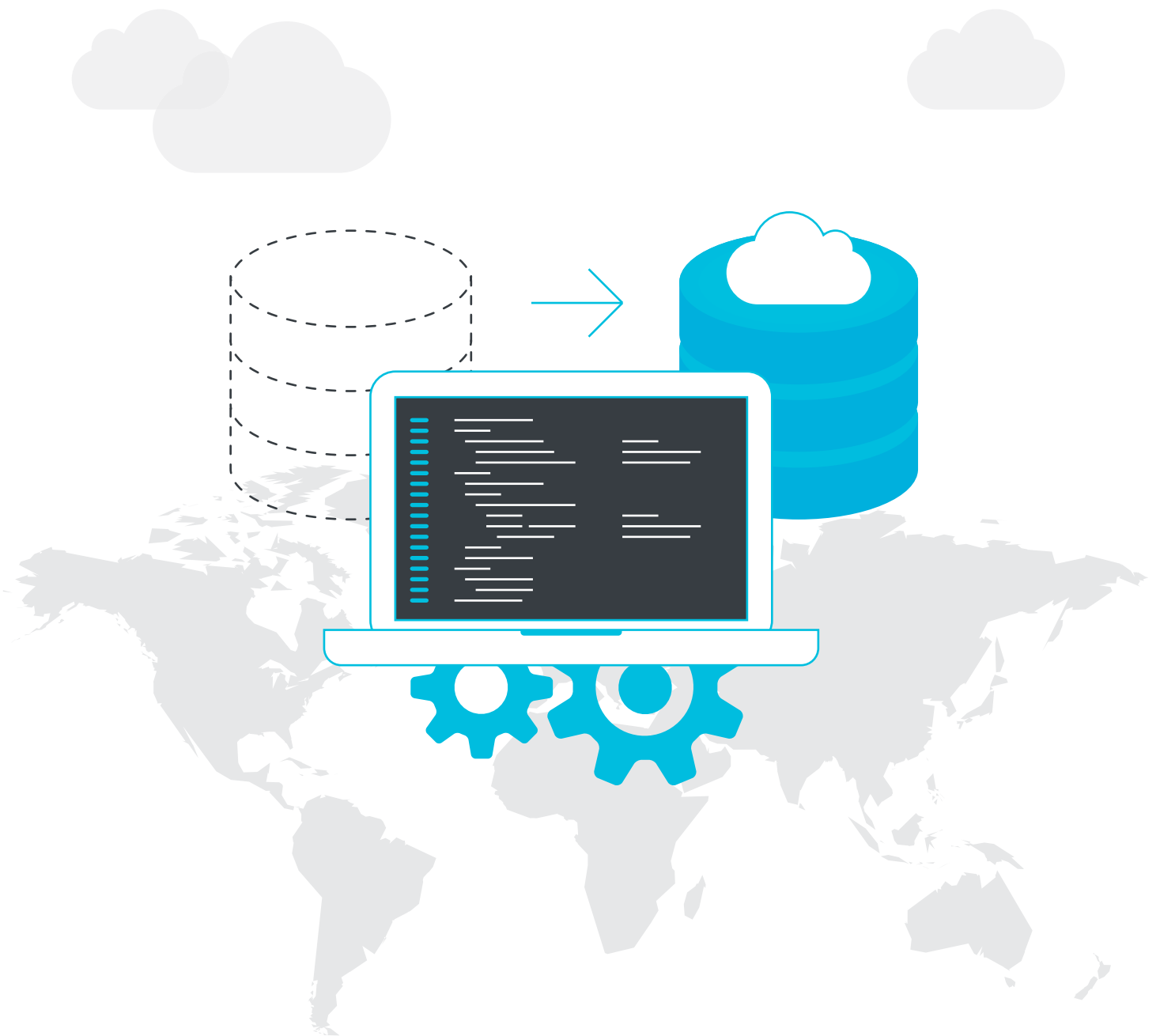


Data Migration Methodology



Content

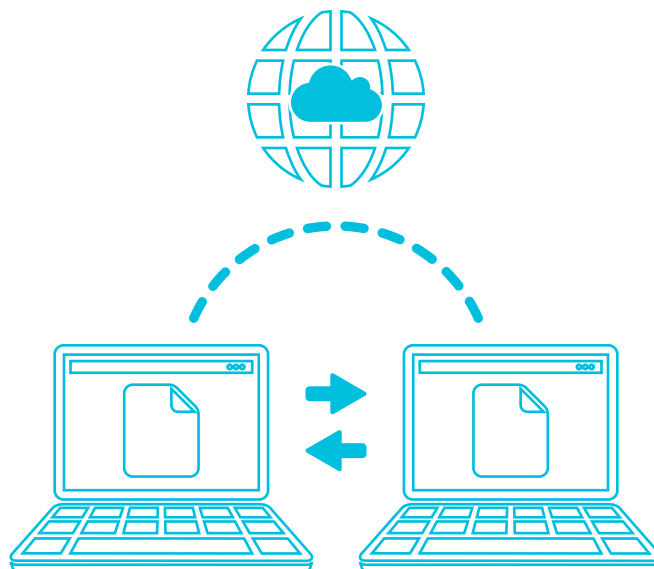
1. Introduction	3
2. Industry Pain Points and Business Risks	4
3. Data Migration Gone Wrong	6
4. Points to Consider Before Migration	7
5. Types of Data Migration	9
6. Data Migration Policy	11
7. Migration Strategy	12
8. Migration Process	14
8.1 System Research	14
8.2 Risk Assessment	16
8.3 Designing Cloud Architecture Solution	18
8.4 System Transformation	19
8.5 Data Migration Verification	19
8.6 Functional/Performance Tests	19
8.7 System Cutover	20
8.8 System Delivery	21
9. Alibaba Migration Platform and Related Solutions	22
10. Conclusion	23

