www.ijraset.com
IC Value: 13.98
Volume 3 Issue VIII, August 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Prime Node: An Optimized System for Data Sharing In Corporate Networks

S.Sandeep[1], V.Padmavathi[2], T.Sujilatha[3]
CSE Dept, GKCE Sullurpeta, Nellore, A.P, India

Abstract: ***In this paper, we implemented a technique called optimized system for sharing data in networks; the amount of sensitive data produced by many administrations is out stepping their storage ability. The management of such large amount of data is quite expensive due to the requirements of high storage capacity and qualified recruits, so the data sharing in network get issue, Ultimately reply to the ever changing business demands and the exterior of Cloud Computing techniques, Best-Peer has develop into cloud-enabled Extended Best-Peer system. Extended Best-Peer, a system which delivers flexible data sharing services for the commercial network applications in the cloud based on Best-Peer a peer-to-peer (P2P) based data management platform. By Combining cloud computing, database, and P2P technologies, Extended Best-Peer achieves its query processing efficiency in a pay-as-you-go manner. We evaluate Extended Best-Peer on Amazon EC2 Cloud platform. The benchmarking results show that Extended Best-Peer better than Hadoop-DB, in presentation when both systems are working to maintain typical corporate network.***

Keywords: Peer-to-peer systems, cloud computing, Map Reduce, query processing, index

I. INTRODUCTION

The original Best-Peer system attempts to exploit peer-to-peer (P2P) technologies for distributed applications. Best-Peer was designed to work as a scalable, sharable, and secure P2P-based Data Management system with full functionalities for building corporate networks in which a group of organizations controlled by different administrative domains collaborate with each other in order to reduce operation cost and improve productivity. Examples of corporate network applications include supply chain management and national healthcare network. Best-Peer provides an effective and efficient way to share data belonging to different

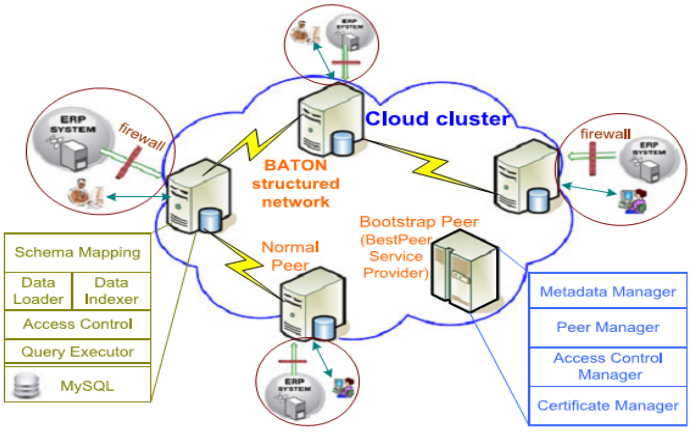


Fig. 1. The BestPeer++ network deployed on Amazon Cloud offering.

organizations and provide enterprise quality query facility, without the need to set up a huge centralized server, As an in time response to the ever changing business demands and the emergence of Cloud Computing techniques, Best-Peer has evolved into its new stage of development the cloud enabled Best-Peer++ system, By integrating cloud computing, database, and P2P technologies, Best-Peer++ achieves its query processing efficiency in a "pay-as-you-go" manner and is a promising approach for corporate network applications,

Specifically, Best-Peer++ is deployed as a service in the cloud. To form a corporate network, companies simply register their sites with the Best-Peer++ service provider, launch Best-Peer++ instances in the cloud and finally export data to those instances for

sharing. Best-Peer++ 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 Best-Peer++ instance 担 hours and storage capacity. The Best-Peer++ service provider elastically scales up the running instances and makes them always available. Notably, Best-Peer++ 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, Best-Peer++ also inherits its predecessor's nice features such as support for semi-automatic schema mapping and data mapping, efficient distributed query processing, effective system load balancing and other functionalities that a corporate network requires.

