ISSN 2347-3983

International Journal of Emerging Trends in Engineering Research, Vol.3. No.10, Pages : 93-98 (2015)
Special Issue of ICACSSE 2015 - Held on October 30, 2015 in St. Ann's College of Engineering & Technology, Chirala, AP, India
http://www.warse.org/IJETER/static/pdf/Issue/icacsse2015sp17.pdf

EFFICIENT PEER TO PEER BASED DESIGN FOR LARGE SCALE DATA PROCESSING PLATFORM

Mr. P.Vandana Sagar[1], Mr. S. Anil Kumar.[2]

[1]*II M.Tech. - II Sem., Dept. of CSE, St. Ann's College of Engineering. & Technology. Chirala, Andhra Pradesh -,523 187 INDIA,*
sagarvandan@gmail.com

[2]*Associate Professor , Dept. of CSE, St. Ann's College of Engg. & Tech., Chirala, A. P, INDIA*
anilkumar.sathupati1980@gmail.com

ABSTRACT:-

The corporate network is usually used for sharing data among the collaborating corporations and facilitating collaboration in a very bound trade sector wherever corporations share a typical interest. It will effectively facilitate the businesses to scale back their operational prices and increase the revenues. However, the inter-company knowledge sharing and process poses distinctive challenges to such an information management system as well as measurability, performance, throughput, and security. During this paper, we have a tendency to gift BestPeer++, a system that delivers elastic knowledge sharing services for company network applications within the cloud supported BestPeer – a peer-to-peer (P2P) primarily based knowledge management platform. By group action cloud computing, database, and P2P technologies into one system, BestPeer++ provides a cheap, versatile and climbable platform for company network applications and delivers knowledge sharing services to participants supported the wide accepted pay-as-you-go business model. We have a tendency to assess BestPeer++ on Amazon EC2 Cloud platform. The benchmarking results show that BestPeer++ outperforms HadoopDB, a recently planned large-scale processing system, in performance once each systems square measure used to handle typical company network workloads. The benchmarking results additionally demonstrate that BestPeer++ achieves close to linear measurability for turnout with relation to the amount of peer nodes.

INTRODUCTION:-

Companies of a similar business sector are usually connected into a company network for collaboration functions. Every company maintains its own web site and by selection shares some of its business knowledge with the others. Samples of such company networks embrace offer chain networks wherever organizations like suppliers, makers, and retailers collaborate with one another to realize their terribly own business goals together with designing production-line, creating acquisition ways and selecting selling solutions. From a technical perspective, the key for the success of a company network is selecting the proper knowledge sharing platform, a system that permits the shared knowledge (stored and maintained by totally different companies) network-wide visible and supports

ISSN 2347-3983

International Journal of Emerging Trends in Engineering Research, Vol.3. No.10, Pages : 93-98 (2015)
Special Issue of ICACSSE 2015 - Held on October 30, 2015 in St. Ann's College of Engineering & Technology, Chirala, AP, India
http://www.warse.org/IJETER/static/pdf/Issue/icacsse2015sp17.pdf

economical analytical queries over that knowledge. Historically, knowledge sharing is achieved by building a centralized knowledge warehouse that sporadically extracts knowledge from the interior production systems (e.g., ERP) of every company for resulting querying. Sadly, such a deposit answer has some deficiencies in real preparation.

From a technical perspective, the key for the success of a company network is selecting the correct knowledge sharing platform, a system that allows the shared knowledge (stored and maintained by completely different companies) network-wide visible and supports economical analytical queries over those knowledge. Historically, knowledge sharing is achieved by building a centralized knowledge warehouse that sporadically extracts knowledge from the interior production systems (e.g., ERP) of every company for succeeding querying. Sadly, such a deposit answer has some deficiencies in real readying.

