ISSN No:-2456-2165

Cause and Effect of Data Migration in Cloud Computing

Sourav Kumar Upadhyay[1]
[1]Assistant Professor, Department of CA & CS,
JRSU, Ranchi, India

Dr. Prakash Kumar[2]
[2]Assistant Professor (HOD), Department of CA & CS,
JRSU, Ranchi, India

Abstract:- The cloud data migration idea includes data warehouses that deal with flat files from many sources and are used for data analysis and validation. In this paper, we describe about Data warehouse and its types, Migration types, Cause and consideration of data migration in cloud and Effect and possible risks Data Migration Cloud.

Keywords:- *Cloud Migration, Data Warehouse, Data Migration in Cloud Computing, Cause and consideration of data migration in cloud, Effect and possible risks in Data Migration Cloud.*

I. INTRODUCTION

An organization can migrate to the cloud in a number of ways. Data and apps can be moved from a nearby on-premises data center to the public cloud using a similar model. Moving data and applications from one cloud platform or provider to another can also be a part of a cloud migration, often known as a cloud-to-cloud transfer. Reverse cloud migration, also known as cloud repatriation or cloud departure, is a third type of migration in which data or applications are transferred from the cloud and back to a local data center. The information is unstructured and kept in many places. Before beginning the validation procedure, the staging area needs the other sources' datasets to be obtained and loaded. The filtered information will all be kept in tabular form in a tabular format in the cloud after the validation procedure. Cloud migration is the process of migrating software, data, or other corporate assets to cloud storage settings. It updates on-premises data to cloud storage and works with validated flat files from various sources.

A. Data Warehouse

Repository is made up of a single, enormous collection of uniform data that has been saved from several sources. All users or clients find it convenient for their market data reports and progress outlines. Data services and repository have a role in data charges. Its environment includes a data repository, a data mart, and metadata. A portion of a data warehouse is known as a "data mart." The data repository's purpose is shown in Figure 1. There are numerous data-related operations in multidimensional space, including extracting, compressing, and manipulating. The turnaround time can then be shortened by processing it right away. Having Depending on the kind of end users and data redundancy, a data warehouse is crucial. The concept of data warehouses has been drastically altered by the fusion of many technologies. We must research and understand the needs of the business, and we must plan an affordable and useful data repository [1].

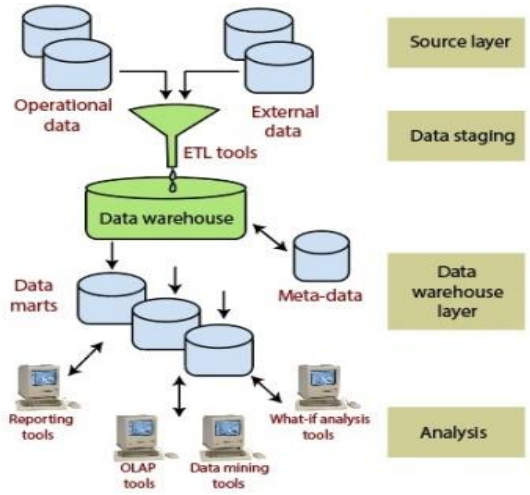

Fig.1. Data Warehouse [2]

B. Kinds of Data Warehouse

B.1. Enterprise Data Warehouse (EDW): An enterprise-wide warehouse that aids in providing decision support services. It presents a uniform method for representing and arranging data. This improves the process of classifying data according to subject and granting access based on data divisions.

