ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

Sharing of Large Scale Data in Corporate Network by BestPeer++ System

Agwan Monika[1], Hirave Rani[2], TikeSonam[3], Ugale Kalyani[4], Swati Abhang[5]

Student of Department of Information Technology Engineering, S.R.E.S. College of Engineering, Kopargaon, India[1,2,3,4]

Professor of Department of Information Technology Engineering, S.R.E.S. College of Engineering, Kopargaon, India[5]

ABSTRACT: Most of the businesses utilize the corporate networks to share the data among the businesses with common interest. BestPeer++ system will help the companies to reduce operational costs and increase the revenues. In today's era, most of the enterprises are migrating their physical infrastructure to cloud based platform to reduce operational cost and achieve best performance of enterprise. BestPeer++ system is the combination of cloud computing, databases and peer to peer based technologies. The system will give the efficiency as pay as you go manner. The system will have security by providing private key and admin authorized to provide access to other user. BestPeer++ system will allow the user to store their data into cloud and access when required using cloud computing. The main approach of the system is to use committed database servers to store data for each business and arrange those database servers through P2P network for data sharing. The main goal of BestPeer++ system is that the companies with same sector should be able to share data within each other securely and efficiently. The main focus is to add data from different companies (peers) at cloud and efficiently, securely retrieve the data from cloud and share that with different companies.

KEYWORDS: Peer-to-peer Systems, Query Processing, Index,Data Integration in Big data (DIB), Storage Efficiency,Schema Mapping.

I. INTRODUCTION

"We are melting in an Ocean of Data, but we need knowledge". Here, the problem is that how to get information or knowledge from large amount of data and thisproblem is solved by using Data Mining.Now-a-days, Data Mining is very popular, fashionable words used. Data Mining is the Extraction or "mine" Knowledge fromlarge amount of Data. The corporate networks are used for sharing information ofsame sectors among the companies with the common interest. The companies are ableto reduce the operational cost as well as increase the revenues by using BestPeer++System. The unique challenges such as performance, throughput, security and scalability of Data Management System issues are processed by using BestPeer++ System.BestPeer++ is an elastic data sharing system which delivers services for corporatenetwork applications in the cloud. The security, income as well as performance of industry are built by BestPeer++ system. Peer-to-peer computing or networking is adistributed application architecture that partitions tasks or workloads between peers.Peer-to-peer network in which interconnected nodes are shared resources amongst eachother without the use of a centralized administrative system. The BestPeer++ corewill be based on the platform-independent logic. BestPeer++ system will consist ofnormal peer and bootstrap peer as software components. The efficient range queryprocessing will support by BATON. It will repartition and redistribute the data automatically from the system. The two algorithms such as Adaptive Query Processingand Bootstrap Daemon will be implemented in the project. The pay-as-you-go queryprocessing will be used by BestPeer++ system. The BestPeer++ system will use tobring the state-of art database techniques into P2P systems. The BestPeer++ system will play an important role of plug and play adapter to easily migrate the wholephysical infrastructure in cloud. The Auto Fail-over and Auto-Scaling will be used tomanaging peer join and peer departure. To maintain the statistics of column valuesthe Histograms will use for query optimization.

DOI: 10.15680/IJIRCCE.2016. 0402233

II. OVERVIEW OF THE BESTPEER++ SYSTEM

In traditional system, An Ecstore supported the automated data partitioning, replication, load balancing, efficient range query and transactional access. The data objects were distributedand replicated in cluster of commodity computer nodes located in the cloud. Balance TreeOverlay Network was used to support efficient range query processing. The automaticrepartition and redistribute of the data was done by BATON [2].RCM is failure resilient, which can tolerate host and network failures that are common in real-world hosting infrastructures [3]. The B+ tree index scheme was for efficient data processing in the cloud. The B+ tree was built for the data set storedin each compute node. A scalable and high-throughputindexing scheme was required to handle large number of concurrent end users. For efficient data processing in the Cloud and to save the maintenance cost, the B+-tree based indexing scheme was used [4]. Ferry was the architecture to expand thedistributed Hash table. Ferry was the overlay structure to build an efficient and scalable platform. Ferry was supported a content-based publish/subscribe scheme with alarge number of event attribute [5]. HadoopDB has some deficiencies in real deployment. In corporate network, the installation of a large-scale centralized data warehouse system required huge hardware/software investments and high maintenance cost. But, most companies are not interested to invest heavily on additional information systems until they can clearly see the potential return on investment (ROI). The companies need to fully customize the access control policy for determining which business partners can see which part of their shared data. The data warehouse solution has not been designed to handle dynamicity.