First, the company network must proportion to support thousands of participants, whereas the installation of a largescale centralized knowledge warehouse system entails nontrivial prices together with immense hardware/software investments (a.k.a Total value of Ownership) and high maintenance value (a.k.a Total value of Operations) [12]. Within the universe, most firms aren't keen to speculate heavily on extra info systems till they'll clearly see the potential come on investment (ROI) [16]. Second, firms wish to totally customise the access management policy to see that business partners will see that a part of their shared knowledge. Sadly, most of the info warehouse solutions fail to supply such flexibilities. Finally, to maximise the revenues, firms usually dynamically alter their business method and should amendment their business partners. Therefore, the participants might be part of and leave the company networks at can. The info warehouse answer has not been designed to handle such dynamicity.

To address the same issues, this paper presents BestPeer++, a cloud enabled information sharing platform designed for company network applications. By desegregation cloud computing, database, and peer-to-peer (P2P) technologies, Best-Peer++ achieves its question process potency and may be a promising approach for company network applications, with the subsequent distinguished options.

- BestPeer++ is deployed as a service in the cloud.To form a corporate network, companies simplyregister their sites with the BestPeer++ service pro-bestPeer++ instances in the cloudand finally export data to those instances for sharing. BestPeer++ adopts the pay-as-you-go businessmodel popularized by cloud computing [5]. Thetotal cost of ownership is therefore substantiallyreduced since companies do not have to buy anyhardware/software in advance. Instead, they payfor what they use in terms of BestPeer++ instance'shours and storage capacity.

- BestPeer++ extends the role-based access management for the inherent distributed surroundings of company networks. Through an internet console interface, firms will simply put together their access

ISSN 2347-3983

International Journal of Emerging Trends in Engineering Research, Vol.3. No.10, Pages : 93-98 (2015)
Special Issue of ICACSSE 2015 - Held on October 30, 2015 in St. Ann's College of Engineering & Technology, Chirala, AP, India
http://www.warse.org/IJETER/static/pdf/Issue/icacsse2015sp17.pdf

management policies and forestall unsought business partners to access their shared information.

- BestPeer++ employs P2P technology to retrieve information between business partners. BestPeer++ instances area unit organized as a structured P2P overlay network named BATON [9]. The info area unit indexed by the table name, column name and information vary for economical retrieval.

- BestPeer++ employs a hybrid style for achieving high performance question process. The foremost work of a company network is easy, low-overhead queries. Such queries generally solely involve querying a really tiny variety of business partners and may be processed briefly time. BestPeer++ is principally optimized for these queries. For rare long analytical tasks, we offer associate degree interface for commerce the information from BestPeer++ to Hadoop and permit users to research those data exploitation MapReduce.

In summary, the main contribution of this paper is the design of BestPeer++ system that provides economical, flexible and scalable solution for corporate network applications. We demonstrate the efficiency of BestPeer++ by benchmarking BestPeer++ against HadoopDb [2], a recently proposed large scale data processing system, over a set of queries designed for information sharing applications.

Best Peer data management platform:-

While acceptable P2P arrangement has not been advised for action applications, the ultimate ambition of BestPeer is to accompany the state-of-art database techniques into P2P systems. In its aboriginal stage, Best-Peer employs baggy arrangement and advice retrieval address to bout columns of altered tables automatically [11]. After defining the mapping functions, queries can be beatific to altered nodes for processing. In its additional stage, BestPeer introduces an alternation of techniques for convalescent concern achievement and aftereffect superior to enhance its adequacy for accumulated arrangement applications. In particular, BestPeer provides able broadcast seek casework with a counterbalanced timberline structured bury arrangement [9] and fractional indexing arrangement [20] for abbreviation the basis size. Moreover, BestPeer develops adaptive accompany concern processing [21] and broadcast online accession [19] techniques to accommodate able concern processing.

BestPeer++, a cloud enabled evolution of BestPeer:-

Now in the endure date of its evolution, BestPeer++ is added with broadcast admission control, assorted types of indexes, and pay-as-you-go concern processing for carrying adaptable abstracts administration casework in the cloud. The software apparatus of BestPeer++ are afar into two parts: amount and adapter. The amount contains all the abstracts administration functionalities and is advised to be belvedere independent. The adapter contains one abstruse adapter which defines the adaptable basement account interface and a set of accurate adapter apparatus which apparatus such an interface

