Data Migration issues in Cloud Computing: A Survey

Mrs. S. M. Barhate

Dr. M. P. Dhore

Abstract- data migration is an issue in focus from the view of cloud computing. Cloud experts stressed upon the importance of data migration from one cloud to another. Considerable amount of progress is being done in the cloud computing right from lock-in, interoperability, standardization, redundancy to data migration. With effective data migration the problems such as cost of migration, time to migrate, and security are the things which need to be seriously thought upon. This paper deals with some of the important data migrations techniques.

Key words: data migration, interoperability, standardization, lock-in.

I. INTRODUCTION TO CLOUD COMPUTING

cloud computing is a new boom in it that moves computing and data away from desktop and portable pcs into huge data centers. It refers to applications delivered as services over the internet and also to the actual cloud infrastructure i.e. The hardware and systems software in data centers that provide these services.

the motivating forces behind cloud computing are the ubiquity of broadband and wireless networking, falling storage costs, and progressive improvements in internet computing software. Cloud-service clients will be able to add more capacity at peak demand, reduce costs, experiment with new services, and remove unneeded capacity, whereas service providers will increase utilization via multiplexing, and allow for larger investments in software and hardware.

currently, the main technical underpinnings of cloud computing infrastructures and services include virtualization, service-oriented software, grid computing technologies, management of large facilities, and power efficiency. Consumers purchase such services in the form of infrastructure-as-a-service (iaas), platform-as-a-service (paas), or software-as-a-service (saas) and sell value-added services (such as utility services) to users. Within the cloud, the laws of probability give service providers great leverage through statistical multiplexing of varying workloads and easier management — a single software installation can cover many users' needs. We can distinguish two different architectural models for clouds: the first one is designed to scale out by providing additional computing instances on

Demand to supply services in the form of saas and paas. The second architectural model is designed to provide data and compute-intensive applications via scaling capacity. In most cases, clouds provide on-demand computing instances or capacities with a "pay-as-you-go" economic model. The cloud

infrastructure can support any computing model compatible with loosely coupled cpu clusters. Organizations can provide hardware for clouds internally (*internal Clouds*), or a third party can provide it externally (*hosted clouds*). A cloud might be restricted to a single organization or group (*private clouds*), available to the general public over the internet (*public clouds*), or shared by multiple groups or organizations (*hybrid clouds*).

II. WHAT IS DATA MIGRATION

the process of transferring data between storage types, formats, or computer systems is known as data migration. Data migration is a very significant factor to consider while stages such as system implementation, upgrade. Data migration is carried out programmatically to achieve an computerized migration, liberating human resources from tedious tasks. Data migration occurs for a variety of reasons, including server or storage equipment replacements, maintenance or upgrades, application migration, website consolidation and data center relocation.

to achieve an effective data migration procedure, data on the old system is mapped to the new system utilization a design for data extraction and data loading. The design relates old data formats to the new system's formats and requirements. The process of data migration may involve many phases but it minimally includes data extraction where data is read from the old system and data loading where data is written to the new system. On being shifted to new system results undergo data verification to determine whether data was precisely translated, is complete, and supports processes in the new system. During verification, there may be a need for a parallel run of both systems to identify areas of disparity and to match erroneous data loss. [1] programmed and manual data cleaning is usually performed in migration to improve data quality, eliminate redundant information, and match the requirements of the new system. there are various phases for data migration like design, extraction, cleansing, load, verification for applications of moderate to high complexity and are repeated several times before the new system is deployed.

the storage of data done on various media in files or databases, and is generated and consumed by software applications which in turn support business processes. The need to transfer and convert data can be driven by multiple business requirements and the approach taken to the migration depends on those requirements.

II. TYPES OF DATA MIGRATION

four major migration categories are proposed on the following basis:

A) *Storage migration*

A business may choose to downsize the physical media to take advantage of more efficient storage technologies thereby having more available physical blocks of data which is usually implemented through <u>virtualization</u> techniques. The data format and content does not get changed in the process and can normally be achieved with minimal or no impact to the layers above.

B) *Database migration*

it may be necessary to move from one <u>database</u> vendor to another, or to upgrade the version of database software being used. In these cases a physical transformation process may be required since the underlying data format can change significantly. This may or may not affect behavior in the applications layer, depending largely on whether the data manipulation language or protocol has changed. For e.g. Applications are written to be skeptic to the database technology so that a change from <u>sybase</u>, <u>mysql</u>, <u>db2</u> or <u>sql server</u> to <u>oracle</u> should only require a testing cycle to be confident that both functional and non-functional performance has not been adversely affected.
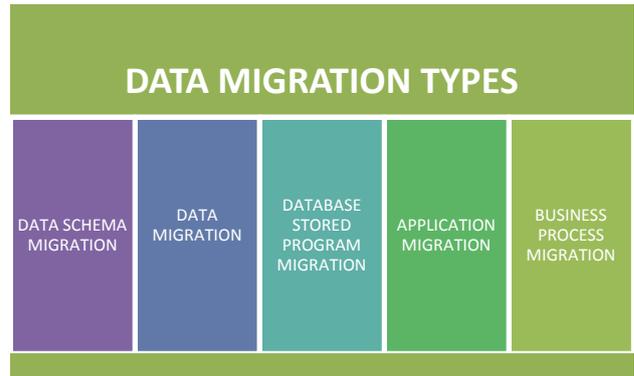
C) *Application migration*

changing platforms will certainly involve substantial transformation as almost every

