Developing Objective Comparison Evaluation Framework for Comprehensive IaaS Assessment Using the AHP Method

Supervisor Dr. A. Abhishta

Supervisor Dr. A. Affinito

PhD Candidate Yasir Haq

October 2024

Chalan, G. (S2940175, Student M-BIT)

UNIVERISTY OF TWENTE | DEPARTMENT OF BUSINESS INFORMATION TECHNOLOGY

FACULTY OF ELECTRICAL ENGINEERING, MATHEMATICS AND COMPUTER SCIENCE

Abstract

In today's cloud computing landscape, Infrastructure-as-a-Service (IaaS) has emerged as one of the most popular solutions for organizations seeking flexible, scalable, and cost-efficient IT infrastructure. However, selecting and evaluating service offerings proves to be complicated and not accessible due to the wide variety of offerings in the marketplace and the solid competitive environment between the cloud service providers. Organizations often find themselves inundated with generalized information and marketing-driven offerings that fail to represent the true nature of the services being provided accurately. In order to tackle this problem, we researched and explored ways for organizations to select the most suitable IaaS providers' products based on set criteria that can involve aspects such as availability, performance, cost, security, scalability, and reliability and produced this artifact as a result.

The thesis presents a comprehensive evaluation framework for IaaS products, built on the Analytic Hierarchy Process (AHP) to structure the decision-making process systematically and objectively. The framework is designed to assess IaaS products based on six primary criteria: Availability, Performance, Scalability, Reliability, Security, and Cost-effectiveness. Each criterion is further broken down into sub-criteria with relevant metrics to measure it, allowing for a granular evaluation of products across different service components specifically catered to IaaS, such as virtualization, storage, networking, and servers. The framework focuses on providing flexibility to the users by allowing them custom weight assignments to each criterion, tailoring the evaluation process to their specific priorities and needs.

A study case was conducted to validate the framework. Ten leading IaaS products were evaluated across five predefined scenarios: Balanced, Performance-Focused, Cost-Sensitive, Security-Centric, and Custom Scenario. Furthermore, we conducted a third-party evaluation with industry experts in cloud services to complement and validate the findings of our case study.

The results, generated using an Excel-based AHP tool developed for this research and the validation process conducted after, show a comprehensive and objective ranking of the IaaS products based on their performance across these scenarios. The tool offers additional features such as scenario customization, consistency checks, and automated scorecards, making it an effective decision-support system.

Challenges encountered during the validation process included data availability for specific metrics, particularly recovery time, network latency, and actual calculations of the total cost of ownership. These limitations were addressed through assumptions and industry benchmarks, though they highlight the need for further real-time testing in future research. Despite these challenges, the proposed framework offers a robust, flexible, and user-friendly approach to evaluating IaaS products, aiding organizations in making informed, data-driven decisions in a rapidly evolving cloud market.

Contents

List of Figures

List of Tables

Chapter 1: Introduction

This chapter begins with a discussion of motivation and an introduction. It also provides essential background information to help understand the scope and context of the problem statement. Additionally, it presents an overview of the research objectives and proposed research questions, followed by a description of the methodology and an outline of the thesis.

1.1 Background

The introduction of cloud computing has revolutionized the provisioning and consumption of technological resources, with Infrastructure as a Service (IaaS) emerging as a critical enabler for organizational agility and digital transformation. IaaS has become a foundational component of this paradigm, offering virtualized computing resources on demand [1]. This service model has allowed businesses to shed the constraints of physical infrastructure, gaining the ability to scale resources dynamically in response to fluctuating demand.

The trajectory of IaaS has paralleled the rapid evolution of the broader technology landscape, embracing advancements in automation, artificial intelligence, and analytics. The proliferation of IaaS providers has led to a competitive market where services are continually enhanced to deliver computing power and sophisticated platform ecosystems that support complex applications and data workloads.[2]

As the spectrum of IaaS offerings expands, incorporating everything from bare-metal servers to managed database services, its role in enterprise IT strategy has grown increasingly central, especially in how it is influenced by the convergence of emerging technologies such as the Internet of Things (IoT), Edge computing, and more recently, Fog computing [2]. These new paradigms are redefining the boundaries of cloud services in general, shifting them toward a more distributed architecture that addresses the new rising limitations and challenges. Nevertheless, IaaS solutions' deployment still spans many industries as a driver for innovation and a backbone for critical operations. Its attractiveness lies in its promise of enhanced flexibility, scalability, and cost savings, enabling businesses to adjust resources dynamically to align with their operational demands.

As enterprises increasingly rely on IaaS for critical operations as the backbone of their core operations, evaluating these services becomes paramount; however, this evaluation needs to be comprehensive and holistic to cover all dimensions of IaaS. These dimensions cover availability, performance, reliability, scalability, security, and cost-effectiveness. The evolution of IaaS is also influenced by technological advancements, demanding continuous adaptation of evaluation criteria to ensure they remain relevant and reflective of the current state of technology. Therefore, it is crucial to underscore the significance of IaaS in the current era of cloud services, setting the stage for a discussion on the necessity of a systematic approach to evaluating and selecting these services, which will be elaborated in the problem statement.

1.2 Problem Statement

While Infrastructure as a Service (IaaS) benefits are manifold, the proliferation of IaaS providers and the consequent diversification of services have introduced significant complexity into the selection process for potential adopters[3]. Businesses face the challenge of comparing different IaaS offerings on aspects beyond cost and basic performance metrics. Therefore, the current landscape requires evaluation across a multi-dimensional space that includes advanced performance metrics, security standards, regulation compliance, service scalability, reliability, and integration capabilities with emerging technologies like IoT and Edge computing.[3]

Moreover, the pace at which the IaaS domain is evolving adds another layer of complexity. The continuous introduction of new technologies and services demands that evaluation metrics be comprehensive and adaptable to change [2]. Existing frameworks for comparing cloud services are often too complex, fall short of capturing the full spectrum of IaaS features, or are not updated frequently enough to reflect the latest industry developments[4]. The lack of a standardized set of objective comparison metrics hinders organizations' ability to make informed decisions that align with their strategic objectives since decisions based on inadequate or subjective information can lead to suboptimal service selection, which potentially results in increased costs, underutilized resources, and misaligned IT capabilities. This leads us to conclude that this gap in the literature and practice signifies a critical need for a robust framework that systematically assesses IaaS offerings with clear and actionable insights for businesses navigating the complex marketplace.

This thesis proposes addressing this gap by developing a set of objective comparison metrics for IaaS assessment. These metrics aim to enable businesses to conduct a comprehensive and balanced evaluation of IaaS-related products, facilitating decision-making that optimizes different scenarios aligned with business goals.

1.3 Research Question

This research aims to develop a set of objective metrics for the comprehensive assessment of IaaS offerings covering availability, performance, reliability, scalability, and cost-effectiveness. To achieve this aim, the following research questions have been formulated:

M.Q *How can we develop objective comparison metrics encompassing IaaS, including availability, performance, reliability, scalability, and cost-effectiveness, to facilitate informed decision-making in cloud adoption?*

This research question is instrumental in guiding the investigative process toward developing a comprehensive and standardized framework for IaaS assessment since it acknowledges the necessity for a multifaceted evaluation that does not solely rely on one or two dimensions but provides a comprehensive approach to understanding the value proposition of IaaS offerings. To explore this question effectively, the following sub-questions have been formulated to dissect the central inquiry into manageable segments, each focusing on a specific aspect of the IaaS evaluation starting by comparing IaaS to other dominant models such PaaS and SaaS, to frameworks used for evaluating and pricing IaaS models in the current business environment. These sub-questions are:

S.Q. 1: Cloud Service Models: *What are the most common cloud service models, and what are their architectures and differences? (SaaS, PaaS, and IaaS)*

S.Q. 2: Cloud Computing Service Metrics: What CCSM */frameworks currently exist, and how do they differ?*

S.Q. 3: Performance, Reliability, Scalability: *How can the existing metrics be linked to measure performance, reliability, and scalability?*

S.Q. 4: Decision-making Framework: *What frameworks can be utilized to establish the objective metrics in informed decision-making in cloud adoption?*

S.Q. 5: *Cost-effectiveness: Which financial models and cost structures can be applied to compare the cost-effectiveness of IaaS providers transparently?*

The intention is to develop metrics that are not only comprehensive but also practical and applicable for businesses at various scales and sectors. By answering these sub-questions, the research will contribute to a more transparent, more systematic approach to IaaS assessment, thus creating a solid base for a robust framework that aids organizations in making more informed and strategic cloud adoption decisions.

1.4 Research Methodology

This research adheres to the Design Science Research Methodology (DSRM) defined by Peffers[5]. Figure 8: Design Science Research Methodology (DSRM) Process Model [5] shows five steps: identification of the problem and motivation, defining objectives of the solution, design and development, demonstration, evaluation, and communication.

1. **Problem identification and motivation***:* This is the first step of the DSRM process. The thesis aims to provide a clear overview of problem identification and motivation, presented in Chapter 1, and supported in Chapter 2.

*2***. Define the objectives for a solution:** This phase involves defining the research objectives for the problem identified in the first phase. These objectives formulate the structure for a proposed solution that addresses the unresolved issues, as discussed in Section 1.3. At this stage, a literature review is also conducted to compile all relevant content related to the sub-research questions, thereby helping to develop a practical framework that defines and sets objective comparison metrics. Chapter 2 thoroughly reviews the available literature and better explains the main problem statement and the needed answers to all sub-research questions SQ1, SQ2, SQ3, and SQ4.

3. **Design and development***:* This phase will involve Designing an Objective Comparison Metrics Framework that will structure the metrics definition, weighting and scoring, data collection and analysis, and finally, validation, as shown in Chapter 4.

*4***. Demonstration and evaluation***:* The developed metrics will be demonstrated by applying them to real-world IaaS offerings. This will involve creating a case study based on several cloud service providers' products to highlight how the metrics work in practical scenarios and how they can be used to compare different IaaS services, as shown in Chapter 5. Moreover, gathering feedback from industry experts, such as TRIMM and Cape Group, to analyze the demonstration outcomes and compare the findings against the set objectives is needed for a transparent evaluation, which will be presented in Chapter 6.

5. **Communication:** The final stage of the research is sharing the findings and results in the thesis document, which presents the developed Framework, its practical applications, and its contributions to the field. This includes documenting the research process and results in the thesis and potentially disseminating the findings through academic publications, presentations, or industry reports. Additionally, we will discuss the implications and limitations of the research findings, as shown in Chapter 7.

1.5 Structure

This thesis is systematically organized into chapters that methodically approach the research question, ensuring a comprehensive exploration of objective comparison metrics for IaaS. The table of contents is as follows:

Chapter 1: Introduction

This opening chapter introduces the thesis, states the problem, poses the research questions, and outlines the document's overall structure.

Chapter 2: Background

The Chapter describes data sources and collection methods, outlines the approach to selecting the resources, explains the sampling strategy, presents the literature review results, discussion, and conclusions, and identifies gaps in the current literature related to objective comparison metrics.

Chapter 3: Research Methodology

Defines and explains the research methodology utilized in the thesis and maps it to the Design science research methodology implemented.

Chapter 4: Designing an Objective Comparison Metric Framework

Defines the requirements, the metrics, and their feasibility, describes the creation of the evaluation framework, discusses data collection and analysis methods, elaborates on weighting and scoring systems, and assesses the metrics' validation and reliability as well as their limitations.

Chapter 5: Implementation & Empirical Analysis

Presents an empirical analysis of the developed comparison metrics, applying them to actual IaaS offerings to evaluate their effectiveness and practicality.

Chapter 6: Validation and Evaluation

Validate the developed metrics with a fictional study case in which ten IaaS products were chosen and evaluated based on real-time metric data collected, in addition to validation by industry experts, to evaluate the effectiveness of the comparison metrics in providing clear, objective, and meaningful assessments of IaaS services.

Chapter 7: Conclusion

Explores the broader implications of the research findings and provides recommendations and future research. Finally, the thesis summarizes the essential findings and contributions and reflects on the research process.

References: A comprehensive list of all the scholarly works cited throughout the thesis.

Appendices: Supplementary material that supports the thesis, including data tables, descriptions of the methodology, scoring scales, and any additional documentation relevant to the research.

Each chapter is designed to provide in-depth analysis and contribute to a holistic understanding of the IaaS assessment framework. The sequential flow from literature review to empirical analysis to implications ensures a logical progression and a strong narrative throughout the thesis. This results in the following table where all chapters contributing to a specific research question can be found:

Research Question:	In which chapter is it answered?
SQ1.	In Chapter 2, Literature Review
SQ2.	In Chapter 2, Literature Review
SQ3.	In Chapter 4 Objective Comparison Metric
SQ4.	In Chapter 4 Objective Comparison Metric
SQ5.	In Chapter 2, Literature Review
MQ1.	In chapters 5, 6, & 7, Empirical Analysis, Validation, & Conclusion

Table 1: Summary of which chapter answers which research question.

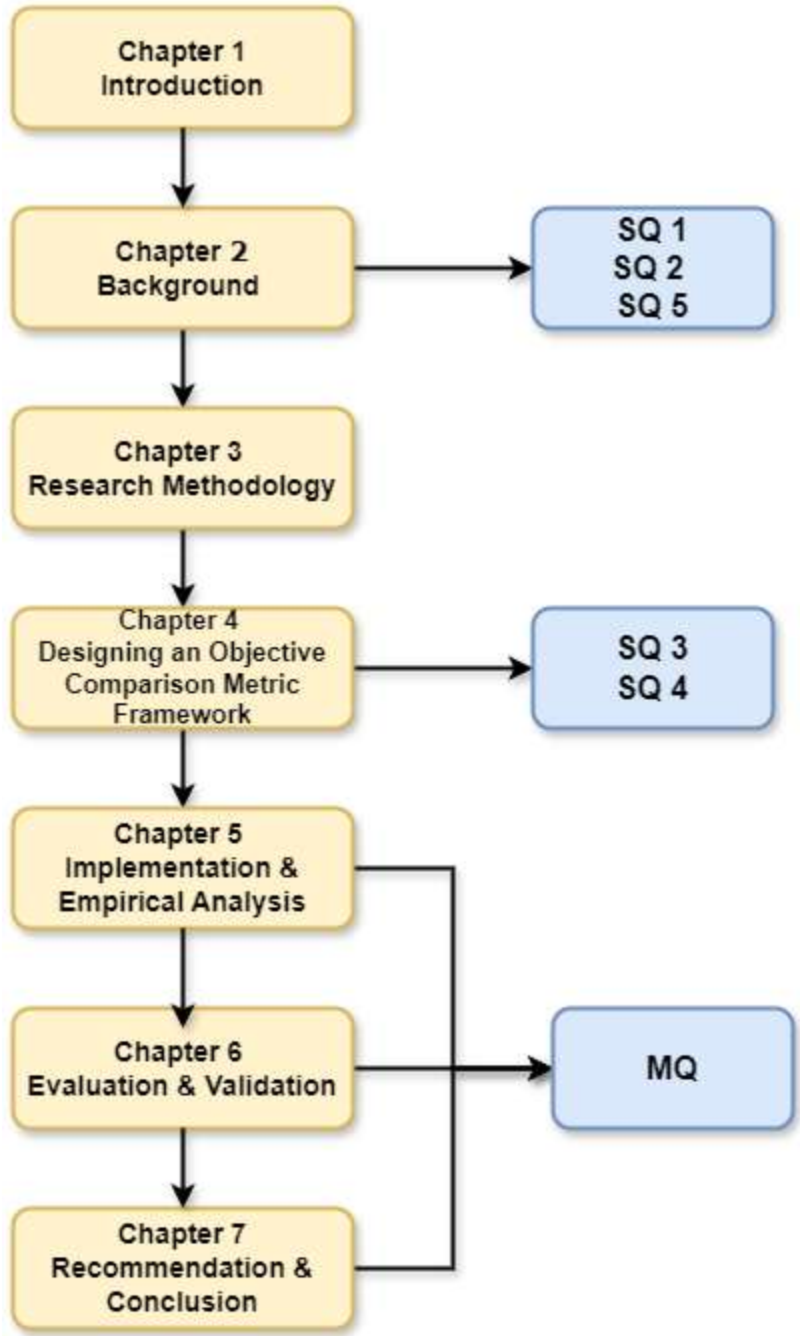


Figure 1:: Research Structure

Chapter 2: Background

This chapter of the paper provides a literature review explaining the methodology used and the research strategy implemented. Moreover, it shows how the literature review is used to answer SQ1, SQ2, SQ3, SQ4, and SQ5 outlined in Chapter 1 and summarizes the main takeaways of the literature review of the proposed research questions.

2.1 Literature Review Methodology

In this section, we outline the approach used to gather, analyze, and synthesize relevant literature for evaluating IaaS providers. The methodology includes identifying key sources, evaluating prior studies on cloud infrastructure, determining the search strategy used, and determining the results. This review is the foundation for understanding industry standards and establishing the evaluation criteria used throughout the research.

2.1.1. Research Questions

The research question and the sub-questions we posed, as seen in Section 1.3, were as follows:

M.Q *How can we develop objective comparison metrics encompassing IaaS, including availability, performance, reliability, scalability, and cost-effectiveness, to facilitate informed decision-making in cloud adoption?*

As discussed, to explore this question effectively, the following sub-questions have been formulated to dissect the main inquiry into manageable segments, each focusing on a specific aspect of the IaaS evaluation:

S.Q. 1: Cloud Service Models: *What are the most common cloud service models, and what are their architectures and differences? (SaaS, PaaS, and IaaS)*

S.Q. 2: Cloud Computing Service Metrics: What CCSM */frameworks currently exist, and how do they differ?*

S.Q. 3: Performance, Reliability, Scalability: *How can the existing metrics be linked to measure performance, reliability, and scalability?*

S.Q. 4: Decision-making Framework: *What frameworks can be utilized to establish the objective metrics in informed decision-making in cloud adoption?*

S.Q. 5: Cost-effectiveness: *Which financial models and cost structures can be applied to compare the cost-effectiveness of IaaS providers transparently?*

The following sections explore the evolution of cloud services and their corresponding financial models, cost structures, and evaluation metrics in relation to these research questions.

2.1.2 Search Strategy

To organize the data sources and collection method, the state-of-the-art matrix (SAM) was used. The SAM analysis method was developed by Beruvides and Omachonu (2001) and defined as follows: "a research mining methodology to develop matrices to partition research information to isolate critical information using statistical methods"[6]. This method divides the literature into categories covering primary theory, secondary theory, empirical studies, and case studies. To link this to our research topic, the SAM for assessment of IaaS offerings focused on cloud services models, their differences, the pricing models used to present these offerings, and the metrics related to availability, performance, reliability, and scalability used to assess these services. Various academic research databases were used to collect the related literature. Databases such as UT library catalogs, Scopus, IEEE Xplore, ResearchGate, Google Scholar, and major cloud services providers white papers. The keywords used in the search queries can be listed as, but not limited to, 'Cloud Services,' 'IaaS performance evaluation,' 'cloud service benchmarking,' 'cloud evaluation metrics,' and 'cloud evaluation frameworks.'

Literature and papers dated from the 2000s to the present were used to keep the research relevant to today's service offerings. Following the SAM methodology structure, the literature was classified into case-specific categories based on the sub-questions established in the section before to include papers classified based on content-wise grouping, which are:

- Define and explain the cloud service model focusing on IaaS.
- Pricing models used to evaluate IaaS offerings.
- Frameworks used for evaluating IaaS services.
- Standard evaluation metrics related to IaaS.
- Decision-making Framework for choosing the suitable metrics.

Therefore, the literature was divided into main categories following the SAM methodology structure and definitions.

Table 2: Operational Definitions for the Categories of the State-of-the-Art-Matrix Analysis defines each category of the SAM method.

Table 2: Operational Definitions for the Categories of the State-of-the-Art-Matrix Analysis

Category	Operational Definition
Primary Framework	A set of guidelines or standards established for evaluating and comparing cloud services. It can also be a collection of pre-developed guidelines, heuristics, and models converted into a framework.
Secondary Framework	Secondary Framework Modified or extended version of a primary framework.

Literature Review	Literature Review: Research has been performed on Cloud Services' definition, properties, and related pricing models.
Opinion Based	A study that includes solely discussion of ideas on Cloud Services Evaluation
Empirical Study	A comparison metric study that includes real users (subjects).
Analytical Study	A comparison metric study that requires expert evaluation where experts put themselves in the position of users and evaluate the system

2.1.3 Results

Based on the SAM methods, the categories were cross-referenced with the main topics we wanted to address in answering our main and sub-questions of the research[6]. The main categories should be sufficient to cover the desired aspects to establish and study the desired objective metrics used for comparison. The table below shows studies coded by the reference list numbers. Note that some of the resource papers used can be found in more than one section as they are relevant to both sections they are placed in.

Table 3: The SAM Matrix

Primary Frameworks	IaaS Evaluation Frameworks	[4], [7], [8], [9], [10]
Secondary Frameworks	IaaS Evaluation Frameworks	[11]
Literature Review	Cloud Services Models Pricing Models Decision-making Framework	[1], [2], [4], [7], [8], [9], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25]
Opinion Based		None
Empirical Evaluation Study	Evaluation metrics related to IaaS	[26], [27], [28], [29]
Analytical Evaluation Study	Evaluation metrics related to IaaS	[3], [20], [21], [27], [28], [30], [31], [32], [33], [34], [35], [36], [37]

Based on the research's data sources, the first two categories (primary and secondary frameworks) were analyzed separately from the other categories to find a trend in framework development for evaluating cloud services in general and assessing comparison metrics used. Figure 1 shows the primary and secondary framework development trends from 2000 to the present.

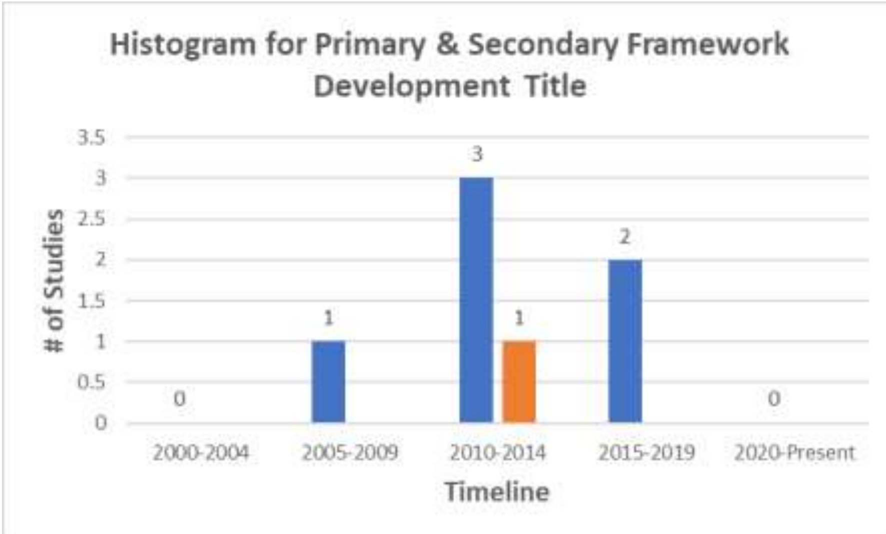


Figure 2: Trends for Primary & Secondary Framework Development

The histogram indicates that most of the primary framework development has been performed between 2010 and 2014, with a declining to nonexistent primary framework development effort in the last five years. This also applies to secondary framework development as well, and we can observe that there are no efforts for development.

The next step of the SAM analysis was determining which evaluation methods were dominantly used or recommended. In the context of our research topics, primarily since it deals with evaluating cloud services or any other technology-related research, the classification between empirical and analytical evaluation studies can be understood as follows:

Empirical evaluation is characterized by direct interaction with the system and its users, providing concrete evidence of usability, performance, and user satisfaction. It involves collecting data through observation or experimentation to validate hypotheses or understand phenomena[6]. In the realm of cloud computing or technology evaluation, this might include User Studies, Questionnaires, and usability testing. Analytical evaluation leverages theoretical knowledge, expert insights, and models to predict issues, assess compliance with standards, or understand potential user experiences involving using theoretical models, expert judgments, or specific criteria to assess a system or technology [6]. This might include Heuristics Evaluation, Design Guidelines, and Performance Modeling. The following figure represents the number of papers listed in Table 3: The SAM Matrix mentioned before. It shows that the use of empirical and analytical methods in resource papers used in the research shows that analytical methods are more dominant, and those empirical methods were used with a couple of papers utilizing both methods.

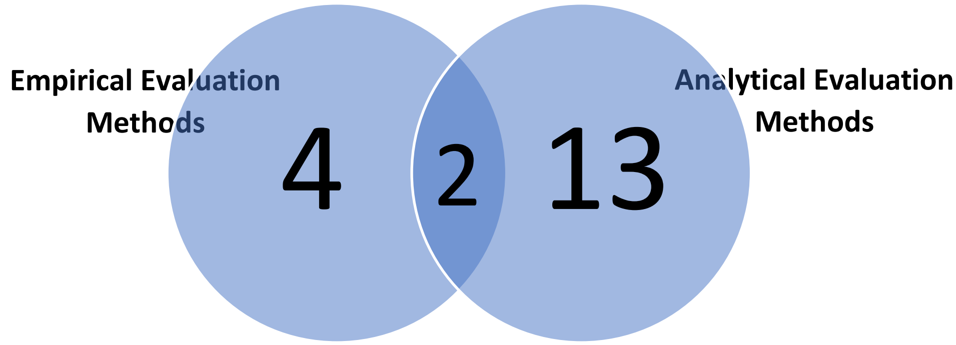


Figure 3: Empirical vs. Analytical Evaluation Methods

2.1.4 Discussion

The methodology section's analysis reveals a trend toward the dominance of analytical methods over empirical ones in cloud service evaluation. This suggests a preference for theoretical and expert judgment approaches in understanding and predicting cloud service performance and user experience. Such a trend underscores the importance of developing robust theoretical models and criteria for assessing IaaS services, highlighting a gap in direct user experience studies. The observed decrease in new framework creation for cloud service evaluation, particularly from 2010 to 2014, with a notable decline thereafter, might be attributed to market saturation or a pivot towards more specialized, niche frameworks, which is more logical as cloud services become more complex and required to address requirements for new emerging technologies that are more specialized and niche. This trend implies that while foundational frameworks have been established, there is a growing need for frameworks that address specific aspects of IaaS evaluation in more depth. Therefore, developing an objective evaluation framework tailored for IaaS could fill this gap, offering nuanced insights into its unique requirements and challenges.