01 Introduction

Cloud computing is a digital and global transformation that has revolutionized how businesses function. With infinite storage available at economical rates, multiple security assurances, and diligent availability, cloud computing makes new inroads towards a different tomorrow.

At the epicenter of cloud computing is data migration. Data migration refers to the process of transferring data between data storage systems. A business usually undertakes data migration to replace or upgrade servers or storage equipment, for website consolidation, to conduct server maintenance or to relocate a data center. Data migration is the core aspect that takes place when a business signs up to the cloud.

In the context of this whitepaper, data migration refers to transferring data from on-premise storage to the cloud. Furthermore, this whitepaper provides answers to questions on data migration methodology. Business leaders, CTOs, Vice Presidents, and technology leads will find this paper helpful in understanding the strategic importance of data migration methodology, pain points, and more importantly, key aspects for a smooth data migration transition.



02 Industry Pain Points and Business Risks

Though data migration facilitates the transition to superior technology, it entails more than what it reveals to the eye. A process as crucial as data migration does not come without its own set of challenges. Data migration as a process leaves organizations with several concerns and potential pain points. Below is a brief description of these points.



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. 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 cloud is 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 "onboarding audits" to reassure prospective customers that their level of security is appropriate. Nonetheless, its level of conviction still needs confirmation.



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. 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.



Vendor Management

From the perspective of businesses, the process of data migration requires businesses to trust their vendor. 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.



Lack of Technical Integration

Data migration often involves various kinds of technologies and data platforms. 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.



Cumbersome Data Cleansing Process

Data cleansing refers to the process of altering data intended for migration. 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.

The criticality of data migration is such that one cannot take any of the pain points mentioned above for granted. Any form of deviation in data matching and verification will lead to failure in migrating the data.

03 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. [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. Putting it in terms of hard numbers, the average budget for a data migration project among surveyed participants was \$875,000 USD, and the average overrun cost was \$268,000 USD!

Furthermore, the post-survey analysis revealed some common factors that resulted in this disparity. Below is a list of these factors:

- Lack of pre-migration planning
- Failure to clearly envision the post-migration environment
- Lack of specialized technical skill set to perform the migration
- Failure to properly scope out the required costs and time
- Improper backup for delayed schedule
- Inefficient project management skills specifically pertaining to migrations

04 What Points to Consider Before Migrating Your Data

Data migration is a crucial step in more ways than one. A CTO or 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.



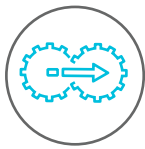
A Continuous Process

Data migration is not a one-time process. Neither is it an open and shut 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.

05 Types of Data Migration

We can broadly categorize data migration into two major categories:



Infrastructure Migration

The term "Infrastructure Migration" refers to the process of migrating all layers of the computing platform along with the applications that support business functionality. 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:

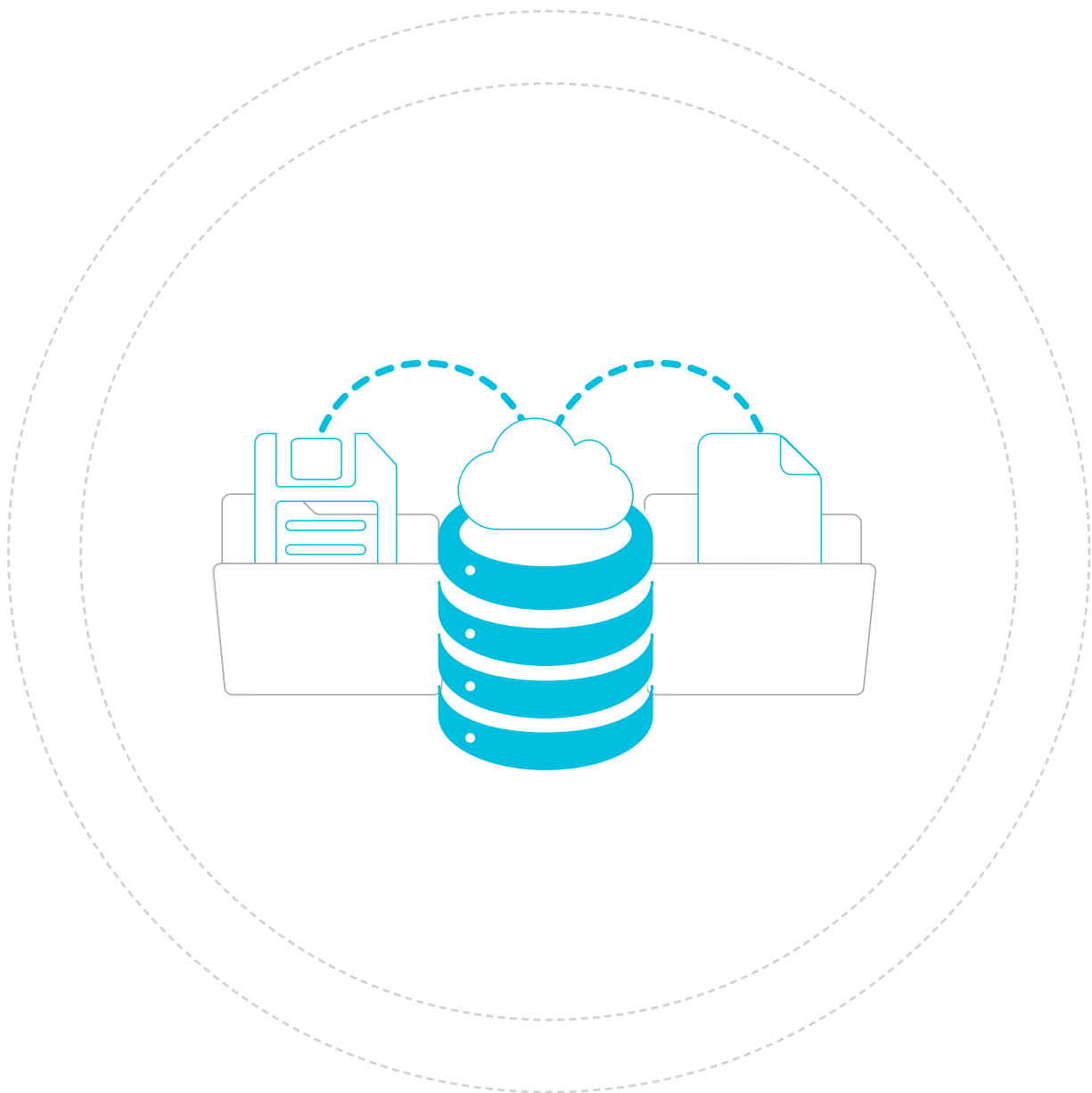
- Applications that support business functionality
- Application infrastructure that supports the applications, such as web servers, application servers, middleware and database technology
- Third-party products provided by ISVs
- Computing and storage platforms, e.g. SAN or attached storage
- Network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



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.

When migrating an application, one must also complete migration of any supporting third-party software. If the software is not available on the new platform, one will need to find and integrate similar software into the application. In case the intensity of integration becomes imtemperate, the migration might start to appear less like re-hosting and more like a re-architecture effort.



06 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. 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. For migration of complicated systems to the cloud, you must adopt customized migration technologies and approaches according to the specific situation. 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.

07 Migration Strategy

The entire data migration process is crucial for any business. A reliable and effective migration strategy starts through proper interaction with your data migration provider. Below are various strategies for effective and efficient data migration.



Refronting

For organizations that wish to change just the data entry portion of the application, instead of rewriting an entire application, refronting is the most suitable migration strategy. It adds a more aesthetic interface to an existing application without changing the functionality. End users will have access to the same data; however, it will visually enhance the entire experience of data accessibility. Note that this does not require the use of expensive terminals, cabling or peripheral interconnects. Development of a browser-based solution is possible whenever desired. Furthermore, web-enabling an application can provide significant cost reduction for the businesses.



Replacement

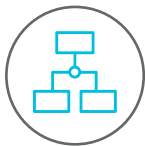
This strategy is ideal for organizations seeking to yield the highest benefit and who do not have an issue with the high degree of the cost involved. Replacement migration strategy decomposes the legacy application into functional building blocks. A Common Off-The-Shelf (COTS) replaces parts of a generic and complex, custom-written legacy application upon fragmenting the legacy. Businesses might find it more beneficial to adopt a new application solution and change the organization's business. Additionally, the replacement data migration strategy can prove to be a quick and minimal risk option for organizations.



Rehosting

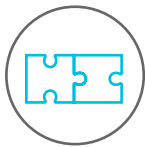
Many organizations want their current business logic and business process to remain unchanged and believe that preserving it is sufficient. For such organizations, rehosting is the ideal migration strategy, which involves moving complete applications from a legacy environment with no change in functionality. Furthermore, it offers the advantage of low development risk and enables quick transfer of familiar legacy applications to a more cost-effective platform. This exhibits lower total cost of ownership (TCO) and a faster return on investment (ROI). Additionally, rehosting is a preferred migrating strategy for organizations that intend to minimize their maintenance and support costs. Also, it

does not change the application or the architecture. However, this would also mean that you might not properly utilize the new technology without some modification of the application. It further proves to be a preferred migrating strategy as it offers the possibility of using cost savings accrued through switching development and runtime environments to fund full re-architecture projects when warranted.



