

Review on Data Efficient Migration for Multi-Server System based on Advantages and disadvantages

Daljeet Kaur¹, Rupinder Kaur²

¹Student (M.Tech, Scholar), Doaba Institute of Engineering and Technology. Kharar

²Assistant Professor, Doaba Institute of Engineering and Technology. Kharar

Abstract - Data migration is a critical process that directly influenced the quality of data management. Data migration had natural on the worth of the data such as, data elements accuracy, and data accessibility, and all data performances. However, many administrations faced all dissimilar challenges with data below the same process. Data is a precious asset for any corporation. So, any unintended transmission of data can be very dangerous for company. In reality, planning is the top most achievement issue for any data migration scheme, independent of underline complexity. Appropriate thorough planning reduces the business influence such as application stoppage, overall performance degradation, and technical incompatibilities, threat for example, semantic risk, completeness risk, data corruption/loss. The first step of the process starts after migrated data from the old/legacy system into the new one; results are subjected to data confirmation, to regulate whether data was exactly translated, completed, and supported processes into the new system. During the verification process, the parallel running strategy of both systems is needed to identify areas of disparity and prevent any wrong ways could lead to data loss. Besides, there is a commonly automated and manual data cleaning is performed in the migration process in order to improve data quality, to eliminate redundant or old data, and match the requirements of the new system.

Keywords - Data Migrations, Data performances, Data management and data elements accuracy.

I. INTRODUCTION

Cloud computing has enabled Database-as-a-Service to provide end users with access to different data and storage models. Unfavorably, the use of different data models introduces the heterogeneity in cloud services which makes difficult for a user to migrate the software from one specific cloud-based storage to another. Furthermore, cloud-based storage infrastructures are responsible for storing and processing large amounts of data in a scalable manner. However, usually, during the initial software development stages, it is difficult for developers to anticipate the growth and storage requirements of an application which in turn compels the developers to design the application using a specific data service[1]. Data migration for big data is a

challenging and resource demanding process. Data Migration is the process of moving data from a system or systems to the new environment. Often, it is a sub-activity of a commercial application placement. Data migration plays a vital role in ensuring business continuity. It needs substantial focus and attention, and discipline, where [2] Data Migration is the process of moving data from a system or systems to the new environment. Often, it is a sub-activity of a commercial application placement. Data migration plays a vital role in ensuring business continuity. It needs substantial focus and attention, and discipline.

- (a) No data is lost due to the data migration process,
- (b) No data is polluted and
- (c) Change to data is governed by endorsed rules.

Data Migration is an important process and much literature has being published. Most of the literature is focused on the area of Data Cleansing[3].

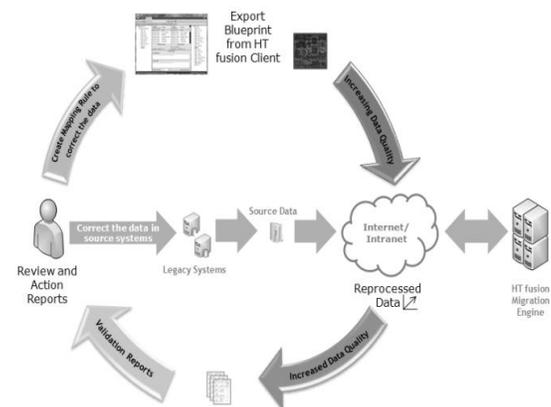


Fig.1: Data Migration Process

II. DATA MIGRATION OVERVIEW

Data transmission can be of dualistic types: first, a simple data amount that is affecting data from source database to destination database without rearrangement and second, data movement. Data movement is the process of transmitting data

between computer storages, types, arrangements, or workstation system. It is the procedure of transmitting data from the base database(s) to a destination database. We called old record as a bequest or source database and this database is transmitted to the original record, named as target or endpoint database. The data migration process develops a difficult test when source and target databases are dissimilar in their internal structures. So, simple import and export processes will not work. Thus data movement process is better to accomplish using programmed ETL (Extract – Transform – Load) tools than doing manually [4].

