Journal of Information Security, 2021, 12, 79-103
https://www.scirp.org/journal/jis
ISSN Online: 2153-1242
ISSN Print: 2153-1234

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)

Abou_el_ela Abdou Hussein

Computer Science Department, Modern Academy-Maddi, ARE, Maddi, Egypt
Email: abo_el_ela_2004@yahoo.com

How to cite this paper: Hussein, A.A. (2021) 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). *Journal of Information Security*, **12**, 79-103.
https://doi.org/10.4236/jis.2021.121004

Received: December 11, 2020
Accepted: January 16, 2021
Published: January 19, 2021

Abstract

Data Migration is a multi-step process that begins with analyzing old data and culminates in data uploading and reconciliation in new applications. With the rapid growth of data, organizations constantly need to migrate data. Data migration can be a complex process as testing must be done to ensure data quality. Migration also can be very costly if best practices are not followed and hidden costs are not identified in the early stage. On the other hand, many organizations today instead of buying IT equipment (hardware and/or software) and managing it themselves, they prefer to buy services from IT service providers. The number of service providers is increasing dramatically and the cloud is becoming the preferred tool for more cloud storage services. However, as more information and personal data are transferred to the cloud, to social media sites, DropBox, Baidu WangPan, etc., data security and privacy issues are questioned. So, academia and industry circles strive to find an effective way to secure data migration in the cloud. Various resolving methods and encryption techniques have been implemented. In this work, we will try to cover many important points in data migration as Strategy, Challenges, Need, methodology, Categories, Risks, and Uses with Cloud computing. Finally, we discuss data migration security and privacy challenge and how to solve this problem by making improvements in it's using with Cloud through suggested proposed model that enhances data security and privacy by gathering Advanced Encryption Standard-256 (ATS256), Data Dispersion Algorithms and Secure Hash Algorithm-512. This model achieves verifiable security ratings and fast execution times.

Keywords

Cloud, Organizations, Migration, Data Quality, Advanced Encryption Standard

1. Introduction

Migration is a process of moving data from one platform/format to another platform/format [1]. It involves migrating data from an old system to the new system without affecting the active applications and ultimately redirecting all I/O activities to the new device. In simple words, it is the process of fetching data from different source systems into a single target system. Data migration is a multi-step process that begins with an analyzing old data and culminates in the loading and normalizing data in new applications. This process involves scrubbing the legacy data, mapping data from the old system to the new system, designing conversion programs, building and testing the conversion programs that perform the conversion, and matching the converter. Data migration could also refer to as the process of making an exact copy of an organization's current data from one device to another device; preferably without disabling or disabling active applications; then redirect all input/output (I/O) activities to the new device [2]. There are a variety of circumstances that may cause an organization to migrate data, including:

- Server or storage technology replacement or upgrade;
- Server or storage consolidation;
- Relocation of the data center;
- Server or storage equipment maintenance, including workload balancing or other performance-related maintenance.

The above scenarios are fairly routine parts of IT operations in organizations of nearly any size. Data migration, as an essential aspect of legacy systems, modernization projects, has been recognized as a challenging task that can lead to project failure as a whole [3] [4] [5]. Industry survey results [4] reveal that the data migration market is rapidly growing and business companies annually invest billions of dollars in data migration tasks ; however, only 16% of projects successfully complete their data migration tasks (*i.e.*, delivered on time and within budget)—64% of data migration projects failed to deliver on time and 37% were over budget. The main reason for overriding time and budget is the lack of a well-defined methodology that can help deal with the complexity of data migration tasks. In general, data migration is the process of transferring data from old data sources of an old system to new data sources of the target system, where the old and new systems have different data structures. There are several issues that may complicate this process greatly. **First**, legacy systems often have a number of heterogeneous data sources designed with different data modeling

tools or their interpretation under various semantics. This requires a thorough understanding of ancient data sources from various aspects, such as explicit or implicit data constraints, interrelationships across different data sources, and data availability. **Second**, legacy systems may contain inaccurate, incomplete, duplicate or inconsistent data. On the other hand, new systems often require additional semantic restrictions on data after it is migrated. Thus, scaling data quality to the level of new systems can be costly and time-consuming. A previous study [6] showed that 62% of data migration projects have significant data quality problems in new systems. **Third**, various data migration tasks such as data identification, validation, and cleansing must be performed frequently in the project and specification changes occur frequently to fix the detected problems. It is estimated [6] that 90% of the initial specifications change and over 25% of the specifications change more than once during the life of a data migration project. These issues highlight the importance of methodologies and best practice approaches that can be used to guide through the process of data migration. We try to introduce various aspects of data migration to be clear for reader and how we solve data migration security and privacy challenge using suggested model. In Section 2 we introduce background about data migration strategy, challenges, need, phases, and policy. In Section 3 we introduce data migration types, categories, methodology, and risks and its solutions. In Section 4 we speak about cloud computing, comparing it with traditional data storage, and existing solutions to secure the cloud. In Section 5 we suggested proposed model that introduces efficient way to secure data migration in the cloud. Finally Section 6 discusses conclusion and future work.

2. Related Background

2.1. Need for Data Migration

In today's world, data migrations for commercial reasons have become common. While replacing the old system is the common cause, some other factors also play an important role in deciding to migrate data to a new environment. Some of them [1]:

- Databases continue to grow exponentially requiring additional storage capacity.
- Businesses are turning to high-end servers.
- To reduce cost and reduce complexity by migrating to consumer and static system.
- Data should be transportable from physical and virtual environments to concepts such as virtualization.
- For clean and accurate consumption data.

The data migration strategy must be designed in an efficient manner so that it enables us to ensure that future purchasing decisions are fully meet for both current and future business and the maximum commercial return on investment.

2.2. Data Migration Strategy

A well-defined data migration strategy should address the legacy data, mapping data from the old system to the challenges of identifying source data, interacting with continuously changing targets, meeting data quality requirements, creating appropriate project methodologies, and developing general migration expertise [1].

The following are the main considerations and inputs for defining a data migration strategy:

- Strategy to ensure the accuracy and completeness of the migrated data post migration.
- Agile principles that let the logical group of data to be migrated iteratively.
- Plans to address the source data quality challenges faced currently as well as data quality expectations of the target systems.
- Design an integrate migration environment with proper checkpoints, controls and audits in place to allow broken accounts/errors to be identified/reported/resolved and fixed.
- A solution to ensure proper reconciliation at various checkpoints to ensure migration is complete.
- A solution to choose the right tools and technologies to meet the complex nature of migration.
- Must be able to handle high volume data during migration.
- Migration development/testing activities must be separated from legacy and target applications.

In brief, the Data Migration strategy will involve the following key steps during end-to-end data migration:

- Identify the Legacy/source data to be migrated.
- Identification any specific configuration data required from Legacy app.
- Classify the process of migration whether manual or automated.
- Profile the legacy data in detail.
- Identify data cleansing areas.
- Map attributes between Legacy and Target systems.
- Determine the data and map to migrate to the historical data warehouse solution (archive).
- Collect and prepare transformation rules.
- Conduct disinfection prior Migration when required.
- Extract the data.
- Data Transfer along with limited clearance or standardization.
- Data loading.
- Reconcile data.

