International Journal of Computing and Artificial Intelligence

E-ISSN: 2707-658X
P-ISSN: 2707-6571
IJCAI 2022; 3(2): 96-101
Received: 18-06-2022
Accepted: 22-07-2022
www.computersciencejournals.com/ijcai

Venkata Nagesh Boddapati
Microsoft, Support Escalation Engineer, United States

Manikanth Sarisa
Ally Financial Inc, Principal Software Engineer, United States

Mohit Surender Reddy
Microsoft, Support Escalation Engineer, United States

Janardhana Rao Sunkara
Siri Info Solutions Inc. Sr. Oracle Database Administrator, United States

Shravan Kumar Rajaram
Microsoft, Support Escalation Engineer, United States

Sanjay Ramdas Bauskar
Pharmavite LLC, Sr. Database Administrator, United States

Kiran Polimetla
Adobe Inc, Adobe Technology Service Department, United States

Corresponding Author:
Venkata Nagesh Boddapati
Microsoft, Support Escalation Engineer, United States

Data migration in the cloud database: A review of vendor solutions and challenges

Venkata Nagesh Boddapati, Manikanth Sarisa, Mohit Surender Reddy, Janardhana Rao Sunkara, Shravan Kumar Rajaram, Sanjay Ramdas Bauskar and Kiran Polimetla

DOI: https://doi.org/10.33545/27076571.2022.v3.i2a.110

Abstract
An important development in information technology, cloud computing allows users to share Internet-based access to pre-con Figured systems and services. While there are many benefits, such cost efficiency and scalability, security is still a big worry for everyone involved. The current practices in authentication have been found to be wanting in providing for the principles of CIA triad; confidentiality, integrity and availability. Data transfer to the cloud is also known as data migration, which takes data from on-premises databases together with other cloud services and which is normally associated with many problems such as data integrity and minimize down time. Additional barriers stem from the continuously maturing cloud environments and different levels of compatibility with the given database structures. This paper focuses on the processes that are involved in data migration and different catalogs of migration including, database migration, data center migration, application migration, business process migration and so on, stressing the significance of planning and implementing these migrations efficiently. The main issues that demand shifting to the cloud are outlined as well as the main approaches that large cloud suppliers such as AWS, Microsoft Azure, and Google Cloud offer. Additionally, potential risks and challenges, such as vendor selection, security concerns, and resource management, are explored. This comprehensive overview highlights the significance of strategic planning and vendor solutions in ensuring successful cloud data migration, while addressing the inherent risks associated with transitioning to cloud-based infrastructures.

Keywords: Database, cloud computing, virtualization, database as a service, data migration, vendors solutions

Introduction
The most popular IT trend right now is "cloud computing," which allows users to access shared configurations of systems and services over the Internet. Cloud security offers a more comprehensive framework for safeguarding data, apps, and the underlying cloud infrastructure. Cloud computing is mostly used by businesses to store and manage massive amounts of data across computers. On the other hand, cloud providers and customers are starting to worry about data security. There has been a failure with the use of conventional authentication mechanisms such as password and key creation. The cloud makes it more difficult to accomplish the three primary aims of CIA: data availability, data integration, and data secrecy. If you want to move data from one cloud service to another, or even just from one database to another, you need to do a cloud migration.

Data migration to the cloud is driven by factors like cost efficiency and scalability, but concerns over security and trust in cloud providers can complicate the process, especially when switching between providers [1]. If a cloud provider discontinues its services, users need to securely transfer data to another provider. The three primary cloud service models include Infrastructure as a Service (IaaS), which manages storage, virtualization, and networking for third-party data centers (e.g., AWS, Google Compute Engine); Platform as a Service (PaaS), which offers a porTable application environment for development and integration (e.g., AWS Elastic Beanstalk, Google App Engine) and Software as a Service (SaaS), providing on-demand software for CRM and business management(e.g., Google Apps, Salesforce).

Deployment can be on the public cloud, private cloud or hybrid cloud with the later known to give better security, availability and costs cutting. The best practices for the migration may differ: direct transfer between clouds can be enabled, while direct download & upload is slower. Main guidelines within data migration into the cloud include planning, analysis of strengths, weaknesses, opportunities, selecting the proper cloud environment and architecture, accurate selection of a cloud provider and partner [2].