A. History of Best Peer

Best-Peer++ is a Open Source, Best-Peer has been constantly developed since 2006. Now in its last stage of development, Best-Peer++ integrates cloud computing, database, and P2P technologies to achieve high query processing efficiency in a “pay-as-you-go” manner. We publish basic layers as open source from which developers can build up their own peer based applications, The source code of Best-Peer++ software's basic components is distributed here. These components work as a infrastructure to establish a peer based corporate network with simple GUIs. Developers are freely to add more features and build up their own applications; The Java documents for these components are available here. As an illustration, here is an example on how to develop a new index function into Best-Peer++ open source, the last two decades have witnessed a growing need for information sharing. With more affordable commodity hardware’s and higher communication bandwidth, data sharing has become an exciting killer application of the Internet, Peer-to-peer (P2P) architecture aims at extending the current distributed computing design to accommodate dynamic resources such as information and computing power. In such environments, the peers are autonomous with highly dynamic behaviors and act as servers and consumers at the same time. These characteristics bring many opportunities for exploitation but with technical challenges to be solved as well.

Now in the last stage of its evolution, Best-Peer++ is enhanced with distributed access control, multiple types of indexes, and pay-as-you-go query processing for delivering elastic data sharing services in the cloud, The software components of Best-Peer++ are separated into two parts: core and adapter. The core contains all the data sharing functionalities and is designed to be platform independent. 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 through APIs provided by specific cloud service providers (e.g., Amazon). We adopt this "two level" design to achieve portability. With appropriate adapters, Best-Peer++ can be ported to any cloud environments (public and private) or even non cloud environment (e.g., on premise data center). The architecture of Best-Peer++ is depicted in Figure 1.

Specifically, highlights of Best-Peer++ are:

Amazon Cloud Adapter: The key idea of Best-Peer++ is to use dedicated database servers to store data for each business and organize those database servers through P2P network for data sharing. The Amazon Cloud Adapter provides an elastic hardware infrastructure for Best-Peer++ to operate on by using Amazon Cloud services.

The Best-Peer++ Core: The Best-Peer++ core contains all platform independent logic, including query processing and P2P overlay. It runs on top of adapter and consists of two software components: bootstrap peer and normal peer.

Adaptive Query Processor: Best-Peer++ 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.

Companies of the same industry are often connected to a corporate network for association purposes. Each company maintains its own site and selectively shares a part of its business data with the others include supply chain networks where organizations such as supplier, manufacturer, and retailer cooperate with each accomplish their own business goals such as planning production-line, making achievement strategies and choose marketing solutions. As per technical perspective, selecting the right data sharing platform for corporate network is very important; a system which enables the pooled data supports capable logical queries over

Volume 3 Issue VIII, August 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

those data, Traditionally, data sharing is achieve by building a centralized data warehouse, which regularly extracts data from the internal production systems (e.g., ERP) of each company for following querying. Such a warehousing solution has some deficiency in real consumption. 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. 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 Best-Peer for corporate Network.

As an in-time response to the ever changing business demands and the appearance of Cloud Computing techniques, Best-Peer has developed into its new stage of development-the cloud-enabled Extended Best-Peer system. The structure of the system shows in fig 2.

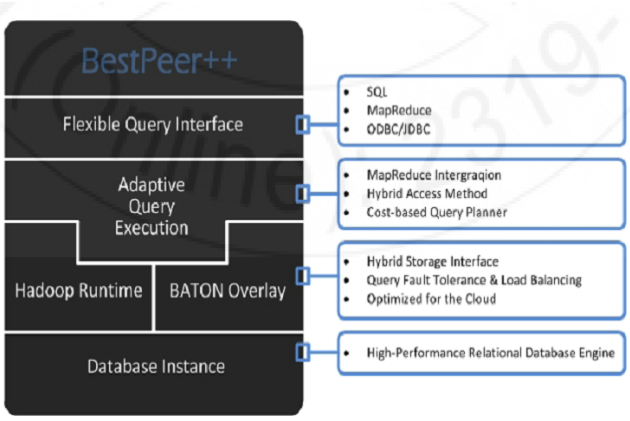


Figure 2: The Structure of Extended Best-Peer System

By integrating cloud computing, database, and P2P technologies, Extended Best-Peer achieves its query processing competence in a pay-as-you-go manner and is a capable approach for corporate network application. In summary, this paper shows that design of Extended Best-Peer system that provides inexpensive, Flexible solutions for corporate network. We demonstrate the competence of Extended Best-Peer by benchmarking Extended Best-Peer against Hadoop-DB, a propose approach for data sharing applications. The results show that for simple queries, the performance of Extended Best-Peer is significantly better than Hadoop-DB, The rest of the paper is organized as follows. Section 2 presents the Literature survey of the Extended Best-Peer system including Existing System and Propose System. We then describe the design of Extended Best-Peer core components, with system architecture in Section 3 and the Section 4 shows that evolution of the performance of system.