Rearchitecting

Businesses may have several applications that lack IT effectiveness and functionality. For organizations with such applications and with minimum time limitations, re-architecting is the most suited technique. It is a customized data migration strategy that uses new programming paradigms and languages. Using this strategy, one can develop applications from scratch on a new platform altogether, enabling businesses to enhance the application functionality. Furthermore, it opens the window to ameliorate the business logic and processes along with the opportunity to modify the developer productivity model. The obstacle to this approach is that it requires additional training for end users, developers and technical staff. Moreover, this approach is the most time consuming and error prone.



Interoperation

This strategy is specifically for the ideal application, which has all the features compatible with the latest technologies. It can prove to be advantageous to incorporate new technologies into the current application stack if it meets business requirements.



Retirement

Changes in technology can eliminate the need for specific functionality in an application or an overall solution. As middleware or third-party products mature, they might cause the functionality implemented to become obsolete. In such instances, legacy utilities or legacy application functionalities become redundant, necessitating their retirement or implementation elsewhere in the solution.

08 Migration Process

The data migration process is an amalgamation of several carefully executed detailed steps. Below is a description of these steps.

8.1 System Research

Thorough research is imperative before the entire data migration process initializes as it allows the data migration team to understand the current system status, existing architecture, and future planning. Furthermore, it helps in understanding if the existing system matches the cloud platform. This is crucial to provide first-hand information on the follow-up system migration scheming and implementation.

The system research stage mainly involves investigations into the application system through questionnaires, interviews, and collection of system data, observation of the application system, as well as other standardized processes and methods. It also includes business research, system architecture research, database research, and application research. Below is a brief description of the various phases of system research.

8.1.1 Business Research

This step involves carrying out basic research and analysis based on the businesses of the application system intended for migration. Additionally, it includes the research and analysis on the business type, users, business use features, and business performance indicators, among other aspects.

It mainly includes the following content:

- System name
- Affiliated to
- System business description and target objects for serviceSystem development/operation status (online, in development, in design, in plan)
- System type (website, OA system, ERP, CRM and so on)

8.1.2 System Architecture Research

It takes into account a comprehensive research and analysis on the overall application system deployment along with a thorough research on the operating system, system operation status quo, system scalability, system data flow, system relevance, and other aspects.

The main content includes:

- Research on peripherals and commercial software demands

- Research on network demands
- Research on transformation planning
- Research on dependencies of various system modules, including whether the system is independent or relies on other systems
- Research on safety requirements
- Resource usage (servers, storage devices, and network bandwidth)
- Whether the system is of the OLAP or OLTP type

8.1.3 Database Research

This phase witnesses collection of basic information. This includes the database version, deployment structure, and data security policy, and the usage information including the existing database capacity, traffic, as well as SQL, and advanced features for technical research and analysis at the database layer.

The main content includes:

- Database vendor/version
- Database architecture (whether it is RAC, or the master/slave architecture)
- Backup policy (cold standby, hot standby, and backup cycle)
- Data capacity and traffic statistics (peak TPS/QPS, case specifications in the database, the number and names of tables with more than 10 million data records, and the number of peak connections).
- SQL collection (the top 50 SQL statements in the database in terms of accesses and slow SQL statements)
- Collection of advanced database features (Oracle/SQL SERVER): Stored procedures, functions, triggers, packages, materialized views, virtual columns, partitions, DBlinks, sequences, full-text indexes, DTS and so on.
- Database character sets

8.1.4 Application Research

This step involves researching the application architecture, usage of middleware, application load, and other information are collected for technical research and analysis at the application layer.

The main content includes:

- **Operating System Architecture:** This refers to analyzing if the application is compatible with the OS to which you are migrating the data.
- **High Availability Design:** This step includes checking if the application design supports the migration process.
- **High Performance Design:** This is to determine if the application can handle heavy load of data.
- **Data Storage Methods:** The application may interact with various types of storage solutions. The questions that may arise during analyzing data storage methods could be:

- Does the system use multiple data sources at the same time?
 - If some files are stored, what is the file storage scheme?
 - What file types does the file storage contain?
 - What is the log file storage method?
 - What is the call method with the database?
- **Compatibility:** Determine if the language used for application development is compatible with the system
 - **Framework and Architecture:** Analyze the framework used for application development and whether the system adopts the B/S architecture or the C/S architecture
 - **Deployment:** How does one deploy the application on the system and what third-party components are used?
 - **Interface:** Does it call any external interface or service and if so, what is the adopted interface protocol type? What is the interface protocol type if it provides services for external calls?
 - **Middleware:** Determine the middleware class, which middleware products are used and does the middleware adopt single-point deployment or cluster deployment?
 - **Additional Information:** How do the third-party components impact the deployment process, are custom extensions used and what are the system performance indicators?

System research is the stepping-stone towards an efficient data migration process. With its in-depth research and analysis, businesses can take necessary steps to initiate their data migration process.

