This study's object is the cloud migration process of information systems (ISs). This paper aims to resolve the task of devising a quantitatively grounded classification of migration strategies while previous approaches relied on conceptual models without empirical validation of industry-specific performance metrics.

Unlike existing categorizations, the proposed approach employs empirical data from 275 successful cloud migration cases, considering cost reduction, performance improvement, migration duration, as well as the number of cloud services used. Missing values are handled by multiple imputations via chained equations (MICE); outliers were removed using the interquartile range criterion, thereby enhancing result reliability. A taxonomy of three strategies – Lift-and-Shift, Re-platforming, and Reengineering – was established.

Quantitative results indicate that Lift-and-Shift was applied in 39.64% of cases with an average cycle of 5.94 months and cost reduction of 40.06%; Re-platforming in 38.55% of cases with 6.10 months and 38.12% cost savings; Reengineering in 21.82% with 6.28 months, 42% cost savings, and 141.66% performance gain. Further analysis revealed an industry dependence in strategy selection: Lift-and-Shift predominated in regulated sectors, whereas Re-platforming and Reengineering were preferred in high tech industries.

The findings could underpin automated decision support systems for planning cloud migration of IS at medium and large enterprises. The quantitative models enable forecasting of temporal and financial indicators based on system scale, technological landscape, and regulatory requirements. Implementation requires acquisition of performance and cost metrics and integration of MICE and outlier detection into pre-migration audits

Keywords: imputation, information systems, quantitative analysis, classification, strategic planning, cloud migration

UDC 004.03; 005.33

DOI: 10.15587/1729-4061.2025.337851

IDENTIFYING THE INDUSTRY-SPECIFIC QUANTITATIVE INDICATORS FOR CLOUD MIGRATION STRATEGY OUTCOMES

Viktor Shutko
Corresponding author
PhD Student*
E-mail: viktor.shutko@nure.ua

Maksym Ievlanov
Doctor of Technical Sciences, Professor*

Ivan Iuriev
PhD, Associate Professor*

*Department of Information Control System
Kharkiv National University of Radio Electronics
Nauky ave., 14, Kharkiv, Ukraine, 61166

Received 02.06.2025
Received in revised form 21.07.2025
Accepted 18.08.2025
Published 29.08.2025

How to Cite: *Shutko, V., Ievlanov, M., Iuriev, I. (2025). Identifying the industry-specific quantitative indicators for cloud migration strategy outcomes. Eastern-European Journal of Enterprise Technologies, 4 (2 (136)), 23–34. https://doi.org/10.15587/1729-4061.2025.337851*

1. Introduction

Cloud migration is described as a structured strategic process of transferring corporate information systems (ISs) from traditional on-premises environments to cloud platforms, aimed at potentially increasing operational efficiency, reducing costs, and strengthening the competitive position of organizations in the market. Careful selection of the optimal migration strategy is a key factor that can significantly affect the achievement of desired results. With proper planning and execution of cloud migration, companies were able to reduce project costs by an average of 40% and reduce system downtime by up to 45%. In the best case scenario, companies achieved a reduction in total cost of ownership by 30–50%. According to some sources, the practical benefit was manifested in a reduction in time to market for new products and services by 45–65%. Against the background of an average increase in operational efficiency by 40% and employee productivity by approximately 35%, development cycles were accelerated by up to 60%. As a result, in the best case scenario, companies that have successfully migrated demonstrate revenue growth that can be 2.5 times higher compared to companies that have not implemented similar measures [1].

The relevance of the topic of strategic planning for cloud migration is due to several factors. First, there is a large body of research on this topic that demonstrate a significant gap between conceptual approaches to strategic planning for IS migration. This may indicate the absence of a single approach for, potentially, contextual, regulatory, corporate, and organizational reasons. Second, studies on this topic indicate that the migration procedure is inextricably linked to a number of security risks. For example, since 2020, the total cost of losses caused by data "leaks" has increased by 15.3% – to USD 4.45 million. The importance of predicting such risks and designing mechanisms for their prevention and minimization in cloud IS migration is almost beyond doubt. Third, the technical complexity of the migration process can lead to undesirable delays (4.5 months in some cases) and budget overruns (indicators reach 2–3% of the annual revenue of the enterprise). This may be due to the industry specificity of IS and requires special tools for planning the stages of IS migration, which forces enterprises to allocate additional resources for the development or adaptation of such tools. Fourth, financial management in the cloud environment is an important planning stage since the absence of the latter leads to an increase in the total cost of ownership (up to 20–30% in the first year) and reduces long-term benefits [1]. Thus, the cloud migration process, in particular its strategic planning, is relevant from an applied point of view since it includes a number of factors that must be taken into account. The diversity and multifaceted nature of these factors allow us to conclude that further research into this area is appropriate.

2. Literature review and problem statement

Currently, there is no single standardized classification of IS cloud migration strategies. Instead, there are several studies aimed at both devising new strategies and modifying existing classifications of IS migration strategies. Thus, in study [2], a categorization of IS migration strategies to the cloud environment was presented according to service delivery models: migration to IaaS (infrastructure as a service), PaaS (platform as a service), and SaaS (software as a service). The latter strategy was divided into three approaches: complete replacement of the system with a cloud solution (replacement), partial replacement of individual functional components (revising), and complete reengineering of the system for cloud architecture (reengineering). This classification allowed the authors to clearly distinguish the level of complexity and the number of necessary changes in IS depending on the chosen migration strategy. However, its disadvantage was the lack of clear criteria for choosing a specific strategy for individual cases, which made it difficult to choose the optimal IS migration strategy in accordance with the needs of the organization and the features of IS implementation.

In [3], an approach was proposed that involved implementing an artificial intelligence (AI)-based superstructure on top of standard cloud IS migration strategies (Rehosting, Refactoring, Rearchitecting, Rebuilding, Retiring). A framework was developed that used machine learning tools to automatically analyze the application portfolio, categorize their suitability for different migration strategies, and also predict resource needs and optimize the planning of migration "waves". However, the practical application of the proposed solution was complicated by the need for high-quality input data and significant requirements for personnel competence in working with AI tools. The lack of adaptation to non-standard architectures and migration scenarios could limit its applied effectiveness.

Work [4] described a classification of IS cloud migration strategies, which involved three common approaches. These included Lift-and-Shift (migration without significant changes), Re-platforming (partial modification using individual cloud services), and Refactoring (complete redesign of applications for cloud technologies). Each of these approaches had its own characteristics in terms of cost, migration speed, level of complexity and resource optimization. The main disadvantage of this approach was its focus on a general comparative assessment. The proposed solution did not contain clear instructions for individual selection of a strategy under the actual conditions of specific organizations. In addition, the complexity and need for significant resources for analysis and implementation made this categorization unsuitable for use without additional adaptations and detailing in specific organizational contexts.