2.3. Data Migration Problems

These problems included, but were not limited to them [2]:

- Prolonged or unexpected downtime.

- Data corruption, data missing or loss.
- Application performance challenges.
- Technical compatibility challenges.

In order to stop these challenges from affecting business operations, the huge majority of data migration projects are typically organized to occur during off-hours, firstly during weekends. However, this can lead to an increase in migration costs as a result of employees' overtime, and it can negatively affect the morale of IT personnel. Moreover, stopping the real systems, even over the weekend, can be drastic. It affects business operations, especially if there are problems fetching systems return online. In fact, the prospective problems with data migration causing some organizations to delay deploying new technology, or even delay purchasing new technology. These delays can be damaging in and of themselves, because older devices may require more practical maintenance, and their performance is generally less and more prone to failure. Most organizations strive to deploy new technology to eliminate such challenges; therefore, the delay in implementing new technology represents a commercial risk. In addition, delaying deployment of a new storage device that has already been purchased or the leased one increases its actual cost, as the company amortizes the cost of both old and new devices and pays rental fees for both old and new devices.

How organizations can reduce the business impacts of data migration-down-

Time, data loss and cost increase? The best approach is to use a consistent, reliable and reproducible methodology for migrations that includes planning, technology implementation and validation.

2.4. Data Migration Process Challenges

Data migration as a process leaves organizations with many potential concerns and pain points [6]. Below is a brief description of these points.

1) Data Security Concerns

For any business organization, data is the most crucial resource. It may consist of business-centric data along with other related data critical for its existence [3]. Any compromise or threat to its security is a risk that businesses would not want to undertake. The same notion spills into migrating data to the cloud. A small hint suggesting that the clouds not secure will make organizations develop cold feet towards migration. Any cloud infrastructure will comprise of patchworks of open source code, which creates security vulnerabilities. Additionally, public clouds are multi-tenant, and such elements as vulnerabilities or defects of a co-subscriber's code could substantially affect other applications. To tackle this concern, many cloud vendors are performing "on boarding audits" to reassure prospective customers that their level of security is appropriate. Nonetheless, its level of conviction still needs confirmation.

2) Poor Knowledge of Source Data

The existence of poor knowledge of the source data is a general trend already

observed over several data migration processes across industries [7]. Issues such as duplicates, spelling errors and erroneous data are always a hindrance to ensuring complete and proper data migration. Often, organizations become complacent and tend to assume that they can configure their data without any complications. However, any mismatch could mean nothing else but the failure of the data migration process.

3) Vendor Management

From the perspective of businesses, the process of data migration requires businesses to trust their vendor [7]. Concerns exist whether technical issues on the vendor's side could affect data security on the cloud. It is therefore imperative that data migration vendors provide SLAs that prioritize the concerns of their clients. Since cloud computing offers a standardized, multi-tenant infrastructure, cloud vendors may not offer the same level of SLAs as IT managers are accustomed to.

4) Lack of Technical Integration

Data migration often involves various kinds of technologies and data platforms [3]. This lack of parity may lead to failure in data transfer between the multiple phases of data migration—analysis, development, testing, and implementation. Such failures not only cause financial repercussions but also compel businesses to re-engage time in the migration of missing data, leading to a loss of precious man-hours.

5) Cumbersome Data Cleansing Process

Data cleansing refers to the process of altering data intended for migration [7]. The mechanism takes into consideration incomplete data, data relevance, data accuracy and data duplication as factors of validation. It focuses on maximizing data accuracy in a system. Additionally, it uses parsing or other relevant methods to omit syntax errors and typographical errors in records. Despite there being cases where data cleansing leads to increase in response time and hampers efficiency, its significance in a fruitful data migration is second to none.

2.5. Data Migration Phases

The following Tables 1-6 describe the different stages of data migration along with the participating groups as well as the output and outputs associated with each of the stages [1].

Table 1. Phase 1—Data assessment.

Key Activities	Key Participating Groups	Deliverables/Outputs
➢ Identification of data sources ➢ Run system extracts and queries ➢ Perform reviews with users on the process data migration ➢ Revise migration scope and validation strategy ➢ Generate work plan with milestone dates	➢ Data migration leads ➢ End/Business users ➢ Program sponsors	➢ Scope document ➢ Strategy document ➢ Work breakdown structure with milestone dates

Table 2. Phase 2—Data cleansing.

Key Activities	Key Participating Groups	Deliverables/Outputs
➢ Analyze and identify if data cleansing is required. ➢ Create worksheets with the steps for data preparation ➢ Clean up of source data ➢ Formatting of unstructured data ➢ Run extracts and queries to determine data quality ➢ Create metrics to capture data volume, peak hours and off-peak hours	➢ Data migration team ➢ Client Information Support team	➢ Cleaned/changed source data that increases successful automated conversion data. ➢ Control metrics and dashboards

Table 3. Phase 3—Test extract and load.

Key Activities	Key Participating Groups	Deliverables/Outputs
➢ Create/verify data element mappings ➢ Run data extracts from current system(s) ➢ Create tables, write scripts and schedule jobs to automate the extraction ➢ Fix any remaining data clean-up related issues ➢ Execution of application specific customizations ➢ Execution of mock migrations ➢ Using bulk loading functionality, load extracts using tools such as SQL loader ➢ Business rules validation and referential integrity checks through data validation. ➢ Exceptions reporting to client team ➢ Perform data validation	➢ Data migration team ➢ Client Support team ➢ Data Administrators team	➢ Source system extraction ➢ Modules, scripts and ETL jobs for data migration ➢ Application populated with transformed data ➢ Various controls points to handle errors, exceptions and alerts.

Table 4. Phase 4—Final extract and load.

Key Activities	Key Participating Groups	Deliverables/Outputs
➢ Execution of final extracts from the current systems ➢ Execution on target database customizations ➢ Conduct application specific customizations ➢ Deploy pilot migrations ➢ Bulk loading of data extracts using ETL tools into target system. ➢ Perform data validation checks including referential integrity validation and business rules ➢ Acknowledge errors and exceptions to client ➢ Perform data validation	➢ Data migration ➢ Support team ➢ Data Administrator group	➢ Extraction of data from source ➢ Data migration modules, ETL jobs and data migration scripts ➢ Control points to identify error and exceptions.

Table 5. Phase 5—Migration validation.

Key Activities	Key Participating Groups	Deliverables/Outputs
➢ Prepare for migration validation reports and metrics related to data movement. ➢ Review and update migration validation reports and metrics ➢ Verify record counts on the new system ➢ Fix any exceptions or unexpected variations on the data. ➢ Validation sign off	➢ Data migration team ➢ Client Information ➢ Support team ➢ Business users	➢ Signed-off migration validation document

Table 6. Phase 6—Post migration activities.