To overcome those problems,BestPeer++ 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++ 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. It can provide affordable store of all company’s data for later use. BestPeer++ system will be based on AmazonEC2 Cloud platform. Both ad-hoc queries and costly analysis queries will handles by BestPeer++system[1].

III. SYSTEM ARCHITECTURE

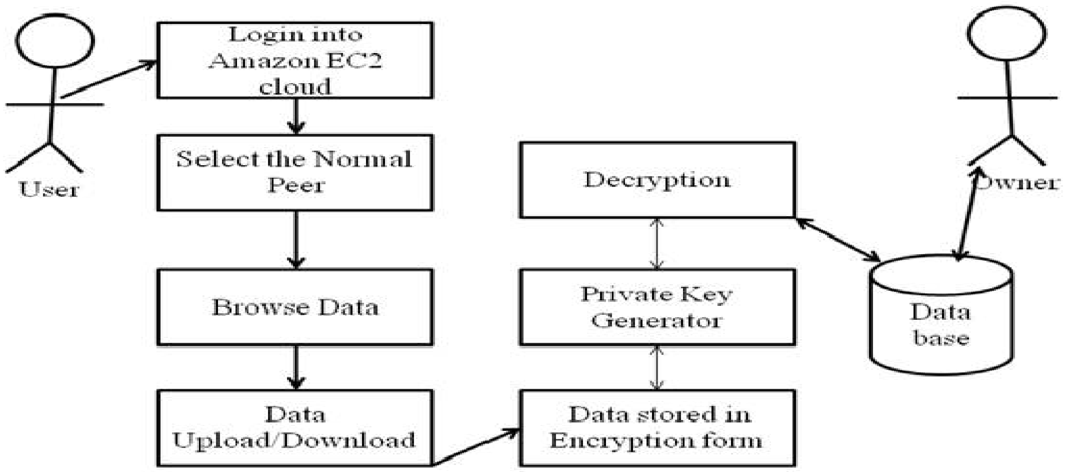

Fig: BestPeer++ System

The design of BestPeer++ system is to provide economical, flexible and scalable solutions for corporate network applications. The unique challenges posed by sharing and processing data in an inter-businesses environment and proposed BestPeer++, a system which delivers elastic data sharing services, by integrating cloud computing, database,

DOI: 10.15680/IJIRCCE.2016. 0402233

and peer-to-peer technologies. BestPeer++ is deployed as a service in the cloud. BestPeer++ achieves its query processing efficiency to handle sharing data within the corporate networks.

IV. METHODOLOGY

4.1. BestPeer++:

BestPeer++ employs a hybrid design for achieving high performance query processing. The major workload of a corporate network is simple, low overheadqueries. Such queries typically only involve querying a very small number of business partners and can be processed in short time. 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 through APIs provided by specific cloud service providers (e.g., Amazon).

4.2. Amazon Cloud Adapter:

Amazon EC2 service is provision of the database server. Each time a new business joins the BestPeer++ network, a dedicated EC2 virtual server is launched for that business. The newlylaunched virtual server (called a BestPeer++ instance)runs dedicated MySQL database softwareand 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.

V. ALGORITHMS

5.1 Auto Fail Over and Auto Scaling Bootstrap Daemon ():-

In addition to managing peer join and peer departure, the bootstrap peer spends most of its running-time on monitoring the health of normal peers and scheduling fail-over and auto-scaling events.

