www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 10 Oct 2015, Page No. 14879-14884

Large scale Data Sharing using BestPeer++ Technique

Prof.S.A.Agrawal [1], Kalyani Pathak[2] , Yogesh Barhe[3] , Chetan Chavan[4] , Shrishailya Bhinge [5]

[1] Department of Computer Engineering ,MMIT,Lohgaon,Pune,
411047 Maharashtra, India
sanjay.agrawal@mmit.edul.in

[2] Department of Computer Engineering ,MMIT,Lohgaon,Pune,
411047 Maharashtra, India
Kalyanipathak12@gmail.com

[3] Department of Computer Engineering ,MMIT,Lohgaon,Pune,
411047 Maharashtra, India
yogbarhe08@gmail.com

[4] Department of Computer Engineering ,MMIT,Lohgaon,Pune,
411047 Maharashtra, India
Chavanchetan90@gmail.com

[5] Department of Computer Engineering ,MMIT,Lohgaon,Pune,
411047 Maharashtra, India
shreebhinge@gmail.com

Abstract:

The corporate network those are working in same sector needs to share information frequently and facilitates collaboration. Certain companies with common interest effectively shares data to reduce operational cost and increases revenues. In this Project, we develop a system which provides functionality among these companies to share data with security , scalability, high performance and high throughput .It also provides elastic data sharing among these companies in a network Based P2P data management with the help of Bootstrap peer. Bootstrap peer acts as a entry point for join and departure of nodes and it also provides access control to those peer and also it manages metadata of whole network .By integrating data base and P2P technology, BestPeer++ gives low cost, flexible and scalable platform for data sharing.

Keywords: P2P Network, Query processing, Database Management System, Corporate Network, Access Control.

1. Introduction

Large scale data sharing is one of the major aspects in corporate networks. For collaboration purpose, companies which has same industry sector are connected into corporate network. Here each company has to maintain its own site and shares portion of its business data with other companies. In this paper, we develop a system "BestPeer++" that enables the shared data visible over networked wide and efficient analytical queries are supported over these data.

Peer-to-Peer is a technique in which each node serves the roles of both clients and servers. As the name suggests that is "Large scale data sharing using BestPeer++ technique". It is a peer to peer system which distribute the load of centralized system into peer and we are using this peer to peer system in order process large scale data of corporate network.

2. Literature Review

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

database techniques into the P2P systems. There are many techniques proposed in order to efficiently process large scale data which has explained below:

[1] S. Wu and J. Li have proposed "Just-in-Time Query Retrieval over Partially Indexed Data on Structured P2P Overlays" Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '08), pp. 279-290, 2008.

It is a Peer-to-Peer based system that helps to Index the Selected Content for Efficient Search. It is not like conventional approache that indexes 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 which is used to facilitate the processing of queries that cannot be fully answered by the tuple-level index. The main limitation is it possibly requires high maintenance cost to maintain the structure.

[2] K.-L. Tan and A. Zhou presented "PeerDB: A P2P-Based System for Distributed Data Sharing," Proc. 19th Int'l Conf. Data Eng., pp. 633-644, 2003.

PeerDB is a peer to peer based database management System which employs information retrieval technique to match columns of different tables. The main issue of unstructured PDBMS is that there is no guarantee for the data retrieval performance and it provides poor quality of result.

[3] S. Jiang and B.C. Ooi have proposed "Distributed Online Aggregation," Proc. VLDB Endowment, vol. 2, no. 1, pp. 443-454, 2009.

In this paper, the on- line aggregation technique extended to a distributed context where sites are maintained in Distributed Hash Table (DHT) network. Distributed Online Aggregation (DOA) scheme works iteratively and produces approximate aggregate answers as follows:

In each iteration small set of random samples are fetched from the data sites and distributed to the processing sites.

At each processing site, local aggregate is computed based on the previously allocated samples.

At a coordinator site, these local aggregates are combined into a global aggregate for further processing.

[4] A. Lakshman and A. Pilchin "Dynamo: Amazon's Highly Available Key-Value Store" Proc. 21st ACM SIGOPS Symp .Operating Systems Principles (SOSP '07), pp. 205-220, 2007.

This paper presents the implementation of Dynamo, which is a highly available key-value storage system that some of Amazon's core services use to provide an always-on experience. The important thing here is that it makes extensive use of application-assisted conflict resolution and object versioning in a manner that provides a novel interface for developers to use.

