CLOUD BASED DATA SHARING PLATFORM USING FUZZY SQL QUERY PROCESSING SYSTEM

[1]G.UmaMaheswari, [2]R.Karthikeyan
[1]PG Scholar, Department of Computer Science and Engineering, Mohamed Sathak Engineering College, Ramanathapuram, India.
[2]HOD, Department of Computer Science and Engineering, Mohamed Sathak Engineering College, Ramanathapuram,India.

Abstract: The companies are used corporate network to shares the information among all participating companies. This technology is mainly used to reduce the operational cost and increase the revenue. The companies have some unique challenges when sharing and processing the data in data management system such as scalability, performance, throughput, security. To solve these problems bestpeer++ is used. In addition, HABE (Hierarchy Attribute Base Encryption) is used for key verification. In this concept, cloud computing, data base and p2p technologies are integrated into one system. It provides flexible, an economical and scalable platform for corporate network application and delivers data sharing services participants based on the widely accepted pay-as-you-go business model. In this concept bestpeer++ is used on Amazon EC2 cloud platform. Each time a new business joins the BestPeer++ network, a dedicated EC2 virtual server is launched for that business.Bootstrap peer is used to manage the bestpeer++. HadoopDB is used and recently proposed large scale data processing system to handle typical corporate network workloads. Fuzzy sql query is used to extract and view the added staff information. The benchmark results demonstrate that BestPeer++ achieves near linear scalability for throughput with respect to the number of peer nodes.

Keywords: BestPeer++ ; Fuzzy Sql; Cloud Computing

1. INTRODUCTION

BestPeer++ is a Syatem to deliver the data sharing platform for corporate network.In order to provide this elastic service a centralized data warehouse has to be built [1]. Unfortunately, such a warehousing solution has some deficiencies in real deployment. The corporate system needs to scale up to thousands of participants while installation of large scale centralized data ware house. It required high cost to maintenance and hardware/software.

In the present study,a new enhaced feature is added to the BestPeer++ system.The Proposed System Comprise of following features.

- BestPeer++ is deployed as a service in the cloud.To form a corporate network, companies simply register their sites with the BestPeer++ service provider,launch BestPeer++ instances in the cloud and finally export data to those instances for sharing.BestPeer++ adopts the pay-as-you-go business model popularized by cloud computing. The total cost of ownership is therefore substantially reduced since companies do not have to buy any hardware/software in advance. Instead, they pay for what they use in terms of BestPeer++ instance's hours and storage capacity.
- BestPeer++ extends the role-based access control for the inherent distributed environment of corporate networks. Through a web console interface, companies can easily configure their access control policies and prevent undesired business partners to access their shared data.
- BestPeer++ employs P2P technology to retrieve data between business partners. BestPeer++ instances are organized as a structured P2P overlay network named BATON. The data are indexed by the table name, column name and data range for efficient retrieval.
- BestPeer++ employs a hybrid design for achieving high performance query processing. The major workload of a corporate network is simple, lowoverhead queries. Such queries typically only involve querying a very small number of business partners and can be processed in short time. BestPeer++ is mainly optimized for these queries. For infrequent time-consuming analytical tasks, we provide an interface for exporting the data from BestPeer++ to Hadoop and allow users to analyze those data using MapReduce.

2. RELATED WORKS

The corporate company has to bulit centralized data warehouse in order to provide data sharing platform. Centralized data ware house take more money to maintenance. It does not provide flexibility to company. The companies are not allowed to determine which business partners can see which part of their shared data[1].

In order to handle large amount of companies data HadoopDB is used.HadoopDB is an combination of parallel DBMS and Hadoop approach.It take the advantage of both parallel DBMS and MapReduce based system.From Parallel DBMS it takes data analysis, achieving the performance and efficiency and from MapReduce based system it added the scalability, fault tolerance, and flexibility features[10].

In a corparate ntwork the data are distributed across the network.For database query processing a massively distributed query engine called PIER is used.PIER is based on overlay network. PIER - a structured query system intended to run at large scale. Due to its design, it can perform queries on heterogeneous data[12].

For providing distributed data sharing platform PeerDB is introduced.It distinguishs itself from P2P system in the following aspects.(i) It support fine-graine content based searching (ii) sharing of data without shared schema (iii) In order to perform operation on peers' site it is enhanced with the power of mobile agents (iv) It is self-configurable[5].

3. EXISTING SYSTEM

The centralized data ware house is used for achieve data sharing. It periodically extracts data from the internal production systems of each company for subsequent querying. Unfortunately, such a warehousing solution has some deficiencies in real deployment. The corporate system needs to scale up to thousands of participants while installation of large scale centralized data ware house. It required high cost to maintenance and hardware/software. Companies are not keen to invest heavily for additional information system. Centralized data ware house is not provide fully customize access control policy to the company.