The findings indicate a critical need for balancing theoretical analysis with empirical research to provide a comprehensive understanding of IaaS offerings. Future research should integrate user-based evaluations to complement the existing analytical frameworks, ensuring a holistic view of cloud service efficacy and user satisfaction.

2.2 Literature Review Results

This section summarizes the results for the MQs defined in Section 1.3. Section 2.2.1 explores the Evolution of Cloud Services from a Historical perspective. Section 2.2.2 answers SQ1 and SQ2, defining the cloud services models, their architecture, role, and functions. Section 2.2.3 answers SQ5 by establishing the financial models and cost structures that can be applied to compare the cost-effectiveness of IaaS providers transparently. Finally, section 2.2.4 answers SQ3 and SQ4 related to metrics used to measure performance, reliability, and scalability and frameworks for establishing the objective metrics for decision-making in cloud adoption. The insights gained from this literature review form the basis for building the IaaS evaluation framework, focusing on performance, cost-effectiveness, scalability, and security.

2.2.1 Evolution of Cloud Services Historical Perspective

As we know it today, cloud computing began taking shape in the late 1990s as it evolved from basic remote applications to highly integrated services. During this period, we have witnessed the rise of the Application Service Provider (ASP) model, which laid the groundwork for modern cloud computing. ASPs offered businesses the ability to remotely host and manage applications, paving the way for what would evolve into cloud-based services [12]. This, in turn, changed the way businesses and individuals interact with technology and established the foundation for modern cloud service models.

The early 2000s marked the period where concepts such as "Software as a Service" (SaaS) gained traction, as it demonstrated the potential for delivering enterprise software over the internet, eliminating the need for local installations [16] [12]. Another milestone was achieved by the mid-2000s when Amazon Web Services (AWS) revolutionized the industry by offering cloud-based Infrastructure through its suite of services, introducing models like Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) [12]. This period was substantial as it solidified the core cloud service models: SaaS, IaaS, and PaaS, with key players like Google and Microsoft expanding their offerings to meet growing demand [12].

As these models proved their ability to provide competitive advantages, especially advantages related to scalability, flexibility, and cost-efficiency, a notable rapid growth in the cloud market emerged in the late 2000s and early 2010s. Major tech companies expanded their cloud services to meet diverse business needs, intensifying competition as businesses began integrating on-premises infrastructure with cloud services to improve performance, security, and compliance [12]. This increased the popularity of multi-cloud strategies as businesses sought to minimize downtime risks by leveraging multiple cloud providers [2]. As business environment requirements keep changing and evolving, limitations to traditional cloud models started emerging, which led to the rise of the Internet of Things (IoT) and other advanced technologies like AI and machine learning in the late 2010s and early 2020s, prompting the development of hybrid computing solutions [2].

Even though the focus has shifted today toward edge computing and sustainable solutions, there is still heavy reliance on traditional cloud computing models. Edge computing addresses latency and bandwidth issues by processing data closer to the source, while sustainability initiatives are reducing the carbon footprint of cloud services by emphasizing energy-efficient data centers and renewable resources [2]. However, it is not feasible for all organizations, especially the cost-sensitive ones.

Nevertheless, the historical evolution of cloud computing illustrates how business needs and technological advancements have shaped the development of cloud services, leading to the emergence of various models like IaaS, PaaS, and SaaS. In the following sections, we will explore these models' roles and functions in cloud computing.

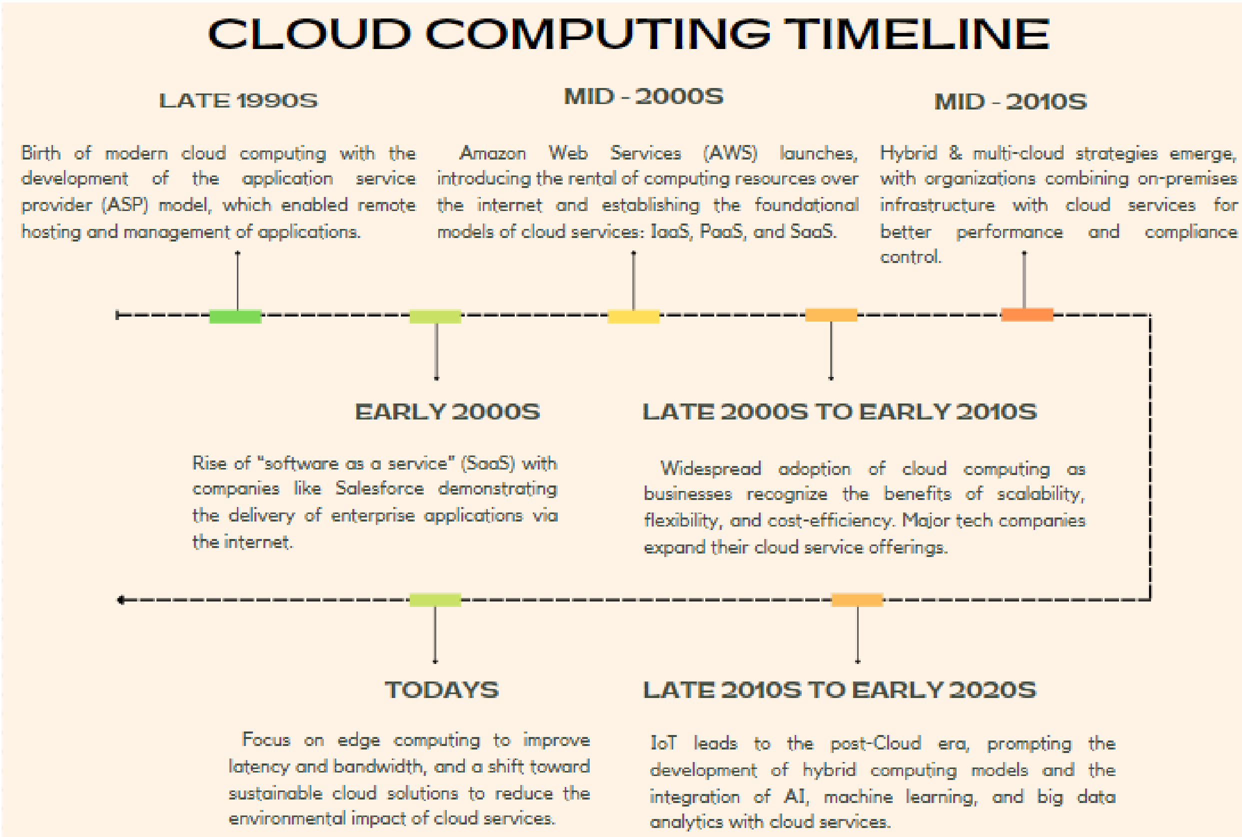


Table 4: Evolution of Cloud Computing [12]

2.2.2 Cloud Service Models

According to the National Institute of Standards and Technology (NIST), cloud computing is characterized by on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service, allowing efficient and flexible resource management. [18]. Moreover, the National Institute of Standards & Technology (NIST), Figure 4: NIST Conceptual Reference Model, describes the main activities of cloud providers and divides them into five categories: service deployment, service orchestration, cloud service management, security, and privacy [7]. These categories will help us understand the core mechanisms of cloud services and will guide us in establishing the comparison metrics we need for IaaS assessment. Nevertheless, as described in the Service orchestration's service layer, various software, platforms, and infrastructural elements are offered on a service basis in the current landscape of digital services. This approach is typically categorized into three core models, as identified by (NIST): Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), and Infrastructure-as-a-Service (IaaS) [7]. Figure 5: Cloud Services Comparison[1] divides cloud services aspects into categories and clearly shows how these models differ in responsibility between the client and the cloud provider. Understanding these service models' differences is crucial, as I will only focus on the IaaS service model. As shown in Figure 5: Cloud Services Comparison[1], the IaaS service model limits the provider's responsibility into four main categories: Virtualization, Networking, Storage, and Servers. Categories will determine how we establish the comparison metric in the coming sections.

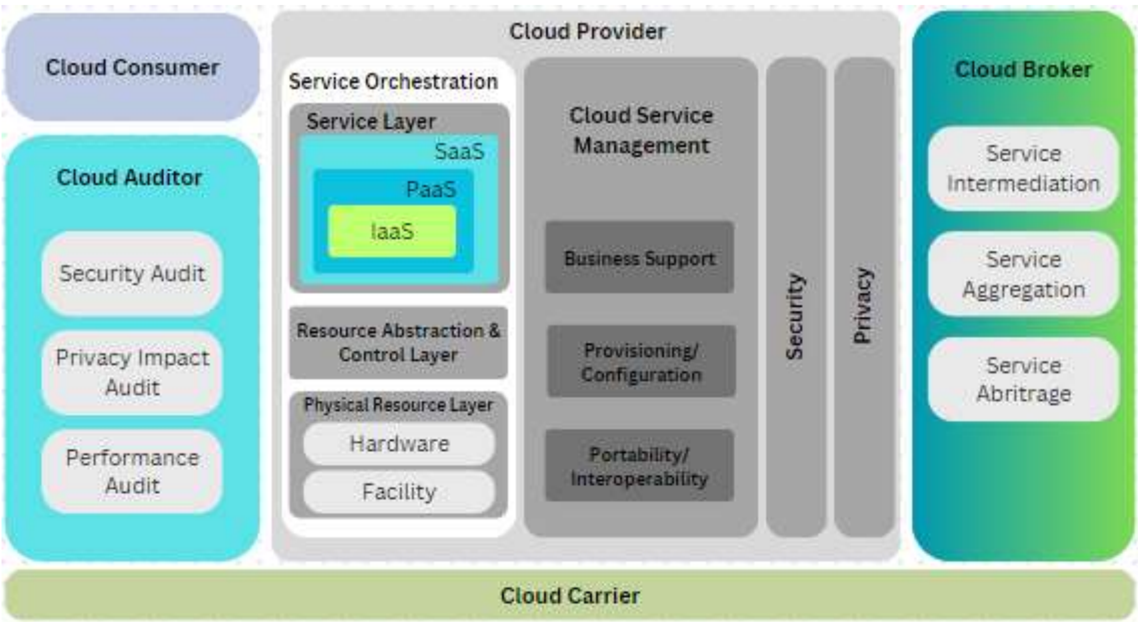


Figure 4: NIST Conceptual Reference Model

2.2.2.1 Software as a Service (SaaS)

Software as a Service (SaaS) is a cloud computing model where applications are hosted by a cloud provider and made accessible to users over the internet. This model can involve an independent software vendor (ISV) contracting a cloud provider for hosting or larger companies like Microsoft hosting their applications, thus eliminating the need for users to manage software updates or underlying infrastructure [7]. It is usually a subscription-based model where payment aligns with usage. Notable SaaS examples include Salesforce for customer relationship management, Google's suite of applications like Gmail and Google Docs, and FreshBooks for invoicing. SaaS stands alongside IaaS and PaaS as a primary category of cloud services, serving a broad range of users, from IT professionals to personal users, with products varying from simple entertainment to advanced IT tools [13].

SaaS operates via a cloud delivery model, where the software provider uses its servers, databases, and computing resources or contracts a cloud provider's data center. Applications are accessible from any network-connected device, typically through web browsers[18]. The SaaS model spares companies from software setup and maintenance, shifting to subscription-based access to ready-made software solutions, as it is linked to the ASP (application service provider) and on-demand computing software delivery models[13]. The software-on-demand aspect of

SaaS allows network-based access to a single copy of an application created for SaaS distribution. New features are uniformly rolled out to all customers, and depending on the SLA, customer data may be stored locally, in the cloud, or both. Organizations can use APIs to integrate SaaS applications with other software, enabling customization and enhanced functionality.

Moreover, SaaS typically uses a multi-tenant architecture, where a single application instance serves multiple customers or tenants[13]. This setup allows all customers to run on the same version and configuration while keeping their data segregated. Multi-tenancy streamlines maintenance, updates, and bug fixes, as changes are made once to the shared instance.

SaaS offers significant advantages such as reduced costs and flexible payment models by eliminating hardware acquisition and maintenance and allowing for predictable budgeting and scalability[1], [13]. It also simplifies IT management with automatic updates and provides easy accessibility from any internet-enabled location, along with customization options for integration with other business applications[13]. However, SaaS also presents challenges like dependency on providers, which can impact service due to disruptions or changes, a lack of control over software versioning, potential difficulties in switching providers due to vendor lock-in and data migration complexities, and security concerns. Despite these challenges, SaaS remains a popular choice for its convenience, scalability, and cost-effectiveness, though it necessitates careful consideration of vendor capabilities and policies.[13]

2.2.2.2 Platform as a Service (PaaS)

Platform as a Service (PaaS) differs from SaaS because it offers the computing platform and environment for running applications without the end-user applications and data. This service model takes care of hosting, operating systems, and Infrastructure, while clients are responsible for the applications and their updates[7]. This eliminates the consumer's need for in-house hardware and software management while offering tools designed for ease of use and convenience. Leading PaaS vendors include Google Cloud, Microsoft Azure, AWS, IBM Cloud, Red Hat OpenShift, VMware (Pivotal) Cloud Foundry, Oracle Cloud Platform, Heroku, OpenStack, Apache Cloud Stack, and Wasabi Cloud Storage [14].

PaaS provides a portion of a company's IT infrastructure for software development via a cloud service provider's hosted infrastructure, typically accessed through a web browser and delivered via public, private, or hybrid clouds[1], [18]. It includes services like application hosting, Java development, development team collaboration, application design, testing and deployment, web service integration, information security, and database integration[14]. PaaS pricing models are usually based on per-use or a flat monthly fee.

PaaS's primary advantage is its flexibility and convenience, allowing users to access much of the Infrastructure and IT services anywhere through a web browser[1][12]. The subscription or per-use payment models enable organizations to reduce the capital expenses of on-premises hardware and software. PaaS products, tailored for software development, provide resources and services that facilitate efficient and collaborative development processes [14]. However, PaaS has its drawbacks. Service availability or resilience can be a concern, as provider outages can significantly impact customers. There is also the risk of vendor lock-in, making migrating services and data between PaaS platforms challenging. Furthermore, internal changes by the PaaS provider, such as discontinuing support for specific programming languages or tools, can disrupt user operations[14].

2.2.2.3 Infrastructure as a Service (IaaS)

Infrastructure as a Service (IaaS) represents the most straightforward level of these service models, where only the essential infrastructure components like networking, storage, servers, and Virtualization are managed by the Cloud Service Provider (CSP). In contrast, the consumer manages everything else, leaving room for customization and personalization of the working environment. They manage the platform and software layers, including the operating system, middleware, runtime, and applications[7]. IaaS offers virtualized computing resources over the internet, allowing clients to install and manage their software stack without the burden of physical hardware. This model is beneficial for its flexibility, cost-effectiveness, and suitability for fluctuating workloads or experimental projects. Popular IaaS providers include AWS, Microsoft Azure, Google Cloud, Digital Ocean, Akamai (Linode), and Alibaba Cloud[15].

IaaS hosts traditional data center infrastructure components (servers, storage, networking hardware, and the virtualization layer) in a cloud-based environment. It supplements these with services like detailed billing, security, load balancing, and storage resilience (backup, replication, recovery)[15]. These services are increasingly policy-driven, allowing for automation and orchestration of critical tasks, such as load balancing for maintaining application performance. Moreover, IaaS operates through a vast area network like the internet. Users can create and manage virtual machines, install operating systems and applications, and handle storage and backups. Providers facilitate monitoring, cost tracking, performance management, and disaster recovery[18].

IaaS's primary benefit is the elimination of capital expenses associated with in-house hardware and software. It is particularly effective for temporary, changing, or experimental workloads, offering a pay-as-you-go model based on actual usage. This model provides scalability and flexibility, which makes it ideal for testing new applications or managing fluctuating workloads [15]. On the other hand, the granular billing model of IaaS can lead to unexpected costs, requiring close monitoring of resource usage. Infrastructure's opaque nature under a provider's management can complicate systems management and monitoring. Dependence on the provider means network issues or downtime can directly impact service availability and performance. The multi-tenant architecture also raises potential issues with resource contention from other users (the noisy neighbor issue)[15].

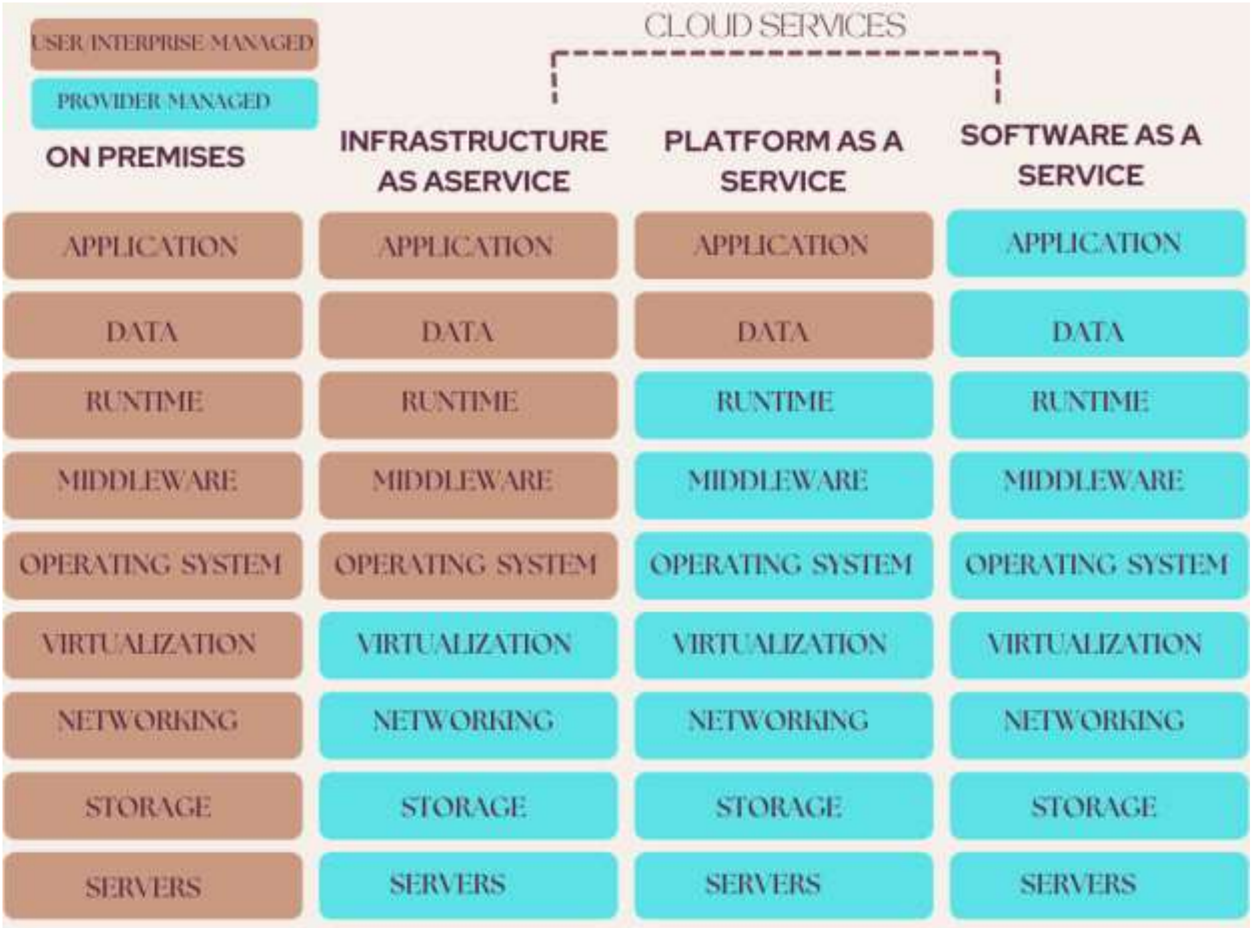


Figure 5: Cloud Services Comparison[1]

2.2.2.4 Discussion

As the cloud computing landscape is defined by these three primary service models: (SaaS (PaaS) and (IaaS), each model caters to different business needs and offers varying levels of control and responsibility. SaaS provides comprehensive services with minimal client management, PaaS offers a balance between control and convenience, and IaaS delivers extensive flexibility with greater client responsibility. Understanding these distinctions is crucial for businesses to make informed decisions about cloud adoption, ensuring alignment with their operational strategies and technical requirements. As established in Figure 5: Cloud Services Comparison[1], The key difference lies in the level of management responsibility the client and provider assumed. Excluding Premises cloud services, SaaS offers the least control but the most convenience, PaaS provides a balance of control and ease of use, and IaaS offers maximum control with corresponding management responsibilities. This knowledge forms the foundation for assessing IaaS services effectively, as it is evident that the focus should be on developing comparison metrics tailored to cover the four main categories that IaaS focuses on: virtualization, networking, storage, and servers. Consequently, understanding how the *availability, performance, reliability, scalability, security, and cost-effectiveness* metrics related to these aspects of the IaaS offering.

In conclusion, it is essential to delve deeper into how these differences inform IaaS assessment metrics when discussing the distinctions between the different cloud service models. As discussed, IaaS, distinct in its provision of fundamental computing resources, presents unique challenges like dynamic resource allocation, performance, storage reliability, speed, and service scalability. Thus, understanding how IaaS differs from SaaS and PaaS, which manage higher-level service layers, is crucial in developing metrics that accurately evaluate IaaS capabilities. We can conclude from this section that metrics for IaaS should focus on the efficiency of resource utilization and elasticity of services in response to workload changes. The robustness of the infrastructure in terms of storage reliability, network performance, and cost efficiency, aspects less emphasized in SaaS or PaaS models, ensuring that metrics not only reflect the specific operational complexities of IaaS but also align with customer expectations and industry standards.

While understanding the differences between the different cloud service models is fundamental, evaluating the cost structures on which these models are built is equally essential. The pricing models service providers use vary significantly and play a crucial role in determining the cost-effectiveness of IaaS solutions. The following section delves into the various pricing models available, comparing their advantages and challenges.

2.2.3 Pricing Models

As we had established earlier, Cloud computing has revolutionized how businesses access and manage digital resources. A crucial aspect of cloud computing is the variety of pricing models available, each catering to different business needs and usage patterns. These models enable organizations to align their cloud expenses with actual usage, ensuring cost efficiency by optimizing their cloud expenses and aligning costs with their actual usage and requirements. Several pricing models are presented in the coming section to give a general overview of how they are structured. The pricing models' analysis will help us understand and answer sub-question six related to cost-effectiveness and understand which cost structures can be applied to compare the cost-effectiveness of IaaS providers transparently.

2.2.3.1 Pay-as-You-Go (PAYG)

Pay-as-You-Go (PAYG) is a flexible pricing model where users pay only for the computing resources they consume without long-term commitments. In the PAYG scheme, customers reserve machine time by paying a flat rent for a specific period plus a time-based fee. The total cost for a job depends on its processing time, with a cost structure that includes both a fixed and a variable component [17]. The average cost per unit of time decreases as the job length increases, reflecting increasing returns of scale for the user. This scheme is financially viable for users only if the job execution time is shorter than a threshold determined by the relationship between the job value and the PAYG fees ([17]. This model is particularly suitable for businesses with fluctuating workload demands, allowing for significant cost savings as users are not required to invest in unused capacity. The PAYG model's cost is typically calculated based on the resources used over a given period, such as the number of active hours a virtual machine is.

2.2.3.2 Reserved Vs. Spot Vs. On-Demand Instances.

Reserved Instances offer users a way to reserve cloud capacity for a predetermined period, usually one to three years, in exchange for a significantly reduced rate compared to on-demand pricing[19]. Users pay an upfront cost for the reservation and benefit from lower hourly rates. This model is ideal for applications with predictable usage patterns, providing cost predictability and savings over time.

On the other hand, Spot Instances allow users to bid for unused cloud computing capacity, often available at lower prices than the standard rates. The pricing for Spot Instances fluctuates based on supply and demand for cloud resources. Spot Market (SM) without interruption costs: In the SM scheme, customers bid for machine time on idle resources by submitting a unit price bid and paying the current spot price, which is lower than or equal to the bid price if the bid is above the spot price [17]. This model suits flexible workloads that tolerate interruptions, such as batch processing jobs. Users can use this cost-effective option to run large-scale computing jobs at a fraction of the typical cost. A user's payoff in this scheme depends on their bid relative to the fluctuating spot price, determined by supply and demand.

Moreover, on-demand instances enable users to pay for computing capacity by the hour or second, with no long-term commitments or upfront payments. The OD scheme is like PAYG but operates without a flat rate for reserving machine time. Instead, there is a probability that a service request may be rejected. The cost for the customer under this scheme is a random variable dependent on the probability of service acceptance and the unit cost [17]. This model offers the most flexibility and is ideal for applications with short-term, spiky, or unpredictable workloads, allowing businesses to scale up or down based on current needs without the risk of over-provisioning or incurring significant upfront costs. However, it tends to be less attractive for high-value jobs or jobs requiring longer machine time, as there is a risk of service request rejection.

2.2.3.3 Custom Pricing

Custom Pricing models are tailored pricing structures designed to fit specific business needs. Usually, when all three pricing, combined PAYG, OD, and SM schemes rules coexist. For example, having the SM scheme without interruption costs complements the PAYG and OD models. This means that jobs not viable under PAYG or OD due to negative expected profits can still achieve non-negative expected profits under the SM model [17]. This complementary nature of SM helps attract a broader range of jobs that would otherwise not engage with the platform if only PAYG and OD were available. Custom pricing is particularly relevant for large enterprises or organizations with unique cloud computing needs that do not fit into standard pricing models.

2.2.3.4 Discussion

The various pricing models in cloud computing, including Pay-as-You-Go, Reserved Instances, Spot Instances, On-Demand Instances, and Custom Pricing, cater to different business needs and usage patterns. These models offer flexibility, scalability, and financial efficiency, enabling businesses to align their cloud expenses with actual usage. By understanding these models, organizations can transparently compare the cost-effectiveness of IaaS providers.