3. SYSTEM ARCHITECTURE

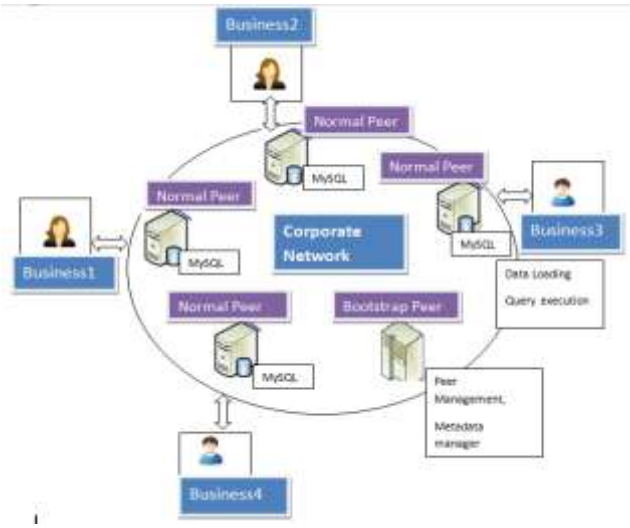
System architecture consists of various parts described as follows:

We are implementing this project by using Java Technology and MySQL database.

Various components of system are:

3.1 Corporate network:

Corporate network is used to share the information among the participating companies (Businesses) and facilitating collaboration in a certain industry sector where companies share a common interest among them. We are creating here a corporate network where each business registers in the corporate network. The network service provider saves this registration information and allows each business or company to share their data with other companies present in the network. The businesses can then upload their data to their local databases. And allow access to this data to corporate network companies



DOI: 10.18535/ijecs/v4i10.44

Fig 1. System Architecture

3.2 Bootstrap Peer :

The bootstrap peer is run by the BestPeer++ service provider. The important functionality of this peer is to manage the BestPeer++ network.

Every normal peer or node wants to join an existing corporate network must first connect to the bootstrap peer. The bootstrap peer authenticates this information. If the join request is permitted by the service provider, the bootstrap peer will put the newly joined peer instance into the peer list of the corporate network.

In addition to managing peer join and peer departure another functionality of bootstrap peer is to monitoring the health of normal peers and scheduling fail-over and auto-scaling events. In this function the bootstrap periodically collects performance metrics of each normal peer.

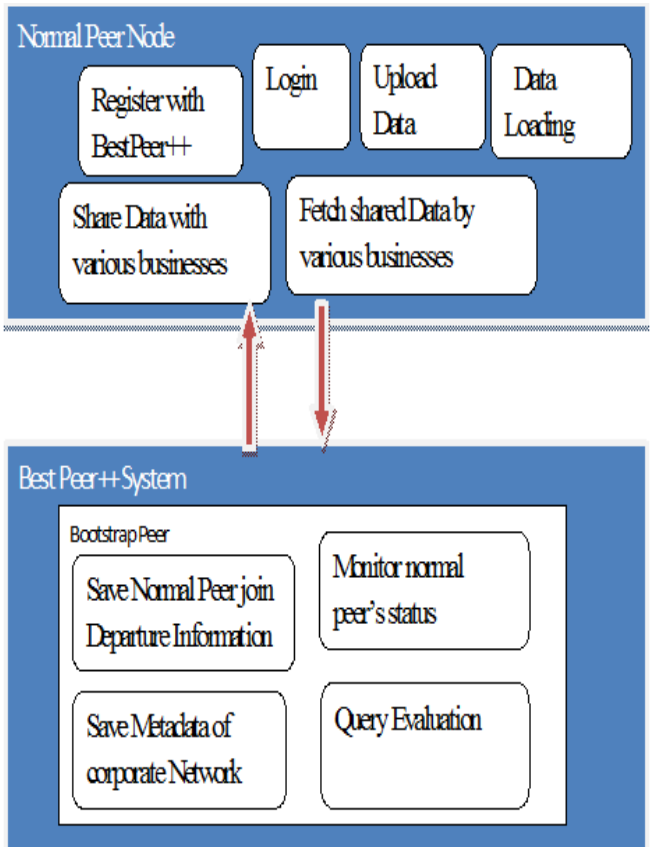


Fig 2. Modular Architecture

3.3 Normal peer:

Each normal peer has some processes like data loading and data indexing. In normal peer there are two data flows first is an offline data flow and an online data flow.