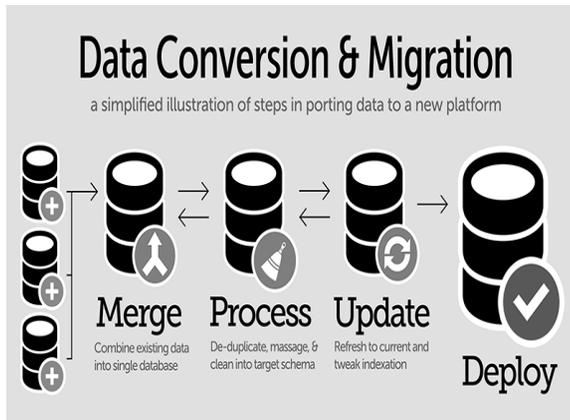


Fig.2: Data Migration Overview

III. USES OF DATA MIGRATION

Multiple technologies and best practices are available for use with data movement, and more than one of these might be useful to a single project. Every data movement will have a distinctive combination of previous systems, plus requirements for the new system and its end users. Then there are the equipment requirements shared to all data migration projects (described next). IT should select machineries based on these features, tempered by a consideration of how often and deeply they will perform data migration and similar projects in the future [5].

(a) Extract, transform, and load (ETL) is the preferred technology for data movement. In a recent Equipment Survey, TDWI requested: “When your organization consolidates or migrates databases, what is the chosen technology for most schemes?” For 41% of respondents, ETL was the preferred technology, ahead of hand-coded keys (27%), and EAI (3.5%) and repetition (11%). Users prefer ETL for its unique capability to handle the exciting requirements of data movement, including terabyte-scale datasets, multi-pass data transformations, deep data interoperability, sketching with data excellencetools and many-tom any data integration capabilities.

(b) Hand coding applications to some, despite absence of output. Studies have shown that toolbased data integration development and maintenance is far more creative (and, more economical, therefore) than hand-coded solutions.³ Yet, hand coding perseveres because designers can’t discourage themselves of it, consultants use it as an excuse to rack up billable hours, and short-sighted directors won’t devote in the close term to get the long-term cost reductions of tool productivity. It’s time for everyone to do the math and identify the economic advantage of tool use at least for data integration projects.

(c) Database repetition is stress-free and available, but misses some requests. All data management professionals are familiar in repetition, and a fair amount of repetition functionality comes at no additional charge with a relational database authorization. This kind of low-end repetition is usually limited to moving data one way without transformation between occurrences of the same database product. Some high-end repetition tools (bought separately) provide bidirectional, transformational, mixed data management, which is required when source and destination databases of different types operate concurrently.

(d) Enterprise application integration (EAI) is not appropriate to data migration. EAI excels at very quickly moving small quantities of information amongst the logic layers of presentations. But EAI tools cannot handle the extreme volume, transformation, profiling, and many-to-one integration, data qualitysupplies of data migrations [5]

IV. APPLICATION OF DATA MIGRATION

Amazon suggests a phase driven method for Cloud application movement, which includes one phase focusing on data migration. The data movement is done in dualistic steps: selection of the Amazon AWS package, and migration of the data. Furthermore, Amazon deliversreferences regarding which of their data and storage facilities best fit for storage a specific kind of data, e.g., Amazon Relational Database(Amazon RDS) Service. Separately from that, Amazon defines three application movement use cases .Advertising and association the Web sites Cardinal asset management answer using batch dispensation pipelinesClaims processing systems using back-end processing workflows. Microsoft identifies the following eight types of applications to be reflected for movement to the Cloud (Azure, 2012)

1. Seas applications
2. Highly-scalable Web sites
3. Corporate applications
4. Business brainpower and data depository applications
5. Social or customer-oriented applications
6. Social (online) games[synopsis]

