A P2P Technology Based on Relational Database in Cloud Computing

Zeenat Kalambkar¹ Akshata Desai² Uma Madake³ Megha Jadhav⁴
¹,²,³,⁴Department of Computer Science and Engineering
¹,²,³,⁴JSPM’s BSIOTR

Abstract—A corporate network is a group of computers or resources, which is connected together in a particular area or in a building, which are all owned by the same institutions or company. This network provides the flexible, economical & Scalable platform for work. The data management system including scalability, flexibility, security. Data Sharing, better performance, no of peers easily added & remove it’s challenging for corporate network but not in distributed. Centralized data processing where data processing supported by one clusters of computers means all data stored on centralized platform, in centralized data processing it’s necessary to handle the computers overload problem, this problem is overcome in using Best Peer++ technology. The Best peer ++ achieves linear scalability throughput with respect to the number of peer nodes, it’s not get overloaded like centralized data processing this problem is overcome using algorithm that is bootstrap and adaptive query processing.

Key words: Cloud computing, Peer to peer system, HadoopDB, Corporate network, Query Processing

I. INTRODUCTION

The no. of companies of the same industry sector are often connected into a corporate network for data sharing & collaboration perspective & each company maintains its own site which is maintained by its own and company shares the selective portion of its business data with the other companies.

The key for the success of a corporate network is selecting the right data sharing platform. The practically, data sharing is retrieved by construct a centralized data warehouse. Unfortunately, such a warehousing solution has have some insufficiency in real deployment. The commercial network desires to improve the support thousands of participant, while the installation of a major centralized data warehouse system is not easy to accomplish & costs expensive including massive hardware/software savings and high preservation cost. But in real world the most of companies are not devoted to invest a lot on extra information systems until they can obviously see the possible return on savings and also companies want to fully modify the access control policy to decide which industry followers can see which part of their collective data. Unfortunately, most of the data warehouse does not provides the flexible solutions. So finally to maximize the income, companies often energetically change their business process as well as they change their company partners.

The business participants may connect and disappear the corporate networks at will and the data warehouse solution has not been developed to handle such dynamicity problems. To overcome those problems, this paper presents BestPeer++ technology, a cloud enabled data distribution stage designed for corporate network. Using the cloud computing, database, peer to peer technologies the BestPeer++ achieves good organization in query processing with a few features that is first, the BestPeer++ is deploy as a service in the cloud.

To form a commercial association, companies only register the sites with the BestPeer++ service provider, in network. Start on BestPeer++ instances in the cloud and at last export data to those instances for distribution data. BestPeer++ adopts the business model which is famous by cloud computing, total cost of possession is therefore significantly reduced since companies do not have to buy any advanced hardware/software. They give for what they utilize in conditions of BestPeer++ instance’s storage space capability and hours. Second, BestPeer++ extending the role-based access control for the inbuilt spread environment of commercial networks for companies. Companies can easily organize their access control policies and avoid undesired business partners to access their collective data. Third, BestPeer++ employs P2P technology to recover data between company partners. BestPeer++ instances are ordered as a structured P2P overlay network i.e BATON. The data are indexing by the table name, column name and data range for efficient recovery. Fourth, BestPeer++ users for achieving high concert processing query. The main workload of a corporate network is simple, low transparency queries. Such queries usually only entail querying a very small number of company partners and can be executed in within time. BestPeer++ is mostly-optimized for these queries; For irregular time-consuming analytical tasks, we offer an interface for exporting the data from BestPeer++ to Hadoop and allow users to evaluate those data.

II. LITERATURE SURVEY

To enhance the usability of conventional peer to peer systems database communities have proposed a series of PDBMS (Peer-to-Peer Database Manage System) by integrating the state-of-art database techniques into the P2P systems. There are many techniques proposed in order to efficiently process large scale data which has explained below

A. “Just-In-Time Query Retrieval Over Partially Indexed Data On Structured P2P Overlays,”:

It is a Peer-based system that Indexes Selected Content for Efficient Search facility. It is not like traditional approaches that index all data and PISCES identifies a subset of tuples to index based on some criteria. Another important addition to this is a coarse-grained range index is built to facilitate the processing of queries that cannot be fully answered by the tuple-level index. The main limitation is the possibly high maintenance cost to maintain the structure

B. “Peerdb: A P2P-Based System For Distributed Data Sharing,”:

PeerDBIs a peer to peer based database management System which employs information retrieval technique to match columns of different tables. The main problem of unstructured PDBMS is that there is no guarantee for the data retrieval performance and result quality.
C. “Hash Tag Recommendation System in A P2p Social Networking Application.”

There are some limitations as to the proposed recommendation methodology. We still have to test the performance of the algorithms in peer-to-peer simulated environment with more number of server nodes. Without we are not able to test the overall popularity.

D. “Query Answering In Peer-To-Peer Systems In Large-Scale Data Processing.”

Clustering-based query is to and the cluster with respect to a specific query object. The applications of clustering-based query exist in many real-life applications which need to and all similar objects with respect to a query in the same cluster. Only a small portion of dataset needs to be transferred over network under certain conditions, while the query result is proved to be consistent with the global clustering. In other words, the query result is guaranteed, while the cost is low.