In study [5], the so-called Holistic Cloud Decision Model was proposed, which provided four consecutive stages of decision-making on the feasibility of implementing cloud services in an organization. The first stage of the model consisted of defining and detailing the functional and non-functional requirements of the organization, including software and hardware specifications, business goals, availability levels, scalability, and reliability. The second stage involved a comprehensive cost assessment that included a comparison of capital (CAPEX) and operating expenses (OPEX) for various deployment scenarios. The assessment was performed using key financial indicators such as NPV (net present value), ROI (return on investment), and BCR (benefit/cost ratio). The third stage of the model aimed at estimating the cost of implementing a cloud infrastructure, where issues of information security, compliance with user requirements, flexibility, risks, and the level of dependence on third-party providers were analyzed. The fourth stage involved analyzing the changes in the organizational structure and processes that accompanied the transition to the cloud, including changes in accounting, distribution of responsibilities, and risk management. The model allowed for a systematic comparison of cloud technology implementation alternatives in terms of both quantitative and qualitative indicators, using expert weighting of factors applying Delphi procedures and the Analytical Hierarchy Process (AHP) method. However, the practical application of the model was complicated by the need to obtain detailed initial data on business requirements and costs, the high labor intensity of the information collection process, as well as significant dependence on the competence and consistency of expert assessments. There were also difficulties in scaling and adapting the model for organizations with atypical or changing business processes, and the subjective factor could negatively affect the objectivity and reproducibility of final decisions.

In [6], a classification of IS cloud migration strategies was introduced based on the k-means clustering method, where three main groups of strategies were distinguished: full migration, replacement, and component-based migration. Full migration involved moving the entire IS or its main components as a single unit to the cloud. Replacement involved a significant restructuring of the IS or its parts with maximum adaptation to the cloud environment, for example, through refactoring or redesign strategies. Component migration involved partial migration of only certain important IS elements or components. However, the main disadvantages of this approach were its conceptual nature and the lack of specific criteria and quantitative metrics that would allow organizations to evaluate and compare different strategies in the context of their own needs. This complicated the practical application of the categorization for choosing the optimal IS cloud migration strategy since no clear recommendations were defined for choosing the appropriate strategy depending on contextual conditions and business requirements.

As part of study [7], an approach to cloud migration planning was proposed, based on the integration of AI methods to reduce the cost of cloud solutions and increase the efficiency of their implementation. A set of AI algorithms was applied, including machine learning methods, in particular reinforcement learning, clustering, and natural language processing. These algorithms automated the assessment of IS readiness for migration and forecasting resource costs. AI also optimized the use of application programming interfaces (APIs) and reduced the time for implementing strategies by ensuring automatic configuration of the cloud environment infrastructure. The study presented detailed pseudocode and scenarios for implementing strategies that had a certain practical value for IT specialists. However, the practical implementation of the proposed solution was complicated by significant requirements for the quality and volume of input data, the complexity of using AI tools with the existing infrastructure. Risks associated with security, interoperability, and data protection in multi-component and hybrid cloud solutions have also made it difficult to use such an approach.

Finally, in [8], a structured methodology for planning and executing migration to S/4HANA Cloud was proposed, combining a choice of three migration strategies. These include Greenfield (implementation with process standardization and reduction of technical debt), Brownfield (migration with preservation of existing configurations), and Hybrid (selective preservation of critical infrastructure with parallel modernization). This allowed the authors to align migration actions with business goals, accelerate deployment, improve the quality of data management, and reduce risks while minimizing operational interruptions. At the same time, a number of unresolved issues were identified, including potential bias on the part of developers, the need for data quality control, and the risk of recommendations quickly losing relevance in the absence of their regular revision. The limited generalizability of the results, due to the industry-specific nature of the sample, may not be universal for all cloud providers.

Based on our review of current literature addressing the classification of IS cloud migration strategies, several general conclusions were drawn. First, analysis revealed that most works offer different approaches to defining, forming, and categorizing strategies. Second, none of the existing approaches has come close to the status of a universal or standardized tool for ensuring the possibility of making decisions on choosing a migration strategy under real organizational conditions. Third, existing approaches do not either contain clear criteria and algorithms for choosing a strategy for specific requirements and characteristics of IS and business, or remain too general and conceptual, complicating practical use. More modern approaches using AI significantly increase the objectivity and speed of preliminary assessment, reduce the influence of the human factor, automate application portfolio analysis and resource forecasting. However, their effectiveness is limited by the need for large volumes of high-quality data, high requirements for personnel expertise, as well as the complexity of integration and adaptation to a specific organizational architecture and IS usage scenarios. The central problem among all is the lack of a single process model, which complicates the assessment and justification of strategic planning and negatively affects the efficiency of the entire migration process. Solving this problem would allow organizations to reduce the time and resources spent on planning and executing migration, as well as minimize the risks of data loss. In addition, it could improve operational performance and strengthen competitiveness through faster implementation of new services [9].

The reasons for the problem remaining unresolved may be the conceptual and methodological inconsistency of approaches in scientific and applied literature. Another reason may be the lack of agreed terms and structural components of the process. The difficulty of unifying practices that have arisen in different organizational and technological contexts also contributes to the persistence of this problem. Such a situation complicates the coordination of assessment criteria, comparison of alternatives and integration of tools to support decision-making at the strategic planning stage. It is in this context that the problem of our study is stated – the lack of a single substantiated classification of the cloud migration strategy of IS.

The results of our review of related sources showed that the state of the problem solution currently remains unsatisfactory; therefore, to solve it, it was advisable to conduct a study of the general issue – the lack of a single categorization of cloud migration strategies of IS. As the shortcomings were analyzed and considered, the research problem was stated as the lack of a quantitatively substantiated classification of cloud migration strategies of IS.

3. The aim and objectives of the study

The purpose of our study is to identify quantitative industry-specific indicators of the results of implementing typical cloud IS migration strategies, namely Lift-and-Shift, Re-platforming, and Reengineering (Re-architecting). Particular attention is paid to determining the most frequently used strategies for each type of business and the corresponding results of their implementation. This will make it possible to assess the results of implemented migration strategies in the relevant industry and compile practical recommendations for choosing the optimal migration strategy for enterprises of the relevant profile, taking into account industry requirements.

To achieve this aim, the following objectives were accomplished:

– to form a correct empirical basis for further research;

– to assess the industry distribution and effectiveness of the results from implementing cloud migration strategies.

4. The study materials and methods

The object of our study is the process of cloud migration of IS. The subject of the study is cloud migration strategies of IS.

The main hypothesis of the study assumes that the descriptions of existing cases of successful cloud migration are complete enough to formulate general conclusions based on their analysis, taking into account the specificity of each migration strategy and allowing us to determine the corresponding quantitative indicators.

To collect data, an analysis was conducted of 339 successful informal cases of migration to the cloud, which were publicly available on the information resources by the most known cloud providers, namely Amazon Web Services, Google Cloud Platform, Microsoft Azure. The order of case analysis was chosen arbitrary.