Application operates on its own specific data model and also interacts with other applications and systems within the enterprise application integration environment. Moreover to allow the application to be sold to the widest possible market, commercial off-the-shelf packages are generally configured for each customer using <u>metadata</u>. Application programming interfaces (apis) may be supplied by vendors to protect the integrity of the data they have to handle.

D) *Business process migration*

business processes operate through a combination of human and application systems actions, often coordinated by business process management tools. When these change they can require the movement of data from one store, database or application to another to reflect the changes to the organization and information about customers, products and operations. Examples of such migration drivers are mergers and acquisitions, business optimization and reorganization to attack new markets of respond to competitive threat. [1]

DATA MIGRATION TYPES

DATA SCHEMA MIGRATION · DATA MIGRATION · DATABASE STORED PROGRAM MIGRATION · APPLICATION MIGRATION · BUSINESS PROCESS MIGRATION

IV. ISSUES IN DATA MIGRATION

the few issues that are considered for cloud data migration are:-

- Security in data migration
- Standardization
- Interoperability
- Cost and complexity for migration
- Time for data migration
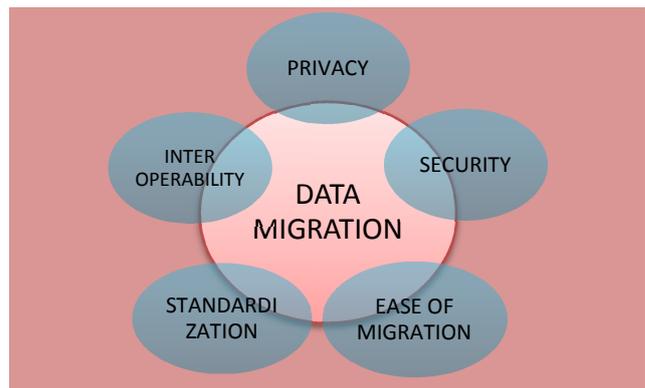
A) *Security in data migration*

cloud utilizes three service models by which various services are provided to end users. The three delivery models are saas, iaas, paas which provide infrastructure resources, platforms, and software to the end users. Due to this a unique security level requirement is needed for each service model.

saas is software security model where software services are deployed remotely and made available to customers on demand. Saas is an emerging deployment model which provides benefits such as operational efficiency and reduced costs. But this model lacks clear visibility of how the data is stored and secured. A great care should be taken for data confidentiality, data locality, data breaches, web application security, network security loss of sensitive data and money in saas. In the saas model, the enterprise data is stored outside the enterprise boundary, at the

saas vendor end. Hence, the saas vendor must adopt additional security checks to ensure data security and prevent breaches due to security vulnerabilities in the application or through malicious employees. This involves the use of strong encryption techniques for data security and **authorization** to control access to data.

Iaas provides with resources such as operating systems, networks etc instead of spending big money on setting up costly infrastructure at data centre. Iaas has completely abstracted the hardware beneath **it** and it allows the free access to consumer without any view of complexity. Iaas only provides basic security such as perimeter firewall, load balancing but application entering to cloud need higher level of security.

paas is the layer above iaas and abstracts everything up to os, middleware etc. Paas offers developer a service that provide complete life cycle management right from planning to deployment and all else is hided from the developer. Metrics should be in place to assess the effectiveness of the application security programs. Some security specific metrics available are **vulnerability scores** and **patch coverage**. These metrics can indicate the quality of application coding. Hackers are likely to attack visible code running in user context thereby attacking infrastructure and perform extensive black box testing. [2]

B) Standardization

data that resides in one cloud provider can be moved to another cloud provider. A standardization effort that supports this use case is cloud data management interface (cdmi). Although soap and rest are not data-specific standards, multiple cloud-storage providers support data and storage-management interfaces that uses simple object access protocol (soap) and representational state transfer (rest). [3]

open cloud standards are considered the eventual solution to issues around application migration and cloud interoperability. We view cloud standards as a collection; this one starts at the low level with something like **ovf** (open virtualization format) that gives you a universal language for describing the metadata and configuration parameters of virtual machines. This would give you the networking between the virtual machines *and* the functions and scale of the environment in which the virtual machines operate.

for example, standardization of vm (virtual machine) image file formats would allow organizations to move workloads from one provider to another or from private to public clouds.

the goal of data standardization is to enable the sharing or exchange of information between multiple parties in a way that guarantees that the interacting parties share the same understanding of what is represented within that information. As more information is being exchanged in different operating environments, the need for defined data standards is becoming more acute in order to coordinate that information exchange in a way that provides the most benefit to all participants.