Finally it maximize the revenues, companies often dynamically adjust their business process and may change their business partners. Centralized data ware house take more money to maintenance. It does not provide flexibility to company. The companies are not allowed to determine which business partners can see which part of their shared data. Hence proposed system is introduced to achieve flexibility and reduce the cost.

In existing system the centralized data ware house is used to achieve data sharing. It does not provide flexibility to user. It increases the maintenance cost and software/hardware investment higher. It does not provide fully customize access control policy to user. It does not allow the companies to determine the business partners can see which part of their shared data.

4. PROPOSED METHOD

4.1BestPeer++

Bestpeer++ is introduced for data sharing among all corporate companies. Bestpeer++ is the combination of data mining, network and cloud. It provide scalable, economical and flexible platform for data sharing. This project is mainly for reducing the cost and increase the revenues. Here data mining is used for data analyzing. Amazon EC2 cloud platform is used in this proposed system. In this concept HadoopDB is used and large scale data processing system to handle typical corporate network workloads. Hadoop data base is used to handle large number of data.

4.2 System Architecture

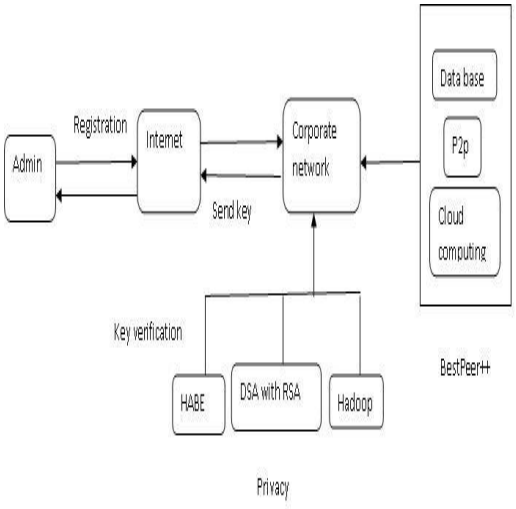


Figure 1: System Architeture for BestPeer++

If the companies are wanted to join itself into the corporate network it must register the company into cloud server through internet. After registration the corporate network send group key and user key to each company. By using group key the company can enter

into the corporate network. The user key is for business login. The corporate network acts as a BestPeer++ which is an integration of cloud computing, database, p2p into one system. DSA with RSA is used for privacy. DSA that is used to generate digital signature. If a staff need to enter the business login , then they use user key.

This user key is verified by HABE. For storing and retriving data to/from cloud server Hadoop is used which implement MapReduce Function.If admin upload the file it is first splitted and then stored in the space alloted by cloud server.If staff ned to access any file in the sence the splitted file is gathered and then merged. All this actions are performed by using MapReduce function. Each time a new business joins the BestPeer++ network, a dedicated EC2 virtual server is launched for that business.Bootstrap peer is used to manage the bestpeer++. In this concept we used HadoopDB and recently proposed large scale data processing system to handle typical corporate network workloads.

5.SYSTEM FLOW DIAGRAM

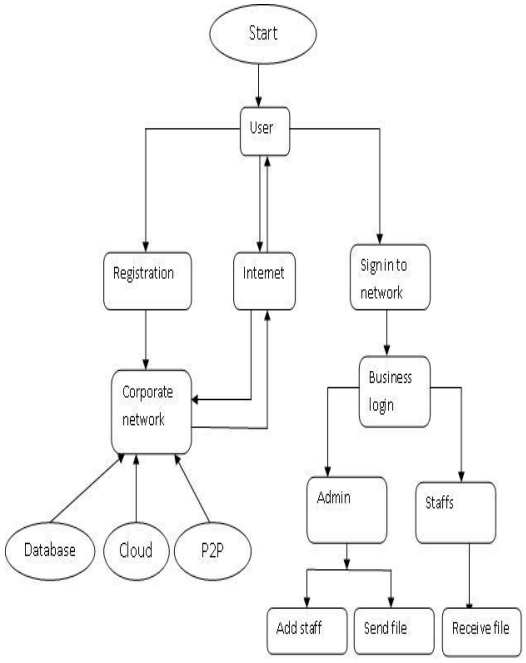


Figure 2: System Flow diagram for Online data Processing

Initailly the owner of the company need to register with the corporate network.through internet they can able to enroll them into the corporat network.after registration the corporate network send them keys.Using the key the owner can able to enter into the network.there are two login one for admin another for staff of that particular company.admin use group key for login into the corporate network.they can able to perform two operations namely add staff and send files to the staff.the staff can able to view the uploaded file.