It should be noted that since the descriptions were informal, the set of metrics was formed as a set of the most frequently mentioned and publicly available metrics among all the studied cases. The main metrics in the dataset were Company name, Company industry, Target cloud, and Migration strategy. Cost reduction, Performance improvement, Number of cloud services used, Migration time, and Source link were also taken into account. Detailed information about the metrics is provided in Table 1.

It should be noted separately that the most widely used conceptual classification model was used to categorize cases by the "Migration Strategy" metric. This model was based on the scope of changes and the depth of integration with cloud services and included Lift-and-Shift, Re-platforming, and Reengineering strategies [4]. The classification was based on the following rules:

1. If the case described a direct migration of existing IS components with minimal or no changes, then the case was categorized as a Lift-and-Shift strategy.

2. If the case description included significant changes to existing IS components or a complete rewrite of the IS from scratch, then the case was categorized as Reengineering.

3. If the case could not be unambiguously categorized as one of the previous strategies (or this was directly stated in the description), then the case was categorized as Re-platforming.

The generated original dataset can be found as an appendix to this study (dataset.xlsx).

Table 1

Detailed information about the metrics of the original dataset

No.	Metric ID	Metric description	Limitation
1	Company name	The actual name of the company specified in the case in question	Valid company name of the enterprise
2	Company industry	Branch of economic activity in which the company carries out its main activity	It is limited to one of the following industries: Agriculture, Big Data, Cryptocurrency, E-commerce, Electronics, Financial Services, Financial Technologies, Government & Public Sector, Healthcare, Hospitality, IT-Services, Insurance, Logistics, Manufacturing, Media & Entertainment, Real Estate, Retail, Software Engineering, Technology, Telecommunications, other
3	Target cloud	Target cloud platform	Limited to one of the following cloud service providers: Amazon Web Services, Google Cloud Platform, Azure, Google Workspace, Multi-cloud
4	Migration strategy	Implemented strategy for migrating to the cloud platform	It is limited to one of the following definitions: Lift-and-Shift, Re-platforming, Reengineering
5	Cost reduction	A quantitative metric that reflects a decrease in the cost of maintaining an information security infrastructure or a decrease in the operating computing costs of the enterprise	Decimal value as a percentage, greater than or equal to 0
6	Performance improvement	Quantitative metric reflecting the growth of the operational productivity of the enterprise's IS after migration to the cloud	Decimal value as a percentage, greater than or equal to 0
7	Number of cloud services used	A quantitative metric that displays the number of cloud services that were used to perform migration to the cloud	Integer, greater than 0
8	Migration time	A quantitative metric that displays the execution time of a migration to the cloud	Decimal value to two decimal places, greater than 0
9	Source link	Link to the specific case from which the metrics were taken	Valid URL

After preliminary data analysis, 60 records with an empty value for the Migration strategy metric were removed from the dataset. Records containing empty values in quantitative indicators were allowed to be analyzed with the available values but were not taken into account when calculating the corresponding missing parameters.

To perform the analysis of the presented dataset, the Python 3.13.3 programming language (Netherlands) was used, involving specialized modules – pandas, sklearn, and matplotlib.

The pandas module for processing and analyzing data in Python was used to obtain convenient data structures and functions for working with tabular information. The main object used was a DataFrame (a two-dimensional table with row and column labels).

From the sklearn machine learning module, a set of effective tools for classification, regression, clustering, dimensionality reduction, and model estimation was used.

Matplotlib, a basic Python data visualization module, allowed for the construction of a variety of charts, as well as flexible customization of the appearance of plots and integration with another Python module used, pandas [10].

Data analysis was performed on a personal computer running Windows 11 (United States) Enterprise version 10.0.26100 Build 2610 with the following hardware:

– CPU – Intel(R) Core (TM) Ultra 7 165H, 1400 MHz, 16 Core(s), 22 Logical Processor(s);

– RAM 32 Gb DDR5.

To improve the accuracy and reliability of analysis, the processing of missing values in the data set was performed using the multiple imputation (MI) method, in particular its implementation based on chain equations (Multiple Imputation by Chained Equations (MICE) [11], which were described by the following formula

$$Y_{j,mis}^{(t+1)} \sim P\left(Y_j \mid Y_{-j}, \theta_j\right), \quad (1)$$

where $Y_{j,mis}^{(t+1)}$ are the imputed values for variable j at $t+1$ iterations, Y_{-j} are all other variables, θ_j are the parameters of the imputation model for Y_j. P is the conditional probability distribution of variable Y_j. The corresponding Python implementation for MICE was IterativeImputer from the sklearn module.

The justification for using this method was the following assumption: the missing values in the data set arose due to the non-disclosure policy of the individual organizations whose successful migration cases were considered. Given the lack of reason to believe that the fact of omission depends on the unknown values of the variables under study, such omissions were categorized as random (Missing at Random, MAR). The use of this approach allowed us to take into account the uncertainty caused by missing data, so the results obtained were interpreted as more reliable and less biased. This allowed us to more accurately reflect the existing patterns in the data set.

In addition, to ensure statistical stability and reliability of further analysis, outliers were identified among four key numerical metrics (Cost Reduction, Performance Improvement, Number of Cloud Services Used, and Migration Time). For this purpose, a classic non-parametric approach based on the interquartile range (IQR) was used, which is recommended for reliable detection of outliers in distributions with potential asymmetry [12]. Here, a quartile is a statistical value that diǃ vides an ordered data set into four equal parts, each containing a quarter of the observations. The first quartile (Q_1) represents the value below which 25% of the data fall, while the third quartile (Q_3) marks the point below which 75% of the data fall.

For each variable Q_1 and Q_3, quartiles and the corresponding interquartile ranges were calculated

$$IQR = Q_3 - Q_1.$$

All values that were outside the following range were considered anomalous

$$Q_1 - 1.5 \cdot IQR;\ Q_3 + 1.5 \cdot IQR. \quad (2)$$

All anomalous observations were subsequently withdrawn from consideration.

5. Results of comparative analysis of quantitative metrics regarding the results of implementing typical cloud migration strategies

5. 1. Results of forming the empirical base

A corresponding Python script was written to preview the first 10 rows of the dataset. At the stage of loading data from the CSV file, parameters were set for encoding 'cp1252' and identification of column headers through header = 0. After loading the dataset, the following informative columns were selected: "Company Name", "Company industry", "Migration Strategy", "Cost Reduction, %", "Performance Improvement, %", "Number of Cloud Services Used", "Migration time, months". The remaining columns were discarded because they did not contain relevant quantitative information within the framework of this study. The selected data were represented using the head (10) function, which allowed us to focus the representation on significant attributes without displaying auxiliary information about the target cloud and the source of the link. Fig. 1 shows the code of the first Python script. The result is given in Table 2.

The original dataset contained missing values in the key information columns "Cost reduction, %", "Performance improvement, %", and "Migration time, months". This significantly limited the possibilities of further analysis and drawing conclusions about the effectiveness of the selected migration strategies. Therefore, it was decided to process the corresponding missing values.