More specifically, in the migration into the cloud databases, there is need for keeping some elements of the database structure consistent, and logical, and not allow for downtime. The cloud environment can have vastly different architecture, response times, and features for security while the migration of data between on-premise or between twelve providers is challenging. Incompatibilities may occur based on the problem of database schema, storage media and data models supported. Also, data migration is a massive transfer of large volumes of data which would need a lot of bandwidth to complete, and can also take a lot of time and even if the process is interrupted accidentally, data can be lost or corrupted. Managing the updates in real-time data becomes even more challenging since it has a direct impact on the process of migration without resulting to service disruptions or decline in the quality of service [3, 4].

A. Contribution of the study

This work will look at the processes, types, challenges, and vendor's solutions of data migration in cloud databases. It aims at providing organisations with knowledge on how best to undertake migration with specificity on how to handle data during change over to cloud. The main contribution of the study are listed below:

This study provides the reader with close understanding of data migration with emphasis on the nature and types of data migration.

- It assesses leading cloud service providers and their migration tools, clarifying available options for seamless cloud transitions.
- The study highlights key challenges and risks in cloud data migration, enabling organizations to proactively address potential issues.
- By outlining migration stages and essential drivers, the study serves as a practical guide for organizations planning cloud data migration projects.

B. Structure of the paper

The paper is structured as follows: Section II provides an overview of data migration in cloud databases. Section III discusses vendor solutions for cloud database migration, while Section IV explores challenges risks of data migration in cloud, machine learning. Section V examines literature review of the paper, and Section VI offers conclusions and future works recommendations.

Overview of data migration in cloud databases

The analysis of legacy data is the first stage in the data migration process, which also includes loading and standardising data into new systems. After the legacy data has been scrubbed, the data may be mapped from the old system to the new system. Next, conversion programs are designed, built, and tested. Finally, the converter is matched. Another definition of data migration is creating a duplicate of an organization's present data on one device and transferring it to another, ideally without stopping any running applications and then rerouting all input/output (I/O) operations to the new device [5, 6].

A. Types of Data Migration

Data migration strategies should be deliberated after careful consideration of the various forms of migration. Migrations that impact the whole system are the most complicated. Furthermore, the following are other considerations that are taken into account:

1. Data Base Migration

The current database is upgraded to the most current version whenever data is migrated from one database resource to another. Take the conversion of the IBM DB2 database to Oracle as an example.

2. Data Center Transfer

It is necessary to transfer all data from the database of the previous data centre to the database of the new data centre when the data centre is relocated.

3. Application Migration

The underlying data must also be transferred to a new application when transferring it, for instance when moving it from a local activity server to the cloud or across cloud domains.

4. Business Process Migration

Depending on the specifics of the process changes brought about by a merger, acquisition, or just general company improvements, data transfers may be necessary to move files from one storage system or app to another [5]. Figure 1 shows the several levels of cloud data transfer, which are as follows

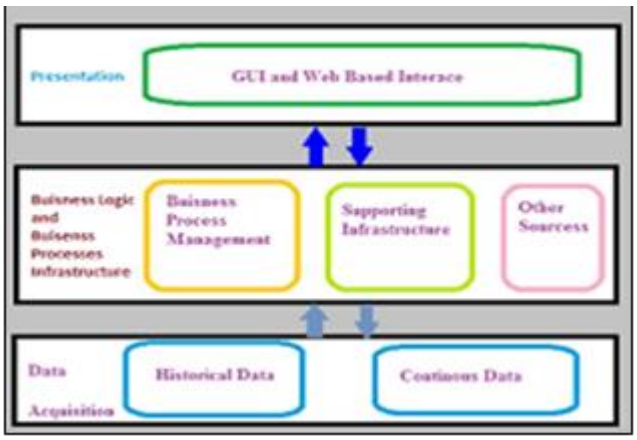

Fig 1: Data Migration in Cloud Layers

B. Need for Data Migration

Data migrations for commercial purposes are rather commonplace nowadays. There are a number of important considerations when opting to transfer data to a new environment, the most prevalent of which being the need to replace an older legacy system. Among these considerations are:

- The need for more storage space is driven by the exponential growth of databases.
- Businesses are elevating their server game to a higher quality.

