A Survey of Techniques used for Dynamic Replication and Migration of Data in Cloud

Sandeep Kaur
Student, Department of Computer Engineering and Technology
Guru Nanak Dev University, Amritsar

Kiranbir Kaur
Assistant Professor, Department of Computer Engineering and Technology
Guru Nanak Dev University, Amritsar

***Abstract*- Cloud computing becomes need of the hour since it provides resources and allow users to store data beyond the capacity of their machines. In case machines are out of capacity, flow control is required to save from data loss. To overcome this problem, data migration to cloud comes to rescue. Performing data migration to cloud can be a daunting task, on the same time it facilitates the IT industry to improve in terms of infrastructure and architecture. The most common driver causing migration of data centre to cloud include: Capital expenditure reduction, reduced costing process, improving scalability and flexibility and improving security of stored data over the cloud. Strategies for data migration are researched over and this paper includes those techniques for evaluating and concluding best possible strategies along with future enhancements.**

KEYWORDS: Cloud Computing, Data Migration, flow control, Data centre

1. INTRODUCTION

Cloud computing provides extensive resources for application that needs them on pay-per-use basis [1]. It is a paradigm for information technology. Cloud provides the services in which resources are extracted from internet using web based tools and direct connection for which the server is not desired [2]. Dedicated resources such as hard drive are not required. It is possible to keep files at remote server using the application of cloud computing. As long as connection with the internet is present, users have access to data and software stored over the cloud. Cloud computing provides mobile access to resources since access to resources is not location dependent [3]. Cloud computing in its essence, take the crunching load from the device you carry around and transfer over the cluster of computers known as cyberspace. Internet is heart and soul of distributed computing supported through cloud computing [4]. With internet access, resources can be accessed from any location remotely. Before discussing data migration techniques that are required for migrating data to cloud, services provided by cloud must be elaborated.

1.1 SERVICES PROVIDED BY CLOUD

Services provided by cloud are the prime reason for its heavy utilization in modern era [5]. SaaS, PaaS and IaaS are the services provided by cloud for providing better and flexible services to the users. These services are described as follows:

1. Software as a Service (SaaS)
   This model provides complete application to the user on demand [6]. It can handle multiple users at a time, but in background, only single application of the service is executed. The single application facilitate the users, they need not to go for any advance payment. Various service provider like Google, SalesForce, Microsoft, etc. provides Saas facility.

2. Platform as a Service (PaaS)
   This model provides environment as a service to user for constructing its own application that run on service provider base [7]. It also provides predefined application server combined with OS to user. Google's App Engine, Force.com, etc. is providing a platform to users.

3. Infrastructure as a Services (IaaS)

This model provides services to fundamental storage and computing capability [8]. In order to manage workload, the shared resources are utilized among various users. To use the services, the user has to install his own software over the infrastructure. Amazon and GoGrid are some examples of IaaS.

In addition to services, there exists additional advantages associated with cloud which are discussed as follows:

1.2 CLOUD COMPUTING ADVANTAGES

Because of extraordinary merits over existing methodologies, use of cloud in present era is enhancing by leaps and bounds [9]. The advantages provided using cloud is listed as under:

- Cloud computing provides service known as SaaS, using which any software can be executed on any machine without requiring specific hardware or software.
- Cloud computing achieves cost effectiveness by allowing the user to access resources on the basis of pay-per-use.
- Mobility is offered in cloud computing using which resources can be accessed from any location using the application of internet.
- Cloud saves space since storage is not on local system, rather data is stored over the remote server.

However, in addition to advantages, there exists some disadvantages associated with cloud computing which are discussed as follows:

1.3 DISADVANTAGES OF CLOUD COMPUTING

With all the advantages including speed, efficiency etc., there comes a risk [10]. These risks includes:

- Security is an important issue while migrating data into cloud.
- Migrating data into the cloud and using it requires application of internet. In case internet is down all the services rendered unusable.
- Encryption protects data but if encryption key used in cloud is lost, then data disappears.
- Sufficient amount of training is required for employees in order to use the cloud.

DATA MIGRATION

From multiple points of view, data migration to a cloud computing condition is an activity in risk administration. Both subjective and quantitative components apply in an examination. The risks must be deliberately adjusted against the accessible protects and expected advantages, with the understanding that responsibility for security stays with the association. Excessively numerous controls can be inefficient and incapable, if the advantages exceed the expenses and related risks. A fitting harmony between the quality of controls and the relative risk related with specific projects and tasks must be guaranteed.