In assessing IaaS providers, the Pay-as-You-Go (PAYG), Reserved Instances, and On-Demand Instances models are the most significant. For instance, PAYG and On-Demand are ideal for fluctuating workloads, while Reserved Instances offer cost benefits for predictable and interruptible workloads. These models relate to how IaaS resources are consumed and billed. The PAYG model aligns well with the scalable and flexible nature of IaaS, while Reserved Instances are suitable for predictable, sustained workloads typical in IaaS environments. On-demand instances offer flexibility for variable demands, which is also characteristic of IaaS usage. On the other hand, Spot Instances and Custom Pricing, while relevant, are more situational. Spot Instances are ideal for flexible, non-critical workloads, and Custom Pricing caters to specific, often larger-scale needs.

It is essential to underline that incorporating pricing models into the IaaS assessment framework involves evaluating how these models intersect with aspects like availability, performance, reliability, scalability, and security, which we focus on incorporating in the assessment metrics and, for instance, examining whether Reserved Instances ensure better availability and reliability than the more flexible but potentially less predictable PAYG or On-Demand models. The assessment should also consider if higher-priced options offer enhanced performance and security features and how different pricing structures support scalability. By considering these factors, the assessment framework can provide a comprehensive view of IaaS providers, evaluating not only cost-effectiveness but also how pricing influences the overall quality and suitability of the service. Therefore, incorporating these models into the assessment framework is crucial to enhance the ability to evaluate IaaS providers based on cost efficiency and financial suitability, ensuring a comprehensive and transparent comparison.

Pricing models provide a financial lens for cloud services, but a more comprehensive evaluation requires objective comparison metrics. These metrics are essential for assessing the performance, scalability, and reliability of IaaS offerings. In the following section, we examine the most relevant frameworks guiding decision-making when assessing IaaS providers' services.

2.2.4 Objective Comparison Metrics

In the decision-making process for transitioning to cloud services, measurable data on capabilities we want to measure to answer our research question, like the quality of service, availability, performance, reliability, scalability, security, and cost-effectiveness, is essential. This data, as defined by the National Institute of Standards and Technology (NIST), forms the cornerstone of a "Measured Service," an essential characteristic of cloud computing [9]. Metrics are vital in evaluating cloud services, offering measurable insights into crucial properties such as availability, performance, and scalability. These standardized metrics help organizations make informed decisions about cloud adoption.

Our primary focus at this stage will be on aspects related to assessing IaaS services effectively; hence, metrics that are tailored to cover the four main categories that IaaS focus on Virtualization, networking, storage, and servers. Consequently, understanding how we can formulate and evaluate metrics measuring availability, performance, reliability, scalability, security, and cost-effectiveness metrics related to the Virtualization, networking, storage, and server aspects of the IaaS offering. To achieve that, we will explore the standard service metrics description established, what model to use to assess existing metrics or create a new one, and finally, other frameworks that we can use to improve the current metric evaluation and creation.

2.2.4.1 Cloud Computing Service Metrics Description

Metrics are instrumental in comprehending specific measurements or measurement types related to cloud service properties, as they offer a standardized approach for both describing these measurements and interpreting their outcomes, enhancing the understanding of the properties themselves [9]. NIST Special Publication 500-307 offers a foundational framework for understanding cloud service metrics, defining critical terms like 'cloud service metric' and 'cloud service property.' As emphasized, these definitions ensure a standardized approach to evaluating cloud services, essential for comparing IaaS providers across various dimensions, such as availability and performance [38]. They believe that employing clear and commonly recognized terminology within a specific domain enhances communication effectiveness among stakeholders, thus minimizing the potential for misinterpreting information and aiding in more efficient information comparison and unification[9]. Moreover, they explain the relationship between a property of cloud services and its corresponding metric as pivotal in understanding the service's characteristics in general; see Figure 6 Relationship between Metric Property. Therefore, we can understand how metrics serve as tools to comprehend and quantify these properties, thereby assessing the service's capabilities. They provided knowledge about aspects of the property through its definition and critical information for verification of measurements and measurement results, as emphasized by the publication.[9]

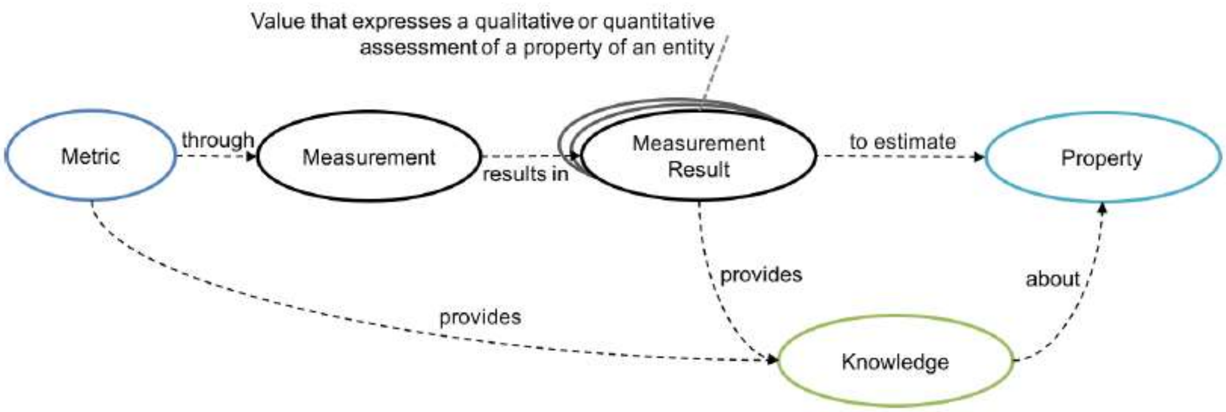


Figure 6 Relationship between Metric Property

Consequently, this enables cloud service providers to effectively articulate measurable aspects of their offerings, as they facilitate a mutual understanding between providers and customers about service deliverables and ensure that cloud service features are quantifiably assessed, aligning with customer requirements.

One last critical point is that the NIST special publication underlines the use of these metrics in different criteria. As emphasized, such metrics can be utilized to assess cloud services. They provided and suggested that different metrics can be used for different processes such as cloud services selection, service agreement, and service verification[9]. We will explore this further in the comings sections and will focus on metrics related to cloud service selections as our primary goal is to focus on the assessment of the adoption stage rather than later stages of the service layer that focus on the service agreement or the service verification after the service has been chosen and adopted.

2.2.4.2 The Cloud Service Metric (CSM) model

The Cloud Service Metric (CSM) model outlined in the NIST Special Publication 500-307 is crucial for understanding and developing meaningful metrics for cloud service evaluation. It breaks down into crucial components: the definition and description of measurement standards, the contextual application of these standards, their practical measurement, and scenario-based usage, incorporating all needed parameters and rules for a clear understanding of specific cloud properties[38]. These metrics are crucial for various functions, such as service selection, agreement enforcement, service monitoring, and accounting. They allow stakeholders to assess, compare, and make informed decisions in scenarios like choosing a cloud service or establishing service agreements, thereby guiding the course of action based on measured cloud service properties. Moreover, the CSM model emphasizes the importance of many characteristics that ensure that the metrics are well-defined and comparable but also adaptable and reusable across different scenarios, making them integral for evaluating cloud services effectively [38].

The CSM Model describes the Cloud Service Metric Ecosystem and focuses predominantly on describing and defining a standard of measurement process[38]. It also introduces several implementations use case scenarios. *Figure 7: CSM Model Scenario Examples [38]* displays several scenario examples of metrics that are defined and evaluated with certain rules and parameters. The CSM Model use cases support several use cases, including those described in this section.

- Express a Description for an Existing Metric (UC1)
- Create a Description for New Metric (UC2)
- Formalize a Metric Description (UC3)
- Generalize a Metric Description (UC4)
- Reuse Metrics Elements Across Metrics (UC5)

Table 27: Saaty's CIr Values for Matrices

Size of Matrix	Random Consistency (CIr)	# of Criteria
1	0	1
2	0	2
3	0.58	3
4	0.90	4
5	1.12	5
6	1.24	6
7	1.32	7
8	1.41	8
9	1.45	9
10	1.49	10

Table 28: Uptime Percentage Scale

Uptime %	Score (1 to 9)	Description
100.00%	9	Products that achieve a perfect availability of 100% receive the highest score of 9, indicating superior reliability with no downtime.
99.9999%	7	Products with an availability of 99.9999% are highly reliable, with minimal downtime, and are awarded a score of 7. This level of availability represents a very high standard, with only a few seconds of downtime per year.
99.999%	5	Availability of 99.999% is also strong, but slightly below the top tier. Products achieving this level receive a score of 5, reflecting their high uptime but with slightly more potential for downtime compared to the top two categories.
99.99%	3	Products with 99.99% availability are still reliable but may experience occasional downtime. These products are scored three, as they offer good performance, though not at the exceptional level of higher categories.
≤99.99%	1	Products with an availability of 99.99% or lower receive the lowest score of 1, indicating that they have significant downtime and are less reliable compared to higher-scoring products.

Table 29: Recovery Feature Scale

Recovery Method	Score (1 to 9)	Description
Backup	*1*	*The least efficient recovery method, requiring manual processes and slow recovery.*
Backup & Restore	*3*	*Slightly better than Backup but still requires manual intervention.*
Auto Backup	*5*	*Automatic backups improve ease of use and reduce downtime, making it more efficient.*
Auto Backup & Restore	*7*	*A highly automated process, with better recovery speed and lower downtime.*
More (Advanced Features)	*9*	*The most advanced method with full automation, fastest recovery speed, and minimal downtime*

Table 30: Latency Performance Scale

Latency (ms)	Score (1 to 9)	Explanation/Use Case
Under twenty ms	*9*	Excellent for real-time applications like gaming, financial trading, and low-latency video streams.
20 - 40 ms	8	Very good for gaming, VoIP, and interactive video calls. Slightly higher but still low enough for most real-time uses.
40 - 60 ms	7	Good for web browsing, video streaming, and VoIP. May cause slight delays in interactive applications.
60 - 80 ms	5	Acceptable for general web use, video streaming, and moderate video conferencing. Noticeable delays in gaming.
80 - 100 ms	2	Borderline acceptable for video conferencing and VoIP. Noticeable delays, especially in interactive content.
Above one hundred ms	1	Poor for real-time applications, VoIP, and gaming. Streaming and general web browsing may still be tolerable.

Table 31: Storage Durability Scale

Durability %	Score (1 to 9)	Explanation/Use Case
99.999999999%	9	Exceptional durability, almost zero data loss risk, suitable for mission-critical storage and archives.
99.99999999%	8	Extremely high durability, excellent for long-term storage of sensitive or high-value data.
99.9999999%	7	Very high durability, typically used in enterprise-class systems where minimal data loss is essential.
99.999999%	6	High durability, ideal for important but not necessarily mission-critical data storage.
99.99999%	5	Reliable durability for general enterprise storage, sufficient for most business applications.
99.9999%	4	Good durability, suitable for backup storage where some minimal risk of data loss is tolerable.

99.999%	**3**	Average durability, acceptable for non-critical data or short-term storage needs.
99.99%	**2**	Lower durability, higher risk of data loss, appropriate for non-essential data storage.
99.9%	**1**	Minimal durability, significant data loss risk, suitable only for temporary or disposable data

Table 32: Data Transfer Fees Scale

Fees after 1st TB	**Score (1 to 9)**	
0/GB	10	Scoring scale was created based on charges for data transfer fees out after the first TB since several providers offer first TB to be fee
$0.01/GB or lower	9	
$0.02 to $0.05/GB	7	
$0.06 to $0.09/GB	5	
$0.10 to $0.12/GB	3	
Higher than $0.12/GB	1	

Table 33: Other Metrics Scales

Metric	**Explanation/Use Case**
CPU Performance	Score is made of the Sysbench and Geekbench CPU test number of operations per second and of the Endurance test number of iterations per hour metric. Converted Score from Third party comparison Platform. Platform Score is a score out of twenty. Tool Score = Portal Score/2
Data Transfer Rates	Score is made of the Sysbench and Fio tests measuring random and sequential storage speeds. Converted Score from Third party comparison Platform. Platform Score is a score out of twenty. Tool Score = Portal Score/2
Network Performance	Score is made of Speed test and Iperf3 network upload and download transfer speeds. Converted Score from Third party comparison Platform. Platform Score is a score out of twenty. Tool Score = Portal Score/2
SLA	Similar Scoring to Uptime Percentage
R Recovery Time	
Recovery Feature	Similar Scoring to Availability Recovery Feature
Auto Scaling	The difference between the Yes and No scores should increase as the weight increases. In other words, the more important a criterion (higher weight), the larger the gap between Yes and No. The range should scale from 1 (small difference for low weight) to 9 (large difference for high weight). **Yes** starts at a base score of 5, and the score increases linearly with the weight (up to 9 at maximum weight, 25%). **No** starts at a lower score (e.g., 1) and the difference from Yes grows with the weight.
Manual Scaling (Vertical or Horizontal)	
Data Encryption Levels	Each product is evaluated based on five features, and the user will indicate whether each feature is available by selecting **Yes or No** (or using checkboxes that return **TRUE or FALSE**). The scoring system assigns a score of 2 points for each feature that is marked Yes

	(**or TRUE**) and zero points for each feature marked **No (or FALSE**). The total score is calculated by summing the points for each feature. Since there are five features, the maximum possible score is ten points (if all features are available), and the minimum score is 0 points (if none of the features are available).
Global Compliance Certifications	The formula is likely used in a compliance certification scoring system where: If the certification count is "10 or more" (indicating a high level of compliance or certification), a fixed score of 10 is assigned. Otherwise, the actual certification number (which could be less than 10) is used as the score.
Cost per Resource Unit	Instead of using a fixed multiplier (like 4 or 8), we can introduce a **dynamic scaling factor** that adjusts based on the difference between the highest and lowest prices. This way, the scoring system is **adaptive**: small price ranges result in smaller score differences, while large price ranges spread the scores more evenly. The formula now adjusts the scoring based on the price range relative to the average price of the products using Dynamic Scaling. This means that: If the price range is small, the score differences will be minimized. If the price range is large, the score differences will be maximized. Multiplying by the Range-to-Average Ratio
Total Cost of Ownership (TCO)	

Table 26:Use Case Description, which provides the definition of each use and its rationale for implementing it. As use one and two are relevant, we believe that use UC4 is most suitable for the thesis research topic as we are trying to create and evaluate general metrics descriptions related to metrics measuring availability, performance, reliability, scalability, security, and cost-effectiveness metrics related to the Virtualization, networking, storage, and servers' aspects of the IaaS offering.

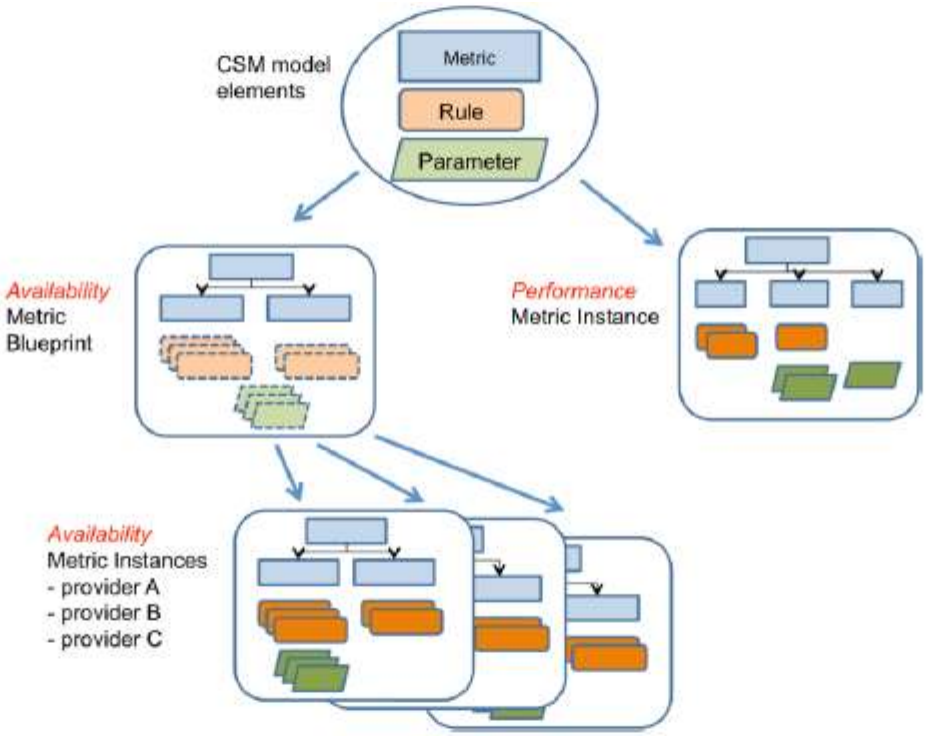


Figure 7: CSM Model Scenario Examples [38]

2.2.4.3 Service Measurement Index (SMI)

The Cloud Services Measurement Initiative Consortium (CSMIC) was established to create universally recognized metrics for evaluating cloud computing services' benefits and risks, as it is clear that the industry is currently in need of a standard evaluation method with the cloud services becoming more complex and more diverse[4]. As a result of this consortium, the work has started on developing the Service Measurement Index (SMI), a standardized framework for consistently measuring and comparing various cloud and non-cloud services[4]. As defined, *"it is a set of business-relevant Key Performance Indicators (KPIs) that provide a standardized method for measuring and comparing a business service regardless of what service is internally provided or sourced from an outside company"*[11]. The SMI focuses on seven major categories, each with several attributes like Accountability, Agility, Assurance, Financials, Performance, Security, Privacy, and Usability, covering 51 attributes[11]. Considering the growing concerns around provider capabilities, costs, security, and the anticipated flexibility cloud services offer, this initiative is especially relevant. However, the SMI initial framework presents measures for these attributes to provide a comprehensive understanding of cloud service evaluations, but it is still in the process of formulating the definition of the KPIs and measures related to several attributes. Some will be service-specific, while others will apply to all services. The SMI framework is a comprehensive framework still being developed and can provide beneficial insight into related metrics for IaaS assessment. However, it is important to highlight that it is still under development and has yet to be validated for more practical use.

Likewise, Bardsiri and Hashemi highlight cloud computing characteristics and cloud services metric evaluation and evaluated metrics associated with three aspects of Cloud services: Performance, Economics, and Security [21]. However, they focus on and emphasize capabilities that reside on top of physical cloud infrastructure related to the application, middleware, and Operating System described in Figure *5*: Cloud Services Comparison[1]. As mentioned before, our focus will be on Virtualization, networking, storage, and servers as these aspects are more specifically related to IaaS offering.

2.2.4.4 Relevant Frameworks & Tools of IaaS Assessment

This section on Relevant Frameworks & Tools for IaaS Assessment focuses on various non-IaaS-specific tools and frameworks that can significantly contribute to validating the evaluation metrics we aim to develop. This includes the Gartner Magic Quadrant for Strategic Cloud Platform Services, offering a comprehensive analysis and comparison of cloud service providers; the Information Technology Infrastructure Library (ITIL), which aligns IT services with business needs; ISO/IEC 20000, a standard demonstrating excellence in IT management; and the Microsoft Cloud Adoption Framework (CAF), guiding organizations in their cloud adoption strategies. While not directly tailored for IaaS assessment, these tools and frameworks provide critical insights and best practices applicable to developing robust IaaS evaluation metrics. The subsequent sections will delve deeper into the specifics of each tool and framework.

Information Technology Infrastructure Library (ITIL) is a set of practices for IT service management (ITSM) that focuses on aligning IT services with the needs of the business[22]. It plays a crucial role in supporting businesses by providing a flexible and stable framework for IT service management, especially in aligning IT goals with business goals, cost tracking and optimization, streamlined service delivery, flexibility, collaboration, and value creation[39]. We can establish the relationship with the intended metrics we are striving to create as the newly updated ITIL version 4 focuses on fostering digital transformation, artificial intelligence, cloud computing, and DevOps. There are five key stages with twenty-six processes in the newly updated version 4. These stages focus on service strategy, service design, service transition, service operation, and continual Service Improvement[23].

Nevertheless, the focus here is the service design stage as it can be related to aspects we are focusing on, especially aspects related to cloud computing assessment. As described, the service design stage focuses on designing services and processes related to service catalog management, availability management, information security management, service level management, and capacity management [23]. These processes can be related and linked to validating the aspects we are trying to define in establishing the evaluation metrics. Dabade also discusses these processes as he explains the service delivery aspects of the ITIL framework and focuses on service level capacity, contingency planning, availability management, and cost management for IT services [22].

As we dive deeper into the IT service management field, we can highlight the importance of the ISO/IEC 20000 to this sector. ISO/IEC 20000 is an international standard that specifies requirements for an IT service management system to establish, implement, maintain, and continually improve a service management system (SMS)[25]. It can be seen as a formal certification demonstrating adherence to ITSM principles while providing a robust foundation for effective and efficient IT service management. It complements any chosen framework for adapting best practices, focusing on aligning IT services with business needs and emphasizing continual improvement.

As we continue, it is imperative to emphasize the importance of the Gartner Magic Quadrant for Strategic Cloud Platform Services, as it provides a well-established definition of strategic cloud platform services and divides them into categories that cover most businesses' needs and requirements in order to choose the right service provider [28]. It is designed to offer vital insights for businesses choosing a Strategic Cloud Platform Service (SCPS) provider, as it enables organizations to assess the strengths and weaknesses of each provider based on criteria crucial to their business [28]. Additionally, it offers information about provider programs and resources, assisting organizations in their digital transformation journey. This comprehensive comparison and evaluation tool is instrumental for businesses in making informed decisions about selecting a cloud service provider that best meets their specific needs. However, Gartner Magic Quadrant for Strategic Cloud Platform Services is not accessible for all businesses as it needs to be purchased and published annually[28]. As Figure 7: Magic Quadrant for Strategic Cloud Platform Services [28] shows below, CSPs are classified into four main categories: Leaders, Challengers, Niche Players, and Visionaries, covering the top 20 CSPs in that year.

Moreover, the quadrant highlights all the listed vendors' strengths and cautions regarding their significant services. Nonetheless, it is important to emphasize that this evaluation has a different kind of evaluation methodology that considers the whole service operations of the CSPs and assesses the whole business rather than looking at a more specific evaluation of the cloud service offering. It can complement a business choice for a CSP selection as it may present additional insight into the selection choice. Insights about aspects that have not necessarily been evaluated in the comparison metrics.

Continuing our exploration of frameworks relevant to IaaS assessment, we now turn to the Microsoft Cloud Adoption Framework (CAF). This framework offers a comprehensive guide for organizations embarking on cloud adoption. Unlike ITIL and ISO/IEC 20000, which focus on service management standards, CAF explicitly addresses the strategies and best practices for successful cloud implementation. It provides tools, documentation, and best practices to help organizations align their cloud strategies with business objectives, optimizing the benefits of cloud services[24]. This framework is beneficial for understanding the organizational and technical considerations in cloud adoption, adding another dimension related to developing the evaluation metrics of IaaS services.

Applying principles from ITIL, ISO/IEC 20000, and Microsoft CAF to IaaS metrics is essential in relevance to metric evaluation establishment. ITIL, emphasizing service delivery, can inform the development of metrics focusing on IaaS service reliability and consistency. This could involve measuring uptime, response times, and incident resolution rates. At the same time, ISO/IEC 20000's IT service management systems standards can guide the creation of metrics for service quality and customer satisfaction in IaaS. Lastly, with its comprehensive cloud strategy guidelines, the CAF can contribute to metrics assessing strategic alignment and cloud adoption efficiency. Together, these frameworks provide a well-rounded approach to developing robust and valid IaaS assessment metrics.

2.2.6 Gaps in the Literature

While the current literature on cloud services is comprehensive, it often needs more focus on Infrastructure as a Service (IaaS). Much of the research and existing frameworks take a broad approach, addressing cloud services without considering the unique aspects of IaaS. This generalized perspective creates a significant gap, as IaaS presents distinct challenges that must be fully addressed in current models. Specifically, the literature often fails to delve deeply into critical areas such as performance optimization, cost comparison, and specialized security needs, which are crucial for IaaS evaluation.

IaaS requires tailored performance metrics due to its foundational role in delivering computing infrastructure. Unlike SaaS or PaaS models, where performance is often managed at the application or platform level, IaaS users need to assess factors such as CPU efficiency, network speed, and storage performance directly. Current frameworks seldom provide sufficient tools to accurately measure or compare these aspects across different providers, leaving a gap in understanding how well infrastructure services perform in real-world scenarios.

Another critical challenge is the need for standardized cost comparison methods tailored for IaaS. While cloud service providers offer pricing calculators, these tools are often influenced by the provider's offerings and may not offer an objective or transparent view. This leads to difficulties in comparing total cost of ownership (TCO) across providers, particularly when factoring in complex pricing structures such as pay-as-you-go models, reserved instances, and spot pricing, and the difficulties most organizations face with hidden costs such as data transfer fees. Moreover, smaller organizations, in particular, may struggle to interpret or apply these models due to the lack of practical guidance.

Moreover, the literature frequently overlooks the specialized security needs associated with IaaS, which differ from SaaS and PaaS models. IaaS clients bear more responsibility for securing their environments, making it crucial to evaluate metrics like data encryption, disaster recovery, and compliance with regulatory standards. Current frameworks either generalize security assessments or fail to address the granular security requirements to protect infrastructure-level services.

A significant gap is the need for validated, practical metrics for organizations evaluating IaaS offerings. Many metrics explored in the literature are too numerous or too few, often influenced by major cloud providers, leading to inconsistencies. The lack of straightforward, applicable metrics complicates the decision-making process for smaller organizations or those new to cloud computing. Additionally, the literature often fails to address how these metrics may differ across providers, making it difficult for users to obtain a true "apples-to-apples" comparison.

Another area for improvement is focusing on post-adoption stages, such as service-level agreements (SLAs) and service verification, rather than the initial selection and adoption of IaaS. Most frameworks reviewed are geared towards managing cloud services after implementation, with little attention paid to the metrics needed during the crucial decision-making phase. This creates a void in the understanding and development of objective metrics that can guide organizations in the initial stages of evaluating IaaS providers.

While these gaps highlight areas where the current literature needs to improve, the emergence of recent technologies, such as edge computing and artificial intelligence (AI), further complicates the landscape. These advancements present additional challenges for IaaS providers, such as handling distributed data and supporting real-time processing, which are not yet fully addressed in the existing frameworks. Future research must focus on developing new IaaS-specific metrics that consider these technologies, ensuring that evaluations remain relevant in an evolving cloud environment.