- To simplify and save costs by switching to a consumable and sTable system
- Concepts like virtualisation rely on data's portability between physical and virtual settings.
- The provision of reliable and clean data for use [7].

C. Key Drivers for Cloud Data Migration

Here are the main factors that contribute to a successful data transfer, as well as how cloud data migration affects it:

Scalability And Capacity : An advantage of using a service provider rather than installing more hardware on your own premises is the ease with which capacity may be increased in response to customer demand [8].

Data Validation and Testing: The most effective strategy for preventing data corruption is to validate all data between the old system and the target system.

- Data Validation.
- Subsets of data Validation.
- Complete data set validation.

Sample Data Validation: involves comparing a randomly selected record from the legacy system to the target system. Sampling is not flawed since it picks little records at random.

Validation Of the Dataset: We should make it a priority to test the migration using this flawless validation approach. Both the old and new systems compare all records in a two-way fashion. This means that the target system compares all records with the old system and the old system compares all records with the target system. In order to overcome the instances required to establish a single database containing both legacy and target system data, it is difficult to conduct such a comparison when two distinct database suppliers are involved [5].

- Project stability
- Data coverage
- Implementation time.
- Efficiency of the Query/Script

D. Migration Process

There are a lot of possible worries and problems that could arise for organisations during data transfer. A process model that the authors have used in practice several times forms the basis of this research. There are a total of fourteen separate steps that make up the four primary stages. The primary steps are:

- Initialization, i.e., setting up the necessary organization
- and infrastructure.
- Development, which refers to creating the actual programs for data movement
- Testing, or confirming that the data and the data transfer routines are accurate, sTable, and have a reasonable execution time
- Cut-Over, which is the process of running the migration programs in order to eventually switch to the destination application.

Vendor solutions for cloud database migration

Cloud database migration is significant for organizations who wish to migrate their databases seamlessly and efficiently into cloud environments, thus requires the solutions from vendors. These solutions are built with the aim of solving challenges that arise with databases migration such as data integrity when moving from one cloud platform to another, avoiding downtime to the highest extent possible, and compatibility of the databases in the new cloud platform.

AWS, Microsoft Azure and Google cloud and other leading cloud service providers have a number of migration tools and services to support this process. For example, AWS includes DMS or Database Migration Service that offers heterogeneous options and offers a high level of automaticity regarding data migration. Likewise, when the users employed the Azure Database Migration service, they are provided with a procedure on how to migrate on-premises databases to Azure, the data integrity and security.\

Key Vendor Solutions

1. AWS has the Database Migration Service (DMS), that is hardware based and is used for databases and covering different platforms and switches to AWS and replicates data constantly with low disruption. Further, the AWS Schema Conversion Tool helps the user in transforming the schemas of databases to the required format, which is many times required for the target database for easier migration.
2. Microsoft Azure offers Azure Database Migration Service which helps users to perform a smooth migration from On-premises database to Azure with little or no downtime factors which makes it convenient for users to improve the consistency of data. In addition, Azure's Data Box is a shipped disk that performs the transfer of large quantities of data securely and quickly, which can suit enterprises with great requirements for transferring big amounts of data.
3. Migrating to Google Cloud SQL is another option from the GCP that is called the Migration Service that is ideal for database migration and the migration of both supported and unsupported varieties is possible. Also, with the Big Query Data Transfer Service, users can easily transfer data from several data sources into Big Query and therefore easily analyze and use the data.
4. Tools like Talend contain detailed Third-Party data integration and migration features which encompass data mapping and transformation features, which goes hand-in-hand with organizing organizational data during migration. Other critical solutions are data management within data migration, integration and control that Informatica offers to ensure transitions for data while maintaining its quality. Finally, Striim enables real-time data integration and replication, helping organizations to transfer data perpetually from one platform to another with low latency and data drift.

Challenges risks of data migration in cloud

Cloud data transfer is not without its difficulties. Security issues pertaining to illegal access and ensuring compliance with data protection laws further exacerbate the situation, necessitating meticulous preparation to minimize risks. Below are some of the challenges that are observed.