Data security is another vital research subject in cloud computing. Since specialist organizations regularly don't approach the physical security arrangement of data focuses, they should depend on the framework supplier to accomplish full data security. Even for a virtual private cloud, the specialist organization can just indicate the security setting remotely, without knowing whether it is completely actualized. The infrastructure supplier, in this specific circumstance, must accomplish the accompanying destinations: (1) secrecy, for secure data access and exchange, and (2) auditability, for bearing witness to whether security setting of utilizations has been altered or not. Classification is typically accomplished by utilizing cryptographic conventions, though auditability can be accomplished by utilizing remote validation strategies.

NEED FOR SECURING DATA MIGRATION PROCESS

Cloud Migration is one of much conversed point, where cloud managers confront extraordinary issues at the season of data migration from an organization’s server to a server that structures cloud somewhere else. The inconveniences

that are confronted must be bring into the light. As cloud acts as an interface through which associations can get to the data in a virtual domain. Subsequently, smooth working of it depends essentially on how very much prepped and educated cloud suppliers are here.

Also, if data migration isn't done appropriately and deliberately, it can offer ascent to issues concerning data and cloud security of organizations benefits that fundamentally contain data. Subsequently, employing cloud suppliers having sound involvement about the field with plentiful information and ranges of abilities ends up crucial for overseeing cloud all the more adequately and effectively.

LIVE VIRTUAL MACHINE

Migration is a method that moves the whole OS and its related application starting with one physical machine then onto the next. The Virtual machines are relocated lively without upsetting the application running on it. The advantages of virtual machine migration incorporate preservation of physical server vitality, stack adjusting among the physical servers and failure resilience in the event of sudden failure. The diverse virtual machine migration methods are as per the following:

(i) Fault Tolerant Migration Techniques

Fault resilience enables the virtual machines to proceed with its activity, even any piece of framework comes up short. This strategy moves the virtual machine starting with one physical server, then onto the next physical server. In view of the forecast of the disappointment happened, fault tolerant migration method is to enhance the accessibility of physical server and keeps away from execution corruption of utilizations.

(ii) Load Balancing Migration Techniques

The Load balancing migration method means to disperse load over the physical servers to enhance the versatility of physical servers in cloud condition. The Load balancing helps in limiting the asset utilization, execution of failover, improving adaptability, staying away from bottlenecks and over provisioning of assets and so forth.

(iii) Energy Efficient Migration Techniques

The power utilization of Data Centre is, for the most part, in view of the usage of the servers and their cooling frameworks. The servers commonly require up to 70 percent level of their most extreme power utilization even at their low use level. Subsequently, there is a requirement for migration procedures that preserves the vitality of servers by ideal asset use.

The work on cloud is still going on and security issues are somewhat tackled but still some work is required to be done [2][11][12][13]. The initial phase of data migration also suffers from this issue of security.

This paper discussed variety of mechanisms associated with data migration in cloud. Rest of the paper is organised as under:

- Section 2 gives brief introduction of the techniques used for data migration
- Section 3 gives issues with current migration techniques
- Section 4 gives comparison table to choose best possible technique for future work. This table also elaborate relative advantages and disadvantages associated with techniques
- Section 5 gives conclusion and future scope
- Section 6 gives references for citation.

2. EXISTING STRATEGIES FOR DATA MIGRATION

The existing migration strategies are applied on legacy systems along with databases [14][15][16]. The migration process is time consuming and huge amount of cost is encountered during migration. Research techniques have

generally concentrated into the cost minimization and time reduction. The strategies associated with data migration are as described below:

2.1 LEGACY MIGRATION

Legacy systems are old system having formats and syntax different from modern day systems [17]. Transferring data from legacy system involves huge amount of time and cost. In general, while migrating from legacy systems, following key factors are required to be tackled:-

- Data location: While migration, it is critical to identify what is to be migrated and where it is located.
- Data Dependency: The interface and other systems depending upon the legacy systems are either required to be shut off or made compatible to new system.
- Legal requirements to keep legacy data available: Even though all the data may not be migrated, but still additional requirement of space may be required in data warehouse for storing additional data corresponding to legal information.
- Infrastructure dependency: The formats required to support legacy information may be platform dependent. The hardware required to support legacy system must also be migrated.

Legacy migration is researched over and concluded that entity relationship model can be used to migrate legacy system in cheap and effective way [18]. The functional requirements are generally shown on the diagram. The EAI (Enterprise Application Integration) tool is commonly used to draw such type of diagram.

