professor, Sri Venkateswara College of Engineering and Technology, Thiruvallur(Dt) in Department of Master of Computer Applications. He received his B.C.A(Computer Applications) from Periyar University, Salem in 2003 and M.C.A(Computer Applications) from Anna University, Chennai. He has awarded Master of philosophy in PRIST University, Thanjavur also has been completed MBA in Alagappa University, Karikuddi and presently pursuing Ph.D (Comp. Science) at Dravidian University, Kuppam, Andrapradesh since 2009. He has published several papers in the area of Semantic web, Information Security, E-Governance, E-Learning, Wireless sensor Networks and Datacenter Virtualization. His area of Interest is Resource sharing in Peer to peer Networking, Data virtualization, E-Governance, E-Society, Internet Security, Information Management System and Data Mining

Dr. C. Suresh Gnana Dhas is serving as Principal, Professor, Holycross College of Engineering Thoothukudi District. He received his B.E Computer Science in National Engineering College. Madrai, and Engineering and M.E computer Science and Engineering in Government college of Technology, Coimbatore. He has awarded Ph.D in computer science and engineering from Magadh University, Bodh-Gaya in 2009. Prior to this he served as a Professor in Computer Science and Engineering Departments, in Various Engineering Colleges. He has written numerous articles, papers, reports mainly concentrating on Artificial Intelligence, Information Retrieval, Network Security, Network Routing, and various models with application to computer network problems. He is being the Member and Reviewer at various Board of Studies from Computer Science and Information Technology. He has published around 40 National papers and 60 International papers in world leading journals.