A Choosing the right cloud vendor

Fundamental research issues include data management and migration, which are never quite as simple as moving data

from traditional systems to the cloud. Selecting the right cloud provider is a challenging process for an organisation, even after doing a SWOT (Strength, Weakness, Opportunities, and Threats) study. Cloud industry heavyweights like Microsoft Azure, Amazon Web Services (AWS), and Google Cloud Platform (GCP) are always trying to find new methods to set themselves apart from the competition. With that in mind, businesses should enquire with cloud providers about data transfer solutions, taking into account factors like vendor lock-in and software portability.

B. Trust deficit about cloud security

Cloud computing has recently improved its information security model, but there are still concerns that people should not put their most sensitive data there due to the potential dangers. All-important parties, including people, organisations, and governments, are affected by this lack of confidence.

C. Departmental downsizing

There is a chance that IT support teams would have to scale down when using a cloud architecture. Downsizing may be in store for IT support departments whose primary responsibilities lie in providing assistance with hardware and organisational matters. The capacity within the IT support department will become superfluous since cloud providers will be responsible for maintaining these portions of the service.

D. Lack of supplementary resources

The deployment of cloud frameworks raises concerns about potential shortages in IT support and sales/marketing department resources. Due to migration, there is a chance that IT support departments may temporarily expand, and support engineers often lack knowledge and expertise with cloud frameworks [9].

Literature Review

This section provides a literature review on data migration in cloud databases, exploring methodologies and frameworks for efficient data transfer across heterogeneous environments. It highlights key strategies, challenges, and advancements, offering insights into the effectiveness of various migration approaches.

In this paper Bansel, González-Vélez and Chis, (2016) to encourage data portability between diverse NoSQL data repositories hosted in the cloud, we suggest a NoSQL data transfer architecture. In order to efficiently map and translate across various NoSQL data stores hosted in the cloud, the suggested method incorporates data standardisation and categorisation phases. Presently, the framework can handle document, columnar, and graph data models. The fact that it is meta-model driven further opens the door for developers to include additional database models into the framework. To facilitate data conversion (from documents to graphs), our method incorporates an online compression mechanism; this allows for a 46% reduction in storage space requirements for graph databases. When compared to the original graph database, the number of nodes in the compressed version is 37% to 55% less [10].

This paper Wang *et al*., (2019) introduces e-MARS, a system that is based on an environmentally conscious market strategy, for the purpose of appropriate data transfer in order to achieve query load balance in a cloud database. The e-MARS concept views cloud databases as a cloudDB market, with data nodes acting as smart traders and query loads as commodities. Traders make their own decisions about query load trading and data migration based on their own knowledge of their local environmental resources, such as processing capability and disc volume. This will bring to a state of balance in the cloud DB industry. With e-MARS, efficiency is greatly improved by experiments run on actual communication data. When it comes to query response time, it's almost 65% faster than HBase Balancer [11].

This paper (Li *et al*., (2016) came up with a standard technique for migrating to private clouds, and summed up the technical concepts and basic procedures for doing so. The subsequent successful completion of a case study in which the suggested approach was used to migrate an enterprise system to the Eucalyptus cloud environment demonstrated the method's viability and validity. Finally, this article offers some technical recommendations for improving system performances based on an investigation of system performances using various migration strategies [12].

This paper, Hsieh *et al*., (2021) creates a system that can handle several host databases in a unified way, which may simplify the stages of the operation and increase the data's security and operational efficiency. Manual, automated, and complex database migration methods are all available. The results of the experiments demonstrate that the advanced database migration approach outperforms the other two methods in terms of efficiency and security when used in a VM environment [13].

This paper Abo Dabowsa *et al*., (2021) recommends an automated method for migrating databases from the MySQL database management system to the popular NoSQL database system MongoDB. This technique can process massive amounts of data stored in RDBs without affecting the data's semantics or instances in any way. The solution processes an existing RDB by extracting its schema and analysing it in an array. Then, it converts the schema with data instances according to the structure of the goal NoSQL database. An approach-based system has been developed. The suggested technique was tested in an experimental research. The results of the experiments demonstrate that the prototype's target database and the target databases created using other approaches were similar and identical [14].