II. LITERATURE SURVEY

This section shows the existing System and its disadvantages then overview of the propose System.

A. Existing System

The original Best-Peer system attempt to develop peer-to-peer (P2P) technologies for Corporate Networks. Best-Peer was designed to work as a scalable, sharable, and secure P2P-based Data Management system with full functionalities 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. Corporate network applications such as supply chain management and national healthcare network. Best-Peer provides an effective and efficient way to share data belong to different association and provide enterprise quality query facility, without the requirement to set up a big centralized server. As per changing business demands and the coming out of Cloud Computing techniques, Best-Peer has developed into its new stage that is cloud-enabled Extended Best-Peer system. Such a warehousing solution has some disadvantages in real consumption, 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.

Disadvantages of Existing System:

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

Solution has not been designed to grip such dynamicity.

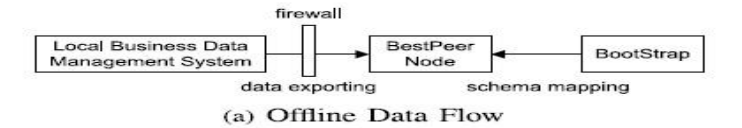
B. Propose System

The main contribution of this paper is the design of Extended Best-Peer system that provides well-organized, elastic and scalable solution for corporate network. The unique challenges pose by sharing and processing data in an inter-businesses environment and designed Extended Best-Peer, a system which give elastic data sharing services, by including cloud computing, database, and peer-to-peer technologies for corporate Network. Best-Peer's product is the Extended Best-Peer Platform, which combines the powerful Map-Reduce processing model with the predictable P2P database technologies. Extended Best-Peer's advanced technology features a hybrid architecture that brings the parallelism of Map-Reduce to the latest development in RDBMS research. Extended Best-Peer is based on our decade's research on P2P database system, and offers an accelerate data processing engine and a more flexible portability via the approval of Map-Reduce framework and Software-as-a-Service (SaaS) paradigm.

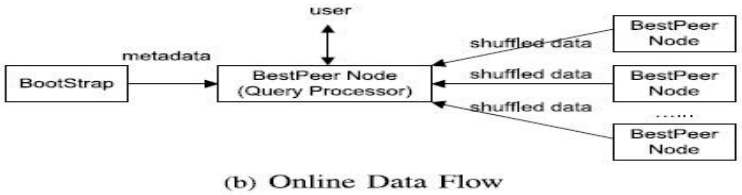
In compare to the "Hadoop Connector" approach employed by many MPP investigative database vendors, Extended Best-Peer uses Hadoop as the parallelization layer to make possible its universal query processing, with each node running a database occasion. Consolidate predictable database query processing and Map-Reduce into a single platform considerably reduces TCO, eliminate performance bottleneck from both mechanism, and allows for richer analytics through expenditure of different data types. Additionally, Extended Best-Peer's combined architecture and supple schema capabilities reduce the complexities associated with rising analytic use cases – including graph analysis, clustering, and classification – while significantly growing show and extent. Explicitly, Extended Best-Peer is deploying as a service in the cloud. To form a corporate network, companies register with the site Extended Best-Peer service provider, initiate Extended Best-Peer instances in the cloud and at last export data to those instances for sharing. Extended Best-Peer adopt the pay-as-you-go business model popularized by cloud computing. The total cost of possession is therefore significantly summary while companies do not have to buy any hardware/software in move on. The Extended Best-Peer service provider elastically grows up the running instance and makes them always available. For occasional sustained analytical tasks, we provide an border for exporting the data from Extended Best-Peer to Hadoop and allow users to analyze those data using Map-Reduce. Extended Best-Peer also inherit its predecessor's nice kind such as support for semi-automatic schema mapping and data mapping, well-organized dispersed query processing, successful system load balancing and other functionalities that a corporate network requires. By combining cloud computing, database, and peer-to-peer (P2P) technologies.