2.2.7 Summary & Conclusion

To summarize, the literature review explored several topics related to cloud services and the evolution of cloud computing, focusing on Infrastructure as a Service (IaaS) and the key factors that affect its evaluation, highlighting the increasing significance of IaaS in modern business operations. The analysis emphasized the differences between cloud service models—SaaS, PaaS, and IaaS—particularly in distributing responsibilities between providers and clients. This distinction is vital because IaaS gives clients greater control and responsibility over their infrastructure, necessitating more specific and detailed evaluation metrics. Additionally, the review examined major pricing models and how this impacts the cost-effectiveness of IaaS offerings. The need for transparent, standardized methods for cost comparison, particularly when assessing total cost of ownership (TCO) in IaaS, emerged as a key finding. Several gaps specific to IaaS were identified and highlighted that need further exploration.

To conclude, the literature review underscores the need for a more focused approach to IaaS evaluation, which moves beyond broad cloud computing frameworks to address the unique characteristics of IaaS offerings. The current literature provides valuable insights into cloud computing models and frameworks but often needs more specificity to evaluate IaaS effectively. This research identifies critical gaps in performance measurement, cost comparison, and security assessments that must be addressed to develop robust, practical metrics for IaaS adoption.

Moving forward, developing, and validating IaaS-specific metrics that offer clear guidance for organizations, particularly those with limited resources, in selecting cloud infrastructure is essential. Furthermore, as emerging technologies like edge computing and AI evolve, these metrics must adapt to reflect the changing landscape of cloud computing, ensuring that they remain relevant and valuable in future evaluations. Therefore, the contribution of the thesis will focus on developing an evaluation framework based on established metrics that measures and assesses the selected aspects we placed for the evaluation framework. The framework will emphasize on utilizing the AHP method for the decision-making process. Chapter 4 will delve deeper on how this framework will be structured and established.

Chapter 3: Research Methodology

This chapter details the methodology employed to achieve the research objectives and answer the research questions outlined in Section 1.4. The first section outlines the overall methodological approach used in this research. Subsequently, we summarize and align our research outline with the research methodology.

3.1 Research Design

This research adopts the Design Science Research Methodology (DSRM) for conducting design science research in information systems [5]. DSRM is particularly suited for research aimed at creating and evaluating IT artifacts to solve identified organizational problems. To link this to the context of the thesis research topic, the artifact will be the objective metrics used to assess IaaS cloud services. The DSRM process includes six distinct activities, as shown in Figure 8: Design Science Research Methodology (DSRM) Process Model [5]:

- Problem Identification and Motivation
- Objectives for a Solution
- Design and Development
- Demonstration
- Evaluation
- Communication

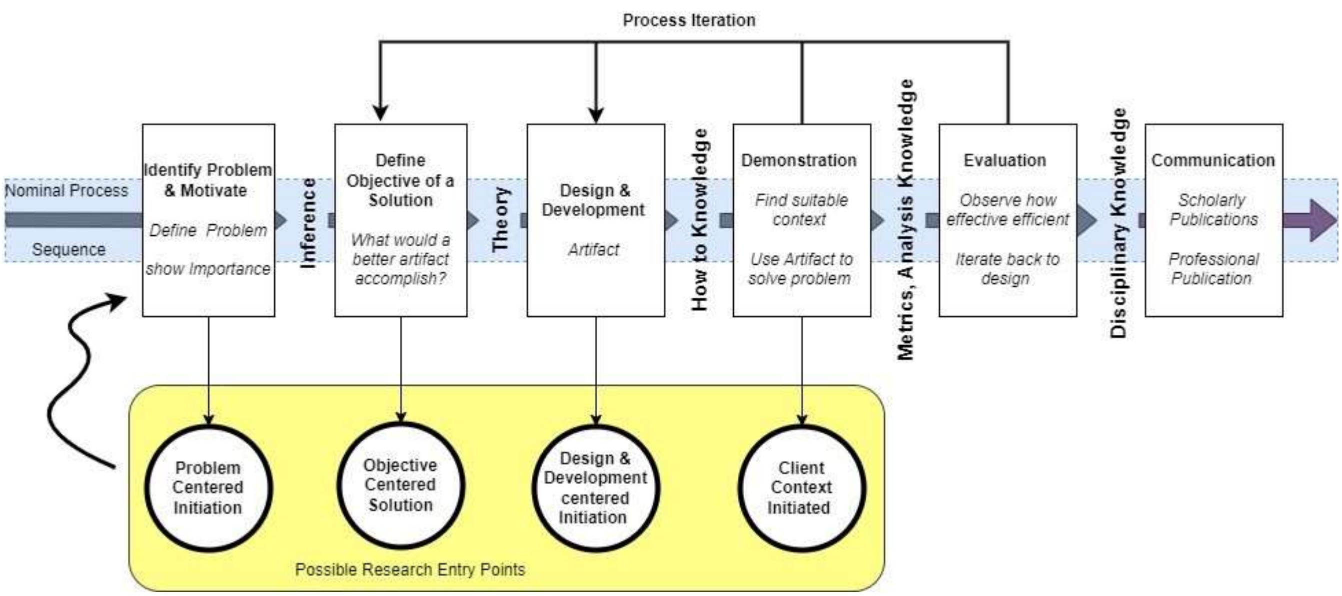


Figure 8: Design Science Research Methodology (DSRM) Process Model [5]

The following section will elaborate on these activities, detailing how they will be applied in developing objective comparison metrics for IaaS. This approach ensures a rigorous, systematic process for creating and validating the research artifacts, aligning with the thesis goals.

3.2 Applying DSRM to IaaS Comparison Metrics Development

Figure 9: Mapping Chapters to DSRM illustrates how DSRM was utilized in the thesis and how it is linked to its chapters.

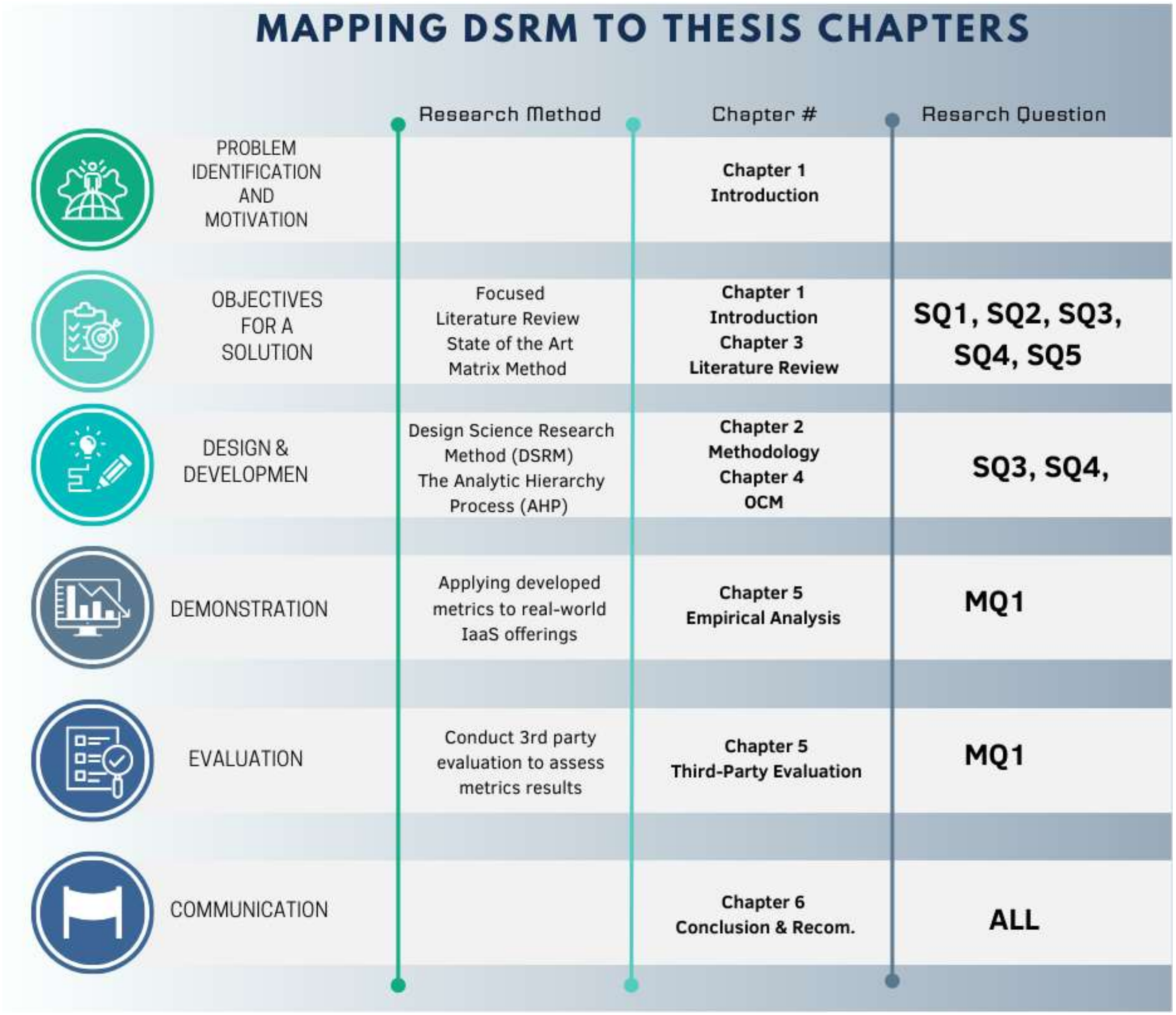


Figure 9: Mapping Chapters to DSRM

3.2.1 Problem Identification and Motivation:

In Chapter 1, we identified the specific challenge of evaluating and comparing cloud services based on several characteristics and established that this research would focus on IaaS services. The need for a comprehensive and objective framework to aid businesses in making informed cloud adoption decisions is crucial, especially since a simple, easy-to-use metric is lacking.

3.2.2 Objectives for a Solution:

In Chapter 1, the research question section, we defined the aim of this research and set the goals of the IaaS comparison metrics. These objectives should address the gaps in current evaluation methods, aiming to provide a more holistic, balanced, and objective approach to assessing IaaS offerings. Therefore, this research focuses on developing the objective metrics and analyzing current service providers' offerings based on these metrics, then validating the results shown in the literature and empirical analysis with a third-party consultant. We performed an extensive focused literature review to answer our primary questions. We derived supporting sub-questions to identify the importance and necessity of bridging the gap between literature and current practice, as discussed in Chapter 2, Literature Review.

3.2.3 Design and Development:

In Chapter 4, Objective Comparison Metrics, we will develop the actual comparison metrics framework based on the supporting literature and address the missing elements to create the objective metrics we strive to achieve. This involves designing the metrics based on identified objectives, ensuring they are comprehensive, objective, and applicable across various IaaS offerings. This stage also includes creating methodologies for data collection and metric application.

3.2.4 Demonstration:

In Chapter 5, the empirical analysis, we will demonstrate the application of the developed metrics to real-world IaaS offerings. This will involve presenting a fictional case study where ten different IaaS products will be evaluated using the framework to highlight how the metrics work in practical scenarios and how they can be used to compare different IaaS products based on different criteria set by an organization.

3.2.5 Evaluation:

In Chapter 6, validation, and evaluation, we will evaluate the effectiveness of the comparison metrics in providing clear, objective, and meaningful assessments of IaaS services by presenting the study case of ten IaaS products. In addition, the tool, and a demo video about it will be presented to industry experts. This step may involve gathering feedback from industry experts to analyze the outcomes of the Demonstration and assess the tool's usability and functionality.

3.2.6 Communication:

In Chapter 7, we will present the findings, the developed framework, its practical applications, and its contributions to the field. This involves documenting your thesis's research process and results and disseminating the findings through academic publications, presentations, or industry reports. It will also include a discussion of the implications and recommendations of the thesis research findings.

3.3 The Analytic Hierarchy Process (AHP)

The Analytic Hierarchy Process (AHP) is a structured technique for organizing and analyzing complex decisions based on mathematics and psychology developed by Thomas L. Saaty in the 1970s [40]. It involves breaking down a decision into a hierarchy of more easily comprehended sub-problems, each of which can be analyzed independently, as these elements of the hierarchy can relate to any aspect of the decision problem—tangible or intangible, carefully measured or estimated, well or poorly understood [41]. As explained, AHP helps to capture both subjective and objective aspects of a decision by using pairwise comparisons and then synthesizing these results to determine the relative weights of the decision criteria and evaluate alternative options [40], [42], [43]. Another critical aspect of the AHP is its structured methodology for assessing the developed metrics. The methodology offers a calibrated numeric scale to assess both quantitative and qualitative performances, in which this scale extends from one-ninth, denoting minimal importance, through one, indicating equal importance, to nine, signifying utmost importance, thus encompassing the full range of comparative analysis [43]. Figure 10: AHP Methodology shows the steps applied in conducting the AHP. These steps will be used as guidelines for assessing and selecting metrics for comparison.

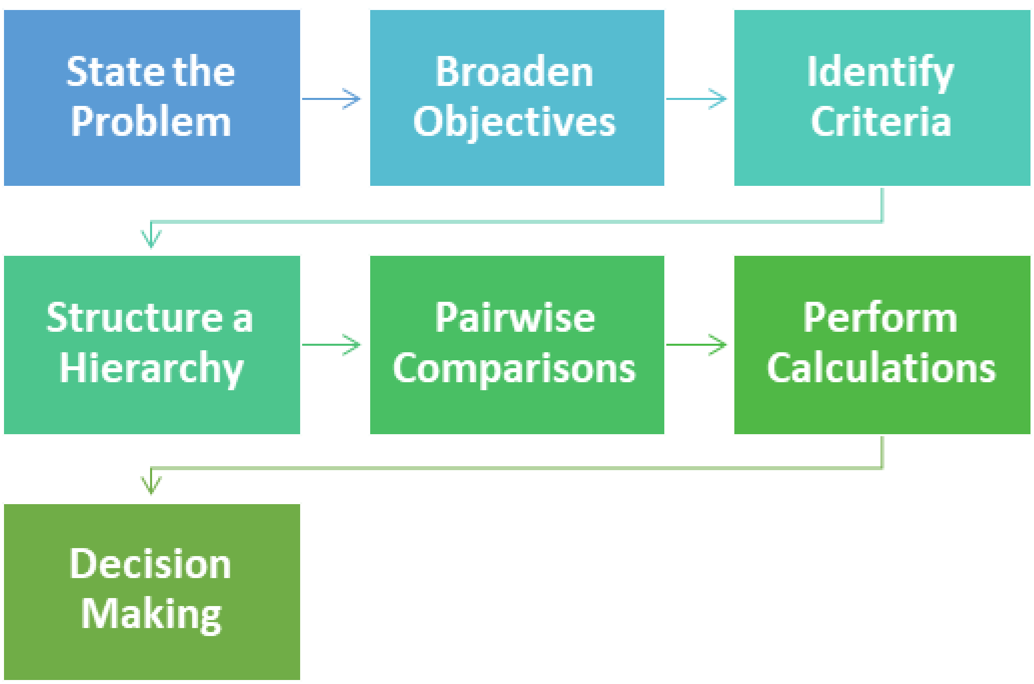


Figure 10: AHP Methodology

The (AHP) choice of method is due to its significant relevance to thesis topics in the IaaS assessment process. As presented, the (AHP) has been utilized in many fields with different themes. Vaidya and Kumar emphasize how it relates to selection, evaluation, benefit-cost analysis, allocations, planning and development, priority and ranking, and decision-making with study cases [43]. Moreover, the (AHP) is a vital methodology for developing IaaS assessment metrics due to its systematic approach to complex decision-making. AHP excels in breaking down multifaceted problems into manageable components, enabling a thorough evaluation of diverse factors, such as cost, performance, availability, scalability, and security. Its pairwise comparisons and a structured hierarchical model allow for a detailed and nuanced assessment, ensuring that all critical aspects of IaaS services are considered. By quantifying subjective judgments, AHP provides a clear, objective framework for comparison, making it an indispensable tool in creating robust and comprehensive IaaS evaluation metrics. Chapter 4, Objective Comparison Metrics, will demonstrate how the AHP is used in more detail.

Chapter 4: Designing an Objective Comparison Metrics Framework

This Chapter introduces an evaluation framework for IaaS providers using the AHP method. It defines the requirements needed to establish the framework and how feasible it is to achieve it. It also outlines the steps for creating the AHP hierarchy, assigning weights to criteria, and performing pairwise comparisons to ensure consistency. Additionally, it explores the feasibility of implementing these metrics in real-world scenarios and addresses potential challenges and limitations in data collection and evaluation processes.

4.1 Requirements

To better understand the evaluation framework and the metrics it will be assessing, it is crucial to establish precise requirements to be measured. We have already established that each of the main functions of IaaS offering needs to be evaluated in relation to the aspects related to these functions to assess the service offerings objectively. Therefore, we can conclude and determine that a clear list of requirements needs to be established and should be as follows:

The first requirement is that the four primary functions: virtualization, storage, networks, and servers must be evaluated in relation to the established aspects and the suitably chosen metric. The framework must evaluate the main components of the IaaS offering we explored earlier. The evaluation must be conducted in relation to the defined aspects of the evaluation framework: Availability, Performance, Scalability, Reliability, Security, and Cost-effectiveness.

Moving to the second requirement, these six main aspects must be assessed using well-established and accepted metrics; the number of metrics for each aspect varies depending on the available metrics. To achieve this requirement, the framework needs to define each of the aspects of availability, performance, scalability, reliability, security, and, finally, cost-effectiveness and the recognized metrics for measuring and evaluating them concerning the component they are measuring, for example, the availability of storage, performance of computing power and virtual machines, failure rates of servers, or cost per dollar on the service.

This brings us to the third requirement, the metrics must be carefully evaluated and established for each of the six evaluated aspects to ensure reliability and validity. Therefore, Availability, Performance, Scalability, Reliability, Security, and Cost-effectiveness metrics must be assessed and chosen based on academic validity and industry benchmarks.

Figure 11: IaaS Evaluation Framework Mechanism illustrates the requirement relationship between the aspects and the components to assess IaaS based on six criteria: Availability, Performance, Reliability, Scalability, Security, and Cost-Effectiveness. Each criterion focuses on key components like Virtualization, Storage, Networks, and Servers to ensure a comprehensive evaluation.

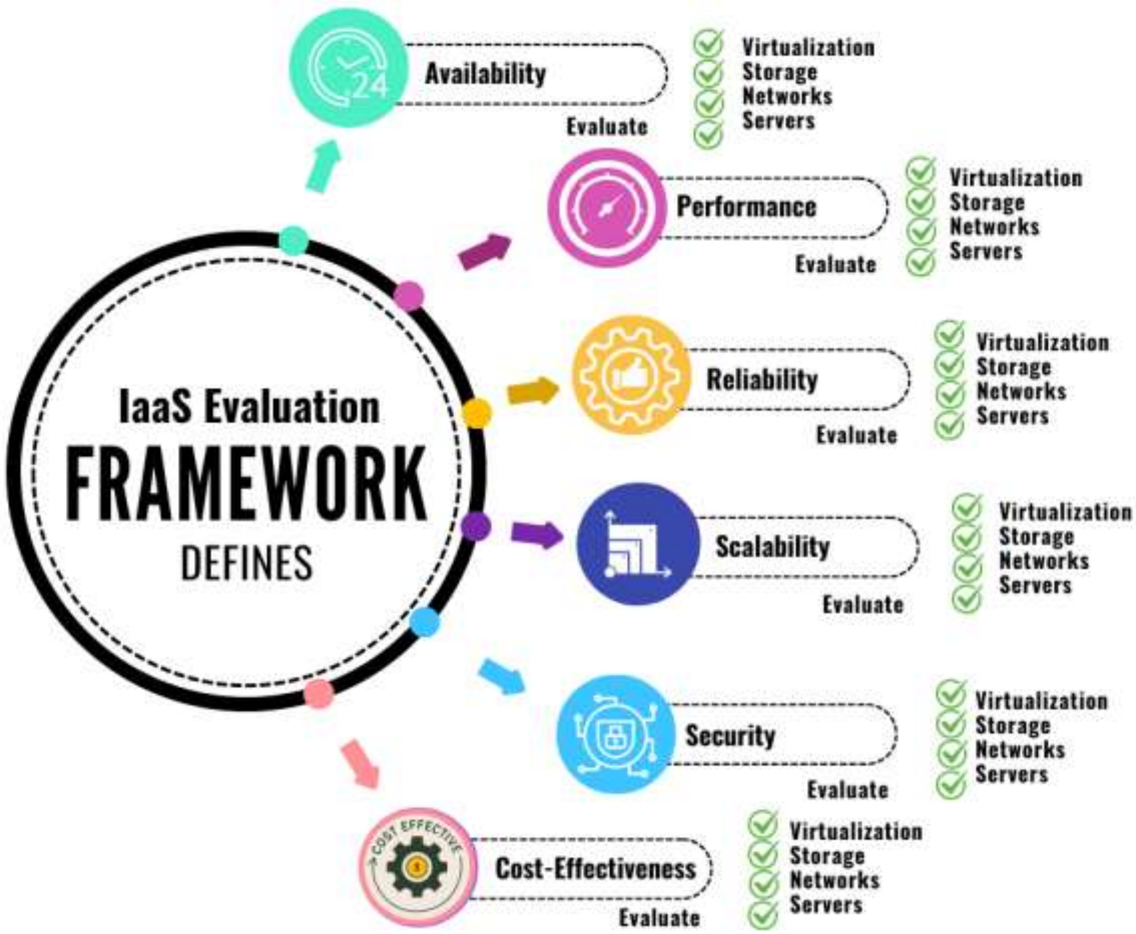


Figure 11: IaaS Evaluation Framework Mechanism

In conclusion, developing a robust framework of IaaS offerings depends on adequately defining the aspects we are measuring and the metrics used for its measurements concerning the IaaS components we are evaluating. As we mentioned, the IaaS components of Virtualization, Storage, Networks, and Servers must be evaluated based on the six critical aspects: Availability, performance, scalability, reliability, security, and cost-effectiveness that we have defined. By doing so, we can assure an objective and comprehensive assessment of the service offerings and provide a detailed and accurate evaluation of the IaaS products. Thus, aligning it with industry standards and customer expectations.

4.2 Metrics Definitions

With the foundation of our evaluation framework firmly established, we will now delve into the specific metrics for each aspect. The following sub-sections of each metric will define the metrics, and their requirements based on benchmarked metrics in the industry. We can establish that the following metrics measure and evaluate each of the properties of the measured aspects. We will dive deeply into each metric in more details, however briefly based on the literature provided for each metric, we can conclude the following:

Table 5: Metrics Area of Focus

Metrics	Focus
Availability Metrics	The uptime and accessibility of virtual machines, storage systems, and network components.
Performance Metrics	The speed and efficiency of computing resources, including CPU performance, memory usage, and data transfer rates.
Scalability Metrics	The ability to dynamically scale resources up or down in response to workload demands.
Reliability metrics	The reliability and dependability of storage systems, including data integrity and fault tolerance.
Security Metrics	The protection of data and applications, including network security, compliance, data encryption, and access controls.
Cost-effectiveness metrics	The financial efficiency of resource utilization, considering factors like CPU hours, storage duration, and bandwidth consumption.

The following sections will explore in more detail each metric category, providing detailed definitions, requirements, and examples of the metrics used to evaluate the respective aspects. This approach will ensure a thorough understanding of the metrics we placed as the cornerstone of the evaluation framework.

