International Journal of Research
International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015
Available at http://internationaljournalofresearch.org

Sharing of Data Services on Cloud Using Bestpeer ++ Strategy

Dr. K.V. Reddy1; B. Rasgna 2 & CH. Laxamana Dayakar3

[1]Professor,Dept of CSE, Marri Laxman Reddy Institute Of Technology & Management, Hyderabad, Telangana

[2] Associate Professor, Dept of CSE, Marri Laxman Reddy Institute Of Technology & Management, Hyderabad, Telangana

[3]M.Tech CSE (PG Scholar), Dept of CSE, Marri Laxman Reddy Institute Of Technology & Management, Hyderabad, Telangana

ABSTRACT

cloud computing has showed up as a popular design in managing world to back up managing large volumetric details using cluster of commodity computer systems.. A cloud computing paradigm dynamically assigns, configures, relocates and de provisions these computing resources as needed. The organizations used sharing data where they need to contribute or they share common interest. As per increasing business trends and maximum used of cloud computing, the new system evolved in new stage of growth towards cloud enabled system. In this system based on peer to peer system develop data sharing service in shared network. This system is the combination of cloud computing, databases and peer to peer based technologies. This system gives the efficiency as pay as you go manner. Here, we present BestPeer++, a system which delivers elastic data sharing services for corporate network applications in the cloud based on BestPeer—a peer-to-peer (P2P) based data management platform. By integrating cloud computing, database, and P2P technologies into one system, BestPeer++provides an economical, flexible and scalable platform for corporate network applications and delivers data sharing services to participants based on the widely accepted pay-as-you-go business model. We evaluate BestPeer++ on Amazon EC2 Cloud platform. The benchmarking results show that BestPeer++ outperforms Hadoop DB, a recently proposed large-scale data processing system, in performance when both systems are employed to handle typical corporate network workloads. The benchmarking results also demonstrate that BestPeer++ achieves near linear scalability for throughput with respect to the number of peer nodes.

I. INTRODUCTION

Cloud computing has become a necessity today when the company plans to increase capacity "or capabilities on the fly without getting to invest new infrastructure, training new individual purchase new license application, etc. based service encompasses any subscription or pay per use which extends the existing IT capabilities of the company, current time through Online. Each company maintains its own site and selectively shares a portion of its business data with the others [1]. Examples of such corporate networks include supply chain networks where organizations such as suppliers, manufacturers, and retailers collaborate with each other to achieve their very own business goals including planning production-line, making acquisition strategies and choosing marketing solutions [2].

International Journal of Research
International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015
Available at http://internationaljournalofresearch.org

From a technical perspective, the key for the success of a corporate network is choosing the right data sharing platform, a system which enables the shared data (stored and maintained by different companies) network-wide visible and supports efficient analytical queries over those data [3]. Usually, centralized data such as Data warehouse is used for data sharing, which extracts data from the internal production systems (e.g., ERP) of each company for following querying. Actually this data warehouse having some deficiency Such as, First, The share data network wants to scope up to support thousands of participants. Second, companies want to fully modify the access control rule to determine which business partners can see which part of their shared data. Most of them failed to overcome such problem. At last to increase the revenue; companies may change their business partners. Therefore, the participants may join and leave the share networks at resolve. This situation cannot be handled by physical data warehouse, to overcome such problem this designs the system for Shared Network for data sharing.

Traditionally, data sharing is achieved by building a centralized data warehouse, which periodically extracts data from the internal production systems (e.g.,ERP) of each company for subsequent querying. Unfortunately, such a warehousing solution has some deficiencies in real deployment. the corporate network needs to scale up to support thousands of participants, while the installation of a large-scale centralized data warehouse system entails nontrivial costs including huge hardware/ software investments and high maintenance cost [4, 5]. BestPeer++ achieves its query processing efficiency and is a promising approach for corporate network applications, with the following distinguished features.

Each company maintains its own site and selectively shares a portion of its business data with the others. Examples of such corporate networks include supply chain networks where organizations such as suppliers, manufacturers, and retailers collaborate with each other to achieve their very own business goals including planning production-line, making acquisition strategies and choosing marketing solutions. Integrating cloud computing, database, and peer-to-peer (P2P) technologies, BestPeer++ achieves its query processing efficiency and is a promising approach for corporate network applications, with the following distinguished features [6].

First, the corporate network needs to extent up to support thousands of participants. In the real world, most companies are not intense to invest heavily on additional information systems until they can clearly see the potential return on investment (ROI) . Second, companies want to fully modify the access control rule to determine which business partners can see which part of their shared data. Most of the data warehouse solutions fail to offer such flexibilities [8]. Finally, to increase the revenues, companies often vigorously adjust their business process and may change their business partners. Therefore, the participants may join and leave the corporate networks at resolve. The data warehouse solution has not been designed to handle such dynamicity. For decrease such problem this paper design Extended BestPeer for corporate Network.