V. RELATED WORK

Yunpeng Chai et.al.,2012 [6] described new energy-efficient technique called Explicit Energy Saving Disk Cooling or EESDC. EESDC suggestively reduces data migration above because of two reasons. First, a set of disks discussed to Explicit Energy Saving Disks was obviously fixed according to temporal system load. Subsequent, all the migrated data in EESDC directly back on extending the idle time of EESD to preserve more energy efficiently. Therefore, the EESDC technique is conducive to saving more energy by quickly accomplishing energy-efficient data layouts without redundant data migrations. They instrument EESDC in a simulated disk system, which is authenticated against a prototype system mechanical by our EESDC. **Yanling Du, Zhenhua Wang et.al.,2012 [7]** planned a hybrid cloud storage explanation in view of high performance, high retreat of private cloud and the large capacity features of public cloud. With the measurable expressing of the real-time property, compassion, decentralization and data access heat of aquatic data, they assumed the model of marine data migration between the hybrid clouds. Temporarily, the data migration method was improved to avoid the restraint of the traditional data migration process which is built just according to the data access heat. **Chadi Kari et.al.,2011 [8]** assumed that each storage node can achieve only one data transfer at a time. A storage node, conversely, can typically handle multiple assignments concurrently and this can reduce the total migration time knowingly. Moreover, storage devices tend to have varied abilities as devices may be additional over time due to storage application growth. In this paper, they consider the assorted data migration problematic, where they assume that each storage node v has different transfer constraint CV , which characterizes how many instantaneous transmissions, can grip. We grow algorithms to minimize the data movement time. **Girts Karnitset.al., 2015 [9]** offered explanation for quick data migration from a personal database into a document worried with database. They have twisted semiautomatically two consistent levels over corporeal data. Users can refine produced logical data model and organize data migration template for each needed article. Data migration structures are realized into relational database browser Digit Browser. Real patients' database was travelled to Clusterpoint database. The existing approach provides means to obtain at smallest proof-of-concept for new manuscript oriented database resolve in a couple of days. **Hui Liu et.al., 2012 [10]** characteristic cloud computing delivery model. In scarce table, one of the most general multi-tenant data storage plans for SaaS, all tenants' data are deposited into spare table and plotted to tenant's logical view by metadata. Throughout the data storage scheme advancement in SaaS, all tenants' data need to be travelled into the new data schema

before it becomes operative to ensure the integrity of the tenants' data. However, the migration is composite and brings overhead workload. Inferior still, it may cause the system unusable. In this paper, they recommend metadata evolution technology. They can understand the mapping from the old data schema to new data scheme smoothly via metadata progress during schema upgrade. So the schema advancement can be completed without data migration.

VI. ADVANTAGE AND DISADVANTAGES OF DATA MIGRATION

Data Migration described various advantages:

1. **Structuring exploratory business workshops:** Every data migration has a learning curve where the project team begins to study extra about how the legacy (and goal) Organizations operate. By starting your workshop with a discovery activity to generate intangible data model you directly focus the attention on data plus you get all parties actively conversing the legacy and goals situations at a level that does not descend into minute technical detail.
2. **Resolving "turf-wars" and politically aware matters early in the scheme:** By generating a more business focused model we can identify where ownership struggles may be taking place. In any movement there are political sensitivities as major business change is often a driver for the movement process. By generating higher-level maps we can quickly ascertain where there are overlaps of data possession that may lead to matters when we wish to retire the legacy environment.
3. **A useful tool for scoping the legacy background and allotting appropriate assets:** By creating high level business objects we get a far more manageable opinion of what data is difficult in our migration. If we are only presented with a physical model, the sheer complexity and absence of information can break the scoping process significantly.
4. **Great for partitioning workload and "chunking" the migration:** Another consequence of the scoping workout is that we are more able to allocate the appropriate resources. With our project cracked down into advanced level subject areas we can identify the necessary capability for each domain and generate more correct plans for each area.
5. **Helps find "data gaps" and "hidden" data stores early in the project:** By creating first conceptual model then a lower-level reasonable models we can classify very quickly where there are gaps in the legacy data stores that might need further examination. For example, our business predictors may indicate that the commercial manage association among a particular facility and a customer segment that does not appear to be found in the physical schemes in scope. Actual often this data is held in a private store in the form of local paper records or spreadsheets etc.