Key Activities	Key Participating Groups	Deliverables/Outputs
➢ Complete documentation on data migration reports, file and manuals. ➢ Data correctness and quality reports ➢ Target system reports and its accuracy ➢ Infrastructure capacity report and dashboards ➢ Sign off	➢ Data migration team ➢ Client Information ➢ Support team ➢ Business users	➢ Exception reports, cross-reference files/manuals ➢ Infrastructure dashboards ➢ Signed-off data migration project closure document

2.6. Seven Steps to Include in Your Data Migration Plan

If you're preparing to replace or upgrade servers, perform server maintenance, or move to a data center, following a data migration plan can simplify the process [8]. Without one, there's a high risk that during the process of moving your data between systems and formats, you'll wind end up with costly downtime, corrupted, lost, and misplaced files, compatibility issues and more. Below are seven steps identified for a successful data migration.

1) Identify the Data Format, Location, and Sensitivity

Determine data format, location and sensitivity before you start the data migration process, determine data you are migrating, what format it is currently in, where it is located, and what format it should be in post-migration phase [8]. By identifying this information, you'll be armed with the knowledge needed to start the project. During this advanced planning process, you may discover potential risks that you'll need to plan for before moving on, or realize that certain security measures must be taken while migrating certain data. This advanced planning step can save you from making a fatal mistake during the actual migration process.

Tip: Choose a method that works for you, whether it's a spreadsheet or whiteboard. Remember, this process is important.

2) Plan the Size and Scope of the Project

Once you understanding the data being transferred, define the scope of the data migration plan [8]. Plan the resources you'll require to use through the migration and create a realistic budget. Perform an advanced analysis of both the source and target system, and write out a flexible project schedule. Consider

whether the data migration will interfere with normal business operations, or subscribe to downtime. You may be able to plan the migration after hours or on weekends to avoid disrupting business continuity.

Tip: Make sure to communicate with key stakeholders about the schedule and potential downtime.

3) Backup All Data

Before migrating, make sure to back up all your data, especially the files you are migrating. If you encounter any problems during migration, such as corrupt, incomplete, or lost files, you'll have the ability to correct the error by restoring the data in its original state.

Tip: Cloud backup is the most secure and safe backup method. Read more about backup strategies.

4) Employee Evaluation and Migration Tool

Data migration can be a big task, especially if you're transferring a large number of files, or if the migration is complex, or you're migrating sensitive information [8]. Refer back to the project size and scope and use this information to determine:

- If your team has the knowledge and skills to complete the project, or if you will need to consult an outside expert.
- If your team has the time and resources available to tackle the project within assigned time frame.
- Who can you bring to help you complete the project? If you've decided that you will use data migration software, re-evaluate its features and flexibility to ensure that it meets the requirements that you need to accomplish the migration.

5) Implement a Data Migration Plan

With your plans to guide you, make sure the correct system permissions are applied to allow the data to be migrated successfully and extract all data that is migrated to the target, from the source system [8]. MAKE SURE TO CLEAN THIS data to protect target system, then convert it to the appropriate format for transfer. Finally, upload your cleaned and de-duplicated data into the target system data migration rules and map you've already laid. Closely monitor data migration during the process, so you can identify any issues that may arise.

6) Final System Test

Once the migration is complete, ensure that there are no problems communicating with the source and target systems [8]. The goal is to ensure all the data migrated is correct, secure, and in the right location. To verify this, perform unit, system, volume, web-based application and batch application tests.

7) Follow-Up and Maintain the Data Migration Plan

Even with testing, it's always possible to go wrong during migration. To account for this, perform a full system and data quality check to ensure everything is correct once the data migration process is complete [8]. If you notice errors, or missing, incomplete, or corrupted data, restore these files from your backup.

Using the seven steps outlined above, you'll be able to successfully transfer

your data from the source system to the target system. Just remember to **backup all data** before you start migrating data. In the event an unexpected issue, you'll be able to undo the damage and restore the critical data your business relies on.

2.7. What Points to Consider Before Migrating Your Data

Data migration is a crucial step in more ways than one [7]. A concerned manager must keep in mind several considerations before deciding to initialize the data migration process. Below are important points that individuals and organizations need to consider before migrating to the cloud.

- **Self-Analysis**

Self-analysis should be the first and most crucial point on the checklist before migrating to the cloud. Organizations often ignore to analyze themselves and their business needs. Questions such as—do we need to migrate? Can we afford it? Can we risk to not have our data on premise? Is the data suitable enough to be migrated to the cloud? The significance of self-analysis is to enable the organization to set business goals and drivers so that the migration process can match their aims and objectives. Additionally, an in-depth rumination will facilitate the qualitative and quantitative analysis of the currently installed base of IT structure.

- **Collect Performance Statistics**

One of the primary reasons for businesses to migrate is to enhance their IT performance. It becomes crucial for organizations to have a clear picture of their current server resources. An unclear analysis would make the organization vulnerable to post-migration performance issues. It is important for organizations to gather performance statistics for physical servers along with CPU usage, memory usage, network throughput and disk input/output. It is advisable to collect six months of data so that the business can identify peak usage requirements and trending data.

- **Roster Physical and Virtual Assets**

To ensure that businesses do not leak costs and spend avoidable time, it is imperative for them to take stock of their specific assets, architecture, and infrastructure size for migration. For physical servers, it is advisable to take note of the server model, operating system or the hypervisor in use. Additionally, one must also take into consideration the number of CPUs and cores, along with the amount of RAM, storage configuration and the amount of storage configuration.

Businesses can analyze the virtual environments by collating the operating system, the number of virtual CPUs, amount of RAM and assigned storage.

- **Categorizing Servers as per Business Need**

Data migration is not only a lengthy process but also brings with it a substantial amount of expenses. One way of avoiding unwanted costs is by matching the right server environment to the workload to which the organization would be migrating.

This will ensure the best capabilities for the job without paying more than needed. Segmenting servers as per their business criticality, business impact and

non-criticality will add to the cause of avoiding unwanted expenses.

- **A Continuous Process**

Data migration is not a one-time process. Nor is it an open and closed case. Future changes are inevitable and CTOs will have to reach out to their vendors to update technology by applying best practice and proven methodologies to deliver high quality, cost-effective solutions.

Considering these important aspects before initializing data migration is crucial. These decisions are significant to ensure that the entire process is coherent with the organization's goals and doesn't take a toll on the organization's finances.

2.8. Why Data Migration Is Performed [9]

- Acquiring and integrating the business unit/organization that leads to the change of process in the organization.
- To enhance efficiency, performance and scalability of program application.
- To adapt new modifications in terms of technology, Market practices, operational efficiency, regulatory requirements to lead to better customer service.
- To reduce operational cost and efficiency by simplifying and eliminating bottlenecks in the application process or when moving different data centers to a group in one location.

2.9. When Data Migration Goes Wrong

With migrating to the cloud increasingly becoming a necessity, it is imperative that organizations pay greater attention to effective data migration [7]. Bloor Research has conducted numerous surveys in recent years regarding data migration. A post in-depth analysis revealed that many organizations went way past their budgets and schedules during data migrations. Not even two-thirds of companies managed to complete their migration projects in time and within the allocated budget. A list of factors that affects them are explained as follows:

- Lack of pre-migration planning.
- Lack of clearly visualizing the post-migration.
- Lack of a set of specialized technical skills to perform migration.
- Failure to determine the appropriate range of costs and time required.
- Improper schedule late backup.
- Inefficient project management skills related to migrations specifically.