This paper Tian *et al*., (2017) suggests a cloud-based data replica transfer technique that is both dynamic and flexible. An improved capacity to adapt to changes in workload, more fault tolerance, and better scalability are all goals of a proposed dynamic scheduling system that is based on workload-based cloud computing. This approach modifies the amount of replicas by modifying the amount of transaction requests that the workload processor watches. By keeping an eye on the workload, we may spot major shifts, which is a baby step towards repartition, and we reach our goal of keeping the partition in excellent shape in the end. The migration approach for dynamic data copies allows for the completion of dynamic data exchange among data nodes. When the workload varies, the experimental findings reveal that the suggested strategy can drastically lower the frequency of distributed transactions [15].

In this paper Rafique *et al*., (2018) provide two further studies that address the aforementioned issues with the three most developed data access middleware systems: Spring

Data, Impetus Kundera, and Playorm. The performance overhead that different platforms bring to the CRUD operations is first evaluated. A second part of the analysis is a comparison of the migration costs with and without these platforms. In spite of their shared architecture, our research reveals that these systems provide quite distinct functionality. These two studies complement each other by illuminating the trade-offs involved in using a data access middleware platform for NoSQL, which is that developers receive portability and ease of migration across heterogeneous data storage in exchange for a performance expense [16].

Table 1: Summarizing the related works on data migration in the cloud database

Study	Focus	Methodology	Key Findings	Limitations	Future Work
Bansel, González-Vélez and Chis, 2016	NoSQL data migration framework for heterogeneous NoSQL repositories	Data standardization, classification, online compression algorithm for data migration	Supports three data models (document, columnar, graph), reduces graph database space by 46%, and reduces nodes by 37%-55%	Limited support for more than three NoSQL data models	Extend framework support to more database models
Wang *et al.*, 2019	e-MARS, a market strategy-based system for data migration	Environment-aware system, query load treated as commodity, data nodes act as intelligent traders	Achieved query load balance, 65% improvement in query response time over HBase Balancer	Complexity in modeling and managing the cloud DB market	Further optimization of trading algorithms for different cloud environments
Li *et al.*, 2016	Private cloud migration methodology	General methods for private cloud migration, specific case of Eucalyptus cloud migration	Feasibility and validity of private cloud migration methods demonstrated	Case study limited to Eucalyptus cloud, lacks diversity of cloud environments	Explore migration strategies across various private and public cloud environments
Hsieh *et al.*, 2021	Managing multiple host databases in a unified manner	Manual, automatic, and advanced migration approaches	Advanced method enhances security and operational efficiency compared to manual and automatic methods	Scalability challenges in larger virtual machine environments	Integrate AI-based optimization for database security and operational efficiency
Abo Dabowsa *et al.*, 2021	Automatic conversion from MySQL to MongoDB	Schema extraction and conversion of RDB to NoSQL (MongoDB)	Successful conversion without data loss, equivalent performance between converted and original databases	Limited to MySQL and MongoDB systems	Extend conversion approach to more relational and NoSQL systems
Tian *et al.*, 2017	Dynamic adaptive migration strategy for data replicas	Dynamic scheduling mechanism based on workload changes	Significant reduction in distributed transaction frequency during workload changes	Limited scalability in cloud environments with high dynamic workloads	Optimize the dynamic scheduling mechanism for larger-scale cloud systems
Rafique *et al.*, 2018	Evaluation of NoSQL data access middleware platforms	Performance evaluation of CRUD operations and migration cost comparison	Middleware platforms (Impetus Kundera, Playorm, Spring Data) offer portability at the cost of performance overhead	Performance trade-offs when adopting middleware platforms	Investigate middleware performance improvements without sacrificing portability

Conclusion & Future Work

Applications that rely heavily on data must undergo data migration. The ever-evolving data landscape is too much for legacy data storage technologies to handle. The failure to adequately address the significance and complexity of data migration initiatives is a common reason of their failure in many organizations. Storage, databases, applications, and business processes are some of the data migration techniques available. In conclusion, data migration to cloud databases is a critical process that requires careful planning and execution to ensure data integrity, security, and minimal downtime. As organizations increasingly rely on cloud services for data management, understanding the complexities of migration and the available vendor solutions is essential. Major cloud providers like AWS, Microsoft Azure, and Google Cloud offer tailored migration tools to facilitate this transition, but challenges remain, including security concerns, vendor selection, and resource allocation. Addressing these challenges will be crucial for organizations aiming to leverage the benefits of cloud computing effectively.