6. **Provides more focused business process analysis:** Opinion is mixed on whether business process analysis should be a part of a data movement project but there is no escaping the fact that a deep understanding of both the legacy and goalsituations is highly beneficial to transporting a "fit-for-purpose" migration. By identifying the entity associations at a more business absorbed level it makes deeper analysis of the business processes far easier and relevant.
7. **Helps to prioritise migration design and build:** Progressive and incremental migrations are now progressively favored over the outdated "big-bang" approach. Advanced migrations need a great deal of focus on what data to migrate so that the businesses can improve to extreme advantage. By creating conceptual and consistent data models we can additionally converse with the business to appreciate which data items are dangerous to achievement on the target policy and which items can be stocked.
8. **Assistances to align target application build with the goal movement design:** Very few migrations have a "firmed-up" target construction at scheme kick-off. Most target schematics[11] can vary considerably throughout the movement build and this can found a severe interval to the movement and obvious risk. By making a common consistent data model that is version controlled and jointly agreed amongst both parties it delivers a much easier means of communication than complex spreadsheets or enterprise documents

Disadvantages

Major outage is required for the restore process and can [12] be disturbing to or effect host performance at 2 distinct times. Process is very slow and does not scale very well when large amounts of tapes are necessary. Method does not scale to big amounts of data. Software certificate fees could be required Method affects server performance. It can develop difficult to achieve in heterogeneous effective system situations. Method becomes hard as the amount and size of situation increases. Method is dependent on system, submission and safety infrastructure (authentication and permissions).

VII. CONCLUSION

Data movement is a tough scheme with high level of risks like period overrun etc. Use of quality ETL tool will minimize the risk of weaknesses in data of destination database. Even though, testing of data for its validation is important and can't be overlooked. Data migration framework had been designed considering the latest model driven

software engineering techniques in order to enable the efficient data classification and translation to address the existing challenges and limitations. We implemented an efficient strategy to establish the effective mapping between source and target cloudbased.

VIII. REFERENCES

- [1]. Bansel, Aryan, HoracioGonz, and Adriana E. Chis. "Cloud-Based NoSQL Data Migration." *2016 24th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP)*. IEEE, 2016.
- [2]. Moskovits, Debra K. *Neotropical birds: ecology and conservation*. No. 598.298 N438. Chicago: University of Chicago Press, 1996.
- [3]. Cohen, Jeffrey, et al. "MAD skills: new analysis practices for big data." *Proceedings of the VLDB Endowment* 2.2 (2009): 1481-1492.
- [4]. Paygude, Priyanka, and P. R. Devale. "Automated data validation testing tool for data migration quality assurance." *Int J Mod Eng Res (IJMER)*(2013): 599-603.
- [5]. Abella, Manolo. "Policies and best practices for management of temporary migration." *paper for* (2006).
- [6]. Chai, Yunpeng, et al. "Efficient data migration to conserve energy in streaming media storage systems." *Parallel and Distributed Systems, IEEE Transactions on* 23.11 (2012): 2081-2093.
- [7]. Du, Yanling, et al. "Study of migration model based on the massive marine data hybrid cloud storage." *Agro-Geoinformatics (Agro-Geoinformatics), 2012 First International Conference on*. IEEE, 2012.
- [8]. Kari, Chadi, Yoo-Ah Kim, and Alexander Russell. "Data migration in heterogeneous storage systems." *Distributed Computing Systems (ICDCS), 2011 31st International Conference on*. IEEE, 2011.
- [9]. Karnitis, Girts, and Guntis Arnicans. "Migration of Relational Database to Document-Oriented Database: Structure Denormalization and Data Transformation." *Computational Intelligence, Communication Systems and Networks (CICSyN), 2015 7th International Conference on*. IEEE, 2015.
- [10]. Liu, Hui, and Shengqi Wu. "Data storage schema upgrade via metadata evolution in SaaS." *Consumer Electronics, Communications and Networks (CECNet), 2012 2nd International Conference on*. IEEE, 2012.
- [11]. <http://www.datamigrationpro.com/data-migration-articles>.
- [12]. <http://1401965.blogspot.in/2013/06/data-migration-methods-advantages.html>.