8.2 Risk Assessment

A detailed assessment of the risks of the actual data migration process follows the research. The risk is gauged based on the conclusion of the research report during the system research phase along with the combination of the architecture characteristics of the cloud platform. Furthermore, the cloud migration team can determine the following risks:

- The feasibility of migrating the system to the cloud (and cloud platform compatibility)
- System eligibility to migrate to the cloud
- Requirement for system transformation or code refactoring
- Transformation difficulty
- Post migration system support

Calculation of the transformation schedule and technical challenges of the migration project is possible through a series of research. Additionally, the cloud migration team can evaluate the risks that emerge during the system migration process and analyze any unsupported features by the cloud platform to work out a targeted solution during the scheming stage. Risk assessment includes several aspects, as shown in the figure below.

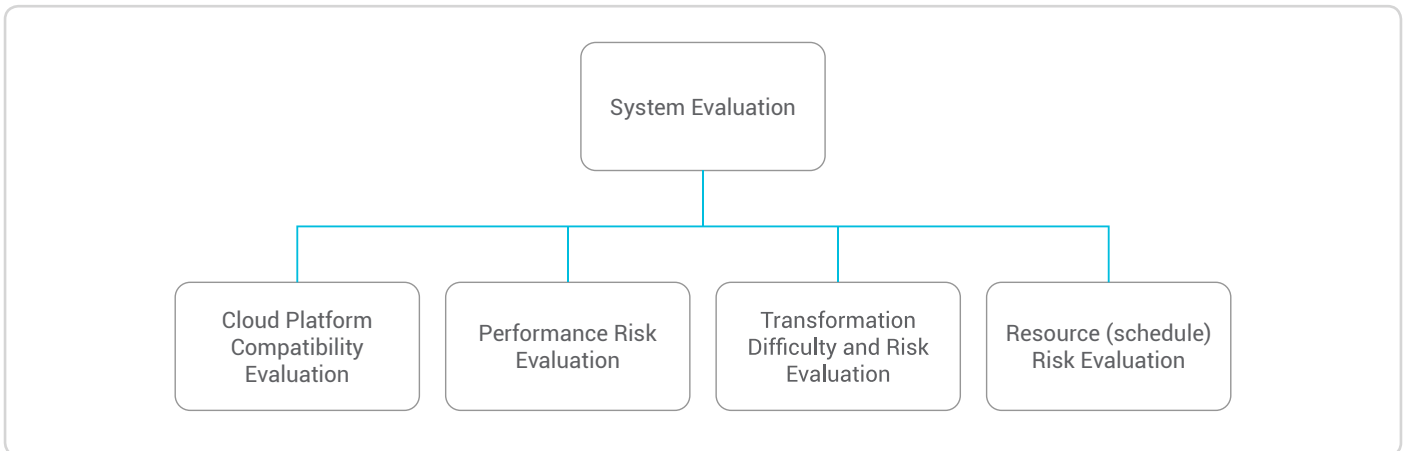


Figure 1. Data Migration System Evaluation

8.2.1 Cloud Platform Compatibility Evaluation

The evaluation involves a thorough study of the actual situation of the application system, the unsupported software, and hardware by the cloud platform. This is done to work out a corresponding solution. Evaluation seeks to achieve the following:

- Specific hardware on the cloud (dongles, leased lines, high-performance graphics cards, and dependencies on some specific IP addresses)
- Whether the cloud network architecture meets the requirements
- Does the cloud security match the security level of the system?

8.2.2 Performance Risk Evaluation

The data migration team can evaluate performance bottlenecks of the existing system to develop an optimization scheme for the application system. This extends to whether there is a need to apply database or table-based splitting, and determining if it requires massive data processing technology.

8.2.3 System Transformation Risk Evaluation

Evaluation of system risks during the transformation process is carried out in accordance with business characteristics, technical features, and cloud platform traits of the existing application system. The following aspects should be considered.

- Whether the application transformation meets the original system design specifications
- Whether the data migration scheme meets the system cutover requirements
- The de-O transformation difficulty
- Whether the transformed module is compatible with the call dependencies of other systems

8.2.4 Resource Risk Evaluation

This step involves evaluating the risks in the cloud migration implementation plan, cloud platform resource preparations, human resources of the cloud migration implementation team, and similar aspects.

8.3 Designing Cloud Architecture Solution

The new architecture of the application system on the cloud, with migration schemes are determined based on system checks and risk assessment results, combined with the characteristics of the cloud platform. One needs to make decisions based on the following cases:

- Whether migrating the system to the cloud must occur directly
- If a series of transformations (such as de-O) are required
- Does the file system need to migrate to the OSS?
- Should the data analysis system be compatible with the cloud platform?
- What are the estimated transformation schedules?

Compared with traditional APP + DB deployment model, the cloud is more suitable for a Server Load Balancer/ECS/RDS combination to achieve high availability (see figure below).