To process missing values in the dataset, the IterativeImputer object from the sklearn module was used, the principle of which was described by formula (1). Table 3 gives the results of processing missing values for the first 10 rows of the dataset. Fig. 2 shows the corresponding Python program code.

```
import pandas as pd
import os

file_path = r'C:\Users\Viktor_Shutko\Desktop\dataset.csv'

df = pd.read_csv(file_path, encoding='cp1252', header=0, low_memory=False)

selected_columns = [
    'Company Name',
    'Company industry',
    'Migration Strategy',
    'Cost Reduction, %',
    'Performance Improvement, %',
    'Number of Cloud Services Used',
    'Migration time, months'
]

display(df[selected_columns].head(10))
```

Fig. 1. Script for reviewing a dataset

Table 2

Overview of dataset.xlsx

Company name	Industry	Migration strategy used	Cost reduction, %	Performance improvement, %	Number of cloud services used	Migration time, months
Capital Express	Manufacturing	Re-platforming	94.0	500.0	9.0	1.0
OneMain Financial	Financial Services	Reengineering	98.8	97.5	5.0	2.0
Altruist	Financial Technology	Reengineering	20.0	NaN	3.0	NaN
M1 Finance	Financial Services	Re-platforming	90.0	NaN	5.0	4.0
Venmo	Financial Services	Re-platforming	90.0	NaN	1.0	NaN
Coinbase	Cryptocurrency	Re-platforming	62.0	50.0	2.0	NaN
ElysianNxt	Financial Technology	Reengineering	NaN	96.0	3.0	NaN
Discover Financial Services	Financial Services	Re-platforming	35.0	NaN	4.0	NaN
Danske Bank	Financial Services	Re-platforming	50.0	50.0	3.0	NaN
Zeta	Financial Services	Reengineering	NaN	NaN	4.0	NaN

Table 3

Result of processing missing values

Company name	Industry	Migration strategy used	Cost reduction, %	Performance improvement, %	Number of cloud services used	Migration time, months
Capital Express	Manufacturing industry	Re-platforming	94.00	500.00	9.0	1.00
OneMain Financial	Financial services	Reengineering	98.80	97.500	5.0	2.00
Altruist	Financial technologies	Reengineering	20.00	0.00	3.0	2.03
M1 Finance	Financial services	Re-platforming	90.00	823.17	5.0	4.00
Venmo	Financial services	Re-platforming	90.00	692.47	1.0	10.52
Coinbase	Cryptocurrency	Re-platforming	62.00	50.00	2.0	18.35
ElysianNxt	Financial technologies	Reengineering	40.10	96.00	3.0	7.55
Discover Financial Services	Financial services	Re-platforming	35.00	41.78	4.0	6.92
Danske Bank	Financial services	Re-platforming	50.00	50.00	3.0	13.15
Zeta	Financial services	Reengineering	39.41	87.49	4.0	7.38

```
from sklearn.experimental import enable_iterative_imputer
from sklearn.impute import IterativeImputer

numeric_cols = [
    'Cost Reduction, %',
    'Performance Improvement, %',
    'Number of Cloud Services Used',
    'Migration time, months'
]

imputer = IterativeImputer(random_state=0, max_iter=50, min_value=0)
df[numeric_cols] = pd.DataFrame(imputer.fit_transform(df[numeric_cols]),
                                columns=numeric_cols)

selected_columns = [
    'Company Name', 'Company industry', 'Migration Strategy',
    'Cost Reduction, %', 'Performance Improvement, %',
    'Number of Cloud Services Used', 'Migration time, months'
]

display(df[selected_columns].head(10))
```

Fig. 2. Script for handling missing values

The next step for preprocessing the datasets was to perform anomaly detection on the numeric columns. The corresponding Python script for anomaly detection using the approach described by formula (2) is shown in Fig. 3. The results of removing the detected anomalies for the metrics Cost Reduction, Performance Improvement, Number of Cloud Services Used, and Migration time are shown in Fig. 4–7, respectively.

Thus, the outliers in all variables were not random and reflected either extreme values in real migration scenarios or possible methodological errors that required separate control. Isolating such outliers was crucial to forming a reliable data set for further analysis.

The Python script to remove all outliers is shown in Fig. 8.

Thus, after pre-processing the dataset, namely, handling missing values, detecting, and removing outliers, a cleaned dataset was generated to perform the next research task. In this case, the number of records in the cleaned dataset decreased from 339 to 275.

```
numeric_cols = [
    'Cost Reduction, %',
    'Performance Improvement, %',
    'Number of Cloud Services Used',
    'Migration time, months'
]
def plot_iqr_anomalies(df, column):
    q1 = df[column].quantile(0.25)
    q3 = df[column].quantile(0.75)
    iqr = q3 - q1
    lower_bound = max(0, q1 - 1.5 * iqr)
    upper_bound = max(0, q3 + 1.5 * iqr)

    outliers = df[(df[column] < lower_bound) | (df[column] > upper_bound)][column]

    plt.figure(figsize=(10, 5))
    sns.boxplot(x=df[column], orient='h')
    plt.axvline(x=lower_bound, color='blue', linestyle='--', label=f'Lower Bound = {lower_bound:.2f}')
    plt.axvline(x=upper_bound, color='red', linestyle='--', label=f'Upper Bound = {upper_bound:.2f}')
    plt.title(f'IQR Boxplot: {column} (Anomalies count: {outliers.count()})')
    plt.xlabel(column)
    plt.legend()
    plt.grid(True)
    plt.tight_layout()
    plt.show()
for col in numeric_cols:
    plot_iqr_anomalies(df, col)
```