II. EXISTING SYSTEM

Existing system was designed to work as a scalable, sharable, and secure P2P-based Data Management system for building corporate networks in which a part of association controlled by different executive domains work together in

order to reduce operation cost and pick up efficiency. Supply chain management and national healthcare network are considered as shared network application. Existing system provides an effective and efficient way to share data among to different association and also provide enterprise quality query facility, without the requirement to set up a big centralized server [7]. Data warehouse used in existing system have some problem. First, the shared network environment, most companies are not dedicated to invest deeply on additional information systems until they can clearly see the potential return on investment (ROI). Second, companies want to completely modify the access control policy to determine which business partners can see which part of their shared data [9].

A solution to detect cheating from owner side as well as CSP side is done through digital signatures. For each file owner attaches digital signature before outsourcing. The CSP first verifies digital signature of owner before storing data on cloud. In case of failed verification, the CSP rejects to store data and asks the owner to resend the correct signature [10]. If the signature is valid, both the file and signature are stored on the cloud servers. The digital signature achieves non-repudiation from the owner side. When an authorized user (the owner) requests to retrieve the data file, the CSP sends file, owner's signature and CSP's signature on (file || owner's signature). The authorized user first verifies the CSP's signature. In case of failed verification, the user asks CSP to re-perform the transmission process. If CSP's signature is valid, the user then verifies owner's signature [11]. If verification fails, this indicates the corruption of data over the cloud servers. The CSP cannot repudiate such corruption for the owner's signature is previously verified and stored by the CSP along with file. Since CSP's signature is attached with the received data, a dishonest owner cannot falsely accuse the CSP regarding data integrity. The above solution increases the storage overhead on cloud as owner's signature is stored along with the file on cloud servers. Moreover, there is an increased computation overhead, CSP has to verify signature of owner before storing file on cloud, and the authorized user verifies two signatures for each received file [12]. First, the corporate network needs to extent support thousands of participants, while the fitting of a large-scale centralized data warehouse system entails nontrivial costs including big hardware/software investments) and high preservation cost. In the environment, most companies are not dedicated to invest deeply on additional information systems until they can clearly see the potential return on investment (ROI). Second, companies want to completely modify the access control policy to determine which business partners can see which part of their shared data [13, 14].

Disadvantages of Existing System:

• Most of the data warehouse solutions fail to present such flexibilities.

• Solution has not been designed to grip such dynamicity.

III. PROPOSED SYSTEM

P2P systems are designed to support sharing of data at a coarse granularity (e.g., files, documents). In its first distinguishes between P2P systems and distributed database systems. Then define P2P distributed data management by looking at three examples (due to space constraints) of how P2P technology can be employed for distributed database applications. This will also serve to

International Journal of Research
International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015
Available at http://internationaljournalofresearch.org

motivate the need for database technology in P2P systems. This system uses the pay-as-you-go business model popularized by cloud computing. By combining cloud computing, database, and peer-to-peer (P2P) technologies containing a cloud development of Best Peer. At the last stage of its development, this System is improved with distributed access control, multiple types of indexes and pay-as you-go query processing for deliver elastic data sharing services in the cloud. The software components of the system are separated into two parts: Core and Adapter. The core having all function of data sharing and it shows that it is platform independent.

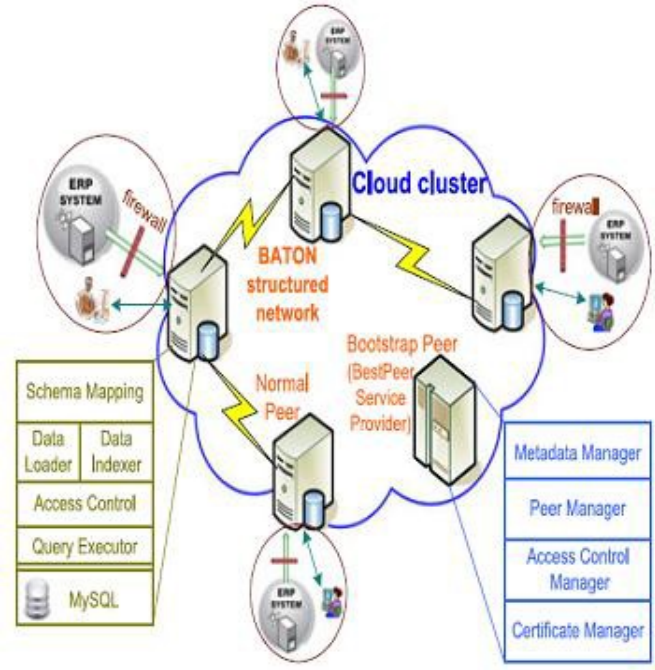

Figure 1: System Architecture

Advantages of Proposed System:

- 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.
- 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.
- BestPeer++ employs P2P technology to retrieve data between business partners. BestPeer++ is a promising solution for efficient data sharing within corporate networks.