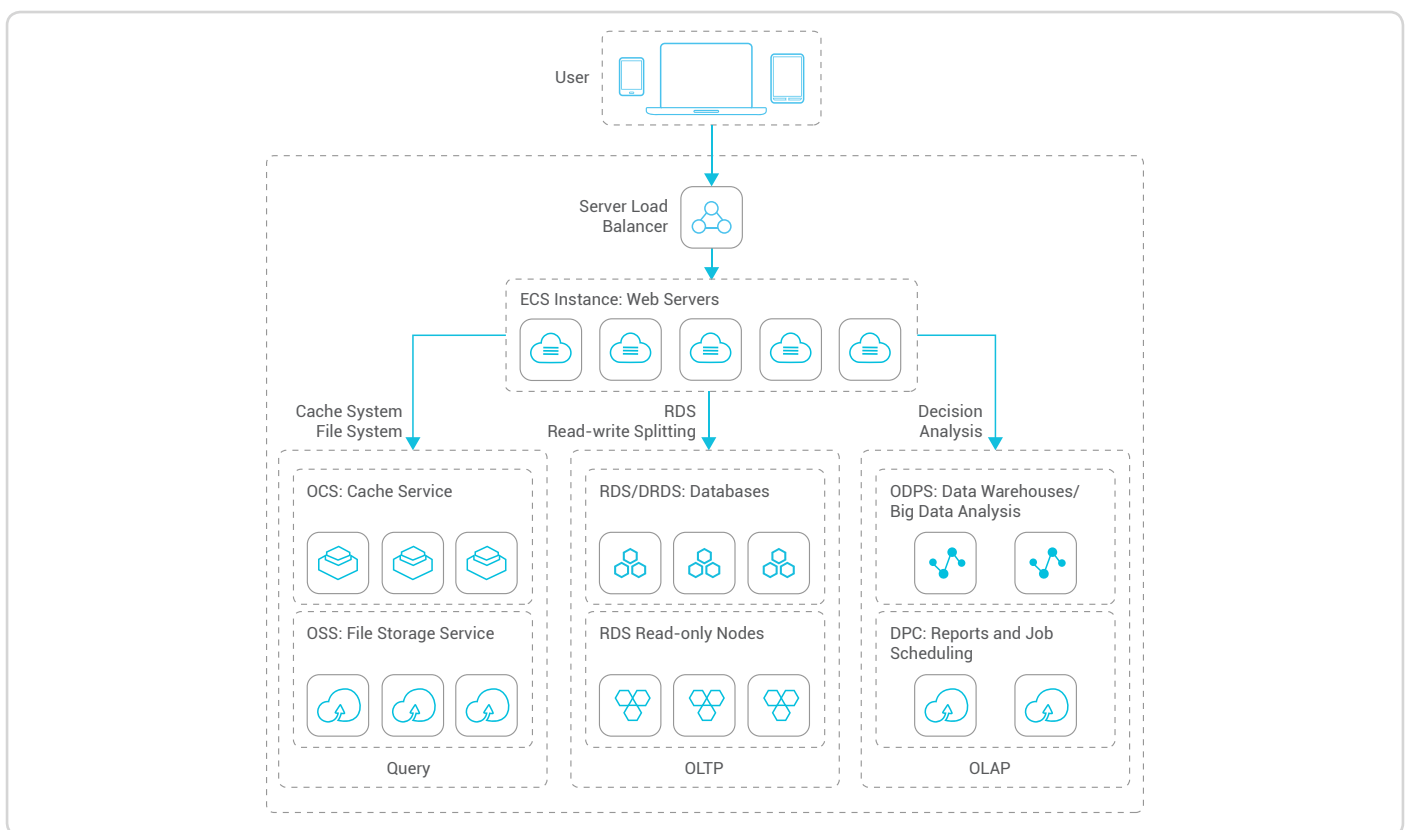


Figure 2. Data Migration Architectural Solution

- **System Deployment Scheme Design:** The cloud platform-based applications and database deployment scheme, based on the application system features, such as availability, stability, and performance requirements, can be output.
- **System Transformation Scheme:** Based on system research and risk assessment results as well as the cloud platform characteristics, the database transformation scheme, application transformation scheme, and application system verification scheme.

8.4 System Transformation

The next step involves transforming, testing and verifying the existing application system and verification based on the system transformation scheme. The process is as shown in the figure below. It mainly includes system architecture transformation, database transformation, application transformation, and system test and verification.