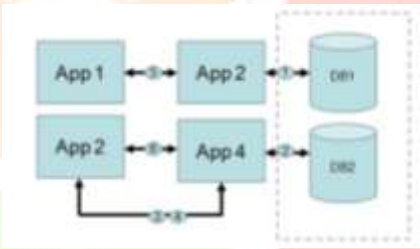

Figure 1: Migration of Legacy system

The circles with numbers indicate interface connecting apps and databases. The apps establish interface that could be command user or graphical in nature for transfer process. The interface resolves the differences of format and then communication is established.

2.2 REVERSE ENGINEERING PROCESS FOR DATA MIGRATION

Reverse engineering process of data migration provides the way of migrating database into the cloud [19]. Space required to store legacy information should be maintained at cloud data centres. Formats and interface are required to be modified, so that database at the destination end can be used as it is. Database query language support is included within the reverse engineering process.

Migration is performed by identifying all the dependencies associated with data being migrated. Once the dependencies are identified, migration process begins. The schema related queries are identified and tackled using the process of reverse engineering. Schema mapping process involves physical-to- conceptual mapping and conceptual-to-external mapping. The transfer process also involves verification process. Verification involves computability checking. The destination or target end is compatible only if number of attributes required for storing data matches with transferred database attributes. Data structure to be migrated could be complex in nature and testing of reverse engineering under complex data structure is an issue which is still to be tackled in the future.

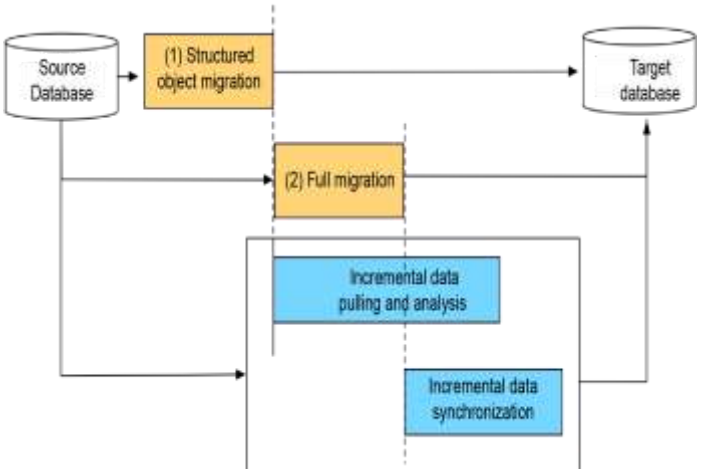

Figure 2: Example of data migration through reverse engineering

Figure 2 shows the steps performed in migration using reverse engineering. Incremental approach used in reverse engineering ensures saving of progress of previous process so that work can be started from where it is left off, hence migration time is reduced using this mechanism.

2.3 MIGRATION USING OOI(Object Oriented Interface)

The object oriented interface provides graphical users interface (GUI) in which user can operate through clicking and double clicking [20]. The main advantage of using such interface is, more concentration is paid towards data security rather than modules or procedures. In order to perform migration keys including primary and foreign are considered. Primary key does not allow replicated data to be migrated again and again, hence lots of space is saved. In addition, foreign key is used to relating multiple tables or relations together. The values of foreign key either matches with the primary key of other table or it should contain null value. Hence only those values which are relevant are retained and rest of the values are neglected causing the storage space to be optimised and therefore, less time is consumed during migration.

2.4 INTEGRATING DATABASE

The database integration allows source and target database to be related and allows transfer by maintaining information in integrated form [21]. Integrity rules are implied in this case. Integrity rules are categorised according to the following:

- Entity Integrity Rule: This rule states that value of the primary key should be unique and it cannot contain null value.
- Referential integrity rule: It states that value of foreign key should either match with the primary key of source or it should contain null value.

Integrating database ensures that valid data is contained within the valid fields. Integrating database also ensure that domain integrity is maintained. The major flaw of this technique is additional attributes required at destination end that cannot be maintained because of space and time constraints.

2.5 SOURCE TO TARGET MAPPING

This is a direct process of migration in which physical view is directly mapped to the target machine [22]. This is a one-step approach, because of which many of integrity constraints may be violated, resulting in ill-mannered database designed. Source to target mapping, hence is performed in emergency only and not in most of the modern day migration mechanisms.