Fig. 3. Python script for detecting outliers

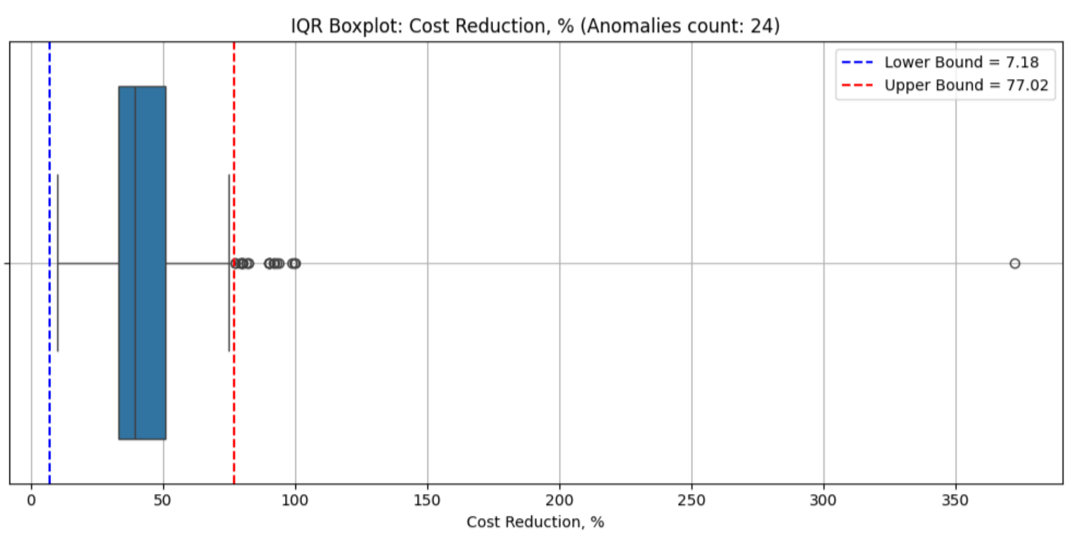

Fig. 4. Outliers for the Cost Reduction metric

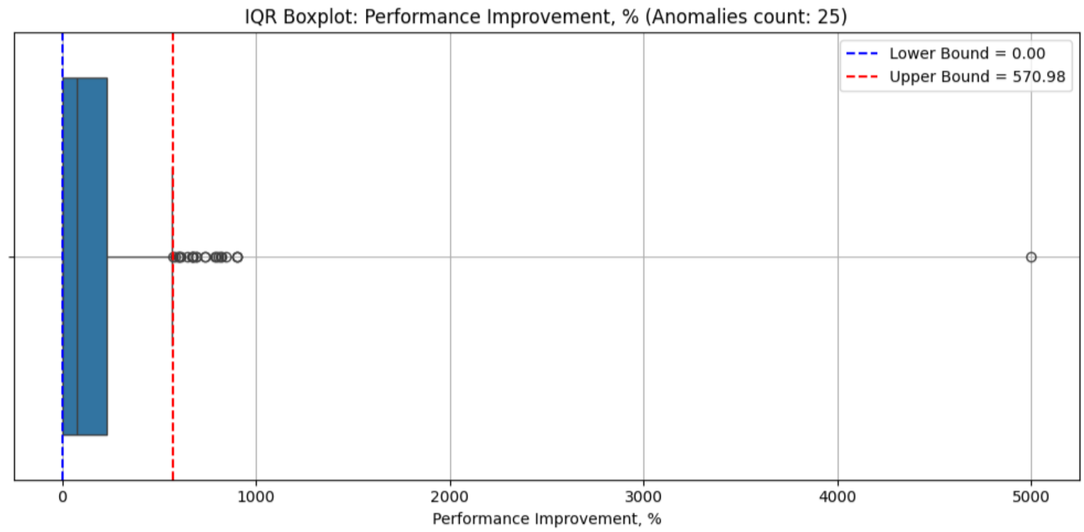

Fig. 5. Outliers for the Performance Improvement metric

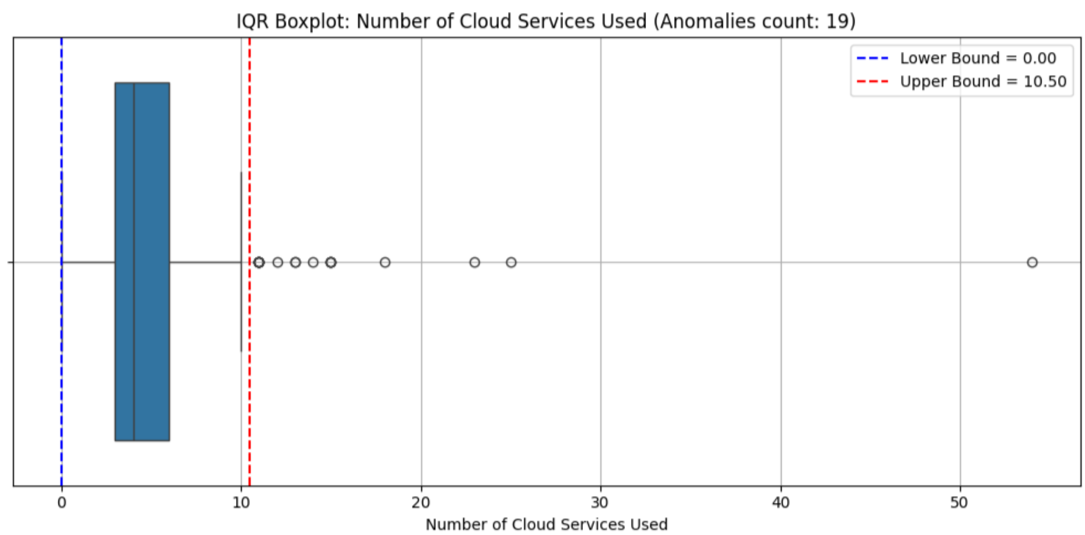

Fig. 6. Outliers for the Number of Cloud Services Used metric

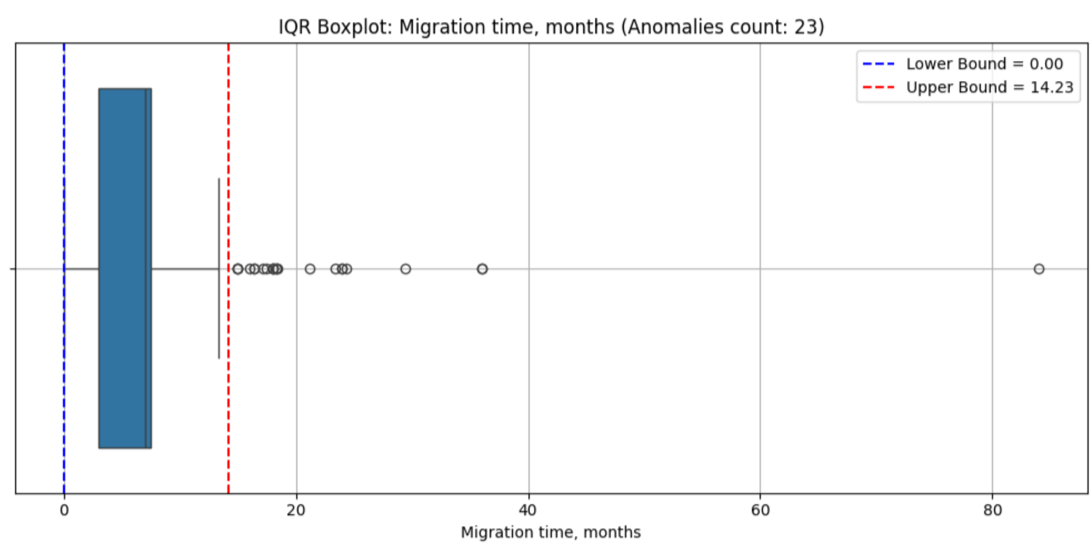

Fig. 7. Outliers for the Migration time metric

```
numeric_cols = [
    'Cost Reduction, %',
    'Performance Improvement, %',
    'Number of Cloud Services Used',
    'Migration time, months'
]
mask = pd.Series(True, index=df.index)

for col in numeric_cols:
    q1 = df[col].quantile(0.25)
    q3 = df[col].quantile(0.75)
    iqr = q3 - q1
    lower = q1 - 1.5 * iqr
    upper = q3 + 1.5 * iqr
    col_mask = df[col].isna() | ((df[col] >= lower) & (df[col] <= upper))
    mask &= col_mask

df_iqr_cleaned = df[mask].reset_index(drop=True)

display(df_iqr_cleaned[numeric_cols])
```

Fig. 8. Python script to remove detected outliers

5. 2. Results of assessing the industry distribution and effectiveness of the results from implementing cloud migration strategies

First, it was determined what share in the total number of cases each migration strategy accounted for. This allowed us to objectively assess the prevalence of different approaches in the analyzed data set and provided a general picture of the distribution of migration strategies in all cases. Fig. 9 shows the corresponding Python script. The results are given in Table 4.

