

Extended Best Peer: A Peer-to-Peer Based System by Corporate Network for Data Sharing

Chandre P.R.¹, Bhavsar Harshada²

Professor, Department of Computer Engineering, Sharadchandra Pawar College of Engineering, Otur, Pune-412409, Maharashtra, India

Department of Computer Engineering, Sharadchandra Pawar College of Engineering, Otur, Pune-412409, Maharashtra, India

Abstract: The corporate network is always used for sharing information where share common among the contribute company and make easy association in a certain industry sector. Ultimately reply to the ever changing business demands and the exterior of Cloud Computing techniques, BestPeer has develop into cloud-enabled Extended BestPeer system. In this paper, we present Extended BestPeer, a system which deliver flexible data sharing services for the commercial network applications in the cloud based on BestPeer a peer-to-peer (P2P) based data management platform. By Combining cloud computing, database, and P2P technologies, Extended BestPeer achieves its query processing efficiency in a pay-as-you-go manner. We evaluate Extended BestPeer on Amazon EC2 Cloud platform. The benchmarking results show that Extended BestPeer better than HadoopDB, in presentation when both systems are working to maintain typical corporate network.

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

1. Introduction

Company of the same industry are often connected to a corporate network for association purposes.[2] 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 enable the pooled data supports capable logical queries over 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.[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) [21]. 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.[3] The data warehouse solution has not been designed to handle such dynamicity. For decrease such problem this paper design Extended BestPeer for corporate Network.

As an in-time response to the ever changing business demands and the appearance of Cloud Computing techniques, BestPeer has developed into its new stage of development-the cloud-enabled Extended BestPeer system.[3] The structure of the system shows in fig 1.

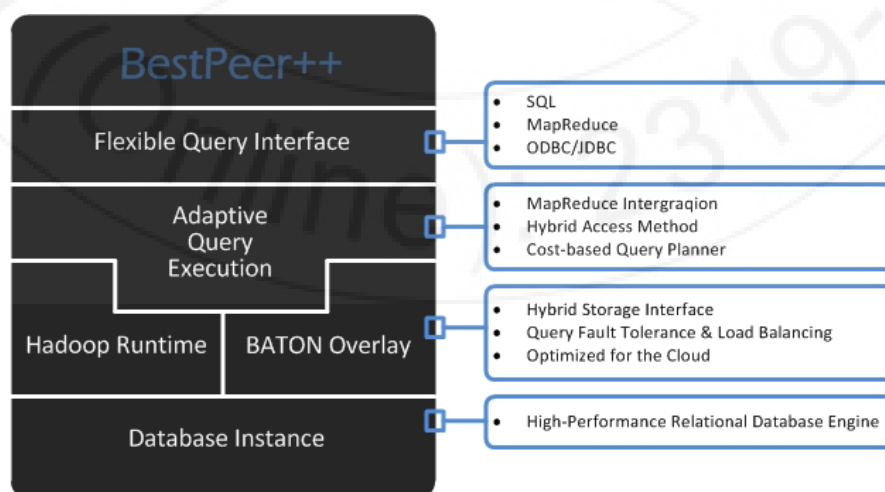


Figure 1: The Structure of Extended BestPeer System

By integrating cloud computing, database, and P2P technologies, Extended BestPeer 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 BestPeer system that provides inexpensive, Flexible solutions for corporate network. We demonstrate the competence of Extended BestPeer by benchmarking Extended BestPeer against HadoopDB [2], a propose approach for data sharing applications. The results show that for simple queries, the performance of Extended BestPeer is significantly better than HadoopDB.

The rest of the paper is organized as follows. Section 2 presents the Literature survey of the Extended BestPeer system including Existing System and Propose System. We then describe the design of Extended BestPeer core components, with system architecture in Section 3 and the Section 4 shows that evolution of the performance of system. Advantages of the system are explain at section 5 and Related work are presented in Section 6, with conclusion in Section 7.

