Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 4, April 2014, pg.131 – 138

RESEARCH ARTICLE

QADR with Energy Consumption for DIA in Cloud

Ms.S.Soundharya[1], Mr.R.Santhosh[2]
PG Scholar[1], Department of Computer Science and Engineering
Assistant Professor[2], Department of Computer Science and Engineering
Karpagam University, Coimbatore, India
Email id: parisha.sounds@gmail.com [1], santhoshrd@gmail.com [2]

Abstract:

In today's IT field scalability on storage resources are provided by cloud computing. Moreover data-intensive applications are also developed in this technology. Occurrence of data Corruption on these DIA (data intensive applications) does not meet the QoS Requirements. So in order to satisfy the requirements, we are comparing two algorithms from existing paper one is HQFR which uses the greedy algorithms. HQFR algorithm cannot minimize the cost of data replication and the count of QoS violated data replicas. These two objectives of QADR problem has been achieved by using MCMF algorithm which provides an optimal solution in polynomial time. But when compared to HQFR algorithm its computation time is high, since we have to consider more number of nodes as it is in cloud environment. To find a solution for this time complexity, Combination of nodes–Technique has been introduced in MCMF. Demonstration of these algorithms has been implemented under Windows Azure environment which provides comparatively good results for data replication. Further this implementation has been extended to concern energy consumption in cloud environment.
Key words: DIA, QoS, HQFR, MCMF, Windows Azure Environment

Introduction:

Cloud Computing delivers essential information for Enterprises from various storage resources without knowing their origin. Some of the issues which are playing a major role in large scale organizations while working under cloud are areas of cloud applications management cloud backup and recovery, cloud interoperability, cloud data analytics, liability issues for data loss on clouds, data integration on clouds, cloud energy consumption, big data on clouds, etc. Out of these issues, we are focusing on data loss and energy consumption on cloud. To endure the data corruption and to provide high data availability, the data replication technique is widely adopted in the cloud computing system. Data Replication not only copies a database but also synchronizes a set of replicas so that changes made to one replica are reflected in all the others. The goal of replication is, it enables many users to work with their own local copy of a database but have the database updated as if they were working on a single, centralized database. Replication is often the most efficient method of database access for database applications where users are geographically widely distributed. Due to limited replication space of a storage node, the data replicas of some applications may be stored in lower performance nodes. Those data replicas which don't satisfy the QoS Requirements of DIA (Data Intensive Applications) are called as QoS-violated data replicas [1] .The count of QoS-violated data replicas is expected to be as small as possible to provide QoS on those applications. This Data Replication technique to cloud Provides benefits such as faster recovery time, Off-site real time data copies, to store and replicate data without external software etc.,

Related Work:

Failures in accessing data under cloud use techniques such as check point and data replication. Occurrences of Name node failure can be tolerated by using this checkpoint technique where the state of the file system namespace has been restored in the disk of NameNode .To protect the stored data blocks in DataNode from failure has been done by Data Replication Technique [1] in the stored data blocks is protected by data replication technique.

K. Shvachko, H. Kuang, S. Radia, and R. Chansler, et.al briefly explained the HDFS where the data replication technique has been extensively adopted. HDFS [3] has master/slave architecture which consists of a single NameNode, a master server that manages the file system namespace and regulates access to files by clients and a number of DataNodes, which manages storage attached to the nodes that they run on. Internally, a file is split into one or more blocks and these blocks are stored in a set of DataNodes. An application can specify the number of replicas of a file and the replication factor can be specified at file creation time and can be changed later.