4.2.1. Availability Metrics

The first aspect we will explore is Availability, focusing on metrics that measure the uptime and accessibility of IaaS resources. As per the articles studied[41], [42], [43], [44], [45], these metrics focus on ensuring the service's availability to users and can significantly impact their satisfaction and operational efficiency. This impact has encouraged many studies to explore the measurement and comparison of availability in virtualized systems, storage, and network components. Kim developed an availability model for virtualized systems, incorporating hardware and software failures and high-availability services[41]. Ford characterized the availability of cloud storage systems, considering factors such as data placement and replication strategies [42]. Nabi surveyed availability solutions in the cloud, identifying a need for a standardized approach to evaluation and comparison[43]. Matos proposed a sensitivity analysis approach to identify critical parameters for improving system availability, focusing on host and application failure rates [44]. Moreover, Amazon provides detailed documentation on measuring and evaluating availability [45]. These studies collectively highlight the complexity of measuring and comparing availability and the need for standardized evaluation methods, and we can conclude the following definition, requirement, and formula:

Table 6: Availability Metric:

Metric	Definition	Requirement	Formula
Uptime Percentage	Measure the percentage of time the IaaS services are operational. This should be based on service uptime over the total required operational time.	Track the total hours of uptime and downtime over a specified period to calculate the availability percentage.	Availability = Service Uptime / Service Uptime + Service Outage
Mean Time Between Failures (MTBF):	The average time between system failures. This metric helps understand the reliability of the infrastructure.	Calculate the total operational time divided by the number of failures over a given period.	MTBF=Total Operational Time / Number of Failures
Mean Time to Repair (MTTR):	The average time taken to repair the system after a failure. It indicates the efficiency of the recovery process	Measure the total time spent on repairs divided by the number of repairs.	MTTR= Total Repair Time / # of Repairs
Recovery Time	The average time required to restore service after a failure, including detection and repair time.	Monitor and log the time from failure detection to the restoration of service.	Recovery Time = Detection Time + Repair Time

4.2.2. Performance Metrics

Next, we will examine Performance metrics, which evaluate the speed and efficiency of computing resources. These metrics capture the attention of most potential users as they are essential for assessing how well the infrastructure meets performance expectations and manages workloads which created a fertile ground for many studies proposing methods for measuring the performance of cloud service in general, and Infrastructure-as-a-Service (IaaS) offerings in particular. Lenk and Ahuja both emphasize the need for accurate performance measurement, with Lenk introducing a method that considers the type of service running in a virtual machine. Ahuja uses system-level benchmarks to compare IaaS cloud services[46], [47]. Sajjad and Stephen on the other hand, both focus on specific performance metrics, with Sajjad evaluating CPU, memory, disk, and network performance in OpenStack and Windows Azure, and Stephen comparing various monitoring tools regarding SLA parameters[48], [49]. These studies collectively highlight the importance of accurate and comprehensive performance measurement in the IaaS context, and we can conclude the following definition and requirement:

Table 7: Performance Metric

Metrics	Definition	Requirement
CPU Performance	Measures the processing power of virtual machines (VMs), typically in terms of CPU cycles per second or CPU utilization.	Monitor and log CPU utilization rates, peak usage times, and average load
Memory Usage	Measures the amount of RAM used by applications and services running on VMs.	Track memory allocation, peak memory usage, and average memory consumption over time.
Data Transfer Rates	Measures the speed of data transfer between storage systems and VMs.	Record data transfer speeds (in Mbps or Gbps) during different operational periods.
Network Latency	The time taken for a data packet to travel from the source to the destination within a network.	Measure and log the time delay (in milliseconds) between sending and receiving data packets.
Network Performance	Measures data transmission efficiency across the network, including bandwidth, jitter, and throughput.	Monitor bandwidth, jitter, and throughput using tools like Iperf; ensure tests are performed under consistent network conditions

4.2.3. Reliability Metrics

Since Reliability metrics ensure the consistency and dependability of the IaaS overall offering, and commonly show a correlation between availability and reliability, we will examine this correlation in more details as we study the metrics related to both. These metrics are crucial for maintaining data integrity and fault tolerance, thus ensuring the availability of the services as well. Academically, various methods have been proposed to measure reliability in IaaS offerings. Alannsary suggests using web server logs to analyze failures and workload [50], while Kozlovszky emphasizes the need for a comprehensive parameter tree to evaluate IaaS cloud systems [51]. Carvalho highlights the importance of integrating dependability into cloud benchmarks [52], and Cotroneo introduces a method for state-based robustness testing of IaaS platforms[53]. These approaches collectively underscore the complexity of measuring reliability in IaaS and the need for a multi-faceted evaluation, and we can conclude the following definition, requirement, and formula:

Table 8: Reliability Metric

Metric	Definition	Requirement	Formula
Failure Rate	The frequency of failures over a specified period. A lower failure rate signifies higher reliability.	Track the number of failures and the total operational time to calculate the failure rate.	Failure Rate= Number of Failures / Total Operational Time
Mean Time Between Failures (MTBF):	The average time between system failures. This metric helps understand the reliability of the infrastructure.	Calculate the total operational time divided by the number of failures over a given period.	MTBF=Total Operational Time / Number of Failures
Mean Time to Repair (MTTR):	The average time taken to repair the system after a failure. It indicates the efficiency of the recovery process	Measure the total time spent on repairs divided by the number of repairs.	MTTR= Total Repair Time / # of Repairs
Recovery Time	The average time required to restore service after a failure, including detection and repair time.	Monitor and log the time from failure detection to the restoration of service.	Recovery Time = Detection Time + Repair Time
Service Reliability Percentage	The percentage of time the system performs without failure	Measure operational time and failures to determine overall reliability.	Service Reliability= (1−Number of Failures / Total Operational Time) ×100

4.2.4 Scalability Metrics

Likewise, scalability metrics measure the infrastructure's ability to dynamically scale resources up or down in response to workload demands. These metrics are vital for ensuring that the system can handle varying loads efficiently, and considered to be the core of IaaS offering as organizations main incentive to utilize cloud services is how easy and fast is it for them to scale their business operations. However, the concept of scalability in IaaS offerings, particularly in cloud computing, is complex and multi-faceted, it involves dynamically adjusting resources in response to workload demands to maintain cost-effectiveness and optimal performance, which can be achieved through various scaling options, such as horizontal and vertical scaling, with different cost and performance implications [54]. A model-driven approach that characterizes the workload and evaluates different scaling options has been proposed to measure scalability [54]. Additionally, more studies were conducted on the use of adaptive autoscaling algorithms, such as Libra, can further enhance the scalability of IaaS offerings by automatically detecting and adjusting resource sets based on workload changes [55], and we can conclude the following definition and requirement:

Table 9: Scalability Metric

Metric	Definition	Requirement
Scale-Up (Vertical Scaling)	Adding more resources (e.g., CPU, memory) to an existing virtual machine (VM).	Measure how service time improves when additional resources are added to a VM.
Scale-Out (Horizontal Scaling)	Adding more instances of VMs to distribute the workload.	Monitor how performance changes as the number of VMs increases.
Combined Scale-Up and Scale-Out	Using both vertical and horizontal scaling to optimize resource allocation and performance.	Analyze cost and performance trade-offs to determine the optimal mix of scaling methods.
Cost Effectiveness	Measuring the efficiency of scaling in terms of resource usage versus performance gains.	Compare the cost of resources used to the performance improvement achieved.
Elasticity	The ability of a system to scale resources up or down quickly in response to changes in demand.	Measure the speed and efficiency of scaling operations

4.2.5 Security Metrics

Moving to Security metrics, which evaluate the measures taken to protect data and applications. These metrics are essential for ensuring compliance with industry standards and maintaining data security; metrics such as Incident Response Time (the time taken to respond to security incidents), Data Encryption Levels (the extent to which data is encrypted both at rest and in transit), and finally Compliance Certifications. The compliance of the IaaS services with industry standards and regulations is the most looked-at aspect and is considered a crucial measure to evaluate the security metrics of the IaaS offerings in a more straightforward, understandable way without diving into the complex layers of evaluating the overall security of the IaaS offering. PCI DSS Level 1 (Payment Card Industry Data Security Standard) [56], SOC 2, a framework for managing and safeguarding data [57, p. 2], ISO/IEC 27001[58], ISO/IEC 27017[59], ISO/IEC 27018[60], ISO/IEC 27701[61] which are standards for information security and privacy management, are the most commonly used certifications for evaluating online services including all cloud offering services. When it comes to data encryption levels, features such as Data-at-Rest Encryption[62], Data-at-Transit Encryption[63], Key Management Services (KMS)[64], Certificate Management Services, and Hardware Security Modules (HSM)[65] are checked if provided by the cloud service providers for the product being evaluated. In conclusion, these features provide a more profound and comprehensive evaluation of the encryption and security mechanisms an IaaS provider uses. Thus, we can conclude the following definition and requirement:

Table 10: Security Metric

Metric	Definition	Requirement
Data-at-Rest Encryption	Evaluate whether the IaaS provider offers encryption to protect sensitive information from unauthorized access if the storage medium is compromised in its infrastructure (e.g., on virtual disks or in storage services).	The provider should offer AES-256 or equivalent encryption for all data at rest.
Data-in-Transit Encryption	Evaluate whether the IaaS provider offers encryption, ensuring that sensitive information is not exposed to attackers while being transmitted across networks, for data as it moves between systems (e.g., between servers, over networks).	The provider should offer TLS/SSL encryption for all data in transit.
Key Management Services (KMS)	Assesses whether the IaaS provider offers a Key Management Service to handle encryption keys securely. This service reduces the risk of crucial exposure and improves security controls by enabling secure key management without manual handling.	The provider should offer a fully managed KMS solution that integrates with cloud services and allows centralized key management.
Certificate Management Services	Check if the IaaS provider offers a Certificate Management Service to issue, renew, and manage digital certificates. A managed service to automate certificate issuance and management is critical for maintaining secure communications.	The provider should offer automated certificate issuance, renewal, and revocation to manage public and internal certificates.
Hardware Security Modules (HSM)	Evaluate whether the IaaS provider offers access to Hardware Security Modules (HSMs) for securely storing and managing cryptographic keys. HSMS provides a highly secure method for managing encryption keys by isolating them from the system and protecting against software-based attacks.	The provider should offer dedicated HSM services that comply with standards like FIPS 140-2 Level 3 or equivalent.

4.2.6 Cost-Effectiveness Metrics

Finally, cost-effectiveness metrics are needed to assess the financial efficiency of resource utilization. These metrics are crucial for ensuring that the IaaS services provide value for money. This is a straightforward metric that all cloud service providers declare to their users for the services they are utilizing. Cost per Resource Unit: the cost associated with each unit of resource (CPU hour, GB of storage, etc.), or total cost of ownership (TCO), the overall cost of using the IaaS services, including initial setup costs, ongoing operational costs such as storage and data transfer fees, and potential costs related to downtime are the most common available metrics for the potential cloud services users.

4.3. Feasibility Analysis

Looking at the requirements of the evaluation framework and the metric definitions, we can directly conclude that in real world scenarios the availability of practical data about these metrics is the next step we need to conduct to achieve a robust practical framework. Thus, the feasibility of implementing the proposed IaaS evaluation metrics depends on the availability of data, the willingness of providers to share necessary information, and the practicality of measuring these metrics in real-world scenarios. Below, we analyze each metric category's preliminary feasibility based on cloud services' available data and the Gartner annual report evaluating service providers. The visual presentation in Figure 12: Feasibility Score lists all the metrics at their feasibility level. Each pie slide represents the feasibility score (Low, Medium/Low, Medium, Medium/High, and High).

High feasibility metrics are easy to measure due to the availability of data and tools by service providers, whitepaper documentation, or third-party analysis tools. On the other hand, the Medium/High feasibility is less so to some extent; hence, the primary challenges here involve consistent monitoring and ensuring accurate measurement. Moreover, Medium feasibility metrics face moderate challenges primarily due to data transparency and provider disclosure; accurate logging and real-time monitoring are essential, and Medium/Low to a greater extent. Finally, low feasibility metrics are challenging to measure due to significant obstacles such as data availability, the complexity of measurement, and, in most cases, the need for a controlled environment. The following section describes the attributes we used to establish the feasibility analysis briefly:

- High Feasibility Metrics that are classified as having high feasibility are those that are easy to measure and implement in the evaluation framework. The data required for these metrics is typically readily available and provided by the service providers, and the tools needed to gather and analyze this data are well-established and straightforward. High feasibility metrics often have minimal associated challenges and can provide reliable and actionable insights with little effort.

- Medium/High Feasibility Metrics with medium/high feasibility are easy to measure but may require additional effort or specific tools. These metrics may present minor challenges, such as occasional data collection difficulties or the need for slightly more advanced analysis techniques; however, these challenges are usually manageable without significant investment in time or resources.

- Medium Feasibility Metrics present moderate challenges in measurement and implementation, as they may require specific tools, more detailed data collection processes, or more complex analysis methods. While not overly challenging to measure, these metrics can necessitate additional effort in setup, data management, and ensuring accuracy. Consistent monitoring and logging are required, and data may need to be aggregated from multiple sources.

- Medium/Low Feasibility Metrics are challenging to measure and may face significant obstacles regarding data availability and accuracy, as they require specialized tools or techniques; moreover, data collection can be inconsistent or incomplete. Implementing these metrics often involves overcoming considerable technical or operational hurdles. In most cases, data transparency issues require detailed and precise logging.

- Low-feasibility metrics are the most difficult to measure and implement, as they often face significant obstacles, such as limited data availability, lack of standardized tools, or high complexity in data collection and analysis. This is due to the requirement of significant resources, advanced technical expertise, and potentially custom-built solutions to measure these metrics. Due to these challenges, these metrics are often reserved for future improvements when the necessary infrastructure and tools are more readily available.

By understanding the feasibility of each metric, we can prioritize efforts and resources to focus on metrics that are easier to measure and significantly impact evaluating IaaS offerings. Thus, we ensure a more efficient and effective evaluation and provide reliable insights into the performance and reliability of the IaaS environment.

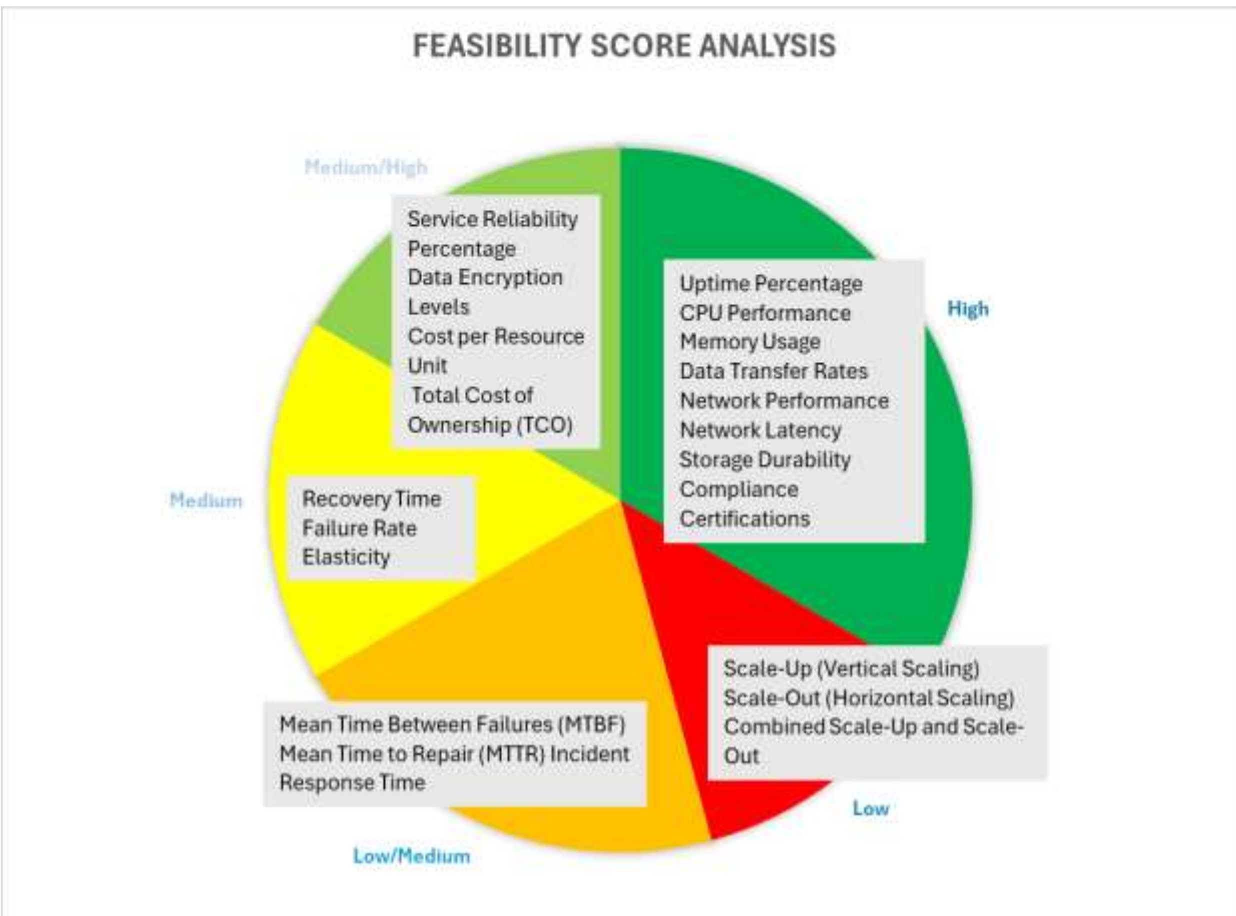


Figure 12: Feasibility Score

When evaluating IaaS services, metrics must be prioritized based on their feasibility and impact. High feasibility metrics should be addressed first because they are easier to measure and significantly influence the overall assessment of IaaS performance. These metrics provide reliable and actionable insights with straightforward measurement processes while being easily attained. Then, Medium feasibility metrics should follow, focusing on enhancing data transparency and ensuring precise logging practices. These metrics might require additional resources or improved methodologies to obtain accurate measurements, but they are still manageable.

On the other hand, Low feasibility metrics should be reserved for future consideration. These metrics present more considerable challenges due to complex measurement requirements or limited data availability, necessitating more sophisticated tools and approaches to capture them effectively. Understanding and addressing these challenges over time is fundamental and can help refine the overall evaluation framework and provide a more comprehensive assessment of IaaS services.

In summary**,** while some metrics are highly feasible due to the availability of data and tools, others pose challenges primarily due to data transparency and providers' willingness to share detailed operational information. We can conclude that Metrics related to uptime, performance, and cost are more accessible; in contrast, reliability, scalability, and security require a different approach, and these metrics can be interpreted to represent the business point of view when evaluating them. Therefore, while developing the IaaS evaluation framework, focusing on impactful and feasible metrics is a priority to ensure a balance between comprehensiveness and practicality.

4.4 Usability of the Evaluation Framework

Finalizing the feasibility analysis allows us to start visualizing how the evaluation framework can be used. The framework's usability is crucial for ensuring that it provides meaningful and actionable insights to users. It encompasses several key aspects, including ease of use, clarity, comprehensiveness, comparability, and applicability. As mentioned in the previous section regarding the feasibility of these metrics, while some are highly feasible, others may need more work to attain, making them impractical for users.

Another important consideration is that even though some metrics are highly feasible, they might need to provide more valuable information to the user due to how they are calculated. In some cases, the outcome is more relevant than the technical aspects. To ensure the relevance and comparability of these metrics, we analyzed how well these metrics measure and evaluate each of the IaaS aspects of virtualization, storage, network, and servers.

Consequently, after some changes and an elimination process, we have refined the list of metrics that will be used. To summarize, we decided to remove the Mean Time Between Failures (MTBF) and Mean Time to Repair (MTTR) related to availability and reliability, as they are already utilized in calculating uptime percentage and might not be needed for simplicity reasons. In other words, MTBF and MTTR are significant metrics used to evaluate availability and reliability; however, from a business's point of view, they are interested in the features related to these metrics. Memory usage, response time, and failure rates can also be removed because users can rely on more straightforward metrics when assessing performance. For the scalability category, due to the low feasibility of metrics such as Scale-Up (Vertical Scaling) and Scale-Out (Horizontal Scaling) and their complex calculations, we decided to adopt the market standard. This means changing these metrics to a categorical metric indicating whether the service provider offers auto-scaling, vertical scaling, or horizontal scaling as a manual service, charges a fee, or if a service change is required to meet scalability needs.

Regarding the security category, a categorical approach is also used as the focus will be on compliance certifications attained by the providers, as they are more relevant to the user, along with encryption levels for aspects such as storage and network. We adopted a list of data encryption features usually standard for acceptable encryption, such as data-at-rest, data-at-transit, critical management services, etc. Finally, cost-effectiveness will be measured by the cost per service paid by the user for the utilized service offering. The data transfer fees that the user gets charged are usually not considered when choosing a service.

Refining these metrics ensures the framework remains user-friendly and provides relevant and actionable insights without unnecessary complexity. The final list of metrics that will be utilized and focused on in the evaluation framework is listed in Table 11: Framework Metric List

Table 11: Framework Metric List

Category	Metric	Type	IaaS Aspect
Availability	Uptime Percentage	Numerical	Compute, Storage, Network
	Recovery Time	Numerical	Compute, Storage, Network
	Recovery Feature	Categorial	Storage
Performance	CPU Performance	Numerical	Compute, Storage, Network
	Data Transfer Rates	Numerical	Storage, Network
	Network Latency	Numerical	Storage, Network
	Network Performance	Numerical	Compute, Storage, Network
Reliability	Recovery Feature	Categorical	Compute, Storage, Network
	Service Reliability Percentage (SLA)	Numerical	Compute, Storage, Network
	Recovery Time	Numerical	Compute, Storage, Network
	Storage Durability	Numerical	Storage
Scalability	Auto Scaling	Categorical	Compute, Storage, Network
	Manual Scaling (Vertical or Horizontal)	Categorical	Compute, Storage, Network
Security	Data Encryption Levels	Categorical	Compute, Storage, Network

	Global Compliance Certifications score of 10	Categorical	Compute, Storage, Network
Cost Effectiveness	Cost per Resource Unit	Numerical	Compute, Storage, Network
	Data Transfer Fees	Numerical	Compute, Storage, Network
	Total cost of Ownership (TCO)	Numerical	Compute, Storage, Network

4.5. Creating the Evaluation Framework

As we are utilizing the Design Science Research Methodology (DSRM) [5] in guiding our research, we are now focused on designing the core artifact of this study: the evaluation framework. As mentioned in Chapter 3, we will employ the Analytic Hierarchy Process (AHP) method to achieve this, as it provides a structured decision-making approach by breaking down the evaluation criteria into a clear and hierarchical model[50]. Through this, we aim to create a robust and reliable framework that supports informed and objective decision-making in selecting IaaS providers.

Basing the theoretical perspective of the AHP method, our AHP-based evaluation framework is structured into a hierarchy comprising three primary levels:

- The overall goal of the decision-making process is to select the most suitable IaaS provider/product.
- The second level includes the primary criteria: Availability, Performance, Scalability, Reliability, Security, and Cost-effectiveness.
- The third level consists of specific sub-criteria under each primary criterion, such as the metrics we associated with each aspect we defined as the primary criteria, and so on.
- The fourth level consists of the choices: We evaluate and assess IaaS products based on the criteria and sub-criteria we have already defined.

Figure 13: Hierarchical structure in the analytic hierarch process (AHP) method [62] is a visual representation of the structure with the aim of the decision on top and branching into the criteria and sub-criteria while connecting with all the available alternatives being evaluated.

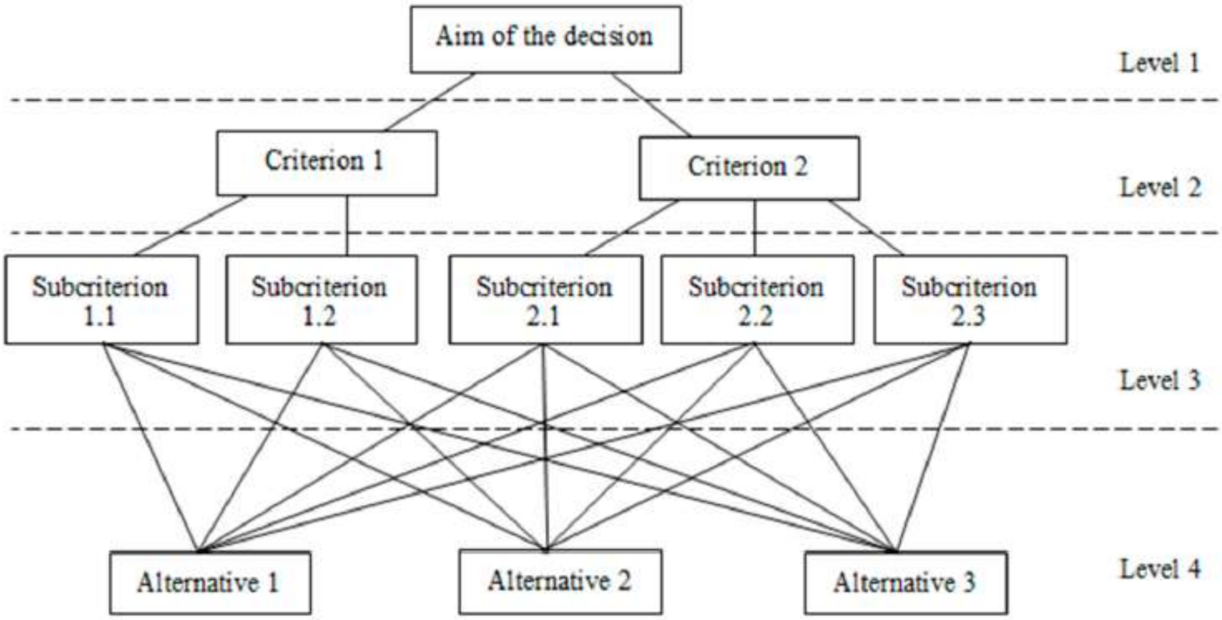


Figure 13: Hierarchical structure in the analytic hierarchy process (AHP) method [62]

4.5.1 Structuring AHP Criteria Hierarchy

The first phase in applying AHP involves constructing a hierarchy representing the decision problem. In our case, the goal at the top level is to select the best IaaS provider. The criteria at the first level include the aspects we already defined to evaluate the IaaS offering: Availability, Performance, Scalability, Reliability, Security, and Cost-effectiveness. Moreover, the Sub-Criteria at the second level are the metrics we chose to evaluate and assess each of the aspects and include:

- Availability: Sub-criteria include Uptime Percentage, Recovery Time, and Recovery Feature
- Performance: Sub-criteria include CPU Performance, Data transfer Rates, Network Latency, and Network Performance.
- Scalability: Sub-criteria include Auto Scaling and scaling for a Fee.
- Reliability: Sub-criteria include Recovery Time, Recovery Feature and SLA.
- Security: Sub-criteria include Data Encryption Levels and Compliance Certifications.
- Cost-effectiveness: Sub-criteria include Cost per Resource Unit, Total Cost of ownership, and Data Transfer Fees.

Table 11: Decision Hierarchy visually represents the structure derived from applying the AHP method. The weights in the hierarchy are not decided on for illustrative purposes; however, these

weights will be adjusted in subsequent sections to reflect the specific requirements of different scenarios.

Table 12: Decision Hierarchy

Level 0	Level 1	Level 2
Select an IaaS Product	Availability	Uptime Percentage
		A Recovery Time
		A Recovery Feature
	Performance	CPU Performance
		Data Transfer Rates
		Network Latency
		Network Performance
	Reliability	R Recovery Time
		R Recovery Feature
		SLA
		Storage Durability
	Scalability	Auto Scaling
		Scaling for a Fee
	Security	Data Encryption Levels
		Compliance Certifications
	Cost-Effectiveness	Cost per Resource Unit
		Data Transfer Fees
		TCO

4.5.2 Pairwise Comparison and Weight Assignment

In the second phase, pairwise comparisons are conducted to evaluate the relative importance of each criterion and sub-criterion. Decision-makers compare each pair of elements using a scale from 1 (equally important) to 9 (extremely more critical)[51]. *Table 13: Saaty's scale of measurement for pair-wise comparisons* illustrates how the scale is defined and explained[10]. These comparisons are then used to calculate the weights for each criterion and sub-criterion, which reflect their relative importance in achieving the overall goal. For example, if availability is deemed more critical than cost-effectiveness, it will have a higher weight.

Using Saaty's scale, each of the defined criteria is evaluated in a pairwise comparison, resulting in a six-by-six matrix of the importance of these criteria.