BestPeer++ employs a hybrid design for achieving high performance query processing. The major workload of a corporate network is simple, low overhead queries. Such queries typically only involve querying a very small number of business partners and can be processed in short time. Best-Peer++ is mainly optimized for these queries. For infrequent time-consuming analytical tasks, we provide an interface for exporting the data from Best- Peer++ to Hadoop and allow users to analyze those data using Map Reduce. The software components of BestPeer++ are separated into two parts: core and adapter.

Core: The core contains all the data sharing functionalities and is designed to be platform independent.

Adapter: The adapter contains one abstract adapter which defines the elastic infrastructure service interface and a set of concrete adapter components which implement such an interface

International Journal of Research
International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015
Available at http://internationaljournalofresearch.org

through APIs provided by specific cloud service providers.

Amazon Cloud Adapter:

Amazon EC2 service to provision the database server. Each time a new business joins the BestPeer++ network, a dedicated EC2 virtual server is launched for that business. The newly launched virtual server (called a BestPeer++ instance) runs dedicated MySQL database software and the BestPeer++ software. The BestPeer++ instance is placed in a separate network security group (i.e., a VPN) to prevent invalid data access. Users can only use BestPeer++ software to submit queries to the network. Amazon relational data service (RDS) to back up and scale each BestPeer++ instance.2 The whole MySQL database is backed up to Amazon's reliable EBS storage devices in a four-minute window. 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. Initially, each BestPeer++ instance is launched as a m1.small EC2 instance (1 virtual core, 1.7 GB memory) with 5 GB storage space. With the growth of business demand, user can scale up to a more powerful EC2 instance (e.g., m1.large instance which has four virtual cores and 7.5 GB memory). In another word, there is no limitation on the resources used. The Amazon Cloud Adapter also provides automatic fail-over service.

Pay-as-you-go Query Processing BestPeer++ provides two services for the participants: the storage service and search service, both of which are charged in a pay-as-you-go model. The pay as-you-go query processing module which offers an optimal performance within the user's budget. The semantics of query processing in the BestPeer++. After data are exported from the local business system into a BestPeer++ instance, the schema mapping rules to transform them into the predefined formats. In this way, given a table T in the global schema, each peer essentially maintains a horizontal partition of it.

IV. CONCLUSION

BestPeer++, a system which delivers elastic data sharing services, by integrating cloud computing, database, and peer-to-peer technologies. The benchmark 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. This is created on Amazon EC2 cloud platform shows that our system can powerfully handle typical workloads in a corporate network and can move near linear query throughput as the number of normal peers grows. Therefore, This System having great capacity for sharing data within shared networks.

REFERENCES

[1]. 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.

[2]. 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.

[3]. Beng Chin Ooi, Yanfeng Shu, "Relational Data Sharing in Peer-based Data Management Systems." *Kian-Lee Tan Sigmod Record special issue on P2P*, 2003.

[4]. B.C. Ooi, K.L. Tan, A.Y. Zhou, C.H. Goh, Y.G. Li, C.Y. Liau, B. Ling, W.S. Ng, Y.F. Shu, X.Y. Wang, M. Zhang " PeerDB: Peering into Personal Databases." *The 2003 ACM SIGMOD Intl. Conf. on Management of Data (Demo)*. (SIGMOD 2003).

[5]. G. Chen, H. T. Vo, S. Wu, B. C. Ooi, T. "A Framework for Supporting DBMS-like Indexes in the Cloud." *Ozsu* VLDB 2011.

[6]. Heng Tao Shen, Yanfeng Shu, and Bei Yu IEEE Trans. Knowl. "Efficient Semantic-Based Content Search in P2P Network." *DataEng.*16(7):813-826(2004)

[7] 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.

[8] 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.

[9] Beng Chin Ooi, Yanfeng Shu, "Relational Data Sharing in Peer-based Data Management Systems." Kian-Lee Tan Sigmod Record special issue on P2P, 2003.

[10] B.C. Ooi, K.L. Tan, A.Y. Zhou, C.H. Goh, Y.G. Li, C.Y. Liau, B. Ling, W.S. Ng, Y.F. Shu, X.Y. Wang, M. Zhang " PeerDB: Peering into Personal Databases." The 2003 ACM SIGMOD Intl. Conf. on Management of Data (Demo). (SIGMOD 2003).

[11] G. Chen, H. T. Vo, S. Wu, B. C. Ooi, T. "A Framework for Supporting DBMS-like Indexes in the Cloud." Ozsu VLDB 2011.

[12] 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.

[13] 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.

[14] H. Garcia-Molina and W.J. Labio, "Efficient Snapshot Differential Algorithms for Data Warehousing," technical report, Stanford Univ., 1996.

[15] Google Inc., "Cloud Computing-What is its Potential Value for Your Company?" White Paper, 2010.

[16] H.V. Jagadish, B.C. Ooi, K.-L. Tan, Q.H. Vu, and R. Zhang, "Speeding up Search in Peer-to-Peer Networks with a Multi-Way Tree Structure," Proc. ACM SIGMOD Int'l Conf. Management of Data, 2006.