A Review on: Storage Database Consolidation Technology

Mayuri D. Kakadiya
Department of Computer Sci.& EngineeringParul
Institute of Technology
Waghodia Road, Vadodara

Pratik A. Patel
Department of Computer Sci.& Engineering
Parul Institute of Technology
Waghodia Road, Vadodara

ABSTRACT

In this paper we are studying the types of data consolidation and enterprise storage. Heterogeneous is one kind of consolidation technique to improve efficiency of result. we are adding one more useful method of cloud computing which is pre-fetching. Generally Consolidation is used with centralized database. In this paper we can apply the consolidation techniques on distributed database with the help of pre-fetching. Pre-fetching is the technique which is used in both approach, centralized as well as in distributed.

Keywords

Consolidation, Pre-fetching, Centralized, Distributed.

1. INTRODUCTION

Cloud computing and cloud storage systems have gained popularity as the most convenient way of transferring information and providing functional tools on the Internet. Like public, private, community hybrid cloud. This four types of cloud are best way to transfer the information via cloud. The idea of "cloud computing" is to combine multiple computers and servers into a single environment designed to solve certain kinds of problems, such as scientific problems or complex calculations. Over time, this structure collects a lot of data, distributed computing and storage nodes. Typically, applications, running in a distributed computing environment, address only one of the data sources. However, when the need arises to access simultaneously multiple sources, difficulties arise because these sources may contain different data and tools of heterogeneous access and also are Storage Database System in the Cloud Data Processing. Distributed at a distance from each other. In addition, for users performing an analysis of historical data, it is convenient to apply to a single source of information, forming a query and get results in the same format. Thus, the main problem of the approach to the storage of information in distributed computing systems is the diversity and remote data sources. The solution is to create a centralized point of access, providing a single interface access to all data sources for cloud computing in real time. It is necessary to choose the most appropriate approach and the corresponding platform that provides a consolidation. But manage this type of database is difficult so we are applying pre-fetching with distributed approach. This distributed approach is explained via below figure 1 & figure 2.

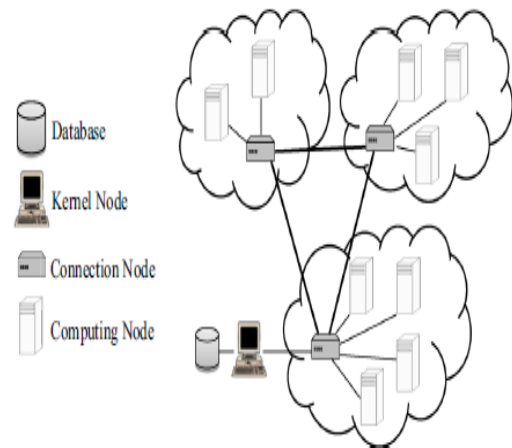Example 1(Distributed approach in data storage consolidation)

Figure 1: Distributed Approach in Data Consolidation [8]

Here in this figure 1 the data base are at remote level, kernel represent the use which also connected with some database, connection node describe the connection between user & remote node, and the end the computing node represent which are available globally. So in this example three different cloud represent the unique centralized approach and connection between them represent the distributed approach.

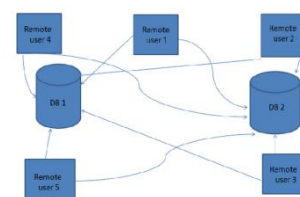Example 2 (Distributed approach in data storage consolidation)

Figure 2: Distributed Approach in Consolidation Technology

Here in this example the 2 databases are available,DB1& DB2. Both databases have some remotes users respectively. But in this example the all users are connected with both database randomly. There is no special kind of cloud are specified like previous example and figure 1.so we can say that both example are explain distributed approach but in its own way. Basis on that different algorithm and methods are used. Algorithm and methods are explain in below section.