Table 13: Saaty's scale of measurement for pair-wise comparisons [10]

Intensity of importance	*Definition*	*Explanation*
1	Equal importance	Two activities contribute equally to the objective
3	Moderate importance	Experience and judgement slightly favour one over another
5	Strong importance	Experience and judgement strongly favour one over another
7	Very strong importance	An activity is strongly favoured and its dominance is demonstrated in practice
9	Absolute importance	The importance of one over another affirmed on the highest possible order
2,4,6,8	Intermediate values	Used to represent compromise between the priorities listed above

4.5.3 Synthesizing Calculations & Results

After determining the weights for each criterion and sub-criterion, the next step is synthesizing the results. The calculation process will generate a pairwise comparison matrix for the Level 1 criteria, the six aspects we identified earlier for assessing IaaS service provider offerings.

Table 14: Pairwise Comparison Matrix is an example of the resulting matrix. The exact process is repeated for each sub-criterion at Level 2, where pairwise comparisons are conducted among the selected metrics for each aspect. After finalizing and validating the matrices, we can move to evaluating the alternatives available for the assessment. This will involve a calculation formula of multiplying the scores assigned to each IaaS provider for a particular sub-criterion by the corresponding weight and summing these products to calculate the overall score for each provider. The providers' products are then ranked based on their overall scores, with the highest score indicating the most suitable provider based on the established criteria weights or selected scenario.

Table 14: Pairwise Comparison Matrix

Select an IaaS Product	**Availability**	**Perforamce**	**Reliability**	**Scalability**	**Securtiy**	**Cost-Effectivness**	**Weights**
Availability	1	-	-	-	-	-	%
Performance	-	1	-	-	-	-	%
Reliability	-	-	1	-	-	-	%
Scalability	-	-	-	1	-	-	%
Security	-	-	-	-	1	-	%
Cost-Effectiveness	-	-	-	-	-	1	%

An essential aspect of the AHP is checking the consistency of the pairwise comparisons to ensure that the judgments made are reliable since it is easily possible to end up with an inconsistent matrix when evaluating criteria on each other. The Consistency Ratio (CR) is calculated, with a CR of 0.1 or less indicating acceptable consistency. If the CR is higher, the comparisons may need to be reviewed and adjusted to improve consistency. Significant inconsistencies in the pairwise comparison matrix often arise due to decision-maker errors, particularly when the rule of complete transitivity in evaluations between compared alternatives is overlooked. [51]Inconsistencies in a pairwise comparison matrix can be assessed by calculating the consistency index (CI) and subsequently determining the consistency ratio (CR) using the appropriate equations.

Calculate the Consistency Index (CI) and Consistency Ratio (CR) to ensure the matrix is consistent. For a matrix of this size (n=6) per the six aspects we are evaluating, the Random Index (RI) is 1.24 based on Saaty's CIr values for matrices [52].

- Calculate Lambda_max (λ_max): Multiply each column of the original matrix by the derived weights and then divide each result by the corresponding weight. Average these values to get λ_max.
- Consistency Index (CI): CI = (λ_max - n) / (n - 1)
- Consistency Ratio (CR): CR = CI / RI

Alonso and Lamata propose a new approach to measuring and improving consistency in AHP, addressing some of the limitations of previous methods, such as the Consistency Index (CI) introduced by Saaty, the original developer of AHP [53].

- CR = (λ_max - n) / ((2.7699*n) – (4.3513 – n))

As mentioned, the matrix will be consistent if the CR is less than 0.1. This framework will use the new approach to calculate the consistency ratio.

Another critical point to emphasize is that the AHP method suffers from issues such as rank reversal and unknown numerical errors, which Tomashevskii addressed and that were also taken into consideration [54].

Finally, some elements of an AHP Excel template allowing multiple inputs and consolidated output for multi-criteria decision-making in corporate enterprises were used as references. The template had elements related to error calculation and the row geometric mean method (RGMM) that we utilized in the AHP tool being developed [55].

4.5.4 Decision Making with the AHP

The final phase involves applying the AHP-based framework to real-world IaaS products and utilizing it to evaluate and rank them. The framework will take the user input for the gathered metric data, do all the necessary calculations based on the chosen weights, and rank the providers' products based on the synthesized scores using actual data from these providers or other third-party evaluation portals. The final results will be validated by comparing them with industry benchmarks and seeking third-party evaluation to ensure the framework's robustness and reliability. Figure 14*: IaaS Evaluation Framework based on the AHP Method* illustrates the structure of the IaaS evaluation framework. It presents the hierarchy of criteria and sub-criteria alongside the selected choices, as defined within the AHP method. This diagram visually represents how the evaluation framework organizes and prioritizes each element to facilitate a systematic assessment of IaaS providers.

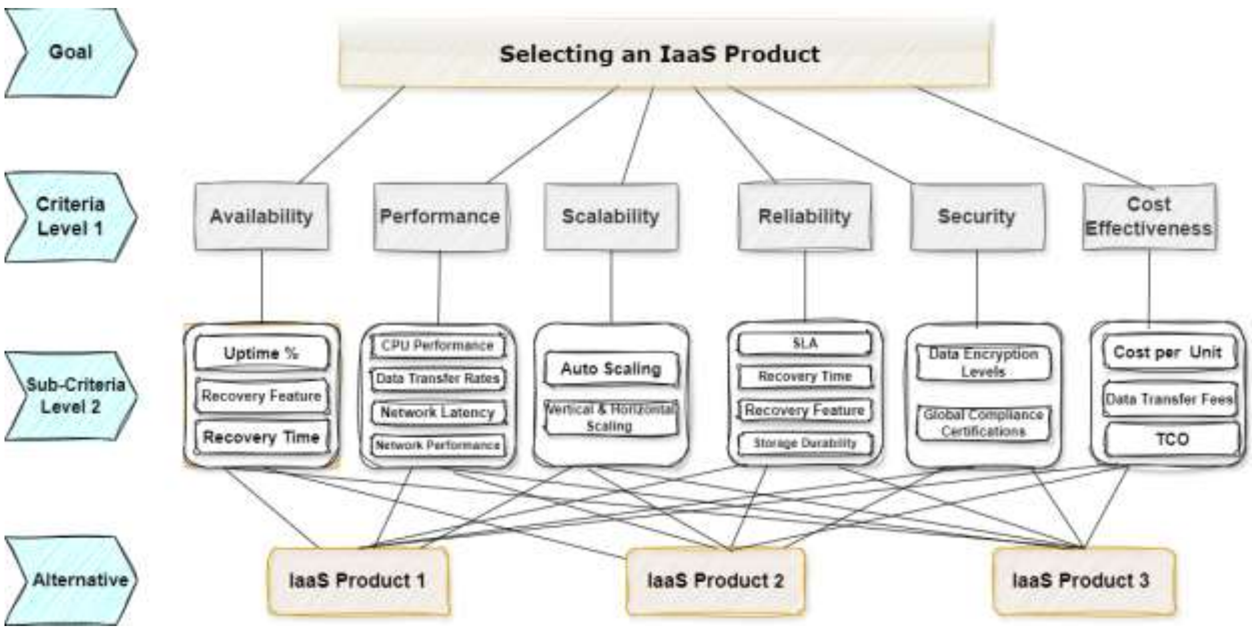


Figure 14: IaaS Evaluation Framework based on the AHP Method

In conclusion, the AHP-based evaluation framework will offer a systematic approach to assessing IaaS providers, balancing multiple criteria and sub-criteria to make informed decisions. The next Chapter will demonstrate the application of this framework to specific IaaS providers and discuss the broader implications of the findings for cloud service selection.

4.6 Limitations

While the evaluation framework provides a structured and comprehensive approach to assessing IaaS providers, specific challenges that affect its real-world application remain and must addressed. The framework's usability is enhanced by its clear organization of criteria and sub-criteria, which allows for an objective comparison of different IaaS products, as well as its adaptability, which will enable it to fit different organizational needs by adjusting the weights and criteria based on specific priorities. However, like any evaluation method, the effectiveness of the framework is dependent on the quality and availability of data. At the same time, the accuracy of weight assignments during the pairwise comparison process is supervised and controlled by the user; the authenticity of the collected data relies on external factors not regulated by the users.

Nevertheless, we must address some of our main limitations while constructing the evaluation framework. The first limitation that must be emphasized is that the framework was constructed assuming that a single participant would conduct the evaluations, primarily for practical reasons and the constraints of the thesis timeline. While this approach simplifies the implementation, it does not fully leverage the AHP method's strength in incorporating multiple viewpoints, creating limitations in turn. Expanding the framework to integrate numerous participations is possible and the ideal approach to solve this issue as it will provide a more comprehensive evaluation. However, it would require additional time and coordination.

Another potential limitation is data quality; in other words, it is crucial to emphasize the quality and reliability of data gathered about the alternatives being evaluated since it plays a pivotal role in ensuring accurate and objective comparison results. This is crucial because evaluation effectiveness heavily depends on it; whether the data is incomplete, outdated, or inaccurate might lead to flawed evaluations and suboptimal decisions. This limitation highlights the importance of using reliable data sources and regularly updating the information within the framework, an external factor that indirectly affects the framework. Organizations must ensure that the data they input into the AHP framework accurately reflects the current state of the IaaS providers' products being evaluated.

Another point worth mentioning is that despite the structured nature of the AHP method, the process of assigning weights to criteria and sub-criteria remains inherently subjective, especially in the case of a single participant. Nevertheless, different decision-makers may have varying perspectives on the relative importance of other criteria, which can lead to different outcomes. This subjectivity can introduce bias into the evaluation process, potentially skewing the results. It is essential to acknowledge that the framework's effectiveness partly depends on the judgment and expertise of the individuals involved in the weight assignment process.

Moreover, the framework was initially created to compare Ten IaaS products, which is sufficient for many use cases but may be limiting for organizations with more options to consider, especially handling products with different specs and hardware options, which would increase its applicability and complexity. Familiarity with the AHP method is also required to avoid errors, especially if the user wants to utilize the tool to its maximum potential by using the fully customizable scenario.

In conclusion, the AHP framework developed in this study is explicitly tailored for evaluating IaaS providers' products, thus limiting its application when evaluating other cloud services or when used in different industries. Since the selected aspects and metrics were concentrated on evaluating IaaS offerings, significant adjustments may be necessary to adapt the framework to other contexts, which could limit its generalizability. Additionally, even if the framework is robust within its intended context, caution should be exercised when applying it to other domains, as the criteria and sub-criteria may need to be redefined to reflect the specific requirements of those applications. Finally, the decision-making individual using the framework must have some basic knowledge of the AHP method to avoid errors and mistakes. While the framework has been designed to be as user-friendly as possible, familiarity with the AHP process will help ensure that the comparisons are accurate and meaningful.

Chapter 5: Implementation & Empirical Analysis

In this chapter, we will delve into the practical implementation of the AHP-based evaluation framework that was developed in the previous chapters, marking the transition from the design phase to the demonstration phase of the Design Science Research Methodology (DSRM), where the focus shifts to finding a suitable context for the created artifact and using it to solve real-world problems. We will explain how the theoretical framework is translated into a functional tool to assist decision-makers in systematically and objectively evaluating IaaS providers' products. Additionally, we will discuss how this tool is applied in a relevant organizational context to address specific decision-making challenges.

5.1 Overview of the AHP-Based Tool

The AHP-based tool has been designed to make the complex decision-making process easier, more accessible, and more manageable for users. Our primary focus in this section is to provide a clear overview of the tool we are presenting, explain its workflow, its core components, key features, the user interface, data input forms needed by the user, and the underlying algorithms that facilitate pairwise comparisons and weight calculations we constructed. The tool is created to guide users through the evaluation process, ensuring that all necessary steps are followed, from inputting the weights of the criteria and sub-criteria to calculating final scores and comparing alternatives.

5.1.1 AHP-Based Tool Description

To describe the created artifact briefly, the AHP-Based Tool is a comprehensive tool designed as an Excel workbook to evaluate, compare, and assist in selecting IaaS products provided by cloud service providers. The tool is structured around the AHP for decision-making with predefined product scenarios based on various factors such as performance, cost, security, and balance. The workbook is structured with multiple sheets; some require user input, and others are just used for calculation and explanation. Each serves a specific purpose in guiding the user through the evaluation process while facilitating the creation of valid tailored scenarios. To provide a clear understanding of the workbook's structure, Table 15: AHP Tool TAB is presented with summaries of each sheet's critical functionalities, ensuring transparency and aiding in both the interpretation of results and the reproducibility of the methodology. The following sections provide a detailed explanation and analysis of selected sheets from the workbook, as these sheets are more complex and need to be thoroughly examined to highlight their purpose, structure, and role in the evaluation process. This comprehensive exploration will ensure a clear understanding of the methodologies and data used, contributing to the robustness and transparency of the research.

The primary aim of this Excel workbook is to provide a structured, data-driven approach that utilizes the AHP method strengths in evaluating and comparing products based on our set criteria and sub-criteria metrics specifically designed to facilitate the objective and straightforward comparison of Infrastructure as a Service (IaaS) products, which typically have complex features and multiple aspects to consider. Moreover, it thrives on helping users make informed decisions by offering predefined scenarios and the flexibility to create custom scenarios tailored to specific needs. Therefore, by consolidating product data, applying scoring methodologies, and providing visual summaries and charts, the workbook should become a valuable decision-support tool for product selection, ensuring that even complex comparisons can be conducted efficiently and precisely.

Table 15: AHP Tool TAB Description

TAB NAME	PURPOSE & DISCRIPTION
START HERE	Serves as the introductory sheet providing users with high-level instructions, an overview of the workbook's purpose, and guidance on navigating through the various sheets. It acts as a roadmap for first-time users.
INSTRUCTIONS	Offers detailed instructions on the effective use of the workbook. This sheet explains the function of each tab, the underlying methodology, and step-by-step instructions for performing the analysis.
Product Data	Contains the raw data for the products being evaluated collected by the user to cover all metrics in the measured aspects. including key attributes such as performance metrics, costs, and security features. This sheet is the foundational data source for all subsequent analysis and comparisons.
Product Scorecard	Consolidates product data into a scoring framework, applying a multi-criteria evaluation methodology. This sheet allows for a comparative analysis of products, highlighting the best alternative choice between the compared products.
Dashboard	Provides a visual summary of the key insights derived from the product data, including charts, graphs, and summary statistics, offering an at-a-glance view of the comparative performance of different products under various scenarios.
Criteria Weights	Allows users to define and adjust the importance of each criterion used by assigning weights to different criteria, users can tailor the analysis to reflect specific priorities or preferences. The weighted criteria are then applied influencing the final scoring and ranking of products, creating a customized evaluation based on user preference.
Custom Scenario	Enables users to create personalized scenarios by adjusting the weights and criteria according to specific needs. This flexibility allows for tailored analysis that aligns with unique preferences and priorities.
Balanced Scenario	Represents a predefined scenario aiming to balance the importance of all measured criteria. This scenario is suitable for general-purpose analysis where no single factor is overwhelmingly prioritized.
Performance-Focused Scenario	Focuses on scenarios where performance and availability is prioritized above other factors. This sheet adjusts the evaluation criteria to highlight products that excel in performance, catering to users with high-performance requirements.
Cost Sensitive Scenario	Tailored for cost-conscious users, this scenario emphasizes minimizing costs while maintaining acceptable levels of other criteria. The sheet recalculates scores to rank products based on cost-effectiveness.
Security-Centric Scenario	Designed for scenarios where security is the top priority. This sheet adjusts the evaluation criteria to emphasize security features, identifying products that offer the highest level of protection.
Product Calculation	Performs the essential calculations needed for the AHP method that support the various scenarios and the product scorecard. This sheet contains the formulas and intermediate data used to compute final scores and rankings, ensuring transparency in the scoring process.
Rating Scale	Defines the scales, thresholds, and criteria used in the evaluation and scoring of products. This sheet serves as a reference for understanding the scoring methodology and how different aspects of the products are assessed primarily focusing on the pairwise comparison matrix scoring scale and Saaty's CIr values for matrices.
Metric Description	Provides detailed explanations of each metric used to evaluate the products, including what each metric measures, its importance, and how it influences the overall product score. Metric definition, requirement, and formula are presented to help the user understand the data he is collecting for the evaluation method.
Calculation Data	Contains the underlying data and calculations that support the various analyses and scenarios within the workbook. It stores intermediate results, such as weighted scores, normalized data, and any other calculations necessary for the Analytic Hierarchy Process (AHP) and other evaluation methodologies used in the workbook.

5.1.2 AHP-Based Tool Key Features

The Excel workbook has several robust features that empower users to evaluate and compare products effectively across various dimensions. Table 16: AHP Tool Key Features lists the key features that make this tool comprehensive and user-friendly. These key features were designed into the tool to make this workbook essential for anyone involved in product selection and evaluation. Whether decision-makers are looking to compare options quickly or dive deep into custom criteria, the workbook provides the flexibility, transparency, and power to make informed decisions confidently.

Table 16: AHP Tool Key Features

FEATURES	FEATURE DISCRIPTION
Multi-Dimensional Product Evaluation	The workbook allows for the evaluation of products based on multiple criteria, including performance, cost, security, and balance. Each of these dimensions can be analyzed separately or in combination, providing a holistic view of each product's strengths and weaknesses.
Scenario-Based Analysis	Users can explore various predefined scenarios such as Balanced, Performance-Focused, Cost-Sensitive, and Security-Centric. Each scenario is tailored to emphasize different aspects of product evaluation, helping users make decisions based on specific priorities.
Custom Scenario Creation	One of the standout features of this workbook is the ability to create custom scenarios. Users can adjust the weights and criteria according to their unique requirements, enabling personalized analysis and product ranking that aligns with individual or organizational goals.
Dynamic Product Scorecard	The Scorecard provides a dynamic, comparative view of all products based on the selected scenario. This feature allows users to quickly assess which products perform best overall and in specific areas, aiding in faster decision-making.
Interactive Dashboard	The workbook includes an interactive dashboard that visually summarizes the key insights derived from the data. With charts, graphs, and summary statistics, the dashboard offers an at-a-glance overview, making complex data more accessible and easier to interpret.
Comprehensive Product Data Integration	The Product Data sheet serves as a centralized repository for all relevant product information. This data is seamlessly integrated into the analysis, ensuring that all evaluations are based on accurate and up-to-date information.
Transparent Calculation Methods	The workbook includes a Product Calculation sheet that transparently displays the underlying formulas and methodologies used to generate scores and rankings. This feature ensures that users understand how results are derived, enhancing trust and credibility in the analysis.
Detailed Rating Scale	The Rating Scale sheet provides a clear and detailed explanation of how different product attributes are rated. This feature helps users comprehend the criteria and thresholds used in the evaluation process, making it easier to understand and justify product rankings.
User Guidance and Support	The workbook is equipped with a START HERE sheet and an INSTRUCTIONS sheet that offer detailed guidance on how to use the tool effectively. These sheets are designed to assist both new and experienced users in navigating the workbook and maximizing its potential.

5.1.3 AHP-Based Tool Workflow

The workflow within the AHP-based tool presents a step-by-step visualization of the evaluation process, ensuring that each critical task is explained and described accurately and efficiently. It illustrates the steps from selecting the scenario, defining the weights, and deciding on these alternatives.

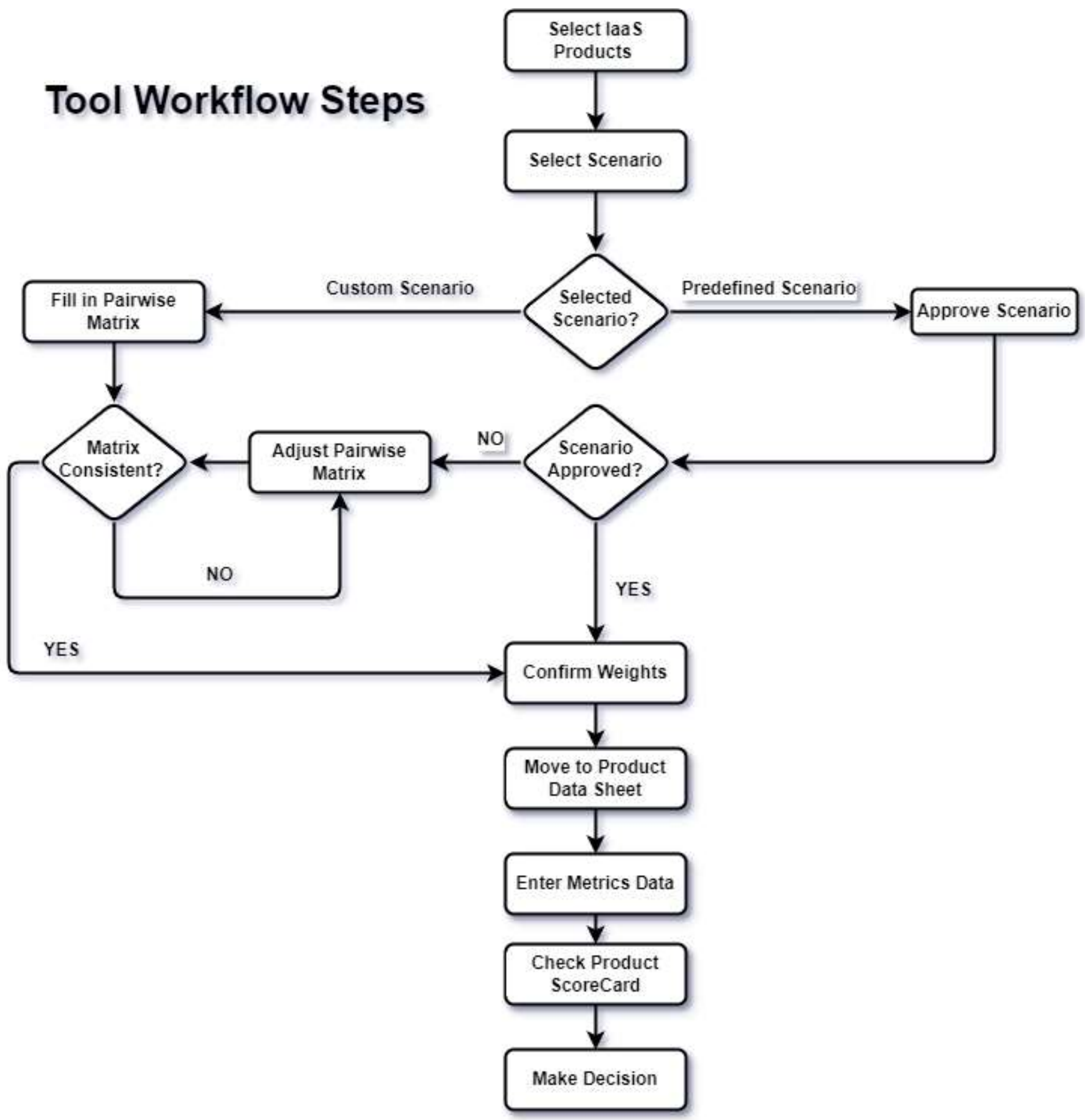


Figure 15: Evaluation Framework Workflow

5.2 Predefined Scenarios

This section presents the predefined scenarios incorporated into the tool, as they are designed to cater to specific priorities, such as performance, security, or cost focused. Each scenario is pre-configured with set criteria and sub-criteria already defined in the evaluation framework to ensure valid consistency across the evaluation process and avoid user confusion as they navigate and conduct their assessments. The main difference in these scenarios is that the pairwise comparison matrix was created, as it is critical to define the weights for the criteria and sub-criteria, which will be explained in more detail in their dedicated sections.

Nevertheless, the predefined scenarios include:

- Balanced Scenario: This scenario assigns equal weight to all six aspects, ensuring no single factor is prioritized over another. It should be ideal for organizations seeking a holistic evaluation without bias toward particular criteria.
- Performance-Focused Scenario: This scenario prioritizes criteria that affect the performance and availability of IaaS providers' products, such as CPU performance, data transfer rates, and throughput.
- Cost-Sensitive Scenario: This scenario focuses on cost efficiency with potential scalability options, which is ideal for organizations looking to optimize their spending on cloud services.
- Security-Centric Scenario: This scenario emphasizes security-related criteria with reliability in mind, primarily focusing on data encryption levels and compliance certifications for organizations where data protection is paramount.

The rationale behind selecting criteria and assigning weights will be explained for each scenario to ensure that users understand the assumptions and priorities embedded in each scenario. As part of the demonstration phase of DSRM, these scenarios will be applied in practical contexts to address specific decision-making challenges within organizations. The real-world application of these scenarios will help validate the framework's effectiveness in solving actual problems decision-makers face.

5.2.1 Balanced Scenario: A Holistic Approach

The Balanced Scenario's primary objective is to provide a comprehensive approach that offers a well-rounded performance score across all key aspects, including performance, availability, reliability, security, cost-effectiveness, and scalability. This scenario is particularly suited for organizations that require a comprehensive evaluation of the cloud service provider's product being evaluated, ensuring that no single factor is disproportionately prioritized over others. This approach distributes weights evenly, allowing for a balanced assessment that reflects the organization's diverse needs, which can be a good starting point for organizations that still need a clear objective regarding their priorities.

Given the focus of this scenario, each criterion is initially treated as equally important, ensuring that no single aspect is prioritized over another. However, users can modify it based on their specific requirements if needed. Consequently, the pairwise comparison process in the Balanced Scenario reflects this even distribution, giving an equal 16.7% weight to each of the six criteria being utilized in the framework, with the option for users to adjust the weights through the AHP tool if they decide that certain factors should be given more or less emphasis.

The weights were calculated using the Analytical Hierarchy Process (AHP) within the tool. Table 17: Balanced Scenario Matrix illustrates the pairwise comparison matrix along with the resulting weights, highlighting the equal assigned weight of 16.7% to each.

Table 17: Balanced Scenario Matrix

Criteria	Availability	Performance	Scalability	Reliability	Security	Cost-Effectiveness
Availability	1	1	1	1	1	1
Performance	1	1	1	1	1	1
Scalability	1	1	1	1	1	1
Reliability	1	1	1	1	1	1
Security	1	1	1	1	1	1
Cost-Effectiveness	1	1	1	1	1	1
Sum	6.00	6.00	6.00	6.00	6.00	6.00
Weights	**16.67%**	**16.67%**	**16.67%**	**16.67%**	**16.67%**	**16.67%**

To conclude, although the Balanced Scenario within the AHP-based evaluation framework is straightforward, it offers a structured and comprehensive approach for organizations that seek an equilibrium between all critical factors when evaluating IaaS providers. Distributing weights evenly across critical criteria facilitates a well-rounded decision-making process when selecting the most suitable IaaS provider products.