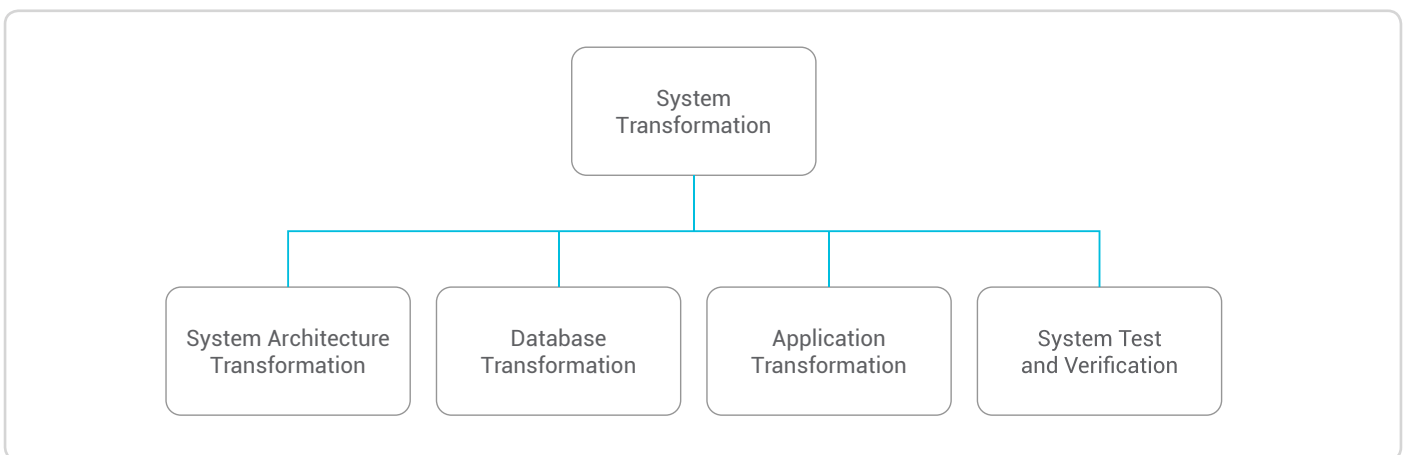


Figure 3. Data Migration System Transformation

8.5 Data Migration Verification

- Migrate the local database to the RDS
- Migrate file systems and videos to OSS
- Use the data comparison tool is used to verify the data consistency between the source database and the target database

8.6 Functional/Performance Tests

Upon verifying the data migration, the business/organization needs to complete the functional, performance, and data integrity verification tests according to the testing cases in the system design. The matching review includes two parts

– the manual review and the tool review.

After the completion of the migration, the application owner conducts a manual review of the system architecture and deployment before initiating the functional tests. This is followed by initiation of the tool review. However, tool review is to be done only if the manual review is errorless. One can then use the migration script to compare and examine the paths included, file lists, and code on the completion of the manual review. After passing the review, the functional test can be activated.

8.7 System Cutover

This phase mainly completes the cutover of the new and old application systems and ensures that the application system migrated to the cloud can run on the cloud platform stably and efficiently. The specific processes include the application and activation of cloud resources, database migration, application migration, and business cutover. Below is a brief description of these processes.