ISSN 2347-3983

International Journal of Emerging Trends in Engineering Research, Vol.3. No.10, Pages : 93-98 (2015)
Special Issue of ICACSSE 2015 - Held on October 30, 2015 in St. Ann's College of Engineering & Technology, Chirala, AP, India
http://www.warse.org/IJETER/static/pdf/Issue/icacsse2015sp17.pdf

through APIs provided by specific billow account providers (e.g., Amazon). We accept this "two-level" architecture to accomplish portability. With adapted adapters, BestPeer++ can be ported to any billow environments (public and private) or even non-cloud ambiance (e.g., on-premise abstracts center). Currently, we accept implemented an adapter for Amazon billow platform. In what follows, we aboriginal present thisadapter and again call the amount components.

Amazon Cloud Adapter:

The key abstraction of BestPeer++ is to use committed database servers to abundance abstracts for anniversary business and adapt those database servers through P2P arrangement for abstracts sharing. The Amazon Cloud Adapter provides an adaptable accoutrements basement for BestPeer++ to accomplish on by application Amazon Cloud services. The basement account that Amazon Cloud Adapter delivers includes launching/terminating committed MySQL database servers and monitoring/backup/auto ascent those servers. We use Amazon EC2 account to accouterment the database server. Anniversary time a new business joins the BestPeer++ network; we barrage a committed EC2 basic server for that business. The anew launched basic server (called a BestPeer++ instance) runs committed MySQL database software and the BestPeer++ software. The BestPeer++ instance is placed in a abstracted arrangement aegis accumulation (i.e., a VPN) to anticipate invalid abstracts access. Users can alone use BestPeer++ software to abide queries to the network. We use Amazon Relational Abstracts Account (RDS) to aback up and calibration anniversary BestPeer++ instance 2. The accomplished MySQL database is backed up to Amazon's reliable EBS accumulator accessories in a four minute window. There will be no account arrest during the action back the advancement operation is performed asynchronously. The ascent arrangement consists of two dimensions: processing and storage. The two ambits can be apart scaled up. Initially, anniversary BestPeer++ instance is launched as a m1.small EC2 instance (1 virtual core, 1.7 GB memory) with 5GB storage space. If the workload grows, the business can scale up the processing and the storage. There is no limitation on the resources used. Finally, the Amazon Cloud Adapter also provides automatic fail-over service. In a BestPeer++ network, a special Best-Peer++ instance (called bootstrap peer) monitors the healthof all other BestPeer++ instances, by querying the Amazon Cloud Watch service. If the bootstrap peer finds another instance fails to respond (e.g., crashed), it calls Amazon Cloud Adapter to perform fail-over for that instance.

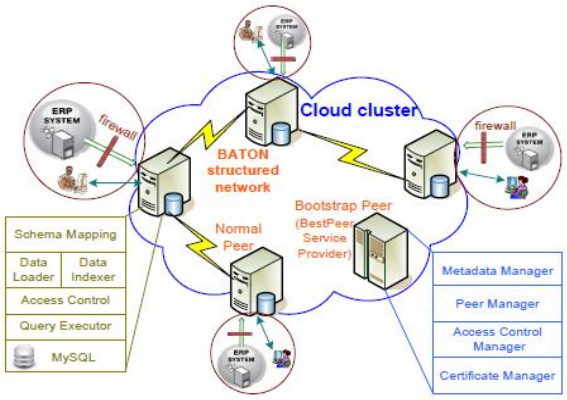

Fig. 1. The BestPeer++ network deployed on Amazon Cloud offering

RELATED WORK:-

To enhance the account of accepted P2P networks, database association accept proposed a alternation of PDBMS (Peer-to-Peer Database Administer System)

by amalgam the state-of-art database techniques into the P2P systems. These PDBMS can be classified as the baggy systems such as PIAZZA [17], Hyperion [15] and PeerDB [11], and the structured systems such as PIER [6]. The works on baggy PDBMS focus on the botheration of mapping amalgamate schemas a part of nodes in the systems. PIAZZA introduces two embodied appearance approaches, namely Local as Appearance (LAV) and Global as Appearance (GAV). PeerDB employs advice retrieval abode to bout columns of altered tables. The capital botheration of baggy PDBMS is that there is no agreement for the abstracts retrieval achievement and aftereffect quality.