Next, for each migration strategy, the average values of four key quantitative indicators were determined – Cost Reduction, Performance Improvement, Number of Cloud Services Used, Migration time. The corresponding Python script is shown in Fig. 10. The calculation results are given in Table 5.

```
import pandas as pd

df_iqr_cleaned['Migration Strategy'] = df_iqr_cleaned['Migration Strategy'].str.strip()

strategies = ['Lift-and-Shift', 'Re-platforming', 'Reengineering']
counts = df_iqr_cleaned['Migration Strategy'] \
    .value_counts() \
    .reindex(strategies, fill_value=0)

subtotal = counts.sum()
percentages = (counts / subtotal * 100).round(2)

summary = pd.DataFrame({
    'Number of cases': counts,
    '% of total': percentages
})
display(summary)
```

Fig. 9. Python script to determine the share of each strategy

Table 4

Share of each strategy in the total data set

Migration strategy	Number of observations	Share of total number of observations, %
Lift-and-Shift	109	39.64
Re-platforming	106	38.55
Reengineering	60	21.82

```
grouped = df_iqr_cleaned.groupby('Migration Strategy')[[
    'Cost Reduction, %',
    'Performance Improvement, %',
    'Number of Cloud Services Used',
    'Migration time, months'
]].mean().reset_index()

grouped.columns = [
    'Migration Strategy',
    'Avg Cost Reduction (%)',
    'Avg Performance Improvement (%)',
    'Avg # Cloud Services Used',
    'Avg Migration Time (months)'
]
cols_to_round = [
    'Avg Cost Reduction (%)',
    'Avg Performance Improvement (%)',
    'Avg Migration Time (months)'
]
grouped[cols_to_round] = grouped[cols_to_round].round(2)
grouped['Avg # Cloud Services Used'] = grouped['Avg # Cloud Services Used'].astype(int)

display(grouped)
```

Fig. 10. Python script for estimating average quantitative indicators of migration strategies

Table 5

Average indicators of migration strategies

Migration strategy	Average cost reduction rate (%)	Average performance improvement rate (%)	Average number of cloud services used	Average migration duration (months)
Lift-and-Shift	40.06	133.04	4	5.94
Re-platforming	38.18	101.17	4	6.19
Reengineering	42.00	141.66	4	6.28

Based on the preliminary results, it was concluded that it was impossible to unambiguously differentiate the quantitative indicators of the respective migration strategies based only on the average values obtained since the values turned out to be as close as possible to each other. That is why it was decided to determine the most common migration strategy in each industry and evaluate its average results.

Initially, the data were grouped by the combination of "Company industry" and "Migration Strategy" based on the entered parameters of frequency, average cost reduction, average performance improvement, average migration duration, and average number of cloud services used. Then, the relative share of each strategy in the corresponding industry was calculated. For each industry, the strategy with the highest frequency of use was selected, after which the results were represented in a standardized form with the number of cloud services used rounded to an integer. The results obtained are given in Table 6.

The corresponding Python script is shown in Fig. 11.

Thus, information was obtained on the dominant strategy in each of the considered industries, as well as the resulting average indicators of the results from implementing the corresponding migration strategy.

Table 6

Sectoral distribution of migration strategies and average indicators of their effectiveness

Industry	Migration strategy	Selection frequency, %	Average cost reduction, %	Average performance improvement, %	Average migration duration, ms	Number of cloud services used
Agriculture	1	100.00	50.00	197.00	8.48	4
Big Data	2	100.00	36.75	49.00	7.45	4
Cryptocurrency	2	100.00	45.12	173.84	7.27	3
E-commerce	1	50.00	24.31	55.00	3.00	3
Electronics	1	50.00	30.00	0.00	5.67	7
Financial Services	2	51.79	37.91	89.49	6.42	4
Financial Technology	3	47.06	44.46	176.27	6.76	3
Public Sector	1	75.00	34.72	48.79	5.33	8
Healthcare	1	45.00	37.49	82.43	6.11	5
Hospitality	1	66.67	35.32	65.00	5.16	5
IT Services	2	55.56	39.94	140.50	4.71	5
Insurance	1	44.44	35.97	49.71	6.95	4
Logistics	1	50.00	34.00	39.65	6.52	4
Manufacturing	2	50.00	36.53	75.51	6.05	4
Media & Entertainment	1	40.00	34.96	55.02	5.30	5
Other	1	46.51	41.02	170.43	5.81	4
Real Estate	1	83.33	50.75	239.69	7.48	4
Retail	1	60.00	44.15	143.11	8.00	3
Software Development	2	40.91	40.43	110.86	6.86	4
Technology	1	55.00	37.11	128.62	4.19	6
Telecommunications	1	75.00	47.47	237.81	5.49	4

```python
import pandas as pd

strategy_mapping = {
    'Lift-and-Shift': 1,
    'Re-platforming': 2,
    'Reengineering': 3
}

df_iqr_cleaned['Migration Strategy'] = df_iqr_cleaned['Migration Strategy'].replace(strategy_mapping)

agg = (df_iqr_cleaned.groupby(['Company industry', 'Migration Strategy'])
        .agg(
            freq=('Migration Strategy', 'size'),
            avg_cost=('Cost Reduction, %', 'mean'),
            avg_perf=('Performance Improvement, %', 'mean'),
            avg_time=('Migration time, months', 'mean'),
            avg_services=('Number of Cloud Services Used', 'mean')).reset_index())

industry_totals = (
    df_iqr_cleaned
    .groupby('Company industry')
    .size()
    .rename('total')
    .reset_index()
)
agg = agg.merge(industry_totals, on='Company industry')
agg['freq_pct'] = agg['freq'] / agg['total'] * 100
```

a

```python
idx = agg.groupby('Company industry')['freq'].idxmax()
top_strategies = (
    agg
    .loc[idx, ['Company industry','Migration Strategy','freq_pct',
        'avg_cost','avg_perf','avg_time','avg_services']]
    .rename(columns={
        'freq_pct': 'Pick rate, %',
        'avg_cost': 'Avg Cost Reduction, %',
        'avg_perf': 'Avg Performance Improvement, %',
        'avg_time': 'Avg Migration Time, months',
        'avg_services': 'Avg Cloud Services Used'
    }).reset_index(drop=True))

cols_to_round = [
    'Avg Cost Reduction, %',
    'Avg Performance Improvement, %',
    'Avg Migration Time, months',
    'Pick rate, %'
]
top_strategies[cols_to_round] = top_strategies[cols_to_round].round(2)
top_strategies['Avg Cloud Services Used'] =
top_strategies['Avg Cloud Services Used'].astype(int)
top_strategies['Migration Strategy'] =
top_strategies['Migration Strategy'].astype(int)

display(top_strategies)
```

b

Fig. 11. Python script for cross-industry assessment of average cloud migration strategy outcomes: *a* – data aggregation; *b* – identification of dominant strategies in each industry