1.1 Intelligent Method for Cloud Computing

According to Chaudhary et al. [12]it is a great tool in improving the competence and precision of decisions made by

intelligent computer programmers. Hence, it is suitable in this work to use the intelligent method because in this work will find the more accurate of prediction data to apply in the real world for reduce the useless data. Besides, it is useful for mobile devices like smart phones, smart cards, and automotive systems [13]. Thus, it is suitable in this research, which looks into mobile-based environment. This all technique have some importance in CC.Below figure 3 describe the techniques.
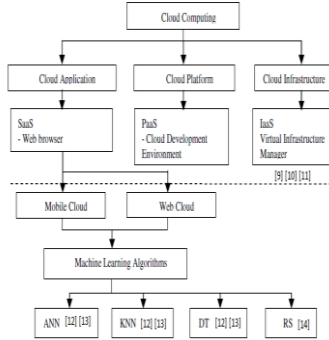
1.2 Table description
Below table 1 represent the description of all intelligent techniques of CC. the benefits and limitations of all techniques are given in this table. It helps us to solve and choose the best one from all of them. It's one kind of smart method or idea to represent your work.in this paper we just summarize the techniques of CC.But in future we can apply any of them to improve our work.

Figure 3: Intelligent Method of CC

Table 1. Intelligent Technique for Prediction of Accuracy

Author(s)	Intelligent Technique	Issues	Benefit	Limitation
Yuan and Yu [16]	ANN	Accurateness of learning result practical with CC	Capable inhandling big dataset for learning	Complex effort to reduce over fitting and it needs a great deal of computational effort
Sulaiman et al. [17]	ANN	Speed performance for mobile devices	To improve the access of social network using mobile devices	Limited in social Network
Sarwar et al. [18]	ANN	Avoiding page faults	Better utility in pre-fetching	For high I/O intensive system only
Rajkumar et al. [15]	ANN	Efficiency and accuracy of Diagnosis	Improves the effectiveness and precision of diagnosis	Research in medical domain alone
Chang et al. [19]	KNN	Performance or accuracy of TV recommendation	Improves the TV programme recommendation system	Limitation in observing the result of the rising number of computing nodes
Wang et al. [20]	KNN	Data mining applications and machine learning problem over cloud computing	Enhances the prediction rating	Weakness in small Data
Liao et al. [21]	DT	Parameter optimization Problem	Feasible to build a precise prediction model for data reduction	Complex decision making systems that is hard to compete
Nagy et al. [22]	DT(c4.5 algorithm)	Accurateness and efficiency of result prediction	Enhances the students' performance and the quality of the education	Scope in education environment alone
Sulaiman et al. [23]	RS	Mobile web caching Performance	Reduces and derives rules for enhanced classification result.	Scope in social network alone
Chimphlee et al. [24]	RS	Web access prediction	Enhances the prediction measure	Researches in web domain alone

2. DATA CONSOLIDATION TECHNOLOGY

Consolidate distributed data sources can be divided into two types:

1. The centralized approach

Data from all external sources are transferred to the central repository and are updated periodically. All users work directly with the central repository.

2. The federated approach

Data is stored directly in the sources, the central link provides transparent redirection of user requests and the formation of the results. In this case, all users can also refer to the central node only which translates requests more data sources. Consolidation of data in distributed heterogeneous systems is an important and challenging task. Out of existing approaches to solving this problem, the most appropriate approach is that of federal databases. Creating and managing such a structure requires the use of specialized software, which in turn must meet a number of requirements for transparency, heterogeneity, security, performance, etc. On market integration software there are a number of solutions from major manufacturers, build on industrial relational database, based on which you can organize a federal structure data access. To select a specific solution, the detailed examination for compliance with the requirements for systems of this type must be made.

3. The Distributed approach

The data from all external sources are not transferred to central repository instead of that we Can work directly or indirectly with different remote repository. This new approach for this paper. The aim for of this paper is how to deal with distributed data base and apply consolidation on it. In this approach we are also applying pre-fetching techniques. Because pre-fetching is feasible. For both type of repository. Centralized database as well as distributed. So in this approach we can combine the technique of both major methods of cloud computing. And at the end will get the result.