2.10. Data Migration Policy

The migration scheme of an existing business system to a cloud platform is a complex process. Businesses need to consider several factors [7]. No organization would want to affect their investment in legacy equipment. Furthermore, it would want to reduce investment waste. Thus, you need to select an appropriate migration approach, based on the type and importance of the system for migrating to the cloud. Specific migration policies include:

- **Migration to the Cloud Platform**

Migrating the business system to IaaS, deploying it to the virtualized resources (such as virtual servers, virtual stores, and virtual networks) on the cloud platform, and additionally employing a stable operation management platform for managing the cloud.

- **Migration after Transformation**

This policy includes transforming the system architecture, operating environment, interfaces, and similar aspects to meet the technical requirements for migration to a cloud platform before the actual migration. This process involves checking, for example: whether the Oracle database needs transformation into the MySQL or SQL Server database.

- **Maintaining the Status Quo**

Businesses may also decide to continue to maintain the current operating environment of the existing business system, including infrastructure until the system retires.

3. Data Migration Types, Categories, Methodology, Risks and Solutions

3.1. Types of Data Migration [9]

- **Data Base Migration**

When you migrate data from one current database resource to another database resource, the current database is updated to the most recent version. For example: IBM DB2 Database to Oracle Database.

- **Data Center Transfer**

When the data center is moved from one site to another site, we need to migrate the data from the old data center database to the target data center database.

- **Application Migration**

When migrate an application, for example migrating from a local activity server to a cloud or from one cloud domain to another, the underlying data must also be migrated to a new application.

- **Business Process Migration**

When a business process changes due to merger, acquisition, or business improvements, depending on the nature of the changes in business processes, the data must be transferred between a different storage system or app or location.

3.2. Data Migration Categories

Data migration could be broadly categorized into two major categories [7].

1) Infrastructure Migration

The term "Infrastructure Migration" mentions the process of migrating all layers of a computing platform along with the applications that support business functions. This type of migration is a more complex exercise that has the potential to significantly impact entire IT operations more than other strategies

would. For example, an infrastructure migration can include changes to the following [7]:

- Applications that support business functionality.
- Application infrastructure that supports the applications, such as web servers, application servers, middleware and database technology.
- Third-party products provided by ISVs.
- Computing and storage platforms, e.g. SAN or attached storage.
- Network infrastructure.
- Facilities infrastructures, such as power, ventilation, and cooling.
- Management policies.
- System monitoring and management tools.
- Locally written scripts to manage applications and data.

2) Application Migration

The term "Application Migration" applies to applications rather than infrastructure. It generally applies to custom-written applications and refers to modifying or normalizing the code of an application. Its objective is to recompile and deploy on a new hardware platform that supports a different operating system (OS). Application migration has an innate association with modifying the code base of an application to allow replication of the functionality provided by Application Programming Interfaces (API) of the existing OS and supporting software products in the new target environment. Application Migration is more of a mechanical effort for making the application compatible with the new environment. It requires the integration of the application with a new development environment, as well as with a new operating system. While source code, scripts and data are shifted, compilers, source code repositories and software tools are replaced by advanced versions that are compatible with the target platform.

3.3. Data Migration Methodology

As illustrated in **Figure 1** data migration methodology could be represented in three phases as follows [2].

1) Plan

IT organizations recognize the critical importance of planning for technology deployment. While the planning amount depends on the size and scope of the migration, the process of planning should totally include [2]:

- Determining the requirements of the migration;
- Identifying the current and future environment;
- Creating and documenting the migration plan.

During the planning stage, it is required to define the hardware or software required to successfully perform the migration. Design requirements include migration engineering, specific hardware and software requirements, migration procedures, and designation and testing plans. When necessary, the IT organization must also obtain any software licenses it needs to perform the migration. Consistent migration planning can help determine where potential problems occur and how to avoid them, and IT professionals can help identify mitigation

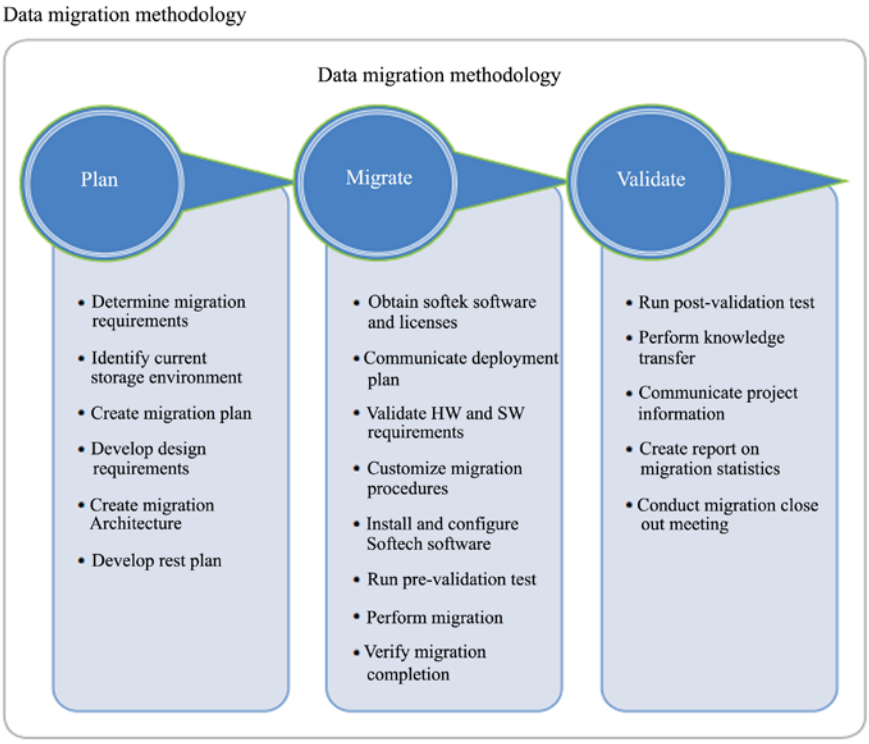

Figure 1. Migration methodology.

strategies if problems are inevitable. Migration planning can also help you decide which data to migrate first, and whether offline applications are taking too long and whether internal and external audiences will be notified about the migration. Table 7 gives an example for migration plan [2].

Correct migration planning involves more than just the IT staff. Business owners should also include the applications and data they migrate-particularly because the IT organization determines how important a particular application or set of data is to business. When planning a migration, it is important to understand design requirements such as migration/replication requirements, schedule, vendors involved, and hardware configuration. When determining the size of data migrations, there are several key elements to consider such as the number of servers, operating system levels, amount of storage, volume managers, types of databases and applications, network speeds, and server clusters.