BestPeer++ offers the offline operation.For performing offline operation thee admin the company need to download the cloud software.If the software is downloaded then the staff under that company can able to access the file even in the offline mode.

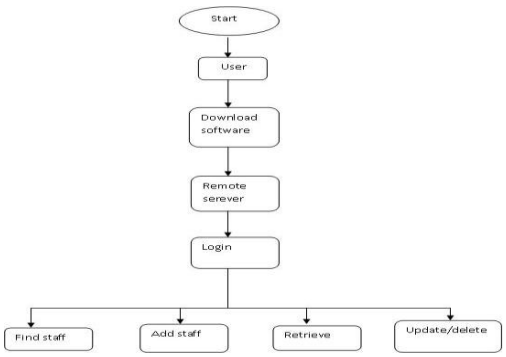


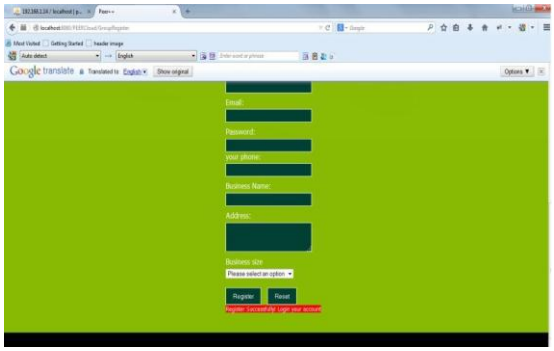
Figure 3: System Flow Diagram for Offline data processing

The modifications are updated once they enter into the business login. In addition the admin can able to retrieve the staff detail in offline mode.

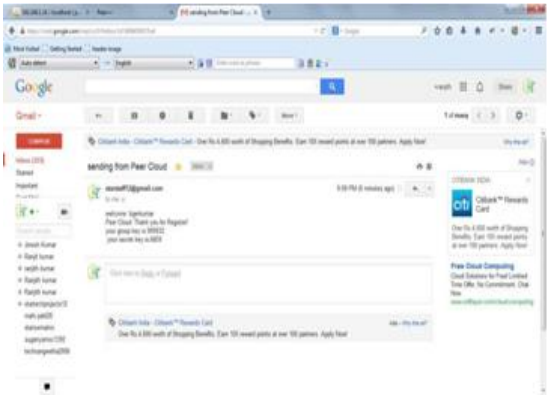
6. EXPERIMENTAL RESULTS

In BestPeer++ the data sharing is done in distributed manner.so initially the database has to be distributed across the corporate network. once database is distributed across the network the server has been started. Fig(a) shows that a company registration with the corporate netwrok. for each register company a new bestpeer++ instance is launched. after the registration process, admin of that particular company recieve a

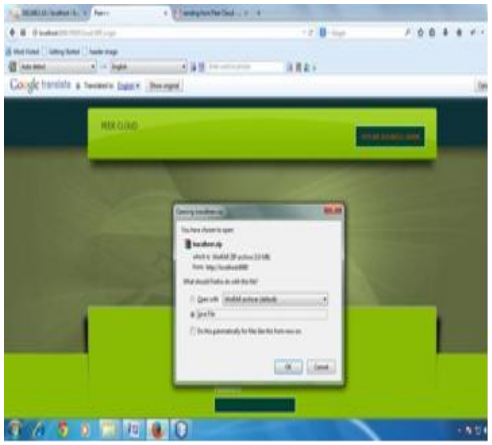
mail carring the keys namely group key and user key which shows in fig(b).After that the admin can add staff detail,share file,view staff details.All these operations can also be performed in offline mode by downloading the software which is shown in fig(c).If any staff misuse their rights



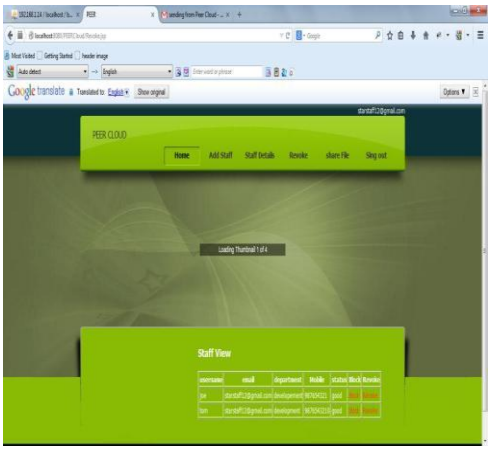
(a)