C) interoperability

the ability to move the services from one cloud to another in a simple way is cloud service portability which can be termed as interoperability. In an interoperable cloud system, different **cloud platforms** should be able to collaborate. The interoperability includes:

➢ Data.
➢ Applications.
➢ Physical or virtual machines.
➢ Various features like provisioning, policy, sla (service level agreement) and qos.

Following are the few existing technologies which can be used for interoperability problem:

- **Intercloud:** the intercloud is an interconnected global "cloud of clouds".
- **Standardization** appears to be a good solution to address the interoperability issue. However, as cloud computing just starts to take off, the interoperability problem has not appeared on the pressing agenda of major industry cloud vendors. For example, neither microsoft nor amazon supports the *unified cloud interface* (uci) project proposed by the cloud computing interoperability forum (ccif) [4].

D) cost and time constraints for data migration

data migration is an ongoing activity at all enterprise it data centres. Traditional data migration consumes considerable resources, and are costly and involves a lot of complexity. Massive public clouds may be more cost effective than large community clouds which may be more cost effective than small private clouds. Cleansing of data becomes very important as cost depend on the size of the data. As the size of the database grows , cost will also grow up. Hence eliminating garbage data from the data from database for cost effective migration is very necessary. [5]

V. DATA MIGRATION TECHNIQUES

A) Pre-migration

pre-migration techniques consist of prior transformation actions before data migration. This step is for the ease of transformation as the data is priorly changed to required format. These actions may include:

- Virtualization
- Data separation
- Server platform upgrades

B) Post-migration

in this the transformation activity the transformation activities are done after completion of migration. The data

centre migration program winds up on successful migration of data to cloud.

CONCLUSION

Data migration is an important event that consumes significant budget and labor. Cloud migration is a very complex procedure. Migration of cloud is vital for achieving real time and updated performance and efficiency. Cloud migration requires careful analysis, planning and execution to ensure cloud solution's compatibility with organizational requirements. In cloud migration we are moving sensitive data to the cloud, so care should be taken about the security and privacy of data. Amazon proposes a phase driven approach for cloud application migration,. The data migration is done in two steps selection of the amazon aws service, and migration of the data. Furthermore, amazon provides recommendations regarding which of their data and storage services best fit for storing a specific type of data, e.g., amazon relational database service. Migration products that maintain continuous data availability during the migration without affecting performance are desirable.

REFERENCES

1. "best practices for data migration", ibm global services, ibm report, june 2007.

2. Bing wu * et al, " the butterfly methodology : a gateway free approach for migrating legacy information", ieee xplore, 16 jan 2009.

3. Rashmi rao, pawan prakash, "improving security for data migration in cloud computing using randomized encryption technique "iosr journal of computer engineering (iosr-jce) e-issn: 2278-0661, p-issn: 2278-8727,volume 11, issue, pp 39-426 (may. - jun. 2013). Www.iosrjournals.org.

4. Prashant pant, sanjeev thakur, "data migration across the clouds", international journal of soft computing and engineering (ijsce) issn: 2231-2307, volume-3, issue-2, may 2013.

5. Virendra singh kushwah, aradhana saxena," a security approach for data migration in cloud", international journal of scientific and research publications, volume 3, issue 5, issn 2250-3153, may 2013.

6. Wei hao, i-ling yen, "dynamic service and data migration in the clouds", 33rd annual ieee international computer software and applications conference, 2009.

7. Divya kapil, e.s.pilli, r.c.joshi "live virtual machine migration techniques: survey and research challenges" .third ieee international advance computing conference(iacc), 2013.

AUTHOR'S PROFILE

Mrs.shweta.m.barhate is assistant professor in computer science, department of electronics & computer science, rtm nagpur university, nagpur. She is pursuing ph.d in computer science and has a teaching experience of 10 years of ug and pg level. Email id- shwetab73@yahoo.com

Dr. Mahendra p. Dhore is associate professor in computer science, department of electronics & computer science, rtm nagpur university, nagpur. He is having teaching experience of more than 18 years at ug & pg level. His research areas include digital image processing, document image analysis, mobile computing & cloud computing. He is member of ieee, iaeng, iacsit, iete, and isca. Email id – mpdhore@ieee.org
mpdhore@rediffmail.com