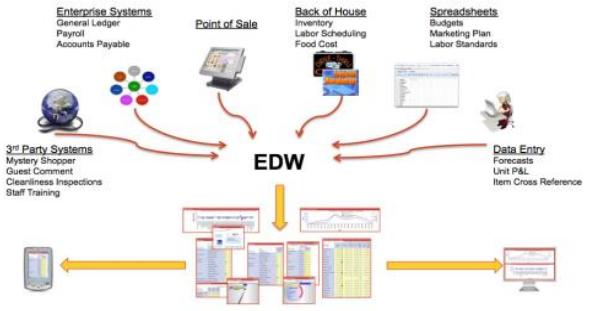

Fig.2. Enterprise Data Warehouse [3]

B.2. Operational Data Store [ODS]: It primarily provides organizational reporting, and neither an OLTP system nor a data warehouse is used to store the data. ODS performs tasks like storing customer or employee data since the data warehouse is always being updated.

ISSN No:-2456-2165

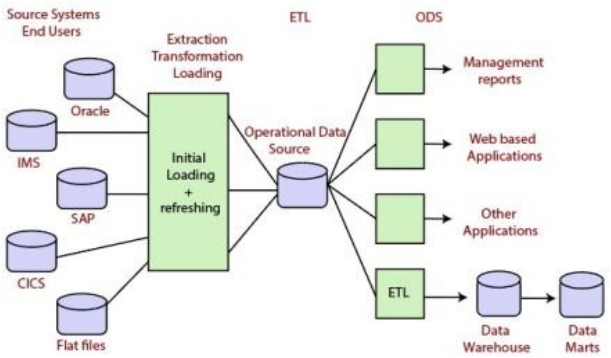

Fig.3. ODS [4]

B.3. Data Mart: Data Mart is a useful subset of the data warehouse. It is specifically made for a certain sales, business, or financial line. Data may be gathered directly from sources or indirectly through various data processing centers by a separate data mart.

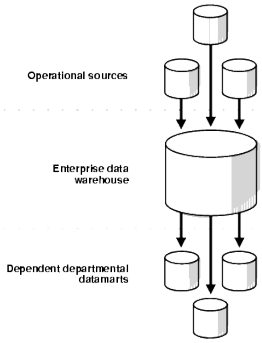

Fig.4. Data Mart [5]

II. MIGRATION OF DATA

Transferring information to a new, updated device or location is a process known as data migration. Data is gathered, scheduled, and converted in order to be permanently transferred from one device's storage to another. Database migration services are growing in popularity as businesses place a greater emphasis on technology advancements and optimization.

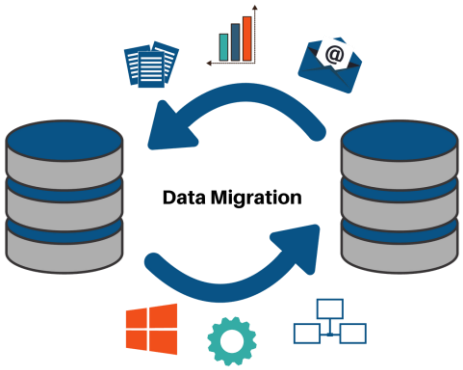

Fig.5. Sketch of Migration [6]

We can see from the diagram that data can be moved across different computer systems or formats. It is one of the most important things to keep in mind when putting a system into place, consolidating, or updating. The introduction of a new data structure or venue frequently causes this. Migration techniques are customized as needed, and hardware and software specs are validated. To make sure that specifications and customizable configurations function as expected, pre-validation testing can also be done. The transfer process, which comprises the crucial operations of data retrieval (reading data from the previous system), and data loading, begins if everything checks okay (writing data to the new system).

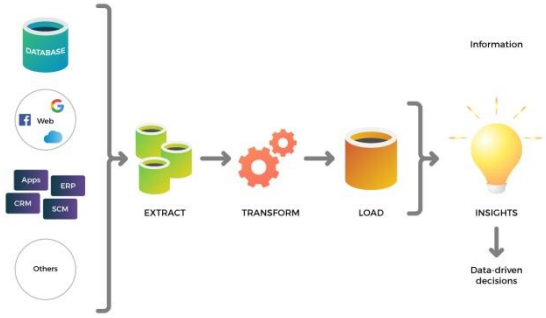

Fig.6. Extract, Transform, Load Sketch [7]

A. After Relocation

After data transfer, the results are examined to determine whether the data was correctly understood, was complete, and adhered to the procedures used by the current system. A concurrent execution of both systems may be necessary during verification to find areas of divergence and avoid losing inaccurate data. Once the migration is certified complete, additional documentation and analysis of the migration process will be finished, and existing systems will be retired. Close-out sessions will put a stop to the transfer procedure.