3. ISSUES WITH DATA MIGRATION TECHNIQUES

The issues and challenges considering in data migration are listed as under [23][24]:

- Reduce capital expenditure: During migration, cost associated with maintenance of records requires huge capital expenditure. Capital expenditure involves purchase of space and hardware resources for maintaining efficiency during migration.
- Reduce ongoing process: The ongoing cost is reduced, since cloud provider gives the option to the user to select the plan on the hourly, daily or monthly basis.
- Improve scalability: The organization is allowed to choose amount of resources which can be increased or decreased depending upon user requirements.
- Improved Speed: This is the prime requirement of every business need. Speed of processing of an application must be enhanced, in case organization is opted for migration. This is not always the case, hence modification to existing techniques are required.
- Security: As more and more users are associated with cloud, hence data presented within cloud is always accompanied with security threat. Stronger form of encryption mechanism such as de-duplication can be used along with data migration for security. It is still an issue and researchers are working towards secure transmission of data within cloud computing.

4. COMPARISON TABLE FROM EXISTING WORK

The existing work is presented in tabular form in this section. Using this table, future work can be derived.

REFERENCE	TECHNIQUE	PARAMETERS	MERITS	DEMERITS
[25]	Butterfly Methodology	(a)Threshold value (b)Legacy Sample Data and Target Sample Data (c)Data Access Allocator (d)Data Transformer and Termination Condition	(a)Clearly defined & strong support given to testing (b)Flexible in terms of use of application tools (c)Gateway-free approach (d)Total duration of migration can be estimated (e)Promotes Parallel activities	(a)Requires thorough understanding of the Legacy &Target system (b)Accurate & concise Sample Database is required (c)Efficient Data-Access-Allocator is required
[26]	Online Algorithms:- [1] Online Lazy Migration (OLM) Algorithm [2] Randomized Fixed Horizon Control (RFHC) Algorithm	(a)Cost (b)Performance bottleneck	(a)Focus is on dynamically generated data rather than on static data (b)Focuses on multiple geo-distributed data centers rather than on single cloud site	Running time and cost of RFHC algorithm may be higher than offline algorithm in some conditions
[27]	[1] Chicken Little Migration Approach [2] Butterfly	(a)Legacy and Target system migration (b)Forward and	(a)In CLM, gateways co-ordinate the migration b/w	(a)complex technical problem regarding

	Methodology	Backward Gateway (in Chicken Little Migration)	legacy & target systems (b)maintains data consistency (c)simultaneous access to both Legacy and Target systems in Butterfly methodology	updation consistency (b)using gateways add complexity in whole process (c)justification of whole process needs to be done
[28]	(a)Data Intensive Applications i.e. Hegira4Cloud (b)NoSQL Database	Offline data migration done by Hegira4Cloud	(a)Data before and after migration is equivalently same (b)Relevant techniques are integrated	(a)Time-consuming development-testing approach (b)Expensive approach in some cases
[29]	Genetic Algorithms	(a)Optimal or near-optimal service migration (b)Service-Oriented Architecture (SOA)	(a)Network traffic congestion is removed (b)Real-time requirements are fulfilled	(a)Managing the service of migration can be a complex task
[5]	(a)Meta-Synthesis Method (b)Maturity model, "ClM3"	Selecting a Cloud Migration Framework with dynamic nature as well as integrity	Phase by phase migration process to detect any shortcoming	Selection procedure of a migration process may be biased sometimes
[7]	(a)Optimal Offline Algorithm (b)Online Algorithms: [1] Deterministic [2] Randomized	Cost	Minimizing the data placement cost for time-varying workload applications by exploiting the price difference b/w storage and network service	High time complexity of the offline algorithm
[8]	No SQLayer Framework	Data migration from relational (i.e., MySQL) to NoSQL DBMS (i.e., MongoDB)	(a)Semantics of original database are preserved (b) No need of modifying queries and application code (c) suitable to handle large volume of data	Overhead generated by the persistence layer makes the framework a less efficient for small volumes of data

Table 1: Comparison of various data migration techniques

From the comparison of existing literature, it is concluded that modification to existing techniques is must to minimize time and cost associated with data migration.

5. CONCLUSION AND FUTURE SCOPE

This paper pays a stress on existing migration techniques for migrating data. The migration of data involves time and space complexity. The time complexity avoids the use of data migration, hence reduces utilization of cloud in modern era. Some modification to existing techniques including redundancy handling, integrity check and priority calculation for migrations are required for improving existing data migration. In case, cost, space and time complexity associated with migration is reduced, then cloud usage in near future could be improved greatly.

6. REFERENCES

[1] D. Singh, J. Singh, and A. Chhabra, "High availability of clouds: Failover strategies for cloud computing using integrated checkpointing algorithms," *Proc. - Int. Conf. Commun. Syst. Netw. Technol. CSNT 2012*, pp. 698–703, 2012.

[2] O. Harfoushi, B. Alfawwaz, N. a. Ghatasheh, R. Obiedat, M. M. Abu-Faraj, and H. Faris, "Data Security Issues and Challenges in Cloud Computing: A Conceptual Analysis and Review," *Commun. Netw.*, vol. 06, no. 01, pp. 15–21, 2014.