2) Migrate

During the migration phase, the IT organization needs to communicate its plans; obtain, setup and design any necessary software; and perform the actual data migration. It is recommended that you run the data validation test before migration, as well as the post-migration validation test. These tests confirm that the data is in the same state after migration as before. The most important part at this stage is clearly the migration itself. As shown above, software technology can simplify this process by improving migration speed, by reducing or eliminating application downtime, and/or by enabling migration during normal business hours, helping the organization get back to work as quickly as possible.

Table 7. Example migration plan.

Action item	Assigned to	Status	Date
Establish a migration management team			
Gather availability and production schedules			
Document change control procedures so they can be incorporated into the migration procedures and plans			
Document the timeline for activities for both hardware changes and the data migration			
Announce the migration at least 30 days before the intended target migration date			
Gather information about the storage server environment and applications (list and/or drawing)			
Work with the storage vendor to understand the new storage configuration			
Create a technical migration team			
Inform the security and compliance groups about the migration			
Schedule a premigration rehearsal that includes all the members of the migration team and a data sampling that will enable the application groups to appropriately conduct the pre- and post-migration verification process			
Follow the required change control process			
Establish a migration status call-in process			
Utilize a migration planning checklist to ensure that all the premigration planning steps have been executed			

3) Validate

After the migration is complete, the IT organization must compile migration statistics and make a report to discover what worked and what didn't and the benefits learned. The report must be propagating with all members of the migration team. These types of reports are important in building a repeatable and consistent process through continuous process enhancement-based on what worked and fix or change what didn't. Moreover, documenting the migration process can help train employees, and simplify or streamline the next migration, which reduces costs and risks.

3.4. The Risks Involved in the Data Migration Process and the Solution to Be Overcome

1) Data Loss Risk

When data is available in the old system but after the migration process, if it is not available in the target system, it is called data loss [9]. Data loss is one of the highest risks in data migration. The Cost involved in correcting the data loss and the cost of work involved due to poor data adds the financial and reputation risk.

The Solution: Reconciliation

There can be folded reconciliation number Reconciliation and primary finan-

cial column reconciliation. Comparing the number of records in the legacy system and the target system will give a fair assessment about data loss during migration [9]. It is not necessary that the number of old system records always match the number of data in the target system; there may be business rules for rejecting records based on certain parameters. If these rules are used, then number of rejected records as well as the number of records in the target system. The laboratory must also validate the reason for refusal according to the business rules.

2) Data Corruption and Data Integrity

The format and content of the data in the old system and the target system differ in comparison with the migration process, and then it is called corrupted data [9]. Due to data migration, anomalies, duplicated data or the presence of meaningless data are associated with data integrity issues. Data corruption and data integrity affect the efficiency of business and totally outweigh the purpose of the migration.

Solution: Data Validation

Validating every data between the legacy system and the target system is the best methodology to avoid data corruption [9]. The following Data validation methodologies are used widely.

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

Sample data validation: includes random selection of the record from the old system and its comparison with the target system. Sampling is not a faulty system as it selects small random records. Records of profiling samples bring more data coverage than random sampling. Subset of data validation: Instead of choosing random sample records, we choose here a subset of records based on the row number such as the first thousand records or ten thousand to fifty thousand records. The only advantage is to select more records and expect more data coverage based on higher likelihood. Full validation of the dataset: It is a perfect validation method we should strive to test the migration. Each record is compared in a bidirectional manner, and each record in the old system is compared against the target system and the target system against the old system. When two different database vendors are involved, such a comparison is impossible, to overcome the instances needed to create a single database with both legacy and target system data. Elements to consider in data validation:

a) Project stability.

b) Data coverage.

c) Implementation time.

d) Efficiency of the Query/Script.

3) Semantics Risks

During migration, sometime the meaning of old column and target column happens to have the same meaning but their unit of measure is different and the meaning of the data is completely changed [9]. It is important to note in this

scenario that the data is not lost or corrupted; migration is successful but not beneficial in terms of purpose.

The Solution

Real time users and subject matter experts should be involved in the feasibility study and such semantic issues should be discovered very early in the project life cycle [9]. The test scope should include test cases to identify inconsistencies and incompatibilities between the carried data and target application criteria. Testers manually compare the objects in the source and target application by looking at the application main screen.

4) Interference Risks

This type of risk arises when all stakeholders use the source application simultaneously during the transition period [9]. For example, if a stakeholder accesses a certain table and closes that table, and if anyone else tries to access that table, they will not be able to do so. In such cases, interference risks arise.

The Solution

This should only be managed at the organizational level and this scenario should be discussed at the time of project planning [9]. One method is to plan for a multiple phantom run that involves all stakeholders and also to plan for a pilot run in a pre-production environment that involves all stakeholders.

4. Cloud Computing

Cloud computing is the availability of computer system resources on demand, especially data storage and computing power, without active management directly by the user.

4.1. Cloud Computing and Its Impacts on Data Migration

Over the past several decades, IT society has been overwhelmed by a new buzzword of "going Cloud" [10]. The basic premise of cloud computing is that consumers (individuals, industry, government, academia and so on) pay for IT services from cloud service providers (CSP). Services offered in cloud computing are generally based on three standard models (Infrastructure-, Platform-, and Software as a Service) defined by the National Institute of Standards and Technology (NIST) [11]. As more cloud-based services available to users, their oceans of data are outsourced in the cloud as well. The cloud becomes, then, the tools of choice for more data storage services.

4.2. Cloud Data Storage versus Traditional Data Storage

Compared to traditional data storage as shown in **Figure 2**, cloud data storage offers several advantages as shown below [10].

1) Lower Costs and Complexity

Nearly six out of 10 providers say that cost reduction is the main goal for customers to use cloud services [10] [12]. Consumers do not need to buy and rely on new equipment. The cloud storage service allows them to start their application

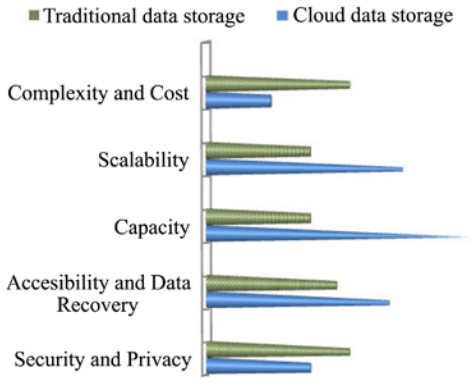

Figure 2. Traditional storage vs. cloud storage.

right away, and the providers charge the users accordingly use of resources (Pay As You Go). Service typically costs a fraction of the cost of implementing an on-site solution.

2) Scalability and Capacity

Since the consumer needs more capacity, the service provider can achieve more scalability, and is much simpler than if you had to add the equipment in your own space [10]. So, you can increase your storage capacity no matter how much space you need.

3) Purposes of Restoration of Archives and Disasters

A cloud storage provider can help an organization enhance the security of its internet services, by preventing loss due to theft, or disaster. Therefore, firms that archive their information in the cloud do not have to worry much about natural disasters.