B. Kinds of Migration

There are four main types of data migrations, and each one requires sufficient planning and verification before implementation.

B.1. Database Migration: Either the functionality of the database is upgraded, or the entire database is passed from one supplier to another. The foundation of every technology we use on a daily basis is a database. It makes sense that SMBs would likewise change database providers, update their programs, or move their databases to the cloud. Data from two separate database engines must be transferred as part of the database migration process.

B.2. Storage Migration: Storage conversion describes the transfer of data from one storage media to another. This entails physically moving data blocks from one type of hardware (such as tapes or discs) to another. Moving data from one storage media, like a hard disc or the cloud, to another is known as data

conversion. Data is transferred from one storage medium to another, to put it another way.

B.3. Business Process Migration: Through this procedure, data, software, and other business-related items are moved from an on-premises data center to a cloud or from one cloud to another. It is focused on a company's operational procedures, particularly outdated or in need of replacement business management instruments. Typically, a merger or takeover causes this to happen.

B.4. Application Migration: If the application vendor needs change, transformation is a necessity. Given that every application relies on a certain data model, it is essential. It entails adapting application programs for a more contemporary environment. It can move a whole application structure from an on-premise IT foundation across clouds or to the cloud.

III. CAUSE AND COSIDERATION OF DATA MIGRATION IN CLOUD

Businesses are under increasing pressure to maximize the value they derive from the data they collect today due to the ever-increasing volume of data being produced. Success in this environment depends more and more on selecting the best environments for your workloads and ensuring sure your data is kept effectively and easily accessible. In an effort to host their applications in the most economical and effective IT environment possible, many businesses are deciding to migrate workloads to the cloud. The process of planning a cloud migration should start with early consideration of choosing the best data migration option. For the introduction of data-intensive technologies like databases, data centers, and data lakes as well as large-scale virtualization programs, data transformation is crucial for upgrading or consolidating server and storage infrastructure. Additionally, data transfer may take place between internal systems and cloud storage, as well as inside HDD or SDD-based systems.

A. Factors to take into account while developing a data migration plan

The better your company manages the less likely you are to incur unforeseen expenditures or unplanned downtime during its data movement, and the less probable it is that your end users will get irritated or inconvenienced both during and after the migration. One should define objectives, create a schedule, and be prepared for any difficulties that may arise. When choosing your strategy for the project, you should primarily take the following three things into account:

A.1. Nature of the work: Tools provided by software vendors that are unique to the type of data being migrated can typically be used to move specialized workloads like databases, backups, or virtual machines (VMs). If you lack access to these resources, you should carefully prepare for any potential outage. For mission-critical workloads, you can transfer data incrementally, testing along the way while retaining the parallel operations of the source and target systems. Alternately, you may arrange a substantial transfer outside of normal business hours (if you can accomplish it the available window).

A.2. Data Volume: On a client-provided storage device, shipping the data to its new storage location is typically the quickest and affordable option when migrating less than 10 terabytes (TB) of data. The most practical and cost-effective alternative for transfers involving bigger volumes of data, such as up to multiple petabytes (PB), may be a specialized data movement tool that your cloud provider offers. Although one could theoretically use online migration for any amount of data, its viability for huge data sets is questionable and constrained by time.

A.3. Quickness of completion: How soon online migrations are finished depends on how much data is being transmitted and how quickly your network connection is. For offline migrations, shipping time must be taken into consideration. If start-to-finish migration speed is your top goal and you have the bandwidth to commit to the migration, online transfer may be the ideal option However, if your migration date is variable and/or you have bandwidth or other networking limitations, offline migration may be your best alternative. [8].