5.2.2 Performance-Focused Scenario: Prioritizing Performance and Availability

In the performance-focused scenario, the primary objective is to prioritize performance-related aspects primarily, followed by availability and reliability. This scenario is particularly suited for organizations where high computational demands, fast data processing, and minimal latency are critical for operations. While performance is the top priority, other aspects are also considered with relatively lower emphasis. The benefits of the Performance-Focused Scenario focus mainly on enhanced computational power and availability assurance. Thus, performance must be prioritized to meet high processing demands and fast data transfer needs while ensuring the service is reliable and available. Although performance is prioritized, the framework still considers all other critical factors, such as security, cost-effectiveness, and scalability, to ensure a comprehensive evaluation.

For the Performance-Focused Scenario, the evaluation framework is structured with the following criteria Weights as presented in

Figure 16: Performance-Focused Scenario

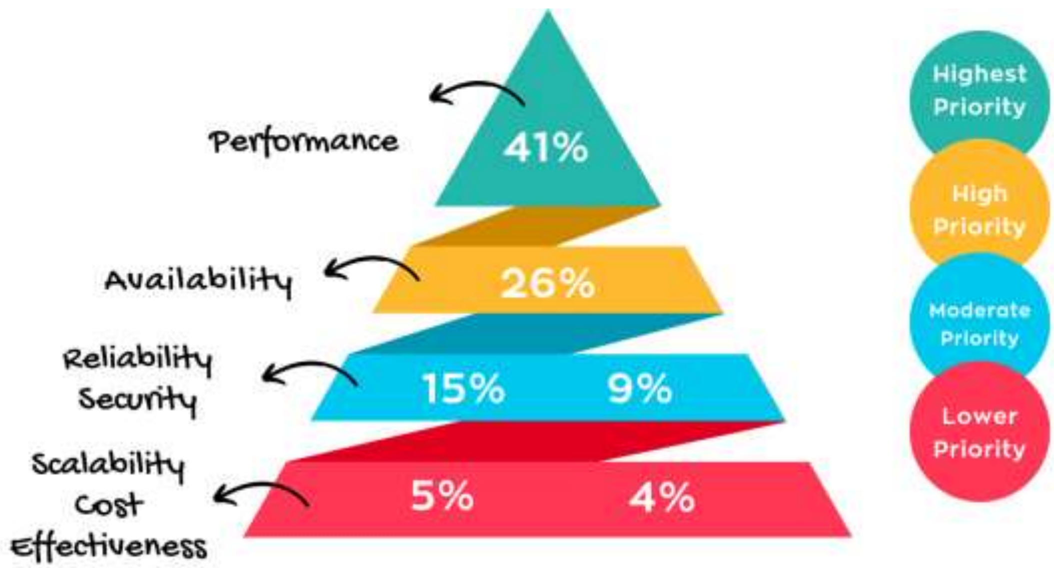


Figure 16: Performance-Focused Scenario

Given this scenario's focus, the highest weights are assigned to Performance and Availability, with moderate weights for Reliability and Security and lower weights for Cost-Effectiveness and Scalability. Consequently, the pairwise comparison process involved comparing each criterion and sub-criterion to determine their relative importance in this scenario, resulting in these assumptions.

- *Performance vs. Availability*: Performance might be considered more important, resulting in a 41% to 26% weight ratio.
- *Performance vs. Reliability*: Performance is likely more critical, leading to a 41% to 15% weight ratio.
- *Performance vs. Security*: Performance is prioritized, but Security still holds some importance, resulting in a 41% to 9% ratio.

The weights were systematically derived by applying the Analytical Hierarchy Process (AHP) using the designated tool. Table 18: Performance-Focused Scenario Matrix displays the pairwise comparison matrix and the corresponding weights, highlighting the outcome of this methodical approach.

Table 18: Performance-Focused Scenario Matrix

Criteria	Availability	Performance	Scalability	Reliability	Security	Cost-Effectiveness
Availability	1	1/2	6	2	4	6
Performance	2	1	8	3	5	8
Scalability	1/6	1/8	1	1/3	1/2	2
Reliability	1/2	1/3	3	1	2	5
Security	1/4	1/5	2	1/2	1	3
Cost-Effectiveness	1/6	1/8	1/2	1/5	1/3	1
Sum	4.08	2.28	20.50	7.03	12.83	25.00
Weights	26.54%	40.90%	5.18%	15.21%	8.59%	3.57%

To conclude, the Performance-Focused Scenario within the AHP-based evaluation framework provides a structured and practical approach for organizations where performance is the primary concern. This scenario simplifies the decision-making process using predefined weights and criteria while ensuring that all relevant factors are considered when selecting the most performant IaaS provider's products.

5.2.3 Cost-Sensitive Scenario: Prioritizing Cost-Effectiveness and Scalability

In this scenario, the focus is on prioritizing cost-effectiveness and scalability, ensuring that the service is affordable and capable of growing with the organization's needs. This scenario suits startups, small to medium-sized enterprises, or organizations with significant budget constraints. The aim is to provide several benefits and advantages targeting budget alignment that prioritizes cost factors aligned with the organization's budget constraints and scalability assurance, ensuring that the service can scale effectively to meet future demands. While cost and scalability are prioritized, the framework still provides a holistic evaluation by considering other critical factors such as availability, performance, reliability, and security to a lower extent, hence a comprehensive evaluation.

The evaluation framework for the Cost-Sensitive Scenario is structured with the following criteria and weights, as shown in Figure 17: Cost-Sensitive Scenario

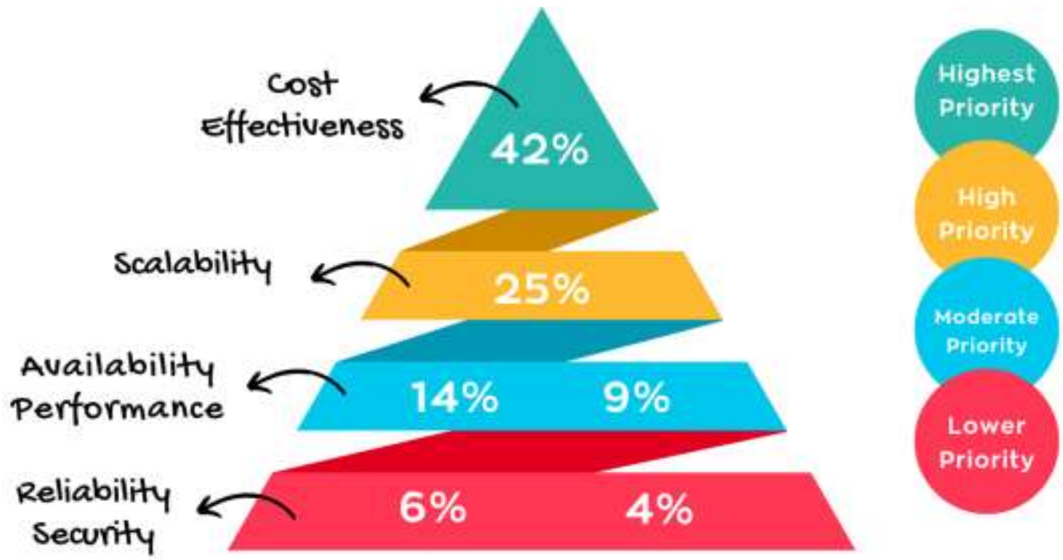


Figure 17: Cost-Sensitive Scenario

Given the focus of this scenario, higher weights are assigned to Cost-Effectiveness and Scalability, with moderate weights for Availability and Performance and lower weights for Reliability and Security. Therefore, the pairwise comparison process involved comparing each criterion and sub-criterion to determine their relative importance in this scenario. For demonstration purposes, the criteria weights were the results of the pairwise comparison that was created to suit the primary goal of the selected scenario. However, the sub-criteria weights were left equal for simplification purposes but can be altered and changed if needed. The following assumptions were considered when creating this scenario:

- *Cost-Effectiveness vs. Scalability*: Cost-Effectiveness might be slightly more critical, resulting in 42% to 25% weights.
- *Scalability vs. Availability*: Scalability could be considered more critical than Availability, leading to weights of 25% to 14%.
- *Availability vs. Performance*: Availability might be considered slightly more important in this scenario, each receiving a weight of 14% and 9%, respectively.
- *Reliability vs. Security*: These are essential but less critical in this scenario, each receiving weight of 6% and 4%.

These weights were carefully determined by implementing the Analytical Hierarchy Process (AHP) using the specified tool. Table 19 Cost Sensitive Scenario Matrix presents the pairwise comparison matrix alongside the resulting weights, reflecting the structured decision-making process.

Table 19 Cost Sensitive Scenario Matrix

Criteria	Availability	Performance	Scalability	Reliability	Security	Cost-Effectiveness
Availability	1	2	1/2	3	3	1/4
Performance	1/2	1	1/4	2	3	1/5
Scalability	2	4	1	4	5	1/2
Reliability	1/3	1/2	1/4	1	2	1/7
Security	1/3	1/3	1/5	1/2	1	1/7
Cost-Effectiveness	4	5	2	7	7	1
Sum	8.17	12.83	4.20	17.50	21.00	2.24
Weights	**13.72%**	**9.09%**	**24.75%**	**5.93%**	**4.24%**	**42.27%**

To conclude, the Cost-Sensitive Scenario within the AHP-based evaluation framework provides a structured and efficient approach for organizations where cost and scalability are the primary concerns. This scenario simplifies the decision-making process using predefined weights and criteria while ensuring that all relevant factors are considered when selecting the most suitable IaaS providers' products.

5.2.4 Security-Centric Scenario: Prioritizing Security Over Other Factors

In the Security-Centric Scenario, the primary focus is giving priority to security-related aspects. Therefore, this scenario is ideal for organizations where data protection, compliance, and Encryption levels are paramount. While security is the top priority, followed by reliability, other factors such as cost-effectiveness, performance, scalability, and availability are also considered with relatively lower emphasis. Ideally, this approach ought to benefit the organization with several advantages, primarily affecting enhanced data protection and reliability assurance by prioritizing security to protect sensitive data and ensure compliance with industry standards and regulations, focusing on minimizing downtime and ensuring that the service provider offers strong reliability guarantees. While security is prioritized, the framework still considers other critical factors such as performance, availability, cost-effectiveness, and scalability, hence a comprehensive evaluation.

The evaluation framework for the Security-Centric Scenario is structured with the following criteria and weights, as shown in Figure 18: Security-Centric Scenario

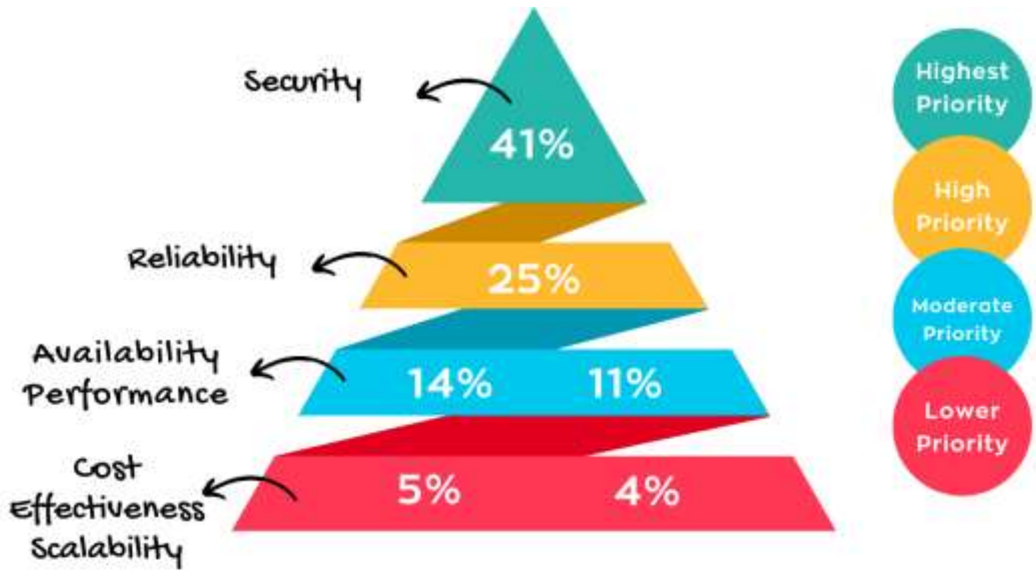


Figure 18: Security-Centric Scenario

Given this scenario's focus, the highest weights are assigned to Security and Reliability, with moderate weights for Availability and Performance and lower weights for Cost-Effectiveness and Scalability. Therefore, the pairwise comparison process, similar to what was done in the other scenarios, involved comparing each criterion and sub-criterion to determine their relative importance in this scenario.

- *Security vs. Reliability*: Security might be considered more important, resulting in a 41% to 25% weight ratio.
- *Security vs. Availability*: Security is significantly more critical, with a weight of 41% compared to 14% for availability.
- *Performance vs. Cost-Effectiveness*: Performance might be slightly more critical than Cost-Effectiveness, leading to a weight ratio of 11% to 5%.

These weights were carefully determined by implementing the Analytical Hierarchy Process (AHP) using the specified tool. Table 20: Security-Centric Scenario Matrix presents the pairwise comparison matrix alongside the resulting weights, reflecting the structured decision-making process.

Table 20: Security-Centric Scenario Matrix

Criteria	Availability	Performance	Scalability	Reliability	Security	Cost-Effectiveness
Availability	1	2	3	1/2	1/4	3
Performance	1/2	1	3	1/3	1/4	3
Scalability	1/3	1/3	1	1/5	1/7	1/2
Reliability	2	3	5	1	1/2	5
Security	4	4	7	2	1	7
Cost-Effectiveness	1/3	1/3	2	1/5	1/7	1
Sum	8.17	10.67	21.00	4.23	2.29	19.50
Weights	**13.90%**	**10.66%**	**4.25%**	**24.59%**	**41.12%**	**5.47%**

To conclude, the Security-Centric Scenario within the AHP-based evaluation framework provides a structured and practical approach for organizations where security is the primary concern. This scenario simplifies the decision-making process using predefined weights and criteria while ensuring that all relevant factors are considered when selecting the most secure and reliable IaaS providers' products.

5.3 Custom Scenarios: Flexibility for Tailored Evaluations

In addition to the predefined scenarios, the tool has the flexibility option to allow users to create custom scenarios tailored to their specific needs. The criteria and sub-criteria are still set as it is the core of the evaluation framework. However, the user can determine the assigned weights for these criteria and their related sub-criteria based on their requirements and needs. The steps needed to create a custom scenario are presented in the following section, guiding users on setting weights for their criteria and sub-criteria and input data for the evaluated alternatives. The flexibility of custom scenarios ensures that the tool can be adapted to a wide range of decision-making contexts, from selecting IaaS providers' products to evaluating other types of services.

5.3.1 The Role of Weight Assignment

Custom scenarios are essential to provide flexibility for organizations that operate in highly specialized environments or have niche priorities. While predefined scenarios like Performance-Focused or Cost-sensitive scenarios are helpful starting points, they may only capture some of the nuances of an organization's needs or offer the flexibility to tackle practical, real-world scenarios. This flexibility in weight assignment is crucial for organizations that may operate in a dynamically changing environment where they need to prioritize various aspects depending on their specific operational requirements, strategic goals, or industry regulations. For example, an organization with stringent security needs might assign a higher weight to the Security criterion, while another organization focused on minimizing costs might prioritize the Cost-Effectiveness criterion. Users can ensure that the evaluation framework aligns closely with their decision-making objectives by adjusting these weights.

5.3.2 Ensuring Validity Through Consistency Checks

One critical aspect of the AHP methodology that plays a crucial role in its validity is ensuring that the weights assigned to the criteria and sub-criteria are consistent. As discussed in section 4.5.3 Synthesizing Calculations & Results, consistency in pairwise comparisons means that if one criterion is judged more important, the relationships between all other comparisons must align logically, and a CR must be of 0.1 or less.

The custom scenario feature within the evaluation framework automates this step and performs all necessary pairwise calculations based on the weights set by the user. As users assign weights, the tool automatically constructs the pairwise comparison matrix and calculates the consistency ratio. The tool will notify the user if the consistency ratio indicates the matrix is inconsistent, meaning the assigned weights do not logically align. A CR of 0.1 or less indicates acceptable consistency. If the CR is higher, the comparisons may need to be reviewed and adjusted to improve consistency. This feedback mechanism ensures that users know of inconsistencies and can adjust their weights to achieve a valid and consistent matrix.

5.3.3 Steps to Create a Custom Scenario

Creating a custom scenario within the AHP-based tool involves primarily three main steps: setting the weight for the criteria and sub-criteria, followed by a consistency check to ensure that the weights are valid, and finally, inputting data and analyzing the results. *Figure 19: Custom Scenario Steps* is a visual representation of these steps.

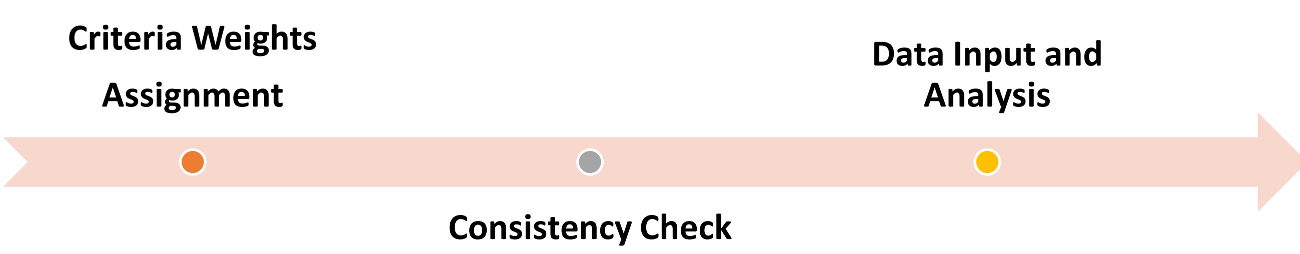


Figure 19: Custom Scenario Steps

1. **Set Weights for Criteria and Sub-Criteria**
   Users begin by filling in the pairwise comparison options, which are used to calculate the assigned weights to the fixed criteria and sub-criteria based on organizational priorities. The tool does not allow for direct weight input, as it deters the proper structure of the AHP method from being used. The tool automatically generates the assigned weights based on the pairwise comparison matrix, ensuring that the relationships between criteria are accurately represented.

2. **Consistency Check**
   After generating the matrix, the tool calculates the consistency ratio to determine whether the assigned weights are consistent. The matrix is valid if the consistency ratio is below the acceptable threshold (typically 0.1). The tool will prompt the user to adjust the weights if it is above the threshold. Users can then refine their weights and re-run the consistency check until a valid matrix is achieved. This iterative process ensures that the final evaluation is both dependable and accurate.

3. **Data Input and Analysis**
   Once a valid consistency ratio is achieved, users can input the relevant data for each IaaS product being evaluated. The tool will use this data with the weights assigned to calculate the final scores for each alternative based on the chosen custom scenario setting. The tool provides a comprehensive analysis, ranking the IaaS providers' products based on the weighted criteria. Users can review the results to make informed decisions that align with their organizational goals.

In conclusion, Custom scenarios within the AHP-based framework offer several key benefits, such as flexibility, allowing users to adjust weights to reflect their specific needs while benefiting from the robust, fixed criteria and sub-criteria. Reliability as the automated consistency checks ensure that the evaluations are logically sound, and the resulting decisions are based on valid and reliable data. Finally, User Empowerment by allowing users to control the weighting process, the framework empowers decision-makers to create scenarios that closely match their operational realities and strategic objectives.

5.4 Integrating the Tool into Organizational Processes

Successful integration into organizational processes is crucial in ensuring that the evaluation framework becomes a valuable resource applicable in real-world scenarios and a functional asset for decision-makers. To attain consistent, objective, and data-driven evaluations aligned with their strategic goals and operational requirements, organizations must go beyond simply using the tool for isolated assessments; it involves embedding the tool into the organization's daily workflows and decision-making processes.

The AHP-based tool, for example, could be strategically integrated into key organizational processes, such as procurement, strategic planning, and project management, to maximize its effectiveness. In a more practical context, the AHP-based tool can play a pivotal role in the procurement process, especially when evaluating and selecting cloud service providers' products, which often involves assessing multiple vendors based on various criteria such as cost, performance, security, or reliability. The tool provides a structured approach to compare these vendors systematically. An additional benefit can be the tool's ability to support long-term strategic planning by aligning cloud service decisions with the organization's future goals. Strategic planning often requires a thorough evaluation of potential investments, where decisions must be made based on current needs and future scalability. Finally, it can assist in evaluating the selected alternatives in the context of specific project requirements, timelines, and budgets as part of an organization's larger IT projects.

Integrating this tool into real-world processes is a crucial aspect of the demonstration phase of DSRM, as It allows us to observe how the framework functions in practice and contributes to solving specific problems within an organizational context.

Chapter 6: Validation & Evaluation

In this chapter, we explore the validation and evaluation of the AHP-based tool, focusing on two essential parts: the Scenario Study Case and Third-Party Validation of the tool. The Scenario Study Case involves collecting real-world data to evaluate ten IaaS products that represent a similar range of services, emphasizing a real-world context where an organization compares a range of products for future implementations. Specifically, the products selected are used for comparable services, such as compute or storage functionalities typically offered by IaaS providers. Additionally, the section will highlight the metric data collection process necessary for the tool to function effectively and provide accurate comparisons, followed by an in-depth evaluation of its overall performance and the accuracy of its final results with third-party domain experts.

6.1 How the AHP Tool Functions in the Evaluation

The AHP tool plays a significant role in evaluating and ranking the IaaS products systemically and structurally. As previously explained, the tool automates pairwise comparisons for each criterion in every scenario, ensuring transparent calculations and an objective comparison process. Using the AHP method, each scenario assigns distinct levels of importance as weights to the six key criteria — Availability, Performance, Scalability, Reliability, Security, and Cost-effectiveness — based on the organization's priorities.

As discussed in more detail in Chapter 5, in the predefined scenarios section, where the Balanced Scenario gave equal weight to criteria, the Performance-Focused Scenario gave a higher weight to performance. Similarly, the Security-Centric Scenario prioritizes security and reliability, ensuring the tool adapts to the specific needs of different organizational contexts.

Moreover, the pairwise comparison process quantifies the relative importance of each criterion, as outlined earlier in the methodology. It translates the choices made by the users in the pairwise comparison phase to percentage weights, which improves the clarity of the choices made. The results of these comparisons are synthesized into a final ranking of the IaaS products, enabling decision-makers to identify the gathered metric data into ratings representing the optimal solution for each predefined or custom scenario. With a clear understanding of how the AHP tool functions in evaluating IaaS products, we move forward with applying it in a real-world scenario and testing its validity in performing its required tasks.

6.2 Scenario Study Case Description

In this scenario, we evaluate ten leading compute IaaS products, chosen based on their presence and ranking of the cloud service providers in the Gartner Magic Quadrant for Cloud Infrastructure and Platform Services. These products represent diverse offerings available in the market and provide scalable computing capacity for various applications, from web hosting to high-performance computing and machine learning. The selected products have similar ranges of capabilities to simulate a realistic decision-making process, where an organization must choose the optimal IaaS computing solution from similar competitors.

For accurate representation purposes, the organization operates in a fast-paced, high-computing environment, requiring flexible, dependable, and high-performance computing infrastructure to support its growing workload demands. We utilized the tool to objectively compare these products based on the main aspects we had already defined in the evaluation framework.

The selected IaaS products simulate real-world decision-making choices, where an organization with growing workload demands seeks to identify the optimal computing solution for their business operations among various cloud service competitors. The organization's key characteristics include workload diversity, scalability, security compliance, and cost efficiency. These factors are familiar challenges faced by organizations seeking a high-performing computing solution. Table 21: IaaS Products Table lists the ten computed IaaS products selected for evaluation.

Table 21: IaaS Products Table

Product	CSP	Product Description
m7a.medium	Amazon Web Services	Amazon EC2 M7a instances, powered by 4th Generation AMD EPYC processors, deliver up to 50% higher performance than M6a instances. These instances support AVX-512, VNNI, and bfloat16, which enable support for more workloads. They use Double Data Rate 5 (DDR5) memory to enable high-speed access to data in memory and deliver 2.25x more memory bandwidth compared to M6a instances.
c3-highcpu-4	Google Compute Engine	Compute-optimized high-CPU instance type with 4 vCPUs and 4 GB RAM, ideal for workloads requiring high compute performance but minimal memory, such as batch processing or web serving.
Premium 4GB	Linode	Compute instance with 4GB of memory. Linode instances are known for simplicity, predictable pricing, and high performance, making them suitable for various workloads, from simple web hosting to more demanding tasks like data processing and machine learning.
BL2.2x4	IBM Cloud	Balanced local storage profiles are primarily for large database workloads that require high I/O performance with low latency. Network performance ranges from standard to premium.
E5 4 threads 8GB	Oracle Cloud	E4-based standard compute. Processor: AMD EPYC 7J13. Base frequency 2.55 GHz, max boost frequency 3.5 GHz.
CPU Optimized 4GB	Digital Ocean	CPU-optimized Drops provide a 2:1 ratio of memory to CPU. This configuration with fast (2.6GHz+) dedicated vCPUs is optimal for applications that demand fast, consistent performance, such as media streaming, gaming, and data analytics. The Premium variant of CPU-optimized Drops also provides up to 10Gbps outbound network speeds and NVMe SSDs.
ecs.c7a.large	Alibaba Cloud	This instance family uses the third-generation SHENLONG architecture to provide predictable and consistent ultra-high performance. It utilizes fast path acceleration

		on chips to improve storage performance, network performance, and computing stability by an order of magnitude.
D2ls v5	Microsoft Azure	The Dlsv5 series virtual machines are based on the 3rd Generation Intel® Xeon® Platinum 8370C (Ice Lake) processor in a hyper-threaded configuration. This custom processor can reach an all-core Turbo clock speed of up to 3.5GHz and features Intel® Turbo Boost Technology 2.0, Intel® Advanced Vector Extensions 512 (Intel® AVX-512) and Intel® Deep Learning Boost. VM sizes provide 2GiBs RAM per vCPU and are optimized for workloads that require less RAM per vCPU than standard VM sizes, such as small-to-medium databases, low-to-medium traffic web servers, virtual desktops, application servers, and more.
b3-16	OVHcloud	General-purpose instances can get projects up and running stress-free and with all the scalability options open. These instances have a balanced ratio of 4 GB RAM per high-end AMD EPYC vCPU to cover nearly all of your standard use cases.
High Frequency 16 GB	Vultr	a VPS option designed for applications requiring high performance. This plan is part of Vultr's High-Frequency server line and is powered by 3GHz+ Intel Xeon CPUs and NVMe SSD storage for fast data access and transfer speeds.