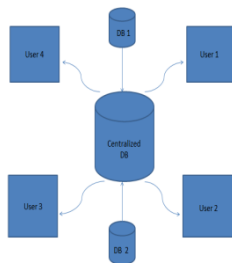## 2.1 Difference between Centralized & Distributed Approach
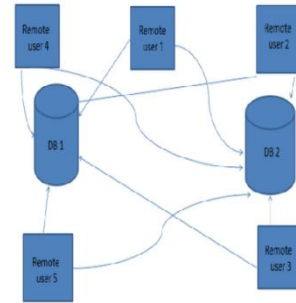

Fig: 4 Centralized Approach

Fig: 5 Distributed Approach

Cloud computing architecture have basic 3 level. SaaS, PaaS,IaaS. This 3 level have different use. Based on uses of this 3 level different technique and some intelligent methods are applied in Cloud computing. So below chart represent the list of intelligent methods of cloud computing for data storage consolidation. For distribution approach we can also use VM. With the help of different VM we can solve the problem of our problem statement. With the help of below figure we can understand the use of VM consolidation. VM consolidation is one type of consolidation techniques.
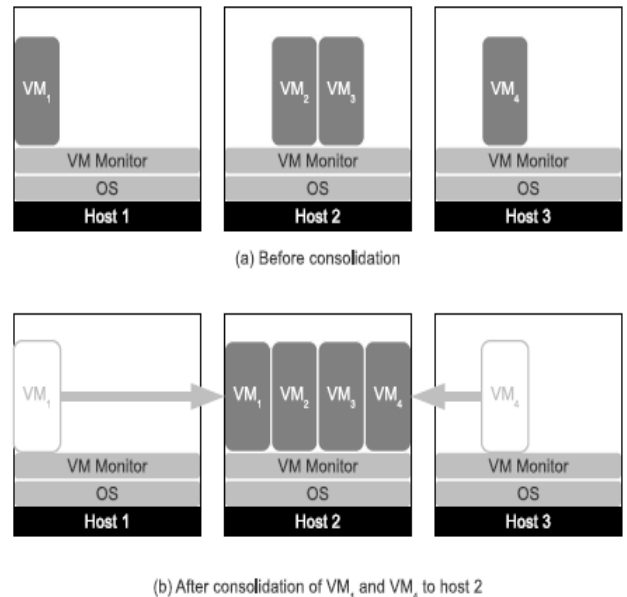

(a) Before consolidation

(b) After consolidation of VM_1 and VM_4 to host 2

Figure 6: VM consolidation technique overview

This figure represents the use of different VM. With the multiple VM data will be distributed and we can access it from anywhere. One replica will be generated.so every user have one VM from where the user can access his/her data. Data storage consolidation will be easy and efficient with this consolidation technique.

For this technique or method of consolidation we need V-MAN.V-MAN, a fully decentralized algorithm for consolidating VMs in large Cloud datacenters. V-MAN can operate on any arbitrary initial allocation of VMs on the Cloud, iteratively producing new allocations that quickly converge towards the one maximizing the number of idle hosts. V-MAN uses a simple gossip protocol to achieve efficiency, scalability and robustness to failures. Simulation experiments indicate that, starting from a random allocation, V-MAN produces an almost-optimal VM placement in just a few rounds; the protocol is intrinsically robust and can cope

with computing nodes being added to or removed from the Cloud.

3. CLOUD COMPUTING WITH PRE-FETCHING

The increasing the demand usage of those services in brought to latency problem. Therefore, the current works have proposed by using the pre-fetching method even though, by using overaggressive of pre-fetching method it will cause the overhead of the storage. With some intelligence of cloud different techniques are used in pre-fetching. According to Chaudhary et al.it is a great tool in improving the competence and precision of decisions made by intelligent computer programmers. Hence, it is suitable in this work to use the intelligent method because in this work will find the more accurate of prediction data to apply in the real world for reduce the useless data. Besides, it is useful for mobile devices like smart phones, smart cards, and automotive systems. Thus, it is suitable in this research, which looks into mobile-based environment.

Earlier days web-fetching technique is used in cloud as intelligence purpose. Web-fetching is the part of web application in cloud computing. SaaS contains the web-fetching technique. But for more accuracy we are proposing the pre-fetching technique with storage database consolidation. Our focus is on Pre-fetching only. SaaS support pre-fetching .all the data are even stored in cloud. So SaaS is a part of cloud computing service. so in this paper we are combine the storage database consolidation with pre-fetching. the use of intelligent method that can select a splitting attribute on compressed data, reduce the complexity of tree, increase the accuracy, and reduce unrelated data. With the help of java or other important language like .NET we can get the result of problem.