B. Suggested tools

There are many solutions available now to make enterprise data migrations easier. These include both licensed and open source tools, as well as vendor-specific solutions that cloud providers offer to assist their clients' migration into their public or private cloud environment. The ideal tools for your project will depend on your data migration approach. The following are some common options (Table 1):

Veeam	To hasten and simplify the transfer of VM-based workloads across hosts and storage environments, *Veeam* provides a Quick Migration tool for VMware vSphere.
Zerto	Zerto provides a unified platform for workload mobility, disaster recovery, and backup that allows migrations of all sizes, from the relocation of a single application to an entire data centre.
Cyberduck	An open-source FTP and SFTP software called Cyberduck can be used to move single files or entire file volumes between systems or into the cloud.
Rclone	Data can be moved to and from cloud object storage using the free source command-line tool Rclone. Large items can be automatically segmented and their components uploaded simultaneously.

Table.1. Tools for Data Migration [8]

ISSN No:-2456-2165

IV. EFFECT AND POSSIBLE RISK OF DATA MIGRATION IN CLOUD

The shift to the cloud has several advantages. Among other benefits, moving your company's activities to the cloud can save money, boost productivity, and provide improved security. You must carefully create a cloud migration project plan if you ultimately decide to migrate to the cloud. Several cloud migration technologies will be helpful throughout this process. But if you're considering moving to the cloud, there are a few risk considerations you should keep in mind. These consist of:

A. Absence of a well-defined cloud migration plan

Without giving it any thought, many people get sucked into the hoopla and rush to switch to the cloud. Before diving headfirst into cloud computing, there are a lot of things you should think about. You should also develop a detailed strategy for transitioning to the cloud. Consider the benefits you want from moving to the cloud as well as your reasons for doing so. Think about the data you want to move to the cloud as well as how much of it. You might want to preserve certain particularly sensitive or important data on-premises. Decide on the amount of storage you need as well as the number of potential cloud providers [9].

B. Security dangers

These are most likely the greatest dangers that businesses moving to cloud computing must deal with. Insecure APIs, unintentional mistakes, malware, external assaults, and more are just a few of the security dangers that come with moving to the cloud, in addition to compliance violations and contractual breaches. You must be aware of these hazards and be prepared to handle them before switching to the cloud.

C. Overspending

Although cloud companies' pricing structures are adaptable, they are frequently challenging to comprehend. This can and occasionally does result in up to 70% of cloud computing costs being wasted. The cost of cloud computing varies, and each supplier will have a varied set of services and costs to provide. The perfect combination can be difficult to decide. You risk wasting a lot of money if you don't make the necessary calculations to determine exactly what you need (Figure 8).

Fig.8. Overspending in the cloud [10]

D. Undesirable delay

An underappreciated danger of cloud migration is increased latency. Your business could be significantly harmed by your app's brief delay. Customers can become frustrated with delay, and it can negatively affect the reputation of your brand. There are several potential remedies for latency problems, but keeping some of your data on-premises may make sense if they fail or are prohibitively expensive.

E. Loss of data

There is always a chance that data will be lost when it is moved to a new location to be stored. You might discover that some of your files are missing, imperfect, or corrupt, whether as a result of different technical problems or human error. Make sure your CSP has alternatives for data backup, restore, and fallback. Having your data backed up by multiple cloud providers is a good idea so you won't have to worry about an individual service going offline unexpectedly. Additionally, it is a smart idea to backup all of your crucial data to a drive.

F. Reduced control and visibility

Performance can be impacted by the very real risk of lack of visibility in the public cloud while moving to the cloud. You have complete control over all of your resources, rules, and infrastructure when your data is housed on-premises. But when utilizing external cloud services, some of these responsibilities are transferred to the cloud service provider (CSP), which might reduce your business's visibility [11-19].

