IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Volume-6, Issue-4 E-ISSN: 2347-2693

Data Migration Techniques within Cloud Computing: A Comprehensive Analysis

Kiranbir kaur[1*], Harpreet kumari [2]

[1,2]Guru Nanak Dev University, Amritsar, India

Available online at: www.ijcseonline.org

Received: 26/Mar/2018, Revised: 06/Apr/2018, Accepted: 19/Apr/2018, Published: 30/Apr/2018

Abstract- Cloud computing is becoming need of the hour for providing resources at pay per use to users. Data migration is the mechanism of transferring data to cloud where it is stored in virtual environment. It is key consideration behind the active data migration process where users storage is preserved. Up gradation or consolidation is accomplished within cloud using the application of data migration. During migration process, parameters are required to be validated. These parameters involve downtime and migration time. As the migration is finished, organization validates the transfer process statistically. The accuracy of data migration process is also questioned by the organization. in case accuracy is low migration is rejected. Data and pre-processing and cleaning facilities improve data quality via removal of unnecessary or repeated data. This paper presents the distinct data migration techniques within cloud used to transfer Users data to data centers for effectively storing and servicing the user. Techniques presented are compared comprehensively for future enhancements.

Keywords-Data migration, techniques, downtime, migration time, accuracy

I. INTRODUCTION

Cloud computing in modern era provides way of using resources without their physical presence at source. The service provided by the cloud is at the front end of computing and internet is at back end. In other words internet is heart and soul of cloud computing. [1]Cloud computing provide mechanism for the users to perform operations that required heavy resources not possessed by them at pay per use basis. With the rapid development of hardware and software cloud computing brings the revolution in the business industry. It provides resources like computational power, storage, computation platform ad applications to user on demand through internet. Some of the cloud providers are Amazon, IBM, Google, Sales force, Microsoft etc. [2]Cloud computing features included resource sharing, multi-tenancy, remote data storage etc. but it challenges the security system to secure, protect and process the data which is the property of the individual, enterprises and governments. Even though, there is no requirement of knowledge or expertise to control the infrastructure of clouds; it is abstract to the user. It is a service of an Internet with high scalability, quality of service, higher throughput and high computing power. [3]Cloud computing providers deploy common online business applications which are accessed from servers through web browser. Data security is the biggest issue in cloud computing and it is not easy to resolve it. In our review paper we will review the different ways to manage the confidentiality of the data. Before discussing migration mechanisms we discuss services provided by cloud along with types of cloud.

1.1 Cloud Services

[4] There exist legion of services associated with cloud. These services are as described below

- IaaS

[5]Infrastructure as services is critical services provided through cloud. virtualised computing resources are provided by the application of IaaS. Internet is key element with which IaaS is accessed. Cast is encountered on the basis of usage.

- PaaS

[6]Platform as a service is another cloud service that enhance the organizational applications. Large number of applications exits that are supposed to execute over the distinct machines. All the applications has distinct requirements in terms of platform. This platform requirement is accomplished using cloud computing. Cost is encountered on the basis of time period for which platform is online.

- SaaS

[7]Software as a service is another critical service supported through cloud. cloud computing host software which can be accessed by users having access to cloud. in other words

machines having limited resources can use SaaS to access software's that they don't possess.

Cloud services are accessed according to user requirements. User can use all the three services simultaneously depending upon requirements. Cost is encountered through pay per use basis. In all the levels of service provider, data migration is required. Data migration techniques are devised depending upon time consumption and reliability.

The techniques associated with data migration is described under the literature survey as under

II. LITERATURE SURVEY

Data migration in cloud indicates transforming data, application and business elements from organization onsite computers to cloud. Sometimes data and other applications are also migrated from one cloud environment to other cloud environment. Techniques associated with cloud migration are listed as under

2.1 FAULT TOLERANCE AND DATA MIGRATION

[1] Fault tolerance is critical since users heavily depending upon the cloud for operation. In case of failures data and other resources stored over the cloud could be lost. in order to overcome this problem, fault tolerance strategies become critical. In fault tolerant data migration strategies, migration of resources starts once prediction of failures initiated. [8]The migration to the fittest machine takes place. The fittest machine is selected on the basis of parameters. These parameters includes downtime, migration time and overall execution time. Overall migration enhance the performance since resources and data is already migrated to fittest physical machine. The rest of the task will initiate at the physical machine. Efficient fault tolerance techniques are divided into following two categories

2.1.1 Proactive fault tolerance
2.1.2 Reactive Fault tolerance

2.1.1 Proactive Scheduling

[9]Proactive fault tolerance techniques are the mechanism in which action is taken before the failure of virtual machines. The entire process is based on the prediction. The proactive fault tolerance techniques include

Software rejuvenation

[10], [11]Periodic reboots are performed under this approach. Every time system reboots, it starts from new state.