4. LITRATURE REVIEW

Most of the procedure focuses on the centralized database only to consolidate some limited data only. But several techniques and algorithms are available in cloud computing which support distributed approaches. Working of that paper, motivation for proposal of that algorithm, advantage and their limitations are briefly describe here. They are as follows:

The problem of Cloud use for computation stands aside both due to the large overheads in parallel libraries in the Cloud and what is more important serious limitations due to file systems peculiarities in virtual cluster architectures. Thus the way of transferring the data to virtual processes is of a vital importance for large scale computations optimization. The main problem of the approach to the storage of information in distributed computing systems is the diversity and remote data sources. the cost of treatment is often increased because of the need to duplicate the data processing and two sets of data[3]. Federated databases access mechanism and management of heterogeneous data hides the features of the reference to a specific data source, but instead provides a single interface, similar to the classical relational databases [3].

A.V. Bogdanov et al. Most applicable approach to creating a platform for the federated database approach is to develop the existing relational database management system and to ensure its interaction with external data sources. This database becomes central to a federal database that stores all the information about data sources, and redirects requests to it [3].

Cloud Computing (CC) provides computing resources for their

users through the Internet that acts as virtual computing resource. It can deploy, share out or change around a computing resource robustly and control the usage of resources every time [4]. When there are a lot of users use request the same data at the same time, the latency will occurs and the users need to wait for a while for access the data. Besides, when the users have kinds of cloud storage, it will take times for the users to recall back which cloud storage they store their current data. This issue is related to management problem and slow down the performance service of the cloud. Hence, it needs some improvements by applying some methods to maintain the quality of cloud service. web pre-fetching helps in increasing the cache hits and reducing the user perceived latency. CC is the technology that is used by the Internet and servers to store data and applications [6]. It allows users to use applications or tools without install the software and they may access their data using any device at any time with Internet access. The increasing the demand usage of those services in brought to latency problem. Therefore, the current works have proposed by using the pre-fetching method even though, by using overaggressive of pre-fetching method it will cause the overhead of the storage. On the basis of size, we have categorized the workloads into six categories and studied the variation of energy consumption by varying the heterogeneity of the workloads.[6].Consolidation of the workload, which consists in allocating the maximum number of VMs in the minimum number of physical machines [7], helps to alleviate the problem.

V-MAN, a fully distributed algorithm for VM consolidation on Cloud systems. V-MAN is based on a simple gossip protocol that does not require any central coordinator or globally shared data structure. V-MAN is completely VM and application agnostic.[25]

5. CONCLUSION

Lots of technique, methods and algorithm are used to improve the efficiency of cloud computing. so many application are used by 1000 of people on single click. so the burden on cloud is increased. to remove this type of complexity of cloud and for faster response in this paper we combined two major concern of cloud computing. In this paper we used consolidation latest methods and also we used pre-fetching technique instead of web-fetching for CC intelligence. with this technique we conclude that within very few seconds user can get the result very easily via mobile computing also. because CC is everywhere.

6. ACKNOWLEDGMENT

7. REFERENCES

[1] Bogdanov, A.: The methodology of Application development for hybrid architectures Computer technologies and applications (4), 543 – 547 (2013).

[2] Larose, D.T.: Discovering Knowledge In Data: An Introduction to Data Mining.

[3] Bogdanov, A.V., Stankova, E.N., Lin :T,K Distributed databases. SPb.:``LETI'', pp.39-43 (2013) 320 A.V.

Bogdanov et al.

[4] Prasath, V., Bharathan, N., Lakshmi, N., Nathiya, M.: Fuzzy Logic In Cloud Computing. Int. J. Eng. Res.

[5] Padhy, P.C., Mishra, S.K.: Cloud Computing: Advance Technique for Corporate Excellence. Int. J. Mech. Eng. Comput. Appl. 1(1), 17–21(2013)

[6] Deep Mann ME (Software Engineering) Computer Science and Engineering Department Thapar University Patiala-147004.Inderveer Chana Associate Professor Computer Science and Engineering Department Thapar University Patiala-147004.