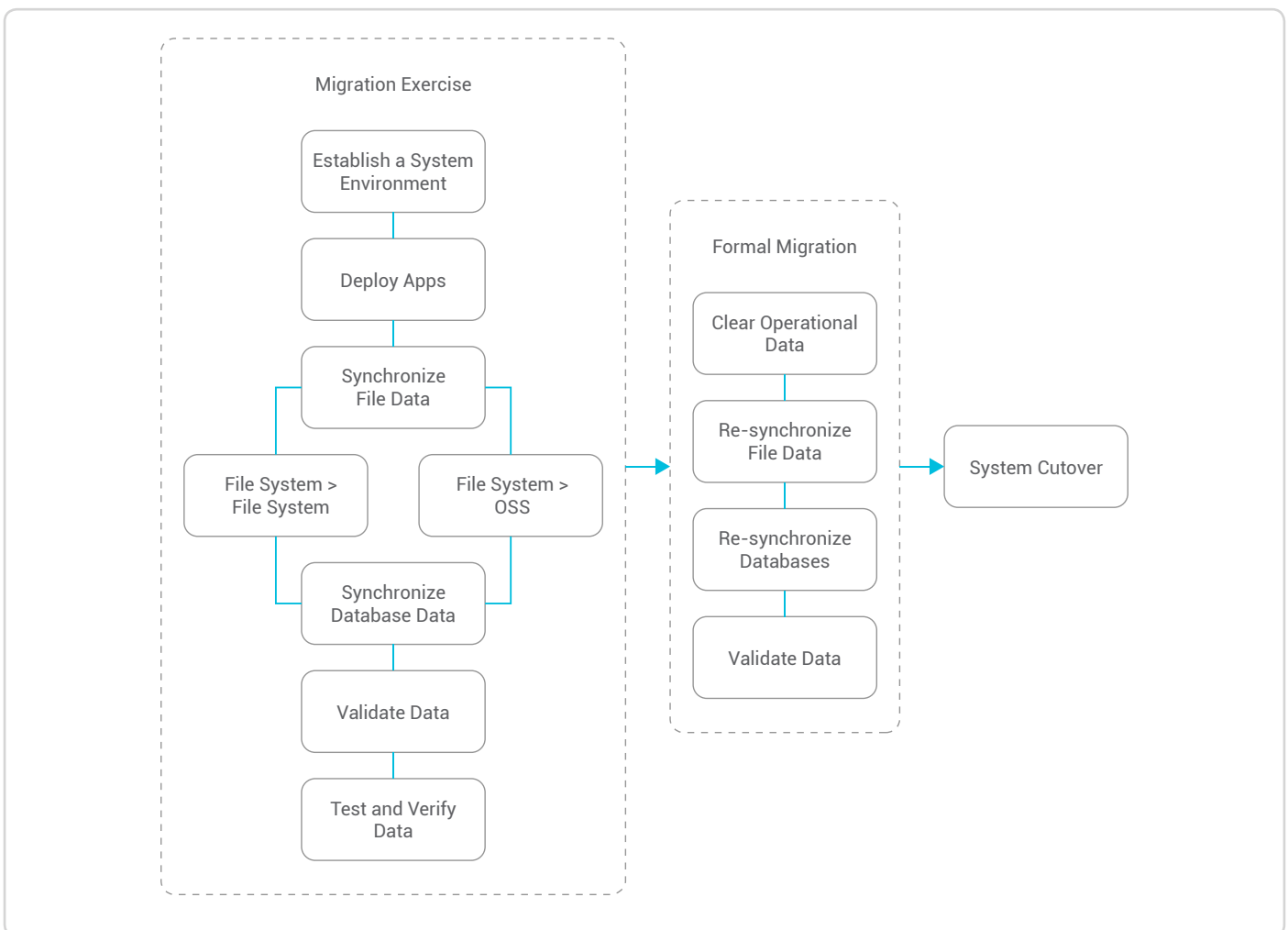


Figure 4. Data Migration System Cutover

- **System Environment Preparation:** The application of cloud product resources required for the application system, preparation of the environment, and preparation of the data migration tools will be completed according to the system requirements.
- **Application Deployment:** Deploy the application to the cloud platform after the application passes the functional and performance tests, according to the application deployment scheme.
- **File/Database Synchronization:** Migrate the transformed database designs, as well as the inventory data and incremental data of the existing application systems, to the cloud platform, and verify the data on the old and new platforms to ensure the correctness of the data on the cloud platform.
- **Business Cutover:** Complete the cutover, verification, and traffic switch for the application system to the cloud platform based on the business cutover scheme after the business cutover schedule has been determined.
- **Rollback Mechanism:** Each system should have a rollback scheme in place, including the application rollback and the database rollback.

8.8 System Delivery

This phase refers to the formal entry of the system into operation, and the later O&M phases after the application system completes system cutover and traffic is successfully switched to the cloud platform.

09 Alibaba Migration Platform and Related Solutions

At Alibaba Cloud, data migration is a service that we take great pride in providing. With our quality service and highly skilled migration team, Alibaba Cloud is well suited to taking care of all your data migration needs.

The AMP (Alibaba Migration Platform) supports data migration between Oracle, MySQL, DRDS (Alibaba Cloud distributed database), and PGAS (PostgreSQL advanced server) databases. In addition to providing data migration, AMP can also help users to migrate structure objects, and provides a consistency check functionality to verify the correctness of migrated data. Businesses can easily migrate data from the source database to the target database, as long as they have configured the connection information of the database and objects intended for migration, and start the task in the AMP console. One can view the migration task status and progress in the AMP console at any time, and stop or delete the migration task as needed once the task is in progress.

AMP provides rich and flexible personalized configuration for migration. It can support a variety of data migration demands of users. Specific features are as follows:

- Support multiple migration granularities. This allows businesses the option to migrate an instance, a database, a table, or a column.
- Facilitate renaming the migration objects, that is, the database name, table name, and column name in the source and target databases can be inconsistent.
- Support inconsistent character sets between the source and target databases. For example, the character set of the source database is GBK, and that of the target database can be UTF-8. However, when the character set of the target database is a subset of that in the source database, it cannot fully guarantee data correctness.
- Full migration that supports migrating part of the data in a table to the target database. You can configure the WHERE condition of a column in the table, and only the part of data that meets the WHERE condition will migrate to the target database.

10 Conclusion

Data migration is undoubtedly a crucial IT process for organizations. Rather than being the small part of a larger project deliverable with less business attention, data migration needs to be thoroughly planned and effectively implemented.

With our vast expertise in data migration, Alibaba Cloud is your ideal data migration partner. To learn more, talk to our experts today.

References

1. <https://www.edq.com/uk/blog/8-hurdles-of-a-data-migration/>
2. <https://www.glowtouch.com/blog/cloud/what-are-the-biggest-challenges-that-occur-with-cloud-migration/>
3. <http://www.techrepublic.com/blog/it-consultant/cloud-computing-migration-issues-what-you-need-to-know/>
4. <http://blog.altran.es/telecomunicaciones-media/five-data-migration-challenges/>
5. <https://blog.rackspace.com/9-considerations-for-a-successful-migration>
6. <http://www.techpageone.co.uk/technology-uk-en/five-common-trends-infrastructure-migration/>
7. <http://www.platformmodernization.org/transvive/Lists/ResearchPapers/DispForm.aspx?ID=1>
8. <http://dataagilitygroup.com/blog/most-data-migration-projects-bust-both-budgets-and-schedules/>