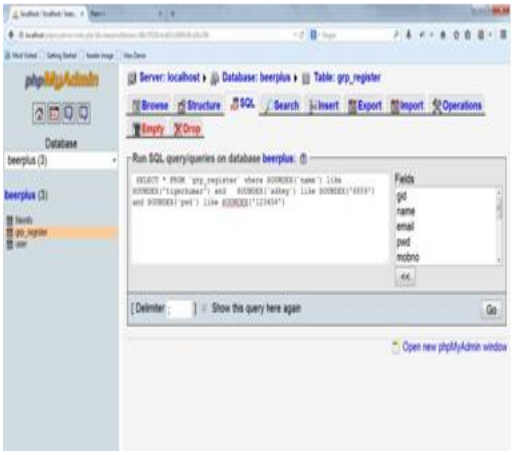
(b)



(c)



(d)



(e)

7. CONCLUSION AND FUTURE WORK

7.1Conclusion

The cloud computing, database and p2p technologies are integrated into one system for data sharing in corporate network. This combined network is called as bestpeer++. Amazon EC2 platform is bestpeer++. It provides efficient data sharing for corporate network. In this proposed system, DSA algorithm, HABE, and hadoop is used. DSA is for providing encryption, HABE for key verification and hadoop for handling large scale data. It also handles typical workload effectively. In addition we used Hierarchy attribute base encryption (HABE) for key verification. Finally bestpeer++ provides efficient data sharing and handles typical workload. This corporate network which offers low cost and increases the company revenue. The benchmarking conducted on Amazon EC2 cloud

platform shows that our system can efficiently handle typical workloads in a corporate network and can deliver near linear query throughput as the number of normal peers grows. Therefore, BestPeer++ is a promising solution for efficient data sharing within corporate networks.

7.2 Future Work

In best peer ++ concept is designed for organization. In corporate sector most of the companies are not interested to maintain its own side, selectively share the portion of this data to others. Best peer can give the solution to this problem. Best peer introduce the technique techniques for improving query performance and result quality to enhance its suitability for corporate network applications. In particular, Best Peer provides efficient distributed search services with a balanced tree structured overlay network and partial indexing scheme for reducing the index size. Best peer++ is a combination of cloud system, database management system, and peer to peer system. In our future work we implement the security concept.

In corporate sector data security is essential. The company confidential data or some data are beneficial to the opposite corporate company; we provide a security to these types of data. For ex, in banking an industry contains account details, deposit and withdraws information, Loan details etc. This information is very confidential and do not access unauthorized person. The content will be leakages by hackers are trying to hack this information while data sharing or otherwise bank workers are copying this information and sold to others. To prevent we can provide a security to this information is view only in bank system. Other system cannot access this information

8. ACKNOWLEDGEMENTS

Words are inadequate in offering thanks to the respective Head of the Institution, Head of the Department and Faculty members for giving valuable advice, guidance, monitoring and constant encouragement for technical support

REFERENCES

[1] Gang Chen, Tianlei Hu, Dawei Jiang, Peng Lu, Kian-Lee Tan, Hoang Tam Vo, and Sai Wu," BestPeer++: A Peer-to-Peer Based Large-Scale Data Processing Platform

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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 Piazza Peer Data Management Project," SIGMOD Record, vol. 32, no. 3, pp. 47-52, 2003.

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

[8] P. Rodr_ıguez-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.

[9] H.V. Jagadish, B.C. Ooi, and Q.H. Vu, "BATON: A Balanced Tree Structure for Peer-to-Peer Networks," Proc. 31st Int'l Conf. Very Large Data Bases (VLDB '05), pp. 661-672, 2005.

[10] A.Abouzeid, K. Bajda-Pawlikowski, D.J. Abadi, A. Rasin, and A.Silberschatz, "HadoopDB: An Architectural Hybrid of MapReduce and DBMS Technologies for Analytical Workloads," Proc.VLDB Endowment, vol. 2, no. 1, pp. 922-933, 2009.

[11] Oracle Inc., "Achieving the Cloud Computing Vision," White Paper, 2010.

[12] 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.

[13] K. Tzoumas, A. Deshpande, and C. S. Jensen. Sharing-aware horizontal partitioning for exploiting correlations during queryprocessing. *PVLDB*, 3(1):542–553, 2010.

[14] H. Kimura, G. Huo, A. Rasin, S. Madden, and S. B. Zdonik. CORADD: Correlation aware database designer for materialized views and indexes. *PVLDB*, 3(1):1103–1113, 2010.

[15] H. Yang et al. Map-Reduce-Merge: Simplified Relational Data Processing on Large Clusters. In SIGMOD, 2007.