[3] X. Chen, "Decentralized Computation Offloading Game for Mobile Cloud Computing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 4, pp. 974–983, Apr. 2015.

[4] J. Liu, F. Zhao, X. Liu, and W. He, "Challenges Towards Elastic Power Management in Internet Data Centers," *2009 29th IEEE Int. Conf. Distrib. Comput. Syst. Work.*, pp. 65–72, 2009.

[5] M. G. Mcgrath, P. Raycroft, and P. R. Brenner, "Intercloud Networks Performance Analysis," pp. 487–492, 2015.

[6] 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.

[7] C. Pahl and I. Centre, "Containerization and the PaaS Cloud," 2015.

[8] B. Nicolae, "BlobCR : Efficient Checkpoint-Restart for HPC Applications on IaaS Clouds using Virtual Disk Image Snapshots."

[9] B. Lin, S. Li, X. Liao, Q. Wu, and S. Yang, "eStor: Energy efficient and resilient data center storage," *Proc. - 2011 Int. Conf. Cloud Serv. Comput. CSC 2011*, pp. 366–371, 2011.

[10] B. Meroufel and G. Belalem, "Adaptive time-based coordinated checkpointing for cloud computing workflows," *Scalable Comput.*, vol. 15, no. 2, pp. 153–168, 2014.

[11] J. I. A. You, Z. Zhong, G. Wang, B. O. Ai, and S. Member, "Security and Reliability Performance Analysis for Cloud Radio Access Networks With Channel Estimation Errors," vol. 2, 2014.

[12] P. You, Y. Peng, W. Liu, and S. Xue, "Security Issues and Solutions in Cloud Computing," 2012.

[13] X. Wu, R. Jiang, and B. Bhargava, "On the Security of Data Access Control for Multiauthority Cloud Storage Systems," pp. 1–14, 2015.

[14] A. Cardoso, F. Moreira, P. Simões, U. Portucalense, R. António, and B. De Almeida, "A Survey of Cloud Computing Migration Issues and Frameworks," vol. 1, pp. 161–170, 2014.

[15] S. Yang, C. Tu, and J. Lin, "Design Issue and Performance Analysis of Data Migration Tool in a Cloud-Based Environment," pp. 749–759.

[16] H. Hammami, H. Brahmi, and I. Brahmi, "A Security Approach for Data Migration in Cloud Computing Based on Human Genetics," pp. 384–396, 2017.

[17] B. Cai, F. Xu, and F. Ye, "Research and Application of Migrating Legacy Systems to The Private Cloud Platform with CloudStack," no. August, pp. 400–404, 2012.

[18] A. Ahmad and M. A. Babar, "A Framework for Architecture-driven Migration of Legacy Systems to Cloud-enabled Software," 2014.

[19] N. Stojanovic, L. Stojanovic, and R. Volz, "A reverse engineering approach for migrating data-intensive web sites to the Semantic Web."

[20] O. Database, S. Monk, H. Campbell, J. Mariani, and B. Eigalal, "Tool Support for Migration from a Relational to an," 1996.

[21] Q. Chen, H. Hu, and J. Xu, "Authenticated Online Data Integration Services," pp. 167–181.

[22] A. Yamanaka, T. Suda, T. Yokohira, Y. Fukushima, and T. Murase, "Destination Selection Algorithm in a Server Migration Service," pp. 15–20, 2012.

[23] R. Nassif and D. Mitchusson, "Issues and approaches for migration/cohabitation between legacy and new systems," pp. 471–474.

[24] S. D. Bhopale, "Cloud Migration Benefits and Its Challenges Issue," pp. 40–45, 2008.

[25] T. Lu, M. Stuart, K. Tang, and X. He, "Clique Migration : Affinity Grouping of Virtual Machines for Inter-Cloud Live Migration," 2014.

[26] O. Rana, "The Costs of Cloud Migration," 2014.

[27] C. O. L. U. Clo, M. N. Ud, and T. Ov, "Options for Cloud," pp. 62–66, 2014.

[28] T. Zeb, A. Ghafoor, A. Shibli, and M. Yousaf, "A Secure Architecture for Inter-cloud Virtual Machine Migration," vol. 1, pp. 24–35, 2015.

[29] M. R. Abid, K. Kaddouri, K. Smith, M. Idriss, E. Ouadghiri, and M. Gerndt, "Virtual Machines ' Load-Balancing in Inter-Clouds," pp. 5–12, 2016.