4) Increase Accessibility and Reliability

The stored files can be accessed from anywhere via an internet connection. However, there are many other profits to use cloud storage services, and there are also some elements that influence cloud adoption such as Usability, Bandwidth, Data Security, Privacy, etc. In fact, the two main barriers to cloud adoption are Security and Privacy. With the exception of these two barriers, the others are outside the scope of research.

4.3. Existing Solutions to Secure the Cloud

These days, many cloud vendors and service providers are leveraging cloud storage models to offer their own security solutions known as Security as a Service (SecaaS) [10]. For any other cloud services, it is possible that the SecaaS model will embrace all of the characteristics of cloud computing and provides users with little or no control over services and tasks. The Cloud Security Alliance (CSA) has identified some of the services that SecaaS provides [13]-among them is encryption as a Service. According to some providers such as Amazon Web Services (AWS) and Dropbox, Encryption as a service is the most straightforward way to maintain data security in a Cloud environment [14]. For example, AWS offers a flexible and scalable security layer for data stored in the cloud by

providing many options for data encrypting at rest from AWS to fully automated encryption solutions [15]. SecaaS offerings can take many forms that cause market confusion and complicated the selection process. Clients are increasingly faced with evaluating SecaaS solutions, which are not working in the workspace, and they need a better understanding of these offerings to assess the type of security issues they address and also to assess the security risks and shared responsibility for the security of the systems for which they are responsible. Although CSPs providers claim that customer data is safe and well-secure, we know, from reports [16] of cloud leaks to celebrities and companies, that there is no complete guarantee of safety when storing sensitive data in a cloud environment unless consumers engage with the cloud in the right way, with the right checks and balances to ensure all security and risk management measures are covered. In addition to SecaaS, academia and industry have developed many of research work and practical methods to secure cloud data storage. Many recent proposed models are based on [17] [18] [19] [20] encryption algorithm such as DES, AES and Elliptic curve cryptography (ECC). Therefore, to ensure data confidentiality before migrating in the cloud, some methods have relied on encrypting data by encrypting the data with AES-128, 192, 256 bits, depending on the file size and data format [20]. 128 and 192 bit encryption is suggested for users to encrypt large and medium data respectively. However, 256 bits key encryption will be done on small data sets, but based on the data's format rather than its size. Another method is to combine Encryption and Obfuscation [21] [22]. Encryption is carried out on the alphanumeric characters and the alphanumeric data type, and the numeric type is obfuscated only [21]. Specifically, encryption helps the user to provide confidentiality for his data when it is transmitted over the network, and data jamming helps the service provider secure the user's data in rest mode [22]. This model provides an enhancement in data confidentiality. A secure personal cloud storage system has been featured that provides plug-and-play convenience for a portable storage device (PSD) [23]. The demonstration of the accessibility and scalability of cloud storage [10], using an information dispersion algorithm (IDA) to meet the dual requirements for data reliability and availability. Its implementation allows slides files to be distributed across multiple CSPs. Therefore, the opponent will need to settle for at least two of the three cloud providers to be able to reconstruct the data. Likewise, two CSPs must be complicit in launching any effective attack. Zhang X. and Wang H. proposed a system architecture that adopts IDAs to secure off-site data storage [24]. The main component of this system is a proxy server. For the process of writing a file to the cloud storage, the user copies the file to the desired folder on the network drive. This file will be stored temporarily on the proxy server; at about the same time, the proxy server will create its own random array to covert the file into multiple slices. In order to further improve the confidentiality of information, pieces of data will be encrypted by the proxy server before they leave the trusted intranet. The resulting data slices will be stored on online cloud drives provided by Amazon, Dropbox, or Rackspace via the proto-

col converter. However, the encryption is good but not sufficient as it is only computationally safe; advances in cryptographic analysis may render what is considered a secure cryptography today ineffective tomorrow [23]. Moreover, the promised capabilities of quantum computing may render current encryption algorithms useless [25].

5. Suggested Model

One big wall when using the cloud is data security. The cloud contains many challenges and issues, but among them, security is a primary concern [10]. Mostly, encryption techniques are preferred for securing data and most of the technologies are outdated [26]. Therefore, to overcome the Security and Privacy issues of cloud storage, there is a suggested model that guarantees data confidentiality, integrity, availability and leakage [10]. The suggested approach chosen consists of data encryption or decryption based on the adoption of the advanced 256-bits key Encryption Standard (AES-256), IDA with C-RS and then the SHA-512 hash algorithm as major components of the process. Detailed suggested model description for both encoding and decoding processes, mathematical background and equations, pseudo codes, and figures are explained in [10].

Experimental and Result Evaluations of Suggested Model

To assess the actual performance, they analyzed the performance of their algorithm by measuring the computation time of the encoding and decoding process for a data file, depending on the size (51 KB to 347,778 KB) and a variability of threshold (m, n); $1 \le m \le n$ and $n \le 256$ [10]. They carry out these experiments on a Windows 7 Operating System (Redmond, WA, USA) with 8-core Intel Xeon E5-1620 (Santa Clara, CA, USA) at 3.50 GHZ with 32 GB of memory. In **Figure 3**, they have shown the computation time of the encoding operation. The encoding time for proposed algorithm mainly depends on the size of the data and the value of the (m, n) threshold.

They obtained the computation time by summing the encryption time of x KB data plus the time to implement IDA with C-RS, and the time to hash and concatenate each slices file. According to suggested model experiment results [10], the maximum encoding time is 14.15 s for a large data file equal to 347,778 KB with 200,254 thresholds. They observed for different (m, n) configurations and data less than 347,778 KB that their encoding algorithm yields the best performance, since the average encoding time is 1.966 s. In **Figure 4**, they have shown the running time of the decoding operation. As the decoding time for x KB data file is equal to the sum of the different steps followed during the process. The maximum computation time is 24.30 s for 347,778 KB file with (2, 3) and the average decoding time in this experiment is 2.907 s. Decoding process provides the highest performance for small and medium size data. However, the time required to decode x KB data depends on the internal validation time and secondly on the rebuild time. Relative to the large volume of data, these two processes get a little computationally expensive, but they are better than the previous works.

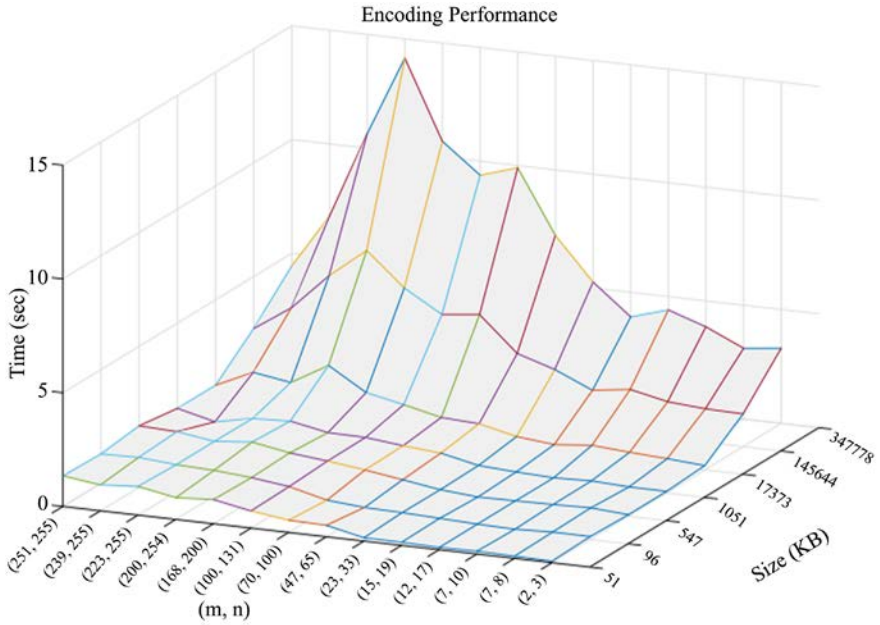

Figure 3. Encoding results.

