IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727,
PP 28-30
www.iosrjournals.org

Best Peer++: A Peer-To-Peer Based Large-Scale Data Processing Platform

Tantak Ashlesha[1], Sugandhi Renuka[2], Thorve Rutuja[3], Mungse Abhilasha[4] G. M. Dahane[5],

[1,2,3,4,5]*(Dept of IT, P.D.V.V.P.C.O.E Ahmednagar, Pune University(MS), India)*

Abstract: *For sharing of information among the participating companies and facilitating collaboration in a certain industry sector where companies share a common interest the corporate network is used. It reduces companies operational costs and increases the revenues. the inter-company data sharing and processing poses unique challenges to such a data management system as like scalability, performance, throughput, and security. In our system we present BestPeer++, a system that provides elastic data sharing services for corporate network applications in the cloud based on BestPeer++ a peer-to-peer (P2P) based data management platform. BestPeer++ provides an economical, flexible and scalable platform for corporate network applications and delivers data sharing services to participated companies based on the widely accepted pay-as-you-go business model by integrating cloud computing, database, and P2P technologies into one system. On Amazon EC2 Cloud platform we evaluate BestPeer++. The benchmarking results show that BestPeer++ performs HadoopDB, a recently proposed large-scale data processing system, in performance when both systems are employed to handle corporate network workloads. The benchmarking results also demonstrate that BestPeer++ gains near linear+ scalability for throughput with respect to the number of peer nodes*

Keywords: *Peer-to-peer systems, cloud computing, MapReduce, query processing, index.*

I. Introduction

COMPANIES of the same industry sector are often connected into a corporate network for collaboration purposes. 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 personal business goals including planning production-line, making acquisition strategies and choosing marketing solutions. 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. 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. First, 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 (a.k.a total cost of ownership) and high maintenance cost (a.k.a total cost of operations) . In the real world, most companies are not keen to invest heavily on additional information systems until they can clearly see the potential return on investment(ROI) . Second, companies want to fully customize the access control policy to determine which business partners can see which part of their shared data. Unfortunately, most of the data warehouse solutions fail to offer such flexibilities. Finally, to maximize the revenue BestPeer++: A Peer-to-Peer Based Large-Scale Data Processing Platform, companies often dynamically adjust their business process and may change their business partners. Therefore, the participants may join and leave the corporate networks.1.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 Best-Peer++ 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 nothave to buy any hardware/software in advance. Instead, they pay for what they use in terms of BestPeer++ instances hours and storage capacity.2.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.3.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 efficientretrieval.4.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, They provide an interface for exporting the data from Best- Peer++ to Hadoop and allow users to analyze those data using MapReduce.

II. Problem Statement

2.1 Existing System

Such a warehousing solution has some deficiencies in real deployment. First, 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 (a.k.a total cost of ownership) and high maintenance cost (a.k.a total cost of operations). In the real world, most companies are not keen to invest heavily on additional information systems until they can clearly see the potential return on investment (ROI). Second, companies want to fully customize 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 offer such flexibilities. Solution has not been designed to handle such dynamicity.

2.2 Proposed System:

The main contribution of this paper is the design of BestPeer++ system that provides economical, 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. The unique challenges posed by sharing and processing data in an inter-businesses environment and proposed BestPeer++: A Peer-to-Peer Based Large-Scale Data Processing Platform BestPeer++, a system which delivers elastic data sharing services, by integrating cloud computing, database, and peer-to-peer technologies.

2.3 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 Best-Peer++ instances 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.

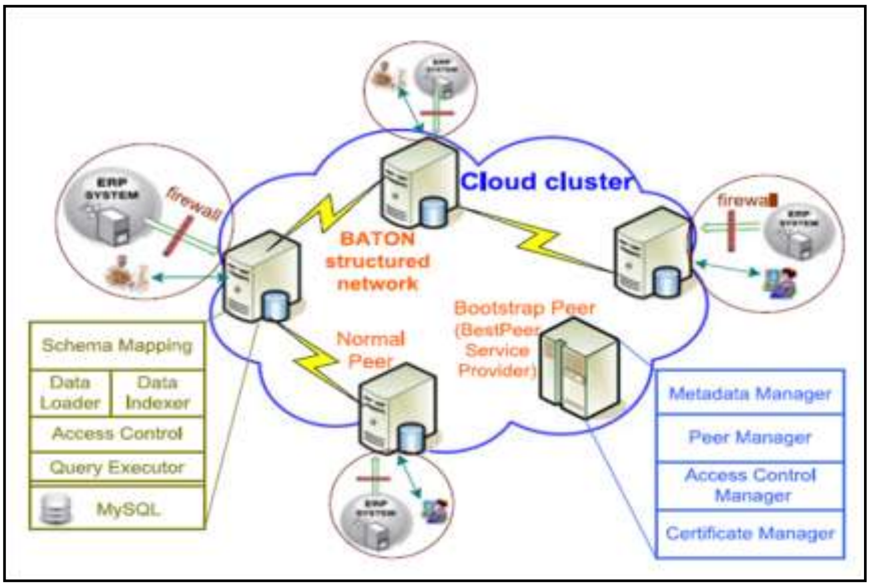

Fig.1. Cloud Cluster

III. Conclusion

We have discussed the unique challenges posed by sharing and processing data in an inter-businesses environment and proposed BestPeer++, a system can delivers elastic data sharing services, by integrating cloud computing, database, and peer-to-peer technologies. The benchmark based 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.

Acknowledgements

Inspiration and guidance are invaluable in every aspect of life, especially in the field of academics, which I have received from our beloved and respected guide **Prof.G.M.Dahane.** and Head of Department **Prof.D.A.Vidhate** who has put his careful guidance through which I can complete my project work. Also I want to express my gratitude to his untiring devotion. He undoubtedly is the member of artistic gallery who is masters in all respect.I wish to express my sincere thanks to the departmental staff members for their support.I would also like thank our Principal **Dr. H.N.KUDAL** for his kind support I would also like to thank to my friends for listening to my ideas, asking questions and providing feedback, suggestions and helping for improving my ideas. At last, I would like to express my gratitude towards my parents and my family members who are inspiration of my life.

References

[1] K. Aberer, A. Datta, and M. Hauswirth,Route “Maintenance Overheads in DHT Overlays”, in 6th Workshop Distrib. Data Struct., 2004.

[2] 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.

[3] C.Batini, M.Lenzerini, and S. Navathe,” A Comparative Analysis of Methodologies for Database Schema Integration”, ACM Computing 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 s3s 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.