Self Healing

[12]It is the mechanism of controlling the failure of instance of process running on multiple virtual machine. The process once recovered is required to be restarted.

Preemptive scheduling

[12], [13]In this scheduling, resources are prompted from the process as the process goes in the state of failure. These resources are assigned to some other optimal process. The time consumption in execution of process is considerably removed. Determining optimal process is critical.

2.1.2 Reactive Scheduling

Reactive fault tolerance is the mechanism in which failure impact is reduced. In other words this techniques is implied only through the application of failure. Techniques under reactive fault tolerance is listed as under

Checkpointing

[14], [15]This is the mechanism in which progress is saved up to the established point. Once the progress reaches that point, it is automatically saved. The progress saved is known as savepoint. This mechanism is implemented as the reactive approach and it consume time in order to perform recovery.

Replication

[16]Task is replicated which is executed on virtual machine. This task is executed on several machines and in case of failure progress is recovered from the earliest finishing job. Replication however is space ineffective mechanism.

2.2 Load Balancing Migration Strategies

[17]This strategy improves the scalability of the server. The servers are analysed for load. The load is presented in terms of data and resources. In case load on the machines increases, load is shifted to the other server having least load. Load balancing strategy for migration is effective enough for tackling issues of downtime and migration time for data migration.

2.3 Energy Efficient Migration Technique

When Data between VM is to be migrated three factors must be considered. 1) Physical Memory image associated with virtual machine 2) network connection and virtual device state 3) Serial cum serial interface(SCSI). In order to optimize migration time, critical point of choosing technique to migrate required being decided. VM migration strategies are considered to be many and one of them is Pre-copy approach[18], [19], [20], [21]. It provides significant improvement in terms of down time. but this approach is not optimal in every situation. E.g. in case of memory intensive job, downtime and migration time increases drastically. Pre-copy approach also may not work properly in case of low bandwidth WAN environment. To resolve the issue, energy efficient strategy to maintain energy efficiency for fault tolerance in data migration across distinct host[22], [23].

All the above listed strategies are efficient enough but still modification for improvement could be desired and migration time can be further optimised using energy efficient mechanism.

2.4 COMPARATIVE ANALYSIS OF DATA MIGRATION TECHNQIUES

The comparative analysis of data migration strategies within cloud presents a way for future enhancements for reducing downtime and migration time during migration process.

Sr No.	Authors name, year	Technique	Description	Benefits	Drawbacks
1.	Jin et al. 2009	adaptive compression of migrated data.	Data being transferred in each round at source node is compressed by their algorithm and decompressed when arrived at target.	Migration time is reduced	overhead increased due to compression algorithm
2.	Liu et al. 2010	Propose Hierarchical copy algorithm	determines number of updated page, threshold and total write interrupt, if write interrupt < threshold than only updated page will be sent to destination .Dirty pages are sent in last iteration.	Migration time is reduced as number of iterations are reduced	Extended monitoring of memory image is required.
3.	Ma et al. presented 2010	improved pre-copy strategy	The pages which are updated frequently are transferred in last round of iteration process only once.	Reduces total migration time by 32.5% and 34% of total data transferred.	The downtime is increased as duplicate pages placed in the last round of the transmission.
4.	zhang et al. 2010	Migration with Data Duplication (MDD)	similar memory pages identified by using hash based finger prints. For further reducing data transfer rate it uses RLE (Run Length Encoding) algorithm.	56.60% decrease in total data transferred.	Indexing by hash fingerprints may cause data inconsistency.
5.	Ma et al. 2012	memory exploration and encoding (ME2) technique	Useful pages are identified then apply compression using run length encoding (RLE) algorithm.	Reduces total migration time and downtime.	Overhead increased due to compression algorithm
6.	hu et al. 2011, Johnson et al.x 2013	time series prediction technique	Used historical statistic of dirty pages it identify the high dirty page in iteration and do not send them repeatedly	Migration time reduced	Only high dirty pages are considered
7.	Jung et al. 2013	VM migration using checkpointing	When a running instance occurs the out-of-bid situation (failure),VM is migrated and it starts execution from saved checkpoint.	rollback time, task waiting time is reduced. Fault tolerance applied by check points.	Cost encounter is high which can be further reduced. Memory utilization is high.
8.	Bangjie Jiang, 2013	Priority-Based Live Migration strategy	Priority is assigned to the application. Dirty pages generated by high priority vm are transferred	Downtime is reduced by 57%	Total migration time is not considered