III. PROPOSED WORK

The following Fig.1. shows the BestPeer++ system. In this The key idea of BestPeer++ is to use dedicated database servers to store data for each business and organize those database servers through P2P network for data sharing. The Cloud Adapter provides an elastic hardware infrastructure for BestPeer++ to operate on by using Cloud services. The infrastructure service that Cloud Adapter delivers includes launching/terminating dedicated MySQL database servers and monitoring/backup/auto-scaling those servers.

![BestPeer++ System Diagram](image)

**Fig. 1: BestPeer ++ System**

We use service to provision the database server. Each time a new business joins the BestPeer++ network. The newly launched virtual server runs a dedicated MySQL database software and the BestPeer++ software. We use relational data service (RDS) to back up and scale each BestPeer++ instance. The whole MySQL database is backed up. In order to provide high availability service, BestPeer++ performs asynchronous back-up operation, and there will be no service interrupt during the back-up process. The scaling scheme of BestPeer++ consists of two dimensions: processing and storage, which scale up independently according to user’s computation requirement.

Finally, the Cloud Adapter also provides automatic fail-over service. In a BestPeer++ network, a special BestPeer++ instance (called bootstrap peer) monitors the health of all other BestPeer++ instances, by querying the Cloud Watch service. If an instance fails to respond to the bootstrap peer (e.g., crashed), Cloud Adapter is called to perform fail-over for that instance. In addition to managing peer join and peer departure, the bootstrap peer spends most of its running-time on monitoring the healthy of normal peers and scheduling fail-over and auto-scaling events.

IV. RELATED WORK

To develop the BestPeer++ system that provides financial, flexible and scalable solutions for corporate network applications. We demonstrate the efficiency of BestPeer++ by benchmarking BestPeer++ against HadoopDB. A Recently proposed large-scale data processing system, over a set of queries designed for data sharing applications. The results show that for simple, low-overhead queries, the performance of BestPeer++ is significantly better than HadoopDB. BestPeer1 data management platform, the ultimate goal of BestPeer is to bring the state-of-art database techniques into P2P systems.

1) In its early stage, BestPeer employs unstructured network and information retrieval technique to match columns of different tables automatically. After defining the mapping functions, queries can be sent to different nodes for processing. In its second stage, BestPeer introduces a series of techniques for improving query performance and result quality to enhance its suitability for corporate network applications.

2) A BestPeer++ network can only have a single bootstrap peer instance which is always launched and maintained by the BestPeer++ service provider, and a set of normal peers. The bootstrap peer is the entry point of the whole network. It has several responsibilities. First, the bootstrap peer serves for various administration purposes, including monitoring and managing normal peers and also scheduling various network management events. Second, the bootstrap peer acts as a central repository for metadata of corporate network.

3) Database community have proposed a series of PDBMS (Peer-to-Peer Database Manage System) by integrating the state-of-art database techniques into the P2P systems. These PDBMS can be classified as the unstructured systems and unstructured system. PeerDB employs information retrieval technique to match columns of different tables. The main problem of unstructured PDBMS is that there is no guarantee for the data retrieval performance and result quality. The structured PDBMS can deliver search service with guaranteed performance.

4) In this system implement the two algorithm that is first, BootStrap Daemon algorithm which is run by the BestPeer++ service provider and it’s main functionality to manage the BestPeer++ network and second, Adaptive Query processing algorithm t functionality of this is to analyze and process the query which is may be low level or high level.

V. EXPERIMENTAL RESULT

The below figures shows the results of working system. The Fig.2,3,4 is the result of BootStrap peer work. In the no. of figures Fig.2 shows the list of no. of companies which are registered in network, then the BootStrap peer provides the
authority to that companies to stay in network and also enable and disable the user when they login or logout.

Fig. 2: New peer request
Fig. 3: Shows how many companies are present or online that information is showed.

Fig. 3: Online peer
Fig. 4: Shows if any company leave the network it not directly discard it maintain into the black list.

Fig. 4: Black list
Fig. 5, 6, 7, 8, 9 shows the working of company or user. Company registered in a network the BootStrap peer provides the authority to that company for working, for accessing any service. Fig. 5. shows the working of DB Manager which is database Manager in which the company upload our data.

Fig. 5: Data upload
Fig. 6: Query browser
Fig. 7: History of the query browser is show in that, it also shows the query size means the query is low level or high level that also shows.

Fig. 8: Using graph the result of query that is low level or high level is viewed as per the timing based. For execution of query is take 8ms (milliseconds) or below the 8ms then result show that query is low level, if takes the time more than 8ms then it is the high level query.
ACKNOWLEDGEMENT

It gives us great pleasure in presenting the preliminary project report on ‘A P2P Technology Based On Relational Database In Cloud Computing’.  

We would like to take this opportunity to thank my internal guide Prof. M.B.Jadhav for giving us all the help and guidance we needed.

We are really grateful to them for their kind support. Their valuable suggestions were very helpful.

In the end our special thanks to Mr. V. H. Chalake for providing various resources such as laboratory with all needed software platforms, continuous Internet connection, for our Project.

We are also grateful to Prof. G. M. Bhandari, Head of Computer Engineering Department, BSIOTR, Wagholi for his indispensable support, suggestions.

REFERENCES