6. Discussion of results based on the comparative analysis of quantitative metrics of the results from implementing cloud migration strategies

During the study, a systematic collection, normalization, and processing of empirical data from 339 successful cases of cloud migration of IS was performed (Fig. 1–8, Tables 2, 3). A cleaned sample was formed, which included 275 observations, for which quantitative indicators such as cost reduction, performance improvement, number of cloud services used, and migration duration were determined. The results show an almost even distribution between the Lift-and-Shift (39.64%) and Re-platforming (38.55%) strategies, reflecting the balance between the speed of implementation and the level of optimization (Fig. 9, Table 4). The Reengineering strategy was used less frequently (21.82%) but provided the highest average cost reduction (42%) and performance improvement (141.66%), indicating the potential for deep architectural transformation (Fig. 10, Table 5). This difference can be explained by the fact that Lift-and-Shift prefers industries with a high level of regulatory requirements and minimal tolerance for change. At the same time, Re-platforming and Reengineering are characteristic of technologically advanced sectors that demonstrate greater flexibility and innovation. Additionally, it was found that the average number of cloud services is the same for all strategies and, apparently, does not act as a differential feature, indicating the dominance of industry or corporate policy in the implementation of migration.

As a result of the cross-sector analysis of cloud migration strategies (Fig. 11, Table 6), a clear relationship was revealed between the chosen strategy and the industry affiliation of the company. This allowed us to formulate a reasoned position on the contextual feasibility of applying each of the approaches, substantiated by empirical data.

In industries with a high degree of regulation, requirements for service continuity and inherited architectural inertia – such as Healthcare (45%), Insurance (44.44%), Government & Public Sector (75%), Logistics (50%) and Telecommunications (75%) – the Lift-and-Shift strategy was preferred (Table 6). In these cases, companies chose a minimal change approach to ensure guaranteed access to critical data, minimize the risk of losing regulatory compliance, and avoid full-scale intervention in the existing infrastructure. At the same time, technologically advanced industries, such as cryptocurrency, IT-Services, Software Engineering, Financial Technologies, and Big Data, demonstrated a clear advantage over Re-platforming or Reengineering strategies.

In particular, in the cryptocurrency sector, 100% of companies implemented Re-platforming, achieving an average performance improvement of 173.84% with cost savings of 45.12%. In the financial technology sector, the Reengineering strategy dominated (47.06%), which allowed it to achieve an average performance improvement of 176.27% (Table 6). The use of these strategies was logical for industries with a high rate of innovation development, where a deep architectural restructuring ensured adaptability, scalability, and technical synergy with the chosen cloud platform.

An interesting pattern was also revealed by industries that were in the transitional phase of technological adaptation – retail (Retail), real estate (Real Estate), technology (Technology). Despite the dominance of the Lift-and-Shift strategy in these sectors, companies achieved significant performance improvement. For example, the real estate industry demonstrated the highest average performance improvement among all analyzed categories (239.69%) with the dominance of the Lift-and-Shift strategy (83.33%) (Table 6). This result suggested that even a relatively technologically simple migration strategy can provide a significant effect in the case of a favorable initial architecture or when transitioning from fragmented on-premises solutions to a single cloud model.

Our results are attributed to the structure of the dataset and the industry context of migrations. The dominance of Lift-and-Shift in sectors with strict regulation and increased requirements for continuity is explained by low tolerance for architectural changes and the need to quickly transfer loads with minimal risk. In contrast, higher average indicators for Reengineering in technologically dynamic industries are quite expected. The management of such enterprises usually recognizes the importance of implementing modern information technologies and is ready to invest appropriate resources in order to obtain long-term benefits. In addition, the normalization and standardization of the initial case descriptions reduced variability and made visible inter-sectoral differences that may remain unnoticed in unsystematized reports. Unlike the classification proposed in [2], where migration strategies were differentiated by IaaS, PaaS, and SaaS service models without quantitative selection criteria, the proposed approach formalizes the differences between strategies. This was made possible by calculating quantitative cross-sectoral indicators, as well as by using MICE imputation to handle missing values and remove outliers to increase the reliability of the calculations. At the same time, unlike the AI-based framework in [3], which required significant amounts of data and highly skilled machine learning specialists, the proposed method verifies key indicators directly on empirical cases. This is made possible by using standardized pre-processing procedures and analytical Python scripts. The difference from the formal mathematical models of comparative assessment in [4] is the use of not only financial metrics, but also performance metrics and migration duration. This is made possible by cross-sectoral analysis of trends and the identification of dominant strategies for each industry, which provides practical recommendations for specific business conditions. In addition, unlike the approach in [5], which relied on subjective weighting through AHP and Delphi procedures, this study used objective quantitative metrics. This is made possible by standardizing the collected indicators and an automated data aggregation procedure. Compared to the k-means clustering approach in [6], which remained conceptual due to the lack of specific thresholds for assessing effectiveness, the resulting approach combines data processing with cross-sector analysis to form generalized recommendations. This is made possible by obtaining numerical criteria for each considered industry from the dataset. Unlike the integration of complex AI algorithms in migration planning in [7], the proposed approach optimizes the choice of strategy through empirical verification on a cleaned dataset without using complex models. This is made possible by a two-step procedure – imputation and anomaly removal and cross-sector analysis of the results. Finally, unlike the S/4HANA Cloud migration approach in [8], the proposed approach is provider-neutral and scalable across economic sectors. This is made possible by using a unified set of quantitative metrics and a cross-sector empirical base. First, it reduces the dependence on a specific cloud platform, which increases the versatility of the proposed approach. Second, it allows for making informed conclusions based on average values in a group of observations in the relevant industry, which reduces the contribution of individual (sometimes anomalous) cases to the formation of general conclusions.

Overall, the advantages of the proposed solution are provided by the combination of cloud provider neutrality, quantitative unification, and transparent data pre-processing procedures. The use of consistent metrics made it possible to make reproducible comparisons between strategies. Imputation of missing values using the MICE method under the MAR assumption and outlier detection based on IQR reduced the impact of incompleteness and anomaly of records. Textual normalization of case descriptions increased the correctness of the assignment to Lift-and-Shift, Re-platforming, or Reengineering. Together, this resulted in industry-specific ranges suitable for strategic planning.