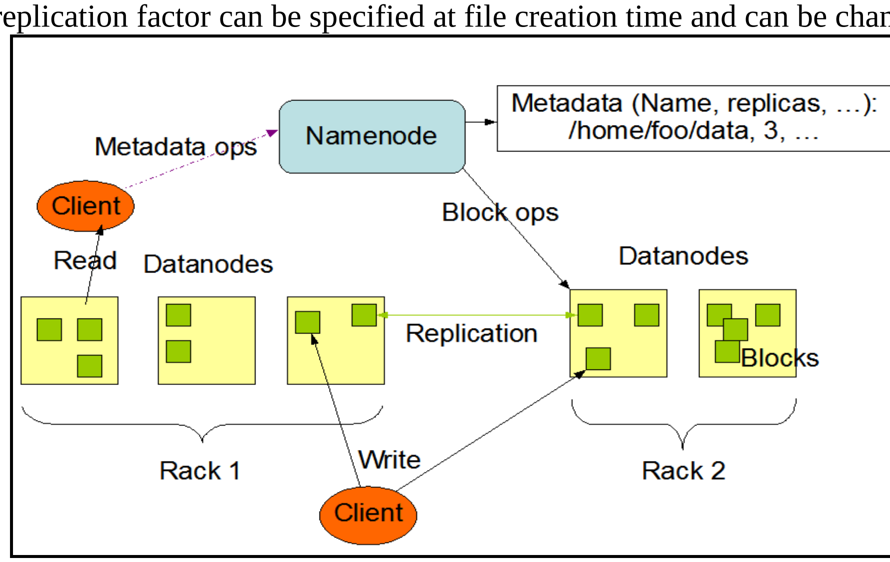

Fig:1 HDFS Architecture

In DataNode, by default, replica factor was taken as two for a single data block. If a data block is written to the DataNode n, the original copy of this data block is stored in the disk of the DataNode n. Two replicas of this data block are stored in two different DataNodes where the rack numbers are different with that of the DataNode n.

A. Gao and L. Diao, et al. discussed about the consistency maintenance problem of data replica in cloud. They proposed lazy update method [6] which improved the data access throughput and reduction in response time

W. Li, Y. Yang, J. Chen, and D. Yuan et.al presented a mechanism which provides data reliability for the replicated data .It is based on proactive replica checking which is cost effective since it reduces storage space consumption.

X. Tang and J. Xu et.al discussed about QADR problem and proved that it is NP-Complete. Without abusing QoS requirements, both the insertion and deletion of data object replicas are done by two algorithms: l-Greedy-Insert and l-Greedy-Delete [8], which results in exceeded execution time.

M. Shorfuzzaman, P. Graham, and R. Eskicioglu et.al presented a distributed replica placement algorithm based on dynamic programming for reducing the execution time and it has been done on data grid systems. It has been designed to satisfy QoS requirements by identifying replication locations, to provide reliability on data and Performance measure.

X. Fu, R. Wang, Y. Wang, and S. Deng et.al addressed replica problem under mobile grid environment for mobile users. They proposed solution by using dynamic programming and binary search problem resulting data availability and high data accessibility.

A. M. Soosai, A. Abdullah, M. Othman, R. Latip, M. N. Sulaiman, and H. Ibrahim, et.al described a strategy called Least Value Replacement (LVR), deals about storage constraints and QoS requirements under data grid. Here the storage limitation problem on replica has been overcome by replacement i.e. listing some information such as future values of files and frequent access. QADR problem considers the replication contention among data blocks because of replication storage limitation. This may cause some data replicas that cannot satisfy the QoS requirements of DIA [1]. Here the problem of violated replicas arises and the previous work doesn't proceed to minimize this violation. If there are many data object replicas to be placed in the server, the replicas of some data objects cannot be stored successfully for the server with limited storage space. In such case, the unsuccessful data object replicas will be put in other servers without QoS satisfaction. We can now undergo the problem of QADR [8] with various algorithms under cloud environment for data intensive applications.

Problem Definition of QADR:

1. Consider a cloud computing system which runs applications and stores data in a set of storage Node S and the functionality of Nodes are based on HDFS.
2. The running application writes a data block b to the disk of r, where $r \in S$, a request has been sent from r to make a number of copies of replica of b to the disks of other nodes. Many concurrent replication requests issued from different nodes at a certain time instant. Each node cannot store too many data replicas from other nodes due to space limitation.
3. The replicated data for block b will be stored at $q(q \in S)$ as dr. This data replica dr is related with RC and AC i .e. Replication cost and access time respectively with desired access time T.