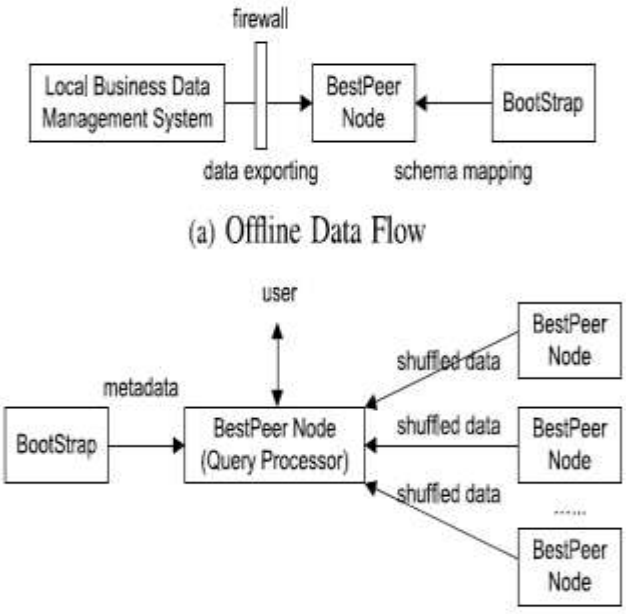


Fig 3. Data Flow Diagrams

In offline data flow, local businesses upload their data to MySQL database and information regarding to this data is sent to bootstrap peers that is the indexing information regarding data tables.

In online data flow, the queries fired by various peers are executed against particular peer node which is having the required data with the help of bootstrap peer which stores the information about data from all peers.

3.3Query Processing:

As we know the data uploaded is present on number of peers so if a peer wants to access data from another peers, it fires the query which goes to bootstrap peer which have metadata of all peers and then it get fired on the peer sharing this data.

3.4Access Control:

In this technique, we are providing an ability to each business to control their data sharing to specific number of companies. And here also adding security to this module by using PKI encryption technique with the help of public and private keys ..

4. ALGORITHM

4.1Bootstrap Daemon Algorithm

```
While true do
 Status S=invokeCloudWatch()
ArrayListpreList= BootStap.getALLpeer()
ArrayListnewPeer=newArraylist()
For i=0 to peerList.size() do
If peerlist.get(i).fails() then
Peer peer=new Peer()
Peer.loadMYSQL.BackUpFromRDS(peerlist.get(i))
newPeer.add(peer)
Bootstrap.setBlackList(peerList.get(i))
      Else
peerList.get(i).overloaded() then
       Peer peer=new peer()
Peer.upscale(peerList.get(i))
Peer.clone(peerList.get(i)).getDB(i))
BootStrap.setBlackList(peerList.get(i)
Newpeer.add(peer)
BootStrap.removeAllPeersInBlackList()
BootStrap.addAllNewPeer(newPeer)
BootStrap.broadcastnetworkStatus()
      Sleep t seconds
```

4.2DESCRIPTION:

4.2.1Auto failover Condition

The bootstrap periodically collects performance metrics of each normal peer. If some peers are malfunctioned, the bootstrap peer will trigger an automatic fail-over event. The automatic fail-over is performed by first launching a new instance from bootstrap peer. Then, the bootstrap peer asks the newly launched instance to perform data- base recovery from the latest database backup stored in bootstrap peer. Finally, the failed peer is put into the blacklist.

4.2.2 Auto Scaling-Up Condition

Similarly, if any normal peer is overloaded (e. g., CPU is over-utilized or free storage space is low), the bootstrap peer triggers an auto-scaling event to either promote the normal peer to a larger instance or allocate more storage spaces.

5. ADVANTAGES

1. Distributed data sharing.
2. It outperforms HadoopDB.
3. Failover and auto-scaling.
4. Uses pay-as-you-go business model.
5 Access control policy.

6. LIMITATIONS

1. Complex queries are time consuming.
2. Require relational databases.

7. CONCLUSION & FUTURE SCOPE

We have discussed problems in sharing and processing data in corporate network and proposed a system BestPeer++, which is used to deliver data sharing facilities by including P2P technology database, Query processing and access control.

To configure a corporate network, companies simply register their sites with the BestPeer++ service provider; launch BestPeer++ instances in the network and finally exports the data to those instances for sharing purpose. BestPeer++ accepts the pay-as-you-go business model popularized by cloud computing.