III. COMPONENT OF PROPOSE SYSTEM

Extended Best-Peer, a cloud enabled evolution of Best-Peer. At the last stage of its development, Extended Best-Peer 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 Extended Best-Peer are separated into two parts: core and adapter. The Architecture is shown in fig. 2. The core contains all the data sharing functionalities and is planned to be platform independent, the adapter contains one abstract adapter which defines the elastic transportation service interface and a set of tangible adapter components which implement such an interface through APIs provided by specific cloud service providers (e.g., Amazon). To achieve portability we developed "two level" design. With appropriate adapters, Extended Best-Peer can be portable to any cloud environments (public and private) or even non-cloud environment (e.g., on -premise data center). We have implemented an adapter for Amazon cloud platform. In what follows, we first present this adapter and then describe the core components



(a) Offline Data Flow



(b) Online Data Flow

Figure 3: Data Flow in Extended Best-Peer

Advantages of Propose System

This system can powerfully handle characteristic workloads in a corporate network.

Extended Best-Peer adopt the pay-as-you-go business model famous by cloud computing. as an optional, what they use of Extended Best-Peer instance's hours and storage capacity they pay for it.

Extended Best-Peer extends the role-based access control for the natural dispersed environment of commercial networks.

Extended Best-Peer employs P2P technology to retrieve data between business partners.

Extended Best-Peer is a great solution for data sharing within corporate networks.

IV. BENCHMARKING

This section shows evolution of the performance and throughput of Extended Best-Peer on Amazon cloud platform. For the performance benchmark, we evaluate the query latency of Best-Peer++ with HadoopDB using five queries selected from typical corporate network applications workloads.[1] For the throughput benchmark, we produce a simple supply-chain network consisting of suppliers and retailers and study the query throughput of the system.

A. Performance Benchmarking

This benchmark compares the performance of Extended Best-Peer with HadoopDB. We choose Hadoop-DB as our benchmark target since it is an alternative promising solution for our problem and adopts architecture similar to ours. Comparing the two systems (i.e., Hadoop-DB and Extended Best-Peer) reveal the performance gap between a general data Warehousing system and a data sharing system specially designed for corporate network applications.

B. Throughput Benchmarking

This section studies the query throughput of Extended Best-Peer. HadoopDB is not designed for high query throughput; therefore, we intentionally skip the results of HadoopDB and only present the results of Extended Best-Peer. We conduct two tiers of benchmark evaluation for the performance and scalability of Extended Best-Peer, respectively

V. CONCLUSION

This paper define exclusive challenges pose by contribution and open-handed out data in an inter-businesses environment and planned Extended Best-Peer, a system which deliver elastic data sharing services, by Containing cloud computing, database, and peer-to-peer technologies. The standard conducted 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, Extended Best-Peer is great solution for capable data sharing within corporate networks.

We plan to further extend the work is to enhance the usability of conventional P2P networks, database community has proposed a series of PDBMS (Peer-to-Peer Database Manage System) by integrating the state-of-art database techniques into the P2P systems, the techniques of PDBMS are also adopted in cloud systems, BestPeer++ is different from the systems based on the Map Reduce and Hadoop framework (e.g., HadoopDB and Hadoop++). Hadoop-based systems are designed to process large-scale data sets in batch mode. They efficiently process aggregate queries by exploiting the parallelism.