V. CONCLUSION

In this paper, we have studied about how data warehouse works the ETL workflow. Data migration concept and its types in detail so that we can conjure the idea of what should be done to minimize the data loss. Later, we have seen that what the various risks are in migrating data in cloud platform. A move to cloud is desirable but one should formulate a good strategy beforehand in order to eliminate any risk that can jeopardize the migration [20-22].

REFERENCES

[1]. N Prasath, J Sreemathy; A New Approach for Cloud Data Migration Technique Using Talend ETL Tool, 7th International Conference on Advanced Computing & Communication Systems (ICACCS), 2021
[2]. Image available at: https://www.javatpoint.com/data-warehouse-architecture
[3]. Image available at: https://blog.mirus.com/what-is-an-enterprise-data-warehouse
[4]. Image available at: https://www.javatpoint.com/data-warehouse-operational-data-stores
[5]. Image available at: https://docs.oracle.com/cd/A81042_01/DOC/server.816/a76994/marts.htm

[6]. Image available at: https://www.auratechnology.com/auranews/blog/data-migrations/
[7]. Image available at: https://blog.bismart.com/en/how-to-choose-the-right-etl-tool
[8]. https://www.ibm.com/cloud/learn/data-migration
[9]. Netanya Karni; Risk in cloud migration, available at: https://www.peerspot.com/articles/risks-in-cloud-migration#:~:text=Migrating%20to%20the%20cloud%20involves,equip%20yourself%20to%20handle%20them.
[10]. Image available at: https://cdn.ttgtmedia.com/rms/onlineImages/cloud_computing-overspending_in_the_cloud-f_desktop.png
[11]. Saranya N, Brindha R; Data migration using ETL Workflow, 7th International Conference on Advanced Computing & Communication Systems (ICACCS), 2021.
[12]. Lei yang; QoS Guaranteed Resource Allocation for Live Virtual Machine Migration in Edge Clouds, School of Software Engineering, South China University of Technology, Guangzhou 510641, China
[13]. Dr. Prakash Kumar, International Journal of Computer Science and Mobile Computing, Vol.6 Issue.12, December- 2017 | ISSN 2320–088X, pp. 157-163.
[14]. Dr. Prakash Kumar, International Journal of Creative Research Thoughts, Volume 6, Issue 2 April 2018 | ISSN: 2320-2882, pp. 428-434.
[15]. Dr. Prakash Kumar, Security Issues in Vehicular network" in **JESMT** Vol. 2 | Issue 1, 2012, ISSN 2231- 1521.
[16]. "Survey on Tools & Technologies used in Semantic web and IOT" **IJCRT,** Vol. 6 | Issue 2, June 2018, ISSN2320-2882,pp. 302-306
[17]. "A Novel Software development life Cycle Model for Developing Software Project", **IJCRT,** Vol. 6 | Issue 2, April 2018, ISSN 2320-2882, pp. 428-434.
[18]. "Security Aspects in Social Networking Model "**IJCRT,** UGC Approved, Vol.6| Issue 1, Jan 2018, ISSN 2320-2882.
[19]. "Cloud Database Services improve using load balancing Technique" in **JETIR** Vol.8| Issue 1, 2021, ISSN 2349-5162, pp.822-826
[20]. Sourav Kumar Upadhyay, Dr S.C. Dutta, Dr. Prakash Kumar; Impact of the Internet Shutdown in Ranchi, Jharkhand: A Survey, 2022. Available at: http://www.jetir.org/view?paper=JETIR2207676
[21]. Purushottam Kumar, Dr. Prakash Kumar; A survey on Load balancing in Cloud Computing. Available at: http://www.jetir.org/view?paper=JETIR2207664
[22]. Sourav Kumar Upadhyay, Dr. S.C. Dutta, Dr. Prakash Kumar; A review on the risk and its countermeasures in cloud environment; JETIR, Volume 9 Issue 8, 2022, ISSN-2349-5162