Future work should focus on developing more advanced migration tools that enhance data security and simplify the migration process across diverse cloud platforms. Research can also explore automated solutions for real-time data validation and integrity checks during migration. Additionally, investigating the long-term impacts of cloud migration on organizational structure and IT resource management will provide insights for effective cloud strategy implementation. Finally, enhancing awareness and education around best practices in cloud migration can help organizations navigate the complexities of this evolving landscape more effectively.

References

1. Ravindranadh K, Kiran MS, Sai Pavan Kumar BD, Priyanka D. Data migration in cloud computing using honey encryption. Int J Eng Technol; c2018. DOI: 10.14419/ijet.v7i2.8.10415.
2. Ellison M, Calinescu R, Paige RF. Evaluating cloud database migration options using workload models. J Cloud Comput; c2018. DOI: 10.1186/s13677-018-0108-5.
3. Kushwaha SG, Pathak P. Review of optimize load balancing algorithms in cloud. Int J Distrib Cloud Comput. 2016;4(2):1-9.
4. Potey M, Digrase M, Deshmukh G, Nerkar M. Database Migration from Structured Database to non-Structured Database. Int J Comput Appl; c2015.
5. Hussein AA. Data migration need, strategy, challenges, methodology, categories, risks, uses with cloud computing, and improvements in its using with cloud

using suggested proposed Model (DMig 1). J Inf Secur; c2021. DOI: 10.4236/jis.2021.121004.
6. Thomas J. enhancing supply chain resilience through cloud-based SCM and advanced machine learning: A Case Study of Logistics. J Emerg Technol Innov Res; c2021.
7. Youn C, Ku CS. Data migration. In: Conf Proc - IEEE Int Conf Syst Man Cybern. 1992;1:1255-1258. DOI: 10.1109/ICSMC.1992.271615.
8. Sighom JRN, Zhang P, You L. Security enhancement for data migration in the cloud. Futur Internet; c2017. DOI: 10.3390/fi9030023.
9. Amin R, Vadlamudi S. Opportunities and Challenges of Data Migration in Cloud. Eng Int. 2021;9(1):41-50. DOI: 10.18034/ei.v9i1.529.
10. Bansel A, Vélez GH, Chis AE. Cloud-Based NoSQL Data Migration. In: Proceedings - 24th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, PDP 2016; c2016. DOI: 10.1109/PDP.2016.111.
11. Wang T, *et al*. An environment-aware market strategy for data allocation and dynamic migration in cloud database. In: Proceedings - International Conference on Data Engineering; c2019. DOI: 10.1109/ICDE.2019.00232.
12. Li Y, Zhang J, Hu Q, Pei J. Research and practice on the theory of private clouds migration. In: International Conference on Signal Processing Proceedings, ICSP; c2016. DOI: 10.1109/ICSP.2016.7878141.
13. Hsieh CH, *et al*. A Migration System of Database for Virtual Machine in Cloud Computing. In: 2021 4th International Conference on Information Communication and Signal Processing, ICICSP 2021; c2021. DOI: 10.1109/ICICSP54369.2021.9611882.
14. Abo Dabowsa NI, Maatuk AM, Elakeili SM, Akhtar Ali M. Converting Relational Database to Document-Oriented NoSQL Cloud Database. In: 2021 IEEE 1st International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering, MI-STA 2021 – Proceedings; c2021. DOI: 10.1109/MI-STA52233.2021.9464488.
15. Tian B, *et al*. A flexible dynamic migration strategy for cloud data replica. In: Proceedings - 2017 IEEE International Conference on Internet of Things, IEEE Green Computing and Communications, IEEE Cyber, Physical and Social Computing, IEEE Smart Data, iThings-GreenCom-CPSCom-SmartData 2017; c2017. DOI: 10.1109/iThings-GreenCom-CPSCom-SmartData.2017.89.
16. Rafique A, Van Landuyt D, Lagaisse B, Joosen W. On the Performance Impact of Data Access Middleware for NoSQL Data Stores: A study of the trade-off between performance and migration cost. IEEE Trans Cloud Comput; c2018. DOI: 10.1109/TCC.2015.2511756.