VI. ACKNOWLEDGEMENT

I like to thank our PRINCIPAL, HOD and OTHER FACULTIES for their valuable comments and helpful suggestions, to make this

paper publish and also they help for my academic growth

REFERENCES

[1] K. Aberer, A. Datta, and M. Hauswirth, "Route Maintenance Overheads in DHT Overlays," in 6th Workshop Distrib. Data Struct., 2004.
[2] A. Abouzeid, K. Bajda-Pawlikowski, D.J. Abadi, A. Rasin, and A. Silberschatz, "HadoopDB: An Architectural Hybrid of MapRe-duce and DBMS Technologies for Analytical Workloads," Proc. VLDB Endowment, vol. 2, no. 1, pp. 922-933, 2009.
[3] C. Batini, M. Lenzerini, and S. Navathe, "A Comparative Analysis of Methodologies for Database Schema Integration," ACM Com-puting Surveys, vol. 18, no. 4, pp. 323-364, 1986.
[4] D. Bermbach and S. Tai, "Eventual Consistency: How Soon is Eventual? An Evaluation of Amazon s3's Consistency Behavior," in Proc. 6th Workshop Middleware Serv. Oriented Comput. (MW4SOC '11), pp. 1:1-1:6, NY, USA, 2011.
[5] B. Cooper, A. Silberstein, E. Tam, R. Ramakrishnan, and R. Sears, "Benchmarking Cloud Serving Systems with YCSB," Proc. First ACM Symp. Cloud Computing, pp. 143-154, 2010.
[6] G. DeCandia, D. Hastorun, M. Jampani, G. Kakulapati, A. Lakshman, A. Pilchin, S. Sivasubramanian, P. Vosshall, and W. Vogels, "Dynamo: Amazon's Highly Available Key-Value Store," Proc. 21st ACM SIGOPS Symp. Operating Systems Princi-ples (SOSP '07), pp. 205-220, 2007.
[7] J. Dittrich, J. Quian_e-Ruiz, A. Jindal, Y. Kargin, V. Setty, and J. Schad, "Hadoop++: Making a Yellow Elephant Run Like a Chee-tah (without it Even Noticing)," Proc. VLDB Endowment, vol. 3, no. 1/2, pp. 515-529, 2010.
[8] H. Garcia-Molina and W.J. Labio, "Efficient Snapshot Differential Algorithms for Data Warehousing," technical report, Stanford Univ., 1996.
[9] Google Inc., "Cloud Computing-What is its Potential Value for Your Company?" White Paper, 2010.
[10] 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.
[11] 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.
[12] http://thanigavelm.blogspot.in/2014/01/latest-technologies-in-computer-science\Internet Traffic Classification in TIE.html
[13] H.V. Jagadish, B.C. Ooi, K.-L. Tan, C. Yu, and R. Zhang, "iDistance: An Adaptive B+-Tree Based Indexing Method for Nearest Neighbor Search," ACM Trans. Database Systems, vol. 30, pp. 364-397, June 2005.
[14] 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.
[15] A. Lakshman and P. Malik, "Cassandra: Structured Storage Sys-tem on a P2P Network," Proc. 28th ACM Symp. Principles of Distrib-uted Computing (PODC '09), p. 5, 2009.
[16] 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.
[17] A. Dainotti, W. De Donato, A. Pescape, and P. Salvo Rossi, "Classification of network traffic via packet-level hidden markov models," IEEE Global Telecommunications Confer-ence, 2008, pp. 1–5.
[18] Oracle Inc., "Achieving the Cloud Computing Vision," White Paper, 2010.
[19] V. Poosala and Y.E. Ioannidis, "Selectivity Estimation without the Attribute Value Independence Assumption," Proc. 23rd Int'l Conf. Very Large Data Bases (VLDB '97), pp. 486-495, 1997.

AUTHORS

Mr S.SANDEEP, Pursuing my M.Tech (CSE) in GOKULAKRISHNA COLLEGE OF ENGINEERING, Sullurpeta , and my area of interest is NETWORKING & CLOUD COMPUTING, E-mail id: sanchi.sandeep59@gmail.com

Mrs. V.PADMAVATHI, ASST PROFESSOR in the Department of CSE at GOKULA KRISHNA COLLEGE OF ENGINEERING, Sullurpeta, and my area of interest is OS, SPM & NETWORKS etc., E-mail id: vanka.padma.vathi@gmail.com

Miss T.SUJILATHA., ASST PROFESSOR in the Department of CSE at GOKULA KRISHNA COLLEGE OF ENGINEERING, Sullurpeta, and my area of interest is networking, Cloud computing, WSN and DMDW etc., E-mail id: illu.suji@gmail.com