			to the target vm after a certain threshold value host vm is suspended and target vm is resumed. Dirty pages of Low priority applications are transferred using stop and copy approach.		
9.	Kim , 2015	Parallel migration approach-breaking chain migration	Vm list is partitioned and parallel migration is performed. Breakup VMs are selected recursively from all the split migration chains until the total migration time can no longer be reduced.	VM relocation time decreased by 21.9–62.0%	1.6 – 5.8% spare PMs required to parallelize the chained migrations

Table 1: Comparison of various data migration techniques used within cloud system.

III. RESEARCH GAP

Virtual machine utilization across data centers grabs significant attention in recent era[25], [26]. In case of faults or deteriorating machine detection, Live VM migration allows workload to be shifted across other optimal VM in some other host hence execution of work originated from source do not suffer. VM migration becomes key mechanism in cluster management including fault tolerance, power management, load balancing and online maintenance[27],[28], [29]. The data migration strategies considered above can be further optimising by identifying the critical and non critical data. In case of data migration, migration and downtime can be further reduced by identifying similar data which is being migrated from source to destination again and again. Also fittest virtual machine is required to be identified for migration. faster and reliable VM once identified, data can be migrated to that machine and parameter optimization can be achieved.

IV. PROBLEM DEFINATION

The existing literatures uses data migration strategies in which size of migrating data is not considered. By transforming such data towards the server machines causes problems since migration time considerably increased by the use of such transformation. Also energy efficiency is considerably be a problem. As size of data being migrated is not monitored hence cost associated with migration is also a problem. Redundancy handling mechanism included with migration process could be a solution to such problem. In general problems associated with existing literature is listed as under

4.1 Migration time can be reduced by avoiding similar data migration
4.2 Load balancing strategy can be implemented to enhance degree of fault tolerance.
4.3 Following live migration enhances utilization of resources and reduces downtime.

V. CONCLUSION AND FUTURE SCOPE

The comparative analysis of various techniques associated with data migration is presented in this paper. Technique used is used to increase the performance of migration. the optimality can be achieved further by identifying maximum power VMs along with critical data. Reliability will be greatly enhanced by identifying critical data. In case of prediction of failure, data migrated to other machines can be accessed using remote machines. Redundancy handling mechanism can be incorporated within data migration to optimize the parameters such as downtime and migration time.

VI. REFERENCES

[1] S. Asif, R. Shah, A. H. Jaikar, and S. Noh, "A Performance Analysis of Precopy , Postcopy and Hybrid Live VM Migration Algorithms in Scientific Cloud Computing Environment," pp. 229–236, 2015.
[2] H. Wang, Z. Kang, and L. Wang, "Performance-Aware Cloud Resource Allocation via Fitness-Enabled Auction," *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 4, pp. 1160–1173, Apr. 2016.
[3] T. Chalermarrewong, T. Achalakul, and S. C. W. See, "The Design of a Fault Management Framework for Cloud," *2012 9th Int. Conf. Electr. Eng. Comput. Telecommun. Inf. Technol.*, pp. 1–4, 2012.
[4] I. P. Egwutuoha, S. Cheny, D. Levy, B. Selic, and R. Calvo, "Energy efficient fault tolerance for high performance computing (HPC) in the cloud," *IEEE Int. Conf. Cloud Comput. CLOUD*, pp. 762–769, 2013.
[5] J. Guitart, M. Macias, K. Djemame, T. Kirkham, M. Jiang, and D. Armstrong, "Risk-driven proactive fault-tolerant operation of IaaS providers," *Proc. Int. Conf. Cloud Comput. Technol. Sci. CloudCom*, vol. 1, pp. 427–432, 2013.
[6] C. Pahl and I. Centre, "Containerization and the PaaS Cloud," 2015.
[7] F. Doelitzscher, A. Sulistio, C. Reich, H. Kuijs, and D. Wolf, "Private cloud for collaboration and e-Learning services : from IaaS to SaaS," pp. 23–42, 2011.
[8] M. E. M. Diouri, O. Gl??ck, and L. Lef??vre, "Towards a