[7] P. Graubner, M. Schmidt, and B.Freisleben, "Energy-efficient virtual Machine consolidation," Professional, vol. 15, no. 2, pp. 28–34, 2013

[8] Kawuu W. Lin1 • Sheng-Hao Chung2 •Chun-Cheng Lin2

[9] Padhy, P.C., Mishra, S.K.: Cloud Computing: Advance Technique for Corporate Excellence. Int. J.Mech.Eng. Comput. Appl. 1(1), 17–21 (2013)

[10] Esseradi, S., Badir, H., Abderrahmane, S., RattroutA.: Mobile Cloud Computing: Current Development and Research Challenges. In: 6th Int. Conf. Inf. Technol., ICIT 2013, pp. 1– 9 (2013)

[11] Neela, K.L., Kavitha, V.: A Survey on Security Issues and Vulnerabilities on Cloud Computing. Int, J. Comput. Sci. Eng. Technol. 4(7), 855–860(2013)

[12] Chaudhary, A., Kolhe, S., Kamal, R.: Machine learning techniques for Mobile Intelligent Systems: A study. In: Ninth Int. Conf. Wirel. Opt. Commun. Networks, pp. 1–5 (2012)

[13] Larose, D.T.: Discovering Knowledge in Data: An Introduction to Data Mining (2005)

[14] Shrivastava, S.K., Tantuway, M.: A Decision Tree Algorithm based on Rough Set Theoryafter Dimensionality Reduction. Int. J. Comput. Appl. 0975 – 8887) 17(7), 29–34 (2011)

[15] Rajkumar, B., Gopikiran, T., Satyanarayana, S.: Neural Network Design in CloudComputing, Int. J, Comput. Trends Technol. 4(2), 63–67 (2013)

[16] Yuan, J., Yu, S.: Privacy Preserving Back-Propagation Neural Network Learning Made Practical with Cloud Computing. IEEE Trans. Parallel Distrib. Syst. 25(1), 212–221(2014)

[17] Sulaiman, S., Shamsuddin, S.M., Abraham, A.: Rough Neuro-PSO Web caching and XML prefetching for accessing Facebook from mobile environment, Rough Neuro-PSO.WebCaching XML Prefetching Access. Faceb. from Mob. Environ. World Congr. Nat.Biol. Inspired Comput., pp. 884–889 (2009)

[18] [Sarwar, S., Ul-Qayyum, Z., Malik, O.A.: CBR and Neural Networks Based Technique for Predictive Prefetching. In: Sidorov, G., Hernández Aguirre, A., Reyes García, C.A.(eds.) MICAI 2010, Part II LNCS, vol. 6438, pp.221–232.Springer,Heidelberg(2010)

[19] Chang, J.-H., Lai, C.-F., Wang, M.-S., Wu, T.-Y.: A cloud-based intelligent TV program recommendation system. Computer Electrical Enginering,1–21 (2013)

[20] Wang, J., Wan, J., Liu, Z., Wang, P.: Data Miningof Mass Storage based on Cloud Computing. In:9th Int. Conf. Grid Coop. Comput.(GCC),pp.426–431(2010)

[21] Liao, S., Hung, T.-H., Nguyen, D., Chou, C., TuC., Zhou, H.: Machine learning- based,prefetchoptimization for data center applications. InProc. Conf. High Perform. Comput. Networking,Storage Anal., Portland, Oregon, pp. 1–10 (2009)

[22] Nagy, H.M., Aly, W.M., Hegazy, O.F.: An Data Mining System for Advising Higher Education.

[23] Chimphlee, S., Salim, N., Salihin, M., NgadimanB,Chimphlee, W., Srinoy, S.: Rough Sets Clustering and Markov model for Web Acces Prediction. In:Proc. Postgrad. Annu. Res Semin., pp. 470–475 (2006)

[24] Sulaiman, S., Shamsuddin, S.M., Abraham, A.: Meaningless to Meaningful Web Log Data for Generation of Web Pre-caching Decision Rules using Rough Set. In: 4th Conf. Data Min. Optim.,vol. 1, pp. 2–4 (2012)

[25] Moreno Marzolla, Ozalp Babaoglu, Fabio Panzieri Universit`a di Bologna, Dipartimento di Scienzedell' Informazione Mura A. Zamboni 7, I- 40127 Bologna, Italy