The previously identified lack of standardized classification and the advantage of conceptual approaches are compensated within the proposed solution by empirically based, transparently obtained industry distributions. Thanks to the unified metrics and cloud provider neutrality, the subjectivity of recommendations is reduced. The availability of quantitative comparable indicators allows for the transition from conceptual models to practical ones, taking into account certain assumptions.

The limitations of our study relate to the nature of the sources and statistical assumptions. The sample was formed from publicly available descriptions of mostly successful migrations, which creates a risk of bias towards positive results; unsuccessful or partially implemented projects are not represented. Imputation of omissions and filtering of outliers are correct provided that the omissions are random and the distributions are statistically stable, and in case of deviations the accuracy of the estimates may decrease. The distribution of cases according to strategies retains the risk of error in hybrid or phased scenarios. Universality is limited by the analyzed industries and available instances: if new strategies appear or changes in the sectoral structure occur, the model needs to be recalibrated and generalizability re-tested.

The shortcomings of the study are due to the limited completeness and depth of the data and the lack of analysis of causal relationships. There is no information on the full scale of IS, the criticality of loads, the level of organizational maturity and regional specificity, which complicates the interpretation of the differences found outside the data set. The cross-sectional nature of the observations does not allow us to separate the effects of the strategy from the accompanying managerial and technological decisions, and also does not cover long-term indicators, such as, for example, the total cost of ownership several years after migration.

Prospects for further scientific development are outlined in several key areas. First of all, it is advisable to expand the empirical base of the study by including open source cases, analytical materials prepared for business needs, as well as reports from consulting companies. A further important step should be to stratify the collected data by enterprise size and regional affiliation, which will allow identifying jurisdictional and structurally determined differences in the application of migration strategies. It is also important to integrate machine learning methods in order to build predictive models capable of generating recommendations for choosing the optimal strategy based on the characteristics of the initial state of the IS. Special attention should be paid to the construction of formalized models for assessing migration-related risks for each of the strategies. Such models should cover not only technical aspects but also organizational factors, in particular, the transformation of the cost structure, the level of professional training of personnel, as well as potential threats to data loss.

7. Conclusions

1. In the course of building an empirical base, a homoe geneous sample of 275 records was formed (a 19% reduction in the original volume), within which missing values were imputed using the MICE method with chain equations, and outliers were removed using the interquartile range criterion. This approach ensured increased data reliability and created a basis for the application of quantitative analysis methods without distortions caused by outliers or missing information.

2. During our analysis and evaluation of the effectiveness of migration strategies, it was determined that Lift-and-Shift was used in 39.64% of cases, Re-platforming – 38.55%, and Re-engineering – 21.82%. The average cost reduction rates were 40.06%, 38.18%, and 42%, respectively, and the average performance improvement was 133.04%, 101.17%, and 141.66%, respectively. Cross-sector analysis revealed a clear connection between the industry and the most used strategy. In Agriculture and Real Estate, Lift-and-Shift was used in 100% and 83.33% of cases with a cost reduction of 50% and 50.75% and a performance improvement of 197% and 239.69%, respectively. In BigData and Cryptocurrency, 100% of companies chose Re-platforming with a cost reduction of 36.75% and 45.12% and a performance improvement of 49% and 173.84%. In Financial Technologies, almost half of the implemented projects (47.06%) used Reengineering with a cost reduction of 44.46% and a performance improvement of 176.27%. In the IT-Services and Software Engineering industries, Re-platforming/Reengineering also dominated (55.56/40.91%, respectively) with a performance improvement of over 110%. The contextual feasibility of applying each strategy was established: Lift-and-Shift dominated in a sector with strict regulatory requirements, while Re-platforming and Re-engineering were prioritized in innovative industries. These results allowed us to empirically substantiate recommendations for choosing a migration strategy depending on the domain specificity of the business.

Conflicts of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study, as well as the results reported in this paper.

Funding

The study was conducted without financial support.

Data availability

The manuscript has associated data in the data warehouse.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

References

1. Mittal, M. (2024). The Great Migration: Understanding the Cloud Revolution in IT. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 10 (6), 2222–2228. https://doi.org/10.32628/cseit2410612423
2. Amin, R., Vadlamudi, S. (2021). Opportunities and Challenges of Data Migration in Cloud. Engineering International, 9 (1), 41–50. https://doi.org/10.18034/ei.v9i1.529
3. Indukuri, A. V. (2025). AI-Powered Cloud Migration: Transforming Enterprise Modernization Strategies. Journal of Computer Science and Technology Studies, 7 (2), 567–575. https://doi.org/10.32996/jcsts.2025.7.2.60
4. Sharma, B. P. (2025). Optimizing Cloud Migration Strategies for Large-Scale Enterprises: A Comparative Analysis of Lift-and-Shift, Replatforming, and Refactoring Approaches. Advances in Theoretical Computation, Algorithmic Foundations, and Emerging Paradigms, 15 (2), 1–14. Available at: https://heilarchive.com/index.php/ATCAEP/article/view/2025-FEB-04
5. Hosseini Shirvani, M., Amin, G. R., Babaeikiadehi, S. (2022). A decision framework for cloud migration: A hybrid approach. IET Software, 16 (6), 603–629. https://doi.org/10.1049/sfw2.12072
6. Aslam, M., Rahim, L. bin A., Watada, J., Hashmani, M. (2018). Clustering-Based Cloud Migration Strategies. Journal of Advanced Computational Intelligence and Intelligent Informatics, 22 (3), 295–305. https://doi.org/10.20965/jaciii.2018.p0295
7. Chanthati, S. R. (2024). Artificial Intelligence-Based Cloud Planning and Migration to Cut the Cost of Cloud Sasibhushan Rao Chanthati. American Journal of Smart Technology and Solutions, 3 (2), 13–24. https://doi.org/10.54536/ajsts.v3i2.3210
8. Ankit, K. G., Apoorva, J. (2025). Efficient Strategies for S/4 HANA Cloud Migration in Large Enterprise Landscapes. International Journal of Innovative Science and Research Technology (IJISRT), 9 (11), 3628–3645. https://doi.org/10.5281/zenodo.14836417
9. Althani, B. (2025). Migration challenges of legacy software to the cloud: a socio-technical perspective. Cogent Business & Management, 12 (1). https://doi.org/10.1080/23311975.2025.2503421
10. Sahoo, K., Samal, A. K., Pramanik, J., Pani, S. K. (2019). Exploratory Data Analysis using Python. International Journal of Innovative Technology and Exploring Engineering, 8 (12), 4727–4735. https://doi.org/10.35940/ijitee.l3591.1081219
11. Buuren, S. van, Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations inR. Journal of Statistical Software, 45 (3). https://doi.org/10.18637/jss.v045.i03
12. Tukey, J. W. (1977). Exploratory Data Analysis. Addison-Wesley Publishing Company. Available at: https://consoleflare.com/blog/wp-content/uploads/2022/09/Exploratory-Data-Analysis-1977-John-Tukey.pdf