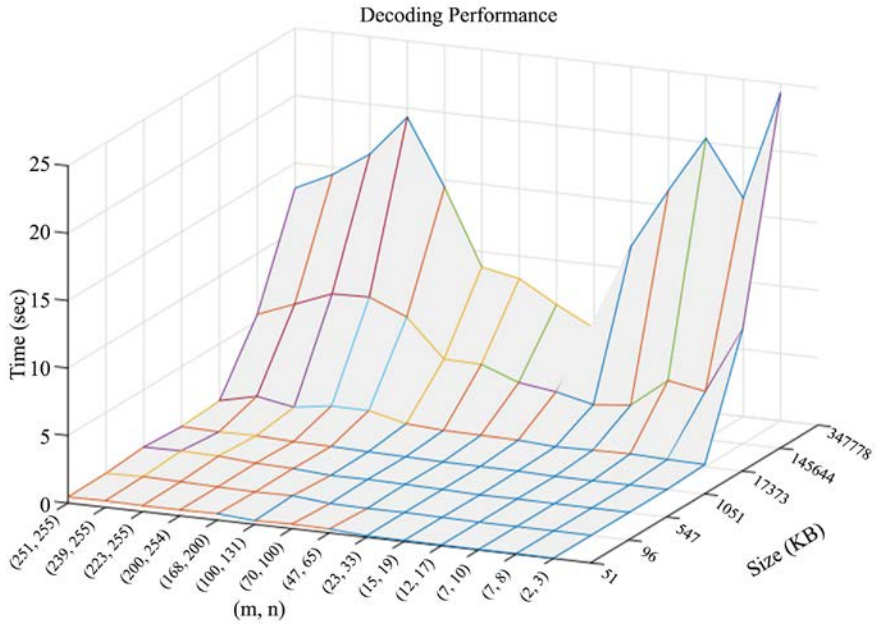

Figure 4. Decoding results.

Depending on the value of (m, n), when the threshold is small, the verification time increases and the rebuilding time decreases. In addition, when the threshold is large enough, the verification time is greatly reduced and the rebuilding time increases. Therefore, they recommend, for the best performance, a medium threshold. This will greatly reduce decoding calculation time. In Table 8 below [10], they show the average execution time for each internal step in the encoding and decoding processes. From table Encryption Takes 0.367 s, Slicing-IDA takes 1.125 s, Hashing and Concatenation takes 0.474 s, Verification takes 1.453 s,

Table 8. Average encoding/decoding time.

Process	Duration
Encryption	0.367
Slicing-IDA	1.125
Hashing and Concatenation	0.474
Verification	1.453
Reconstruction-IDA	1.157
Decryption	0.297

Reconstruction-IDA takes 1.157 s, and Decryption takes 0.297 s. It is quite evident that the encoding time is mainly estimated through the slicing execution time, since it requires more parameters and calculations than the other internal processes.

Encryption is everywhere; most of the products collected today support an encryption algorithm and especially AES. Even those who support other algorithms tend to recommend AES [10] [27]. However; most of cryptographic algorithms (AES, RSA, etc.) are not safe against quantum computing [28]. In the following lines, they are looking specifically for products that encrypt files, not full-disk solutions like VeraCrypt. Boxcrypor: is file-encryption software designed and built specifically for cloud use, with support for all major cloud-storage providers. The user has the ability to encrypt, decrypt, Share, add and remove files from cloud storage directly through its end device. Moreover, there is the possibility to protect files by encrypting them locally before publishing them to the cloud-storage provider, ensuring security and privacy of file-data throughout the process. **Boxcryptor** uses AES-256 and RSA cryptography algorithms. It ensures data accessibility, confidentiality and privacy protection.

SecureDoc CloudSync: designed by WinMagic, this is an enterprise endpoint encryption solution that encrypts data on endpoint devices before storing it in the cloud, providing an extra layer of security that is provided by a storage service provider [10]. SecureDoc CloudSync uses the Advanced Encryption Standard-New Instruction (AES-NI) 256-bit encryption. It provides data confidentiality and data privacy but does not prevent eavesdropping. Regarding these two products and their deployment, it is quite evident that they provide a faster execution time than the proposed methodology since both are cryptographic, which corresponds to layer 1 of the proposed algorithm (AES-256). However, their use guarantees the confidentiality and privacy of the data but does not guarantee the integrity and availability of the data. However, implementation enhances CIA data (Confidentiality-Integrity-Availability) [29].

Confidentiality is achieved by combining AES-256 and IDA, integrity is achieved by linking each particular chip with its corresponding SHA-512 hash code, and, ultimately, data availability is achieved by dispersing the resulting data segments (IDA) [10].

6. Conclusion and Future Work

Cloud computing is a multi-sharing environment, in which resources are shared. Threats can be happened from anywhere; inside or outside the common environment. Deciding whether to migrate or retain sensitive data in the workspace is one of the most important decisions faced by personal users, as well as small and medium-sized enterprises. We have described throughout this paper a number of key points starting from explaining why we need data migration, its strategy, data migration problems and challenges, data migration phases, when performing data migration and when don't, data migration policy, data migration types, categories, methodology, risks and solutions, cloud computing and its impacts on data migration, solutions to secure cloud, and finally we suggested to overcome the Security and Privacy issues of cloud storage and risks associated with cloud data storage and examining current ways to mitigate data security and privacy threats. The suggested model based on the combination of AES-256, IDAs and SHA-512 consists of encoding and decoding data on premise and guarantees data confidentiality, integrity, availability and leakage. Suggested model consists of encoding operation that includes: AES-256 Encryption to ensure data confidentiality, IDA with Cauchy Reed-Solomon code to break the encrypted data into n slices such that we can recover from m and then SHA-512 Hashing algorithm for signature and decoding operation that includes a verification process to check data slice integrity, IDAs to reconstruct the encrypted data from m slices, and, finally, the decryption process to recover the original data. Suggested model achieves far a greater degree of security and also better performance for small and large data files. Future work, which is already in progress, will take title named (DMig 2) referring to complete searching in the same field data migration because, as it is clear from the title of current paper called (DMig 1) that means first paper in the field of data migration.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Sarmah, S.S. (2018) Data Migration. *Science and Technology*, **8**, 1-10.