4. If there is any data corruption and the original data b cannot be read successfully, the node r tries to get the data from the replication which is stored in q as dr. And the dr is termed as a QoS violated data replica if AC > T.
5. The QADR Problem in cloud tries to minimize the data replication cost and the count of QoS violated data replica using optimal replica replacement strategy to achieve the objective.

Replication Algorithms

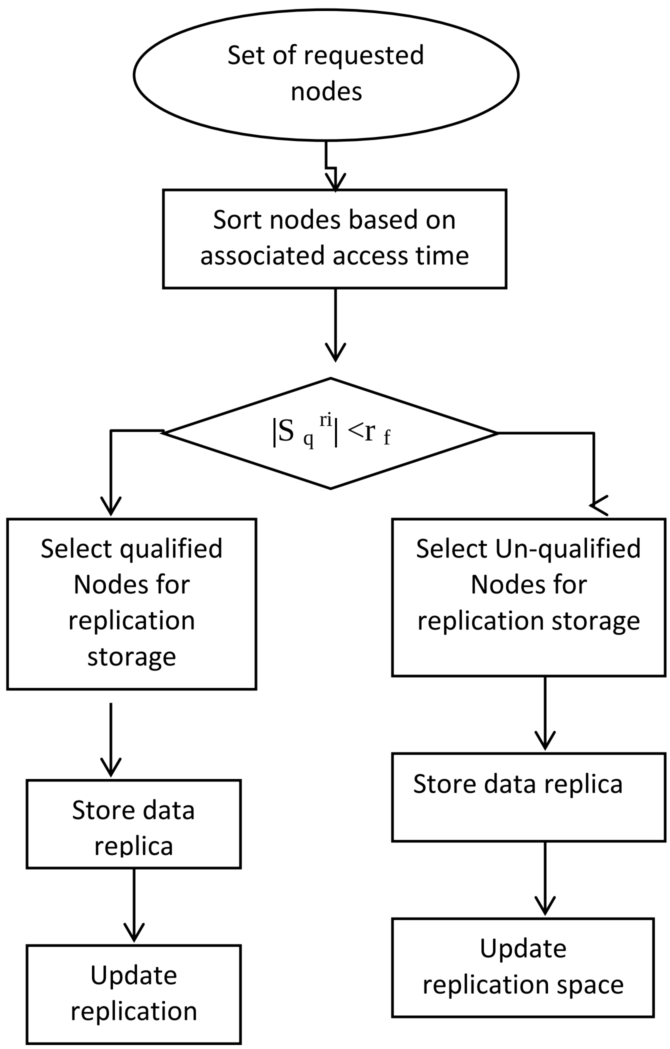

Fig 2: HQFR Algorithm

Optimal replica replacement:

In optimal replica replacement algorithm the set of requested nodes which is indicated as S_n and the output will be QoS-violated data replicas. It is easy to obtain the MCMF solution of the network flow graph by applying existing polynomial- time MCMF algorithm. In this we are using node combination techniques such as rack-based combination and equivalent state combination to avoid the large computational time.

Performance Evaluation:

We used real time public cloud to evaluate the performance of the replication algorithms in a large scale cloud computing system. We are experimenting our proposed algorithm under this cloud with 6 to 10 nodes. Windows Azure is Microsoft's application platform for the public cloud. Windows Azure is used to build a web application that runs and stores its data in Windows Azure datacenters [15]. Windows Azure provides several different ways to store and manage data. By HQFR algorithm we cannot minimize the replication cost [1] and number of QoS violated data replicas. To overcome this we used MCMFR algorithm which gives comparatively good result for obtaining the objectives but the computational time is high. By using C-MCMFR algorithm [1] the nodes are combined to reduce the computational time and the objectives are also achieved. In addition to replica allocation, *Energy Efficient Storage Node Identification Technique (EESNIT)*[13]is proposed to reduce energy consumption in transport and switching process involved.

The proposed system is evaluated by considering the parameters such as replication cost, average recovery time, Qos violation ratio [12] and execution time.