For monitoring and controlling peers the algorithm is as follows:

```
Step 1: while truedo
Step 2:identify the network status by
        calling invokeCloudWatch ()
        function
         Status S: =invokeCloudWatch()
Step 3: Declare ArrayList for pList and
newPeer
ArrayList PList: =
BootStrap.getAllPeer ()
ArrayList newPeer: = new ArrayList ()
Step 4: for i: = 0 to pList.size ()then
Step 5: if pList.get (i).fails ()then
        Peer peer: = new Peer ()
        peer.loadMySQLBackUpFromRDS
        (pList.get (i))
        newPeer.add(peer)
        BootStrap.setBlackList(pList.get(i))
Step 6: else
        If pList.get (i).overloaded ()then
        Peer peer: = new Peer ()
        peer.upScale(pList.get(i))
        peer.clone(pList.get(i).getDB())
        BootStrap.setBlackList(pList.get(i))
        newPeer.add(peer)
```

Step 7: BootStrap.removeAllPeersInBlackList ()
Step 8: BootStrap.addAllNewPeers (newPeer)
Step 9: BootStrap.broadcastNetworkStatus()
Step 10: sleep T seconds.

5.2 Adaptive Query Processing:

When a query is submitted, the query planner retrieves related histogram and index information from the bootstrap node, analyzes the query and constructs a processing graph for the query. Then the costs of both the P2P engine and Map Reduce engine are predicted based on the histograms and runtime parameters of the cost models. The query planner compares the costs between two methods and executes the one with lower cost.

Input: Query Q
Output: Query configuration on a specific query engine

```
TableSet S←TableParser (Q);
Cost Cmin ← MAX_VALUE;
QueryPlan Target←null;
QueryPlanSet QS← ϕ;
foreach Table  T∈Sdo
//Generate Processing Graphs Rooted on T
GraphSet GS =GraphGen (T);
//Iterate through all processing
Graph rooted on T
foreachGraph G∈GS do
QueryPlan P1=P2P PlanGen(G);
QueryPlan P2 =MapredPlanGen(G);
QS=QS ∩ {P1};
QS=QS ∩ {P2};
foreach QueryPlan P ∈QS do
If CostEst (P) <Cmin then
Cmin=CostEst (P);
Target= P;
Return Target
```

VI. CONCLUSION

The BestPeer++ System had efficiently distributed the query processing as wellas supported the system load balance in the corporate network. It had defined theexclusive challenges faced by sharing and processing of data in inter-business environment. The BestPeer++ system had reduced the operational as well as maintenancecost by increasing the Return on Invest. It can efficiently handle typical workload in acorporate network.

REFERENCES

[1] Gang Chen, Tianlei Hu, Dawei Jiang, Peng Lu, Kian-Lee Tan, Hoang Tam Vo,andSaiWu, "BestPeer++: A Peer-to-Peer Based Large-Scale Data Processing Platform", IEEE Transactions On Knowledge And Data EngineeringVol. 26, No. 6, June 2014.

[2] H.T. Vo, C. Chen, and B.C. Ooi, "Towards Elastic Transactional Cloud Storagewith Range Query Support", Proc. VLDB Endowment, vol. 3, no. 1, pp. 506-517,2010.

[3] Yongmin Tan, Student Member, IEEE, Vinay Venkatesh, and XiaohuiGu, Senior Member, IEEE, "Resilient Self-Compressive Monitoring for Large-Scale Hosting Infrastructures", IEEE Transactions On Parallel And Distributed Systems, Vol. 24, NO., 3 MARCH 2013.

[4] Sai Wu, Dawei Jiang, Beng Chin Ooi and Kun-lung Wu, "Efficient B-tree BasedIndexing for Cloud Data Processing", Proceedings of the VLDB Endowment, Vol.3,No. 1, 2010. And Automation ICTA05, Thessaloniki, Greece, 365-370, pg.15-16.

[5] Yingwu Zhu and Yiming Hu, "Ferry: A P2P-Based Architecture for Content-Based Publish/Subscribe Services", vol. 18, no. 5, May 2007.