The structured PDBMS can bear seek account with affirmed performance. The capital affair is the possibly top aliment cost [1]. To abode this problem, fractional indexing arrangement [20] is proposed to abate the basis size. Moreover, adaptive concern processing [21] and online accession [9] techniques accept aswell been alien to advance concern performance.

The techniques of PDBMS are also adopted in cloud sys-tems. In Dynamo [6], Cassandra [14], and ecStore [24], similar data dissemination and routing strategy is appliedto manage the large-scale data

CONCLUSION:-

We accept discussed the different challenges fairish by administration and processing abstracts in an inter-businesses ambiance and proposed BestPeer++, an arrangement which delivers adaptable abstracts administration services, by amalgam billow computing, database, and peer-to-peer technologies. The criterion conducted on Amazon EC2 billow belvedere shows that our arrangement can calmly handle archetypal workloads in an accumulated arrangement and can bear abreast beeline concern throughput as the amount of accustomed aeon grows. Therefore, BestPeer++ is a able band-aid for able abstracts administration aural accumulated networks.

REFERENCES:-

[1] K. Aberer, A. Datta, and M. Hauswirth.RouteMaintenance Overheads in DHT Overlays.In the 6th Workshop on Distributed Data and Structures, 2004.

[2] 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. PVLDB, 2(1):922–933, 2009.

[3] 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. In SOSP, pages 205–220, 2007.

[4] H. Garcia-Molina and W. J. Labio.Efficient Snapshot Differential Algorithms for Data Warehousing.Technical report, Stanford, CA, USA, 1996.

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

[6] R. Huebsch, J. M. Hellerstein, N. Lanham, B. T. Loo, S. Shenker, and I. Stoica. Querying the Internet with PIER. In VLDB, pages 321–332, 2003.

[7] 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.In SIGMOD, 2006.

[8]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 Syst., 30:364–397, June 2005.

[9] H. V. Jagadish, B. C. Ooi, and Q. H. Vu. BATON: A Balanced Tree Structure for Peer-to-Peer Networks. In VLDB, pages 661–672, 2005.

[10] A. Lakshman and P. Malik. Cassandra: structured storage system on a p2p network. In PODC, pages 5–5, 2009.

[11] W. S. Ng, B. C. Ooi, K.-L. Tan, and A. Zhou. PeerDB: A P2P-based System for Distributed Data Sharing. In ICDE, pages 633–644, 2003.

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

[13] V. Poosala and Y. E. Ioannidis.Selectivity estimation without the attribute value independence assumption. In VLDB, pages 486–495, 1997.

[14] M. O. Rabin. Fingerprinting by Random Polynomials, 1981.Harvard Aiken Computational Laboratory TR-15-81.

[15] 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.In VLDB, pages 1291–1294, 2005.

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

[17] 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, 32(3):47–52, 2003.

[18] H. T. Vo, C. Chen, and B. C. Ooi. Towards elastic transactional cloud storage with range query support. PVLDB, 3(1):506–517, 2010.

[19] S. Wu, S. Jiang, B. C. Ooi, and K.-L.Tan.Distributed online aggregation. PVLDB, 2(1):443–454, 2009.

[20] S. Wu, J. Li, B. C. Ooi, and K.-L.Tan.Just-in-time query retrieval over partially indexed data on structured p2p overlays. In SIGMOD, pages 279–290, 2008.

[21] S. Wu, Q. H. Vu, J. Li, and K.-L.Tan. Adaptive multi-join query processing in pdbms. In ICDE, pages 1239–1242, 2009.

AUTHORS :

Mr. P.VandanaSagar Studying II M.Tech (CSE) in St. Ann's College of Engineering & Technology, Chirala,He completed B.Tech.(CSE) in 2013 in Bapatla Engineering College, Bapatla.

Mr.S.Anil Kumar is presently working as asAssociate Professor, Department of Computer science & Engineering in St. Ann's College of Engineering and Technology, Chirala.He Completed M.Tech. in CSE. He guided many U.G. & P.G projects. He has 10 Years of TeachingExperiance.