2. Literature Survey

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

2.1. Existing System

The original BestPeer system attempt to develop peer-to-peer (P2P) technologies for Corporate Network. BestPeer 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.[6] corporate network applications such as supply chain management and national healthcare network. BestPeer 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. [4] As per changing business demands and the coming out of Cloud Computing techniques, BestPeer has developed into its new stage that is cloud-enabled Extended BestPeer 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 (a.k.a total cost of ownership) and high preservation cost (a.k.a total cost of operations) [6]. In the enviroment, most companies are not dedicated to invest deeply on additional information systems until they can clearly see the potential return on investment (ROI) [12].

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.

2.2. Propose System

The main contribution of this paper is the design of Extended BestPeer 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 BestPeer, a system which give elastic data sharing services, by including cloud computing, database, and peer-to-peer technologies for corporate Network. BestPeer's product is the Extended BestPeer Platform, which combines the powerful MapReduce processing model with the predictable P2P database technologies. Extended BestPeer's advanced technology features a hybrid architecture that brings the parallelism of MapReduce to the latest development in RDBMS research.[4] Extended BestPeer 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 MapReduce framework and Software-as-a-Service(SaaS) paradigm.

In compare to the "Hadoop Connector" approach employed by many MPP investigative database vendor, Extended BestPeer uses Hadoop as the parallelization layer to make possible its universal query processing, with each node running a database occasion.[5] consolidate predictable database query processing and MapReduce 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 BestPeer'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 BestPeer is deploy as a service in the cloud. To form a corporate network, companies register with the site Extended BestPeer service provider, initiate Extended BestPeer instances in the cloud and at last export data to those instances for sharing. Extended BestPeer 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 BestPeer service provider elastically grow up the running instance and makes them always available. For occasional sustained analytical tasks, we provide an border for exporting the data from Extended BestPeer to Hadoop and allow users to analyze those data using MapReduce.[3] Extended BestPeer 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[8].

3. Component of Propose System

Extended BestPeer, a cloud enabled evolution of BestPeer. At the last stage of its development, Extended BestPeer 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.[6] The software components of Extended BestPeer 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.

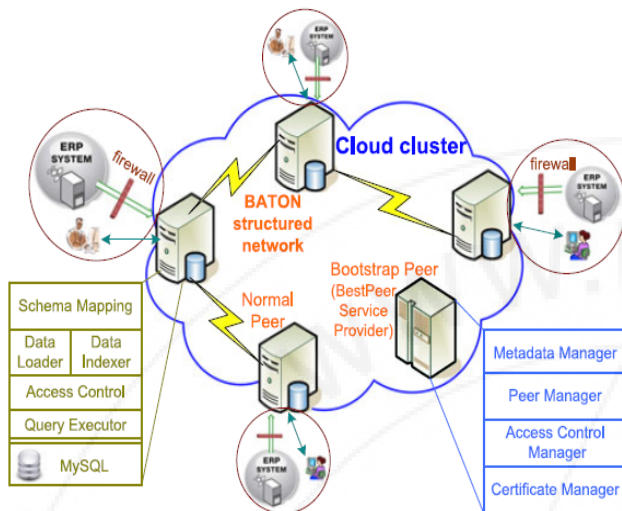


Figure 2: The Extended BestPeer network deployed on Amazon Cloud Offering

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 BestPeer 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[6].

Specifically, highlights of Extended BestPeer are:

- a) Amazon Cloud Adapter: The main approach of Extended BestPeer is to use dedicated database servers to store data for each business and arrange those database servers through P2P network for data sharing. The Amazon Cloud Adapter provides an elastic hardware infrastructure for Extended BestPeer to operate on by using Amazon Cloud services.
- b) The Extended BestPeer Core: The Extended BestPeer 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.
 - The bootstrap peer is run by the Extended BestPeer service provider and main functionality is to manage the Extended BestPeer network of bootstrap peer.
 - The normal peer software having five components such as schema mapping, data loader, data indexer, access control and query executor. As shown in Fig. 3, it define two data flows inside the normal peer as an offline data flow and an online data flow. The data are extract periodically by a data loader from the business production system to the normal peer instance in offline data flow.