Replication cost [1] is the cost associated with replica deployment, dynamic replica creation and tear down at n candidate locations. Given client access patterns, replica failure patterns, network partition patterns, a required consistency level and a target level of availability, the minimal replication cost is the lower bound on a system's replication cost. The lower the replication cost better is the system performance.

If the requested node cannot read a data block from its disk due to data corruption, recovery time is the time taken by it to retrieve one replica of the data bock from another node. Shorter the recovery time better is the efficiency.

NODES	HQFR	MCMF	EQADR
8	0.82	0.6	0.709
12	0.912	0.862	0.817
16	1.013	0.96	0.92
20	1.138	1.086	1.04
24	1.262	1.4	1.12

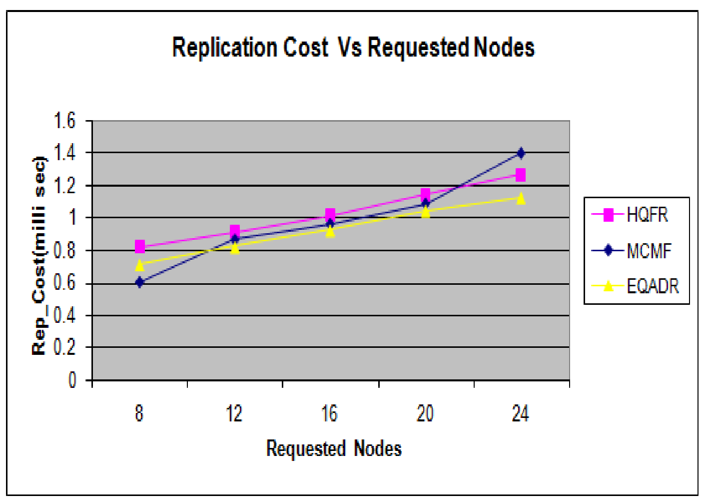

Fig 3: Replication Cost Vs Requested nodes

QoS violation ratio [1] is the ratio of total number of QoS-violated data block replicas to total number of data block replicas.

It is the time taken to complete the replication process which involves disk I/O for processing of large volume of data, replica identification and allocation

	HQFR	MCMF	EQADR
8	3.1	1.87	1.34
12	2.37	1.6	1.28
16	2.04	1.56	1.13
20	1.63	1.36	1.086
24	1.8	1.56	1.05

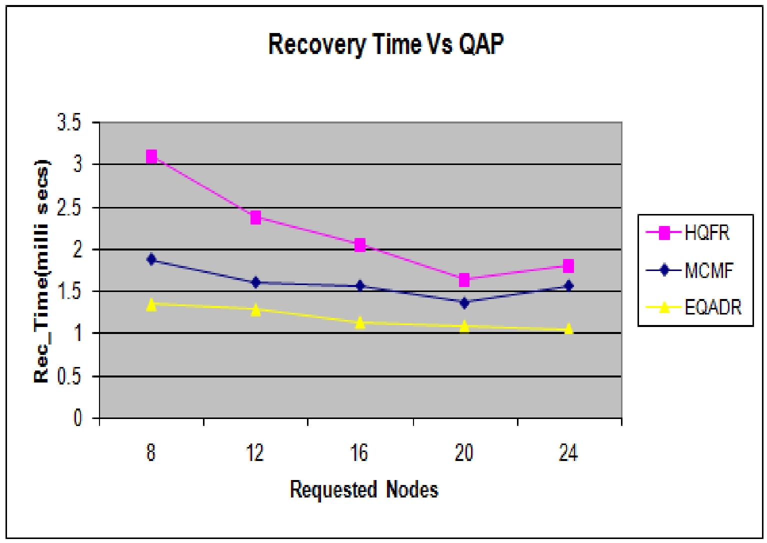

Fig 4: Recovery Time Vs Requested Nodes

Network-based cloud computing is rapidly expanding as an alternative to conventional office-based computing. As cloud computing becomes more widespread, the energy consumption of the network and computing resources that underpin the cloud will grow. This is happening at a time when there is increasing attention being paid to the need to manage energy consumption across the entire information and communications technology (ICT) sector. While data center energy use has received much attention recently, there has been less attention paid to the energy consumption of the transmission and switching networks that are key to connecting users to the cloud. Cloud computing can enable more energy-efficient use of computing power, especially when the computing tasks are of low intensity or infrequent and easily accessible.

Conclusion & Future enhancements

We have proceeded the QADR problem under real time public cloud and we solve the minimum objectives of the existing system. And also we used a technique called EESNIT to enable more computing power. Nowadays in ICT sector utilizes entire information with minimum energy consumption. So this technique is more useful for the users those who are working under cloud as it contains many storage nodes. Here we solved the problem of QoS requirement of an data intensive application [1].

In the future we can try to impose security for individual data which is going to get replicated. This algorithm will be also extended to use mobile service.

References

[1]QoS-Aware Data Replication for Data Intensive Applications in Cloud Computing Systems Jenn-Wei Lin, Chien-Hung Chen, and J. Morris Chang

[2] (2013) Apache Hadoop Project. [Online]. Available: http://hadoop. apache.org

[3]E. Pinheiro, W.-D. Weber, and L. A. Barroso, "Failure Trends in a Large Disk Drive Population," in Proc. 5th USENIX Conf. File and Storage Technologies, Feb. 2007, pp. 17–28.

[4]F. Wang, J. Qiu, J. Yang, B. Dong, X. Li, and Y. Li, "Hadoop High Availability through Metadata Replication," in Proc. 1st Int. Workshop Cloud Data Manage., 2009, pp. 37–44.

[5] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The Hadoop Distributed File System," in Proc. IEEE 26th Symp. Mass Storage Systems and Technologies (MSST), Jun. 2010, pp. 1–10.

[6]A. Gao and L. Diao, "Lazy Update Propagation for Data Replication in Cloud Computing," in Proc. 2010 5th Int. Conf. Pervasive Computing and Applications (ICPCA), Dec. 2010, pp. 250–254.

[7] W. Li, Y. Yang, J. Chen, and D. Yuan, "A Cost-Effective Mechanism for Cloud Data Reliability Management Based on Proactive Replica Checking," in Proc. 2012 12th IEEE/ACM Int. Symp. Cluster, Cloud and Grid Computing (CCGrid), May 2012, pp. 564–571.

[8]X. Tang and J. Xu, "QoS-Aware Replica Placement for Content Dis- tribution," IEEE Trans. Parallel and Distrib. Syst., vol. 16, no. 10, pp. 921–932, Oct. 2005.

[9]S. Ghemawat, H. Gobioff, and S.-T. Leung, "The Google File System," in Proc. 19th ACM Symp. Operating Systems Principles, vol. 37, no. 5, Dec. 2003, pp. 29–43.

[10]IEEE Standard for Local and Metropolitan Area Networks: Media Access Control (MAC) Bridges, IEEE 802.1D Std., 2004.

[11] M. Shorfuzzaman, P. Graham, and R. Eskicioglu, "QoS-Aware Dis- tributed Replica Placement in Hierarchical Data Grids," in Proc. 2011 IEEE Int. Conf. Advanced Inform. Networking and Applicat., Mar. 2011, pp. 291–299.

[12]H. Wang, P. Liu, and J.-J. Wu, "A QoS-Aware Heuristic Algorithm for Replica Placement," in Proc. 7th IEEE/ACM Int. Conf. Grid Computing, Sep. 2006, pp. 96–103.

[13] X. Fu, R. Wang, Y. Wang, and S. Deng, "A Replica Placement Algorithm in Mobile Grid Environments," in Proc. Int. Conf. Embedded Software and Systems (ICESS '09), May 2009, pp. 601–606.

[14]A. M. Soosai, A. Abdullah, M. Othman, R. Latip, M. N. Sulaiman, and H. Ibrahim, "Dynamic Replica Replacement Strategy in Data Grid," in Proc. 2012 8th Int. Conf. Computing Technology and Inform. Manage. (ICCM), Apr. 2012, pp. 578–584.

[15]http://www.windowsazure.com/en-us/documentation/articles/fundamentals-introduction-to-Windows-Azure/