So that, BestPeer++ is the effective solution for data sharing among companies.

ACKNOWLEGEMENT

It gives us great pleasure in presenting the preliminary project report on 'Large scale data sharing using BestPeer++ technique'.

We would like to thank our internal guide Prof. S .A. Agrawal 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. He saved us lots of time by preventing us from choosing to implement algorithms that were overlay complicated or impractical. He bought wisdom and sophistication that was much needed for this project.

We are also grateful to Prof.P.M.Daflapurkar, Head of Computer Engineering Department, M.M.I.T. Lohgaon, Pune, for his indispensable support, suggestions.

In the end our special thanks to D.R.Patil for providing various resources such as laboratory with all needed software platforms, continuous Internet connection, for Our Project.

REFERENCES

[1] I. Tatarinov, Z.G. Ives, J. Madhavan, A.Y. Halevy, D. Suciu, N.N.Dalvi, X. Dong, Y. Kadiyska, G. Miklau, and P. Mork, "The PiazzaPeer Data Management Project," SIGMOD Record, vol. 32, no. 3,pp. 47-52, 2003.

[2] W.S. Ng, B.C. Ooi, K.-L. Tan, and A. Zhou, "PeerDB: A P2P-Based System for Distributed Data Sharing," Proc. 19th Int'l Conf. Data Eng., pp. 633-644, 2003.

[3] R. Huebsch, J.M. Hellerstein, N. Lanham, B.T. Loo, S. Shenker, and I. Stoica, "Querying the Internet with PIER," Proc. 29th Int'l Conf. Very Large Data Bases, pp. 321-332, 2003.

[4] Saepio Technologies Inc., "The Enterprise Marketing management Strategy Guide," White Paper, 2010.

[5] Thusoo, J. Sarma, N. Jain, Z. Shao, P. Chakka, S. Anthony, H. Liu, P. Wyckoff, and R. Murthy, "HIVE: A Warehousing Solution over a Map-Reduce Framework," Proc. VLDB Endowment, vol. 2, no. 2, pp. 1626-1629, 2009.

[6] H.T. Vo, C. Chen, and B.C. Ooi, "Towards Elastic Transactional Cloud Storage with Range Query Support," Proc. VLDB Endowment, vol. 3, no. 1, pp. 506-517, 2010.

[7] S.Wu, S. Jiang, B.C. Ooi, and K.-L. Tan, "Distributed Online Aggregation," Proc. VLDB Endowment, vol. 2, no. 1, pp. 443-454, 2009.

[8] S.Wu, J. Li, B.C. Ooi, and K.-L. Tan, "Just-in-Time Query Retrieval over Partially Indexed Data on Structured P2P Overlays," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '08), pp. 279-290, 2008.

[9] S. Wu, Q.H. Vu, J. Li, and K.-L.Tan, "Adaptive Multi-Join Query Processing in PDBMS," Proc. IEEE Int'l Conf. Data Eng. (ICDE '09), pp. 1239-1242, 2009.

[10] P. Rodrıguez-Gianolli, M. Garzetti, L. Jiang, A. Kementsietsidis, I. Kiringa, M. Masud, R.J. Miller, and J. Mylopoulos, "Data Sharing in the Hyperion Peer Database System," Proc. Int'l Conf. Very Large Data Bases, pp. 1291-1294, 2005.

[11] J. Dittrich, J. Quian_e-Ruiz, A. Jindal, Y. Kargin, V. Setty, and J. Schad, "Hadoop++: Making a Yellow Elephant Run Like a Cheetah (without it Even Noticing)," Proc. VLDB Endowment, vol. 3,no. 1/2, pp. 515-529, 2010.

Author Profile

Prof.S.A.Agrawal
Assistant Professor
Computer Engineering Dept.
MMIT Lohgaon ,pune

Ms.Kalyani Pathak
Student
Computer Engineering Dept.
MMIT Lohgaon ,pune

Mr.Yogesh Barhe

DOI: 10.18535/ijecs/v4i10.44

Student
Computer Engineering Dept.
MMIT Lohgaon ,pune

Mr Chetan Chavan
Student
Computer Engineering Dept.
MMIT Lohgaon ,pune

Mr.Shrishailya Bhinge
Computer Engineering Dept.
MMIT Lohgaon ,pune