These products will be evaluated across all six key criteria of the evaluation framework: availability, performance, scalability, reliability, security, and cost-effectiveness. The evaluation will focus on gathering the identified and chosen metric data to assess these aspects.

This evaluation aims to determine which compute IaaS product best aligns with the organization's operational and business objectives, ensuring the fulfillment of its requirements. The organization will be able to evaluate all products across all the set predefined scenarios in the tool and a custom scenario, which will focus on different criteria from the set scenarios for showing usability and flexibility options. Therefore, the products will be evaluated based on:

- Balanced Scenario assigning equal weight to all six aspects, ensuring no single factor is prioritized over another without bias towards any particular criteria.
- Performance-Focused Scenario: Prioritizing criteria that affect the performance and availability of IaaS providers' products, such as CPU performance, data transfer rates, and network performance.
- Cost-Sensitive Scenario focuses on cost-effectiveness and scalability, ideally looking to optimize spending on cloud services.
- Security-Centric Scenario: This scenario emphasizes security-related criteria with reliability in mind, including data encryption levels and compliance certifications, for organizations where data protection is paramount.
- Custom Scenario: Emphasizing Performance and Security for optimal results and heavy compliance requirements.

Since some of the defined metrics in the evaluation framework require actual data, such as genuine quotations, to assess the total cost of ownership and recovery time to measure downtime and service restoration in practice, the organization will need to engage directly with the cloud providers to obtain this information. We have made the best and most realistic assumptions for this case study where actual data could not be obtained. These assumptions are based on industry standards, publicly available pricing information, and known performance benchmarks. Other metrics, such as performance, scalability, and security-related data, have been collected from reliable sources, including provider documentation, third-party reports, and case studies. The detailed data collection process for these metrics will be explained in the next section. With the products and evaluation criteria established, we now delve into the data collection process, which is critical for an objective evaluation.

6.3 Scenario Study Case: Data Collection for Metric Evaluation

The data collection process for evaluating the four IaaS alternatives involved gathering qualitative and quantitative metrics. The sources of this data were as follows:

- Provider Documentation: Technical documentation, service level agreements (SLAs), and performance reports published by each IaaS provider.
- Third-Party Benchmarks: Independent performance tests, security audits, and cost analyses conducted by industry experts.
- Case Studies: Feedback from existing cases of the IaaS providers' products, gathered from online reviews and detailed case studies published by clients.
- Internal Company Needs and Use Cases: Input from the organization's internal stakeholders, identifying specific performance, cost, security, and scalability requirements based on real-world needs.

The evaluation of the collected metric data and the scoring scaling are explained in detail in the appendix, the scaling tables, and the tool Excel separate sheet for the rating scale.

6.4 Study Case Results

This section summarizes the results generated by applying the AHP-based evaluation framework to the ten selected IaaS products. Each product was evaluated under the five predefined scenarios (Balanced, Performance-Focused, Cost-Sensitive, Security-Centric, and Custom which is Performance & Security focused scenario) to reflect different organizational priorities. The results for each scenario are presented below in Table 22: Study Case Results, along with the ranking of products and insights drawn from the findings. The following visual representations, including heatmaps and clustered bar charts, provide a comparative view of each product's performance across the various scenarios, making it easier for decision-makers to identify trends and key differentiators.

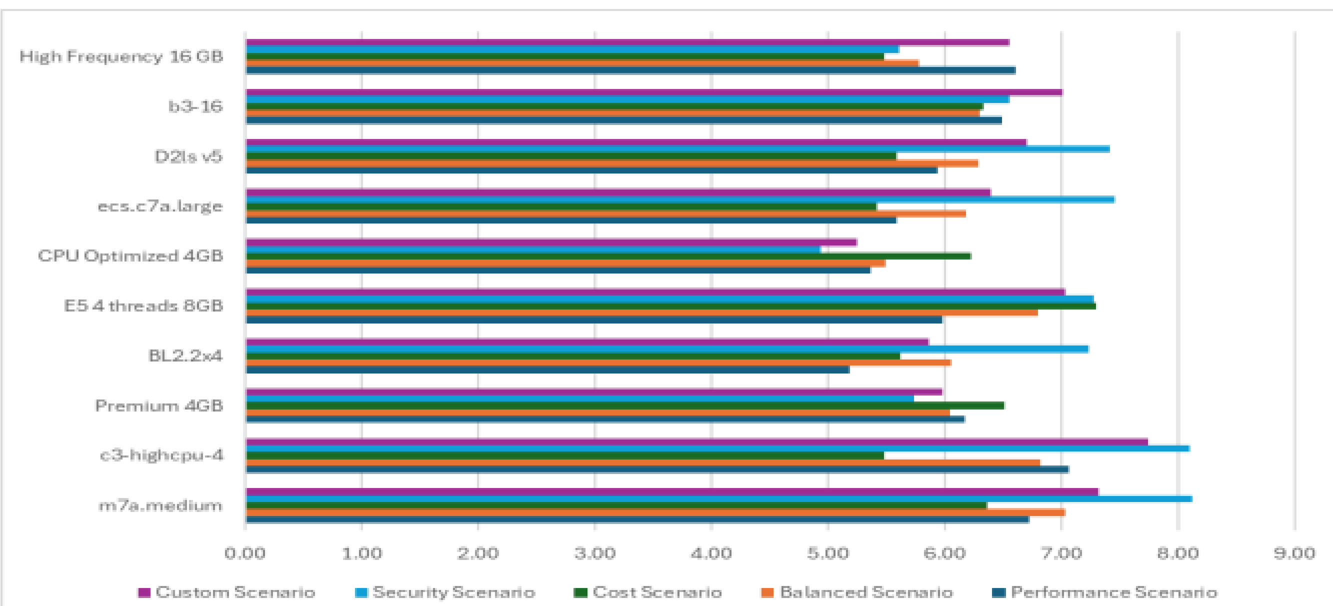
Table 22: Study Case Results

Product	Service Provider	Performance Scenario	Balanced Scenario	Cost Scenario	Security Scenario	Custom Scenario
m7a.medium	Amazon Web Services	6.73	7.04	6.37	8.12	7.32
c3-highcpu-4	Google Compute	7.07	6.82	5.48	8.10	7.74
Premium 4GB	Linode	6.17	6.05	6.51	5.74	5.98
BL2.2x4	IBM Cloud	5.18	6.06	5.62	7.24	5.86
E5 4 threads 8GB	Oracle Cloud	5.98	6.80	7.30	7.28	7.04
CPU Optimized 4GB	Digital Ocean	5.36	5.49	6.22	4.94	5.25
ecs.c7a.large	Alibaba Cloud	5.59	6.18	5.42	7.46	6.40
D2ls v5	Microsoft Azure	5.94	6.29	5.59	7.42	6.71
b3-16	OVHcloud	6.49	6.30	6.34	6.56	7.02
High Freq 16 GB	Vultr	6.61	5.78	5.48	5.61	6.56

The heatmap in the table above visualizes the evaluation scores of ten IaaS products across five scenarios: Performance, Balanced, Cost, Security, and Custom. The green represents higher scores, indicating better performance in that particular scenario. By examining the heatmap, decision-makers can quickly identify which products excel in different criteria, with key strengths being more easily distinguishable through color gradients from green to red.

Figure 20: Study Case Results shows the clustered bar chart comparing the ten IaaS products across five scenarios. Each product is represented by a separate set of bars, allowing for a side-by-side performance comparison across scenarios.

Figure 20: Study Case Results



6.4.1 Balanced Scenario Results

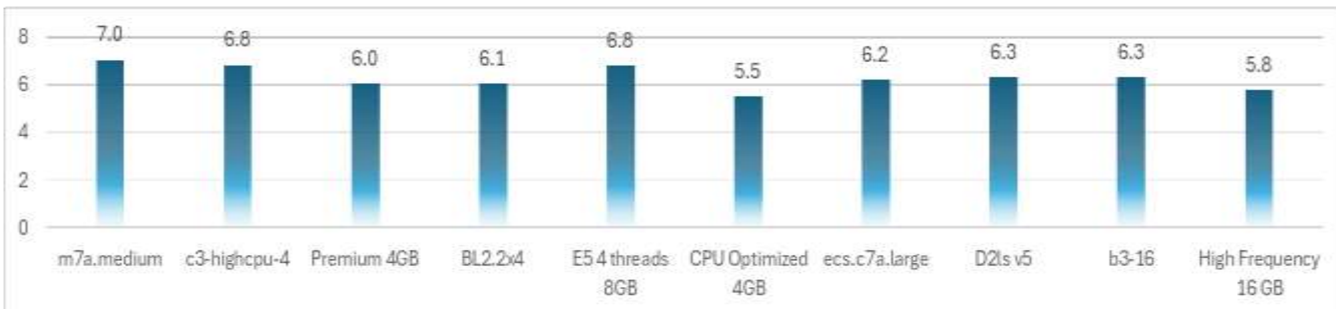
In this scenario, all six criteria were assigned equal weight, ensuring no single aspect was prioritized over another. This scenario is designed for organizations seeking an overall balanced performance from their IaaS product.

Top Performing Products:

- Amazon EC2 m7a.medium (7.04) and Google Compute Engine (6.82) stood out in this scenario, offering a well-rounded performance across all six criteria.

Insights: Amazon EC2 m7a.medium performed the best, balancing performance, security, and cost-effectiveness. Google Compute Engine offers a comparable alternative with overall solid performance. Visual Representation:

Figure 21: Balanced Scenario Results



6.4.2 Performance-Focused Scenario Results

This scenario prioritizes performance and availability, emphasizing cost and security less. It is ideal for organizations with high computational demands and minimal tolerance for downtime.

Top Performing Products:

- Google Compute Engine (7.07) emerged as the top performer in this scenario, excelling in performance-based metrics such as CPU performance and data transfer rates.
- Amazon EC2 m7a.medium (6.73) followed closely, demonstrating solid performance across various high-computational tasks.

Insights: Google Compute Engine is ideal for organizations prioritizing high performance, particularly in compute-heavy environments. Visual Representation:

Figure 22: Performance-Focused Scenario Results

6.4.3 Cost-Sensitive Scenario Results

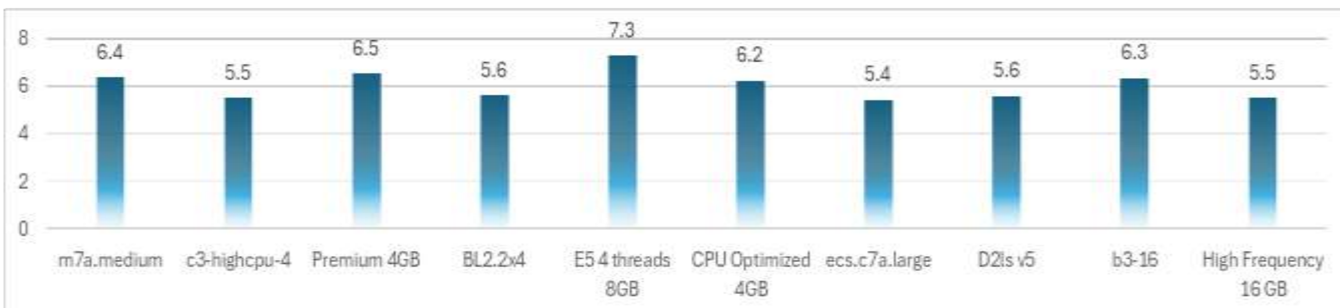
The Cost-Sensitive Scenario prioritizes cost-effectiveness and scalability, making it suitable for organizations looking to optimize cloud service spending while ensuring scalability.

Top Performing Products:

- Oracle Cloud (7.30) and Linode (6.51) performed the best in this cost-driven scenario, offering cost-effective solutions with scalability.

Insights: Oracle Cloud emerged as the top choice for organizations seeking cost-effective cloud solutions while maintaining scalability, with Linode as another strong contender for budget-conscious organizations. Visual Representation:

Figure 23: Cost-Sensitive Scenario Results



6.4.4 Security-Centric Scenario Results

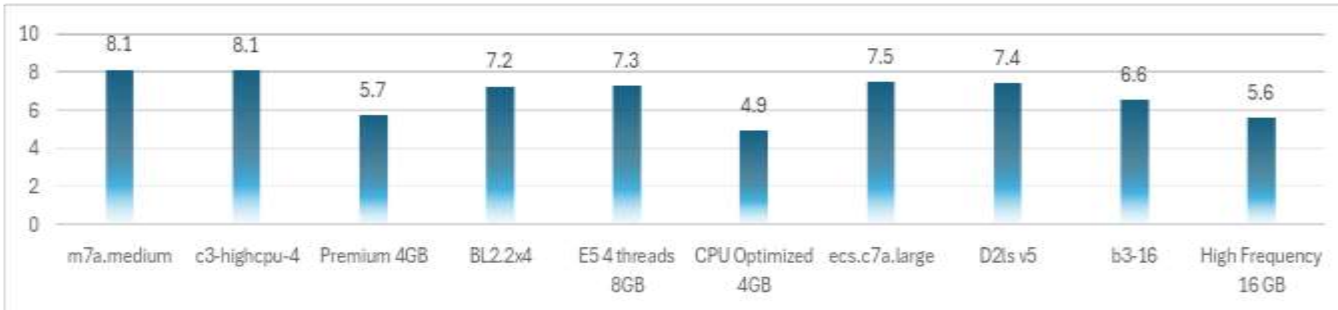
This scenario prioritizes security and reliability, making it suitable for organizations that handle sensitive data and require compliance with industry security standards.

Top Performing Products:

- Amazon EC2 m7a.medium (8.12) and Google Compute Engine (8.10) dominated this scenario, prioritizing security, and compliance.

Insights: Amazon EC2 and Google Compute Engine offer comprehensive security features, including strong encryption and compliance certifications, making them ideal for organizations with stringent security needs. Visual Representation:

Figure 24: Security-Centric Scenario Results



6.4.5 Custom Scenario Results

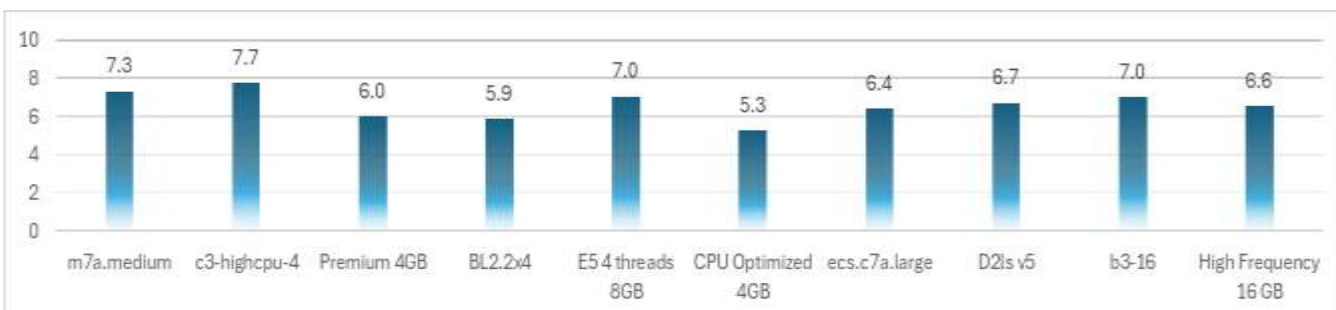
In this custom scenario, performance and security were prioritized based on the organization's need for high computational power and strict compliance requirements.

Top Performing Products:

- Google Compute Engine (7.74) and Amazon EC2 m7a.medium (7.32) excelled in this scenario, focusing on performance and security.

Insights: Google Compute Engine emerged as the top choice, offering the best combination of performance and security, closely followed by Amazon EC2. Visual Representation:

Figure 25: Custom Scenario Results



6.4.6 Conclusion

After applying the AHP-based tool to evaluate the 10 IaaS products across various predefined and custom scenarios, it is evident that various products excel in different areas depending on the organization's priorities. Google Compute Engine consistently performed well in performance and security-related scenarios, making it a top contender for high-performance, secure workloads. Amazon EC2 m7a.medium was the most well-rounded option, excelling in balanced and security-centric scenarios. Oracle Cloud and Linode offer the best value for organizations focused on cost-efficiency without compromising scalability or performance.

6.5 Study Case Data Validation and Challenges

During the validation phase of the study, various challenges arose related to data availability and accuracy. While some metrics were straightforward to collect using official documentation, SLAs, and white papers from the IaaS providers, other metrics required assumptions due to limited data availability or the inherent complexity of real-world conditions.

6.5.1 Data Sources and Validation

The primary data sources used for this evaluation included the official websites, SLAs, and white papers from the respective cloud service providers. To further validate some of the results, we utilized third-party sources like VPSBenchmarks[56], TechRadar[57], and Geekbench Cloud[58]. These sources provided additional benchmarks and independent performance analyses to evaluate each IaaS product comprehensively. The detailed logic behind the rating scale and evaluation criteria is explained in the appendix and embedded within the AHP Excel tool.

6.5.2 Assumptions for Certain Metrics

Several metrics posed unique challenges during the data collection process, leading to further assumptions being made. Specifically:

For a more accurate calculation of TCO, actual quotes from service providers would have been necessary. However, given the constraints of this study, we relied on provider price calculators to make reasonable assumptions. The estimates were generated based on the Pay-As-You-Go pricing model for one year, with the following assumptions:

- Resource Costs: Monthly cost of the selected instance type.
- Data Transfer Fees: Assumed 10 TB of outbound data transfer per month.
- Storage Fees: We assumed 5 TB of block storage (SSD), which allowed for a rough estimate of TCO, though real-world quotations may produce more precise results.
- Storage Fees were based on block storage (SSD), commonly used for high-performance computing workloads due to its faster read/write speeds than traditional HDDs.

Table 23: Product TCO DATA is a visual representation of how the total cost of ownership was calculated.

Table 23: Product TCO DATA

Provider	Cost per Resource (Year)	Data Transfer Fees (Year)	Storage Fees (Year)	Total TCO (Year)	Storage Type	Reference
Amazon EC2	$519.96	$10,800	$6,000	$17,319.96	Amazon EBS General Purpose SSD (gp2)	AWS Pricing[1]
Google Compute Engine	$1,673.40	$10,200	$10,200	$22,073.40	Standard Persistent Disks (SSD)	Google Cloud Pricing[2]
Linode	$516.00	$600	$6,000	$7,116.00	Block Storage (Standard SSD)	Linode Pricing[3]
IBM Cloud	$752.16	$10,800	$6,000	$17,552.16	Standard SSD Block Storage	IBM Cloud Pricing[4]
Oracle Cloud	$702.48	$0	$1,500	$2,202.48	Block Volume (SSD)	Oracle Cloud Pricing[5]
DigitalOcean	$564.00	$1,200	$6,000	$7,764.00	Block Storage (Standard SSD)	DigitalOcean Pricing[6]
Alibaba Cloud	$699.84	$13,200	$6,000	$19,899.84	Enhanced SSD	Alibaba Cloud Pricing[7]
Microsoft Azure	$744.60	$9,600	$6,000	$16,344.60	Standard SSD Managed Disks	Azure Pricing[8]
OVHcloud	$867.84	$1,200	$4,800	$6,867.84	Block Storage (Standard SSD)	OVHcloud Pricing[9]
Vultr	$1,152.00	$1,200	$6,000	$8,352.00	Block Storage (Standard SSD)	Vultr Pricing[10]

Moreover, accurate recovery time data for IaaS products is rarely provided directly, as it depends on infrastructure, system configurations, and the tools used for disaster recovery. To estimate recovery times, we relied on SLAs, benchmarks, and disaster recovery features provided by each vendor.

[1] https://calculator.aws/#/
[2] https://cloud.google.com/products/calculator
[3] https://www.linode.com/pricing
[4] https://www.ibm.com/flashsystem/pricing/nl-en?utm_content=SRCWW&p1=Search&p4=43700077616024396&p5=p&p9=58700008515054796&gad_source=1&gclid=CjwKCAjwufq2BhAmEiwAnZqw8jGf48EAJ57wb9XNMkPn_ObGpDX807krcSQWmc7MyITc8jSoNQ4KYxoC8EkQAvD_BwE&gclsrc=aw.ds
[5] https://www.oracle.com/cloud/cost-estimator.html
[6] https://www.digitalocean.com/pricing/calculator?utm_campaign=emea_brand_kw_en_cpc&utm_adgroup=digitalocean_pricing_exact&_keyword=digitalocean%20pricing&_device=c&_adposition=&utm_content=conversion&utm_medium=cpc&utm_source=google&gad_source=1&gclid=CjwKCAjwufq2BhAmEiwAnZqw8uLqtUc6OvfQgCZiwkZSe772a9nIfvguwZJG2edQvsDrR2ord5LL8BoCqSMQAvD_BwE#cart=%5B%5D
[7] https://www.alibabacloud.com/en/pricing?_p_lc=1&spm=a3c0i.7938564.6791778070.101.184e441eMsO6fd
[8] https://azure.microsoft.com/en-us/pricing/calculator/
[9] https://www.ovhcloud.com/en/public-cloud/prices/
[10] https://www.vultr.com/pricing/

The following scoring scale was developed to evaluate recovery times:

- 5: Best recovery times with solid disaster recovery features and high availability, likely under 5 minutes.
- 4: Good recovery times, provisioning times under 10 minutes, and some disaster recovery tools.
- 3: Average recovery times around 10–30 minutes with limited recovery features.
- 2: Slower recovery times requiring manual intervention.
- 1: Little or no disaster recovery features, long recovery times expected.

Moreover, finally, calculating network latency accurately required real-time testing with specific configurations, which was outside the scope of this study. However, to simulate network latency, we used data from the VPSBenchmarks Web Server Latency metric[56]. While this benchmark focused primarily on web server performance, it was a rough proxy for latency under high load conditions in IaaS environments. Although it did not directly measure network latency, it provides an estimated view of performance in network-heavy applications. While assumptions helped address some data gaps, additional challenges were encountered during data collection and comparison.

6.5.3 Additional Challenges in Data Collection and Comparisons

While some metrics were relatively easy to collect, several challenges persisted throughout the data collection process. For example, there were notable data gaps in key metrics like Recovery Time and Network Latency, which depend heavily on real-time conditions and complex configurations. These metrics, which depend on real-time conditions and complex configurations, were challenging to assess. The lack of standardized testing across providers made it challenging to gather reliable data. While third-party benchmarks were valid, they were not fully aligned with real-world conditions specific to each IaaS offering.

Another challenge is that many metrics rely on publicly available data, which may only partially represent performance or cost under specific usage patterns. For instance, metrics like TCO would benefit from direct quotes from providers, which were not feasible to collect during this study. This reliance on generalized public data introduces potential inaccuracies in the evaluation process.

Additionally, while the AHP tool strives to maintain objectivity, some metrics, such as those involving recovery times and certain cost elements, require assumptions based on available data and benchmarks. As a result, the evaluation process included elements of estimation and approximation, which may vary depending on specific organizational requirements or real-time data.

The main challenges in validating the study case were data availability and accuracy, particularly for recovery time, network latency, and TCO-required estimations based on assumptions, benchmarks, and available tools. Difficulty in Direct Comparisons: while the study compared products across standardized metrics, not all products provided sufficient detail for direct, like-for-like comparisons. Reliability of public data since the evaluation relied on publicly available data, which may introduce discrepancies when compared to real-world scenarios. It is also important to mention that although the AHP tool aimed to reduce subjectivity, some assumptions were necessary due to data limitations, potentially affecting the objectivity of specific evaluations. Thus, future improvements in the validation process could include direct engagement with IaaS providers for access to more precise data and conducting real-time performance tests for more accurate latency and recovery metrics.

6.6 Third-Party Validation Using the Technology Acceptance Model (TAM)

In order to finalize the evaluation stage of the followed DSRM framework, we formulated a third-party validation to validate the usability and effectiveness of the AHP tool development. We decided to utilize the Technology Acceptance Model (TAM) as a guiding framework for the validation. To discuss TAM briefly, it was introduced by Fred Davis in 1989 as a theoretical framework used to explain and predict user behavior concerning the adoption of information technology. The model suggests and focuses on two primary factors—Perceived Usefulness (PU) and Perceived Ease of Use (PEOU), and how they significantly influence an individual's decision to use a system[75]. In other words, these constructs help determine whether users find the tool beneficial and easy to interact with, suggesting that ease of use influences perceived Usefulness, affecting the intention to use the technology and ultimately determining actual system usage [75]. This model has become a widely adopted tool for assessing the acceptance of various technologies across multiple domains, thus proving a valid choice to utilize in the validation process.

To gather industry expert opinions on the tool, a demonstration of the tool through a pre-recorded demo video and an interactive presentation that explained the tool's core features and functionality was presented. After familiarizing themselves with the tool, the experts were asked to test the AHP tool themselves, allowing them to interact with the system firsthand. Finally, the validation was concluded with a TAM-based survey to capture data across the three critical areas emphasized in the TAM model, alongside adding open-ended questions, adding a qualitative dimension to the feedback. As mentioned before, our main target is professionals considered industry experts experienced with cloud services and Infrastructure as a Service (IaaS) evaluation, thus ensuring that the feedback reflects practical, real-world application of the AHP tool. We tried to validate with experts from different areas of IaaS, from consultants in the fields to sales and after-sales, and finally, technical after-sales people. Nevertheless, the survey was designed to capture feedback on how effectively the AHP tool supports decision-making processes in IaaS provider selection and whether the tool is intuitive and user-friendly. The following section will discuss the methodology used in the survey and how it was constructed.

6.6.1 Survey Methodology for TAM Validation

To discuss the methodology used for TAM validation, A survey was conducted using the framework, and a short interview was conducted afterward to assess the open-ended part of the assessment process of the industry expert's validation of the AHP tool. It was designed to evaluate the tool's perceived Usefulness, ease of use, and potential for future adoption, with