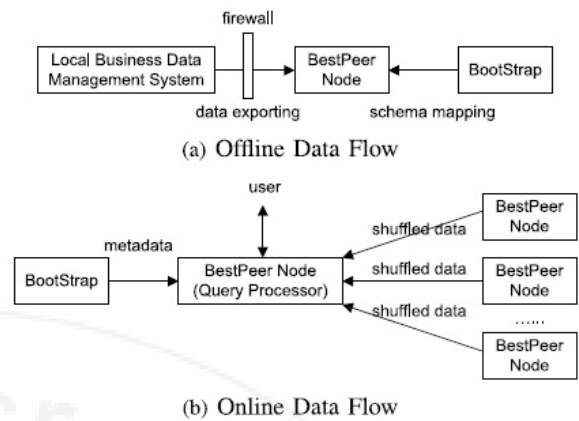


Figure 3: Data Flow in Extended BestPeer

- Adaptive Query Processor: Extended BestPeer employs a amalgam design for achieve high act query processing. The major workload of a corporate network is simple, low-overhead queries.[4] This queries only occupy querying a very small number of business partners and can be processed in small moment. Extended BestPeer is mainly optimized for these queries. For rare time-consuming analytical tasks, we provide an interface for exporting the data from Extended BestPeer to Hadoop and allow users to analyze those data using MapReduce.

4. Advantages of Propose System

- This system can powerfully handle characteristic workloads in a corporate network .
- Extended BestPeer adopt the pay-as-you-go business model famous by cloud computing. as an optional, what they use of Extended BestPeer instance's hours and storage capacity they pay for it.
- Extended BestPeer extend the role-based access control for the natural dispersed environment of commercial networks.
- Extended BestPeer employs P2P technology to retrieve data between business partners.
- Extended BestPeer is a great solution for data sharing within corporate networks.

5. Benchmarking

This section shows evolution of the performance and throughput of Extended BestPeer 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 BestPeer with HadoopDB. We choose HadoopDB 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., HadoopDB and Extended BestPeer) 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 BestPeer. HadoopDB is not designed for high query throughput, therefore, we intentionally skip the results of HadoopDB and only present the results of Extended BestPeer. We conduct two tiers of benchmark evaluation for the performance and scalability of Extended BestPeer, respectively.

6. Conclusion

This paper define exclusive challenges pose by contribution and open-handed out data in an inter-businesses environment and planned Extended BestPeer, a system which deliver elastic data sharing services, by Containing cloud computing, database, and peer-to-peer technologies.[5] 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 BestPeer is great solution for capable data sharing within corporate networks.

References

- [1] Gang Chen, Tianlei Hu, Dawei Jiang, Peng Lu, Kian-Lee Tan, Hoang Tam Vo, and Sai Wu, "Extended BestPeer: A Peer-to-Peer Based Large-Scale Data Processing Platform", VOL. 26, NO. 6, JUNE 2014.
- [2] 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.
- [3] 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.
- [4] S. Wu, S. Jiang, B.C. Ooi, and K.-L. Tan, "Distributed Online Aggregation," Proc. VLDB Endowment, vol. 2, no. 1, pp. 443-454, 2009.
- [5] 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.
- [6] 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.
- [7] Beng Chin Ooi, Yanfeng Shu, "Relational Data Sharing in Peer-based Data Management Systems." Kian-Lee Tan Sigmod Record special issue on P2P, 2003.
- [8] B.C. Ooi, K.L. Tan, A.Y. Zhou, C.H. Goh, Y.G. Li, C.Y. Liao, 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).
- [9] G. Chen, H. T. Vo, S. Wu, B. C. Ooi, T. "A Framework for Supporting DBMS-like Indexes in the Cloud." Ozsu VLDB 2011.

- [10] Sai Wu, Dawei Jiang, Beng Chin Ooi, Kun Lun Wu" Efficient B+-tree Based Indexing for Cloud Data Processing VLDB 2010.
- [11] Heng Tao Shen, Yanfeng Shu, and Bei Yu IEEE Trans. Knowl. "Efficient Semantic-Based Content Search in P2P Network." Data Eng. 16(7): 813-826 (2004)