[2] Best Practices for Data Migration (2007) Methodologies for Planning, Designing, Migrating and Validating Data Migration. IBM Global Technology Services.

[3] Thalheim, B. and Wang, Q. (2013) Data Migration: A Theoretical Perspective. *The Journal of Data Knowledge Engineering*, **87**, 260-278. https://doi.org/10.1016/j.datak.2012.12.003

[4] Howard, P. and Potter, C. (2007) Data Migration in Global 2000, Research, Forecasts and Survey Results. A Survey Paper by Bloor Research. http://www.bloorresearch.com/research/survey/876/data_migrtaion_survey.html

[5] Hull, R. (1984) Relative Information Capacity of Simple Relational Database Schemata. In: *Proceedings of Principles of Database Systems*, ACM, New York, 97-109.

https://doi.org/10.1145/588011.588027

[6] (2004) Rapid Application Development (RAD) for Data Migration = White Paper Solutions by Premier International. http://www.premier-international.com/pdf/Applaud_White_Paper.pdf

[7] Alibaba Cloud. Data Migration Methodology. aliyun.com, Alibaba Migration Platform and Related Solutions.

[8] Nordic Backup (2017) 7 Steps to Include in Your Data Migration Plan. https://medium.com/@NordicBackup/7-steps-to-include-in-your-data-migration-plan-42d571dc01ab

[9] Chellamuthu, P. (2014) Data Migration Challenges and Solution for Successful Implementation.

[10] Sighom, J.R.N., Zhang, P. and You, L. (2017) Security Enhancement for Data Migration in the Cloud. *Future Internet*, **9**, 23. https://doi.org/10.3390/fi9030023

[11] NIST (2017) Cloud Computing Program-29 July 2016: Cloud Computing. https://www.nist.Gov/programs-projects/cloud-computing

[12] Wolfgang, G. (2017) Cost Reduction Remains Chief Reason to Adopt Cloud, Confusion Still Apparent. http://www.tomsitpro.com/articles/cloud_survey-kpmg-tech_adoption-provider-itsecurity,1-803.html

[13] Jaeger, B. (2016) Security as a Service Working Group, Defined Categories of Security as a Service (Preview)-Continuous Monitoring as a Service. Cloud Security Alliance 2016. https://downloads.cloudsecurityalliance.org/assets/research/security-as-a-service/csa-categories-securities-prep.pdf

[14] Wall, M. (2016) Can We Trust Cloud Providers to Keep Our Data Safe? http://www.bbc.com/news/business-36151754

[15] (2016) White Paper. Securing Sensitive Data within Amazon Web Services Ec2 and Ebs: Challenges and the Solutions to Protecting Data within the AWS Cloud. Copyright 2013 Vormetric. http://go.thalesesecurity.com/rs/480-LWA-970/images/wp-securing-data-within-AWS.pdf

[16] Rancourt, C. (2017) Celebrities Hacked: "Are Your Personal Photos Safe in the Cloud?" http://www.nextadvisor.com/blog/2014/09/02/celebrities-hacked-personal-photos-cloud-safe

[17] Ahmadi, M., Moghaddam, F.F., Jam, A.J., Gholizadeh, S. and Eslami, M. (2014) A 3-Level Re-Encryption Model to Ensure Data Protection in Cloud Computing Environments. *Proceedings of the IEEE Conference on System, Process & Control*, Kuala Lumpur, 12-14 December 2014. https://doi.org/10.1109/SPC.2014.7086226

[18] Surv, N., Wanve, B., Kamble, R., Patil, S. and Katti, J. (2015) Framework for Client Side AES Encryption Technique in Cloud Computing. *Proceedings of the IEEE International Advance Computing Conference*, Bangalore, 12-13 June 2015, 525-528. https://doi.org/10.1109/IADCC.2015.7154763

[19] Singh, S. and Kumar, V. (2015) Secured User's Authentication and Private Data Storage-Access Scheme in Cloud Computing Using Elliptic Curve Cryptography. *Proceedings of the* 2015 2*nd International Conference on Computing for Sustainable Global Development*, New Delhi, 11-13 March 2015.

[20] Raj, G., Kesireddi, R.C. and Gupta, S. (2015) Enhancement of Security Mechanism for Confidential Data Using AES-128, 192 and 256 Bit Encryption in Cloud. *Pro-*

ceedings of the 2015 1*st International Conference on Next Generation Computing Technologies*, Dehradun, 4-5 September 2015.
https://doi.org/10.1109/NGCT.2015.7375144

[21] Arockiam, L. and Monikandan, S. (2014) Efficient Cloud Storage Confidentiality to Ensure Data Security. *Proceedings of the International Conference on Computer Communication and Informatics*, Coimbatore, 3-5 January 2014, 1-5.
https://doi.org/10.1109/ICCCI.2014.6921762

[22] Suthar, K. and Patel, J. (2015) EncryScation. A Novel Framework for Cloud IAAS, DAAS Security Using Encryption and Obfuscation Techniques. *Proceedings of the* 2015 5*th Nirma University International Conference on Engineering* (*NUiCONE*), Ahmedabad, 26-28 November 2015.
https://doi.org/10.1109/NUICONE.2015.7449636

[23] Mar, K.K., Law, C.Y. and Chin, V. (2015) Secure Personal Cloud Storage. *Proceedings of the* 10*th International Conference for Internet Technology and Secured Transactions* (*ICITST*-2015), London, 14-16 December 2015.
https://doi.org/10.1109/ICITST.2015.7412068

[24] Zhang, X. and Wang, H. (2013) A Study of the Use of IDAs in Cloud Storage. *International Journal of Future Computer and Communication*, **212**, 67-70.
https://doi.org/10.7763/IJFCC.2013.V2.123

[25] Rich, S. and Gellman, B. (2016) NSA Seeks to Build Quantum Computer That Could Crack Most Types of Encryption.
https://www.washingtonpost.com/world/national-security/nsa-seeks-to-build-quantum-computer-that-could-crack-most-types-of-encryption/2014/01/02/8fff297e-7195-11e3-8def-a33011492df2_story.html?utm_term=.217ad6b56479

[26] Balamurugan, S., Sathyanarayana, S. and Manikandasaran, S.S. (2016) ESSAO: Enhanced Security Service Algorithm Using Data Obfuscation Technique to Protect Data in Public Cloud Storage. *Indian Journal of Science and Technology*, **9**, 1-6.
https://doi.org/10.17485/ijst/2016/v9i17/90229

[27] Rubenking, N.J. (2017) The Best Encryption Software of 2017.
http://www.pcmag.com/article/347066/the-best-encryption-software-of-2016

[28] Mishra, B. and Jena, D. (2016) Securing Files in the Cloud. *Proceedings of the* 2016 *IEEE International Conference on Cloud Computing in Emerging Markets*, Bangalore, 19-21 October 2016. https://doi.org/10.1109/CCEM.2016.016

[29] Stallings, W. (2005) Cryptography and Network Security Principles and Practices. 6th Edition, Prentice Hall, Upper Saddle River.