novel smart and energy-aware service-oriented manager for extreme-scale applications," *2012 Int. Green Comput. Conf. IGCC 2012*, 2012.
[9] and R. B. Liu, Jialei, S. Wang, Ao Zhou, S.A P Kumar, "Using Proactive Fault - Tolerance Approach to Enhance Cloud Service Reliability," *IEEE Trans. Cloud Comput.*, pp. 1–13, 2016.
[10] M. S. Bruneo, Dario, S. Distefano, F. Longo, A. Puliafito, "Workload-based software rejuvenation in cloud systems.," vol. 62, no. 6, pp. 1072–1085.
[11] and J. T. L. Silva, J. Alonso, "Using Virtualization to Improve Software Rejuvenation," *IEEE Trans. Comput.*, vol. 58, no. 11, pp. 1525–1538, 2009.
[12] P. Gupta and S. Banga, "Topic - Review of Cloud Computing in Fault Tolerant Environment With Efficient Energy Consumption," vol. 1, no. 4, pp. 251–254, 2013.
[13] A. Kumar, "An Efficient Ch heckpointing Approach h for Fault Tolerance in Tim me Critical Systems wi ith Energy Minimization," pp. 704–707, 2015.
[14] R. Baldoni, J. M. Hélary, A. Mostefaoui, and M. Raynal, "On modeling consistent checkpoints and the domino effect in distributed systems," *Rapp. Rech. Natl. Rech. En Inform. En Autom.*, 1995.
[15] M. Salehi, M. K. Tavana, S. Rehman, S. Member, M. Shafique, and A. Ejlali, "Two-State Checkpointing for Energy-Efficient Fault Tolerance in Hard Real-Time Systems," pp. 1–12, 2016.
[16] W. Lang, J. M. Patel, and J. F. Naughton, "On energy management, load balancing and replication," *ACM SIGMOD Rec.*, vol. 38, no. 4, p. 35, 2010.
[17] J. Zhao, K. Yang, X. Wei, Y. Ding, L. Hu, and G. Xu, "A Heuristic Clustering-Based Task Deployment Approach for Load Balancing Using Bayes Theorem in Cloud Environment," *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 2, pp. 305–316, Feb. 2016.
[18] F. Ma, F. Liu, and Z. Liu, "Live Virtual Machine Migration based on Improved Pre-copy Approach," pp. 230–233, 2010.
[19] Y. Zhong, J. Xu, Q. Li, H. Zhang, and F. Liu, "Memory State Transfer Optimization for Pre-copy based Live VM Migration," pp. 290–293, 2014.
[20] D. Kapil, E. S. Pilli, and R. C. Joshi, "Live virtual machine migration techniques: Survey and research challenges," in *2013 3rd IEEE International Advance Computing Conference (IACC)*, 2013, pp. 963–969.
[21] M. R. Desai, "Efficient Virtual Machine Migration in Cloud Computing," no. Vm, pp. 1015–1019, 2015.
[22] Y. Liu and W. Wei, "A Replication-Based Mechanism for Fault Tolerance in MapReduce Framework," *Math. Probl. Eng.*, vol. 2015, 2015.
[23] A. Elghirani, R. Subrata, A. Y. Zomaya, and A. Al Mazari, "Performance enhancement through hybrid replication and genetic algorithm co-scheduling in data grids," *AICCSA 08 - 6th IEEE/ACS Int. Conf. Comput. Syst. Appl.*, pp. 436–443, 2008.
[24] D. Jung, S. Chin, K. S. Chung, and H. Yu, "VM Migration for Fault Tolerance in Spot Instance Based Cloud Computing," pp. 142–151, 2013.
[25] A. Gupta, U. Mandal, P. Chowdhury, M. Tornatore, and B. Mukherjee, "Cost-Efficient Live VM Migration Based on Varying Electricity Cost in Optical Cloud Networks," pp. 4–6, 2014.
[26] S. Rajput and A. C. Computing, "International Journal of Advanced Research in Computer Science and Software Engineering Live-VM Migration Policies , Attacks & Security – A Survey," vol. 4, no. 2, pp. 366–373, 2014.
[27] Z. Li, "Optimizing VM Live Migration Strategy Based On Migration Time Cost Modeling," pp. 99–109, 2016.
[28] V. Verroios and M. Roussopoulos, "Time-Constrained Live VM Migration in Share-Nothing IaaS -Clouds," 2014.
[29] S. Zhang, Z. Qian, Z. Luo, J. Wu, and S. Lu, "Burstiness-Aware Resource Reservation for Server Consolidation in Computing Clouds," *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 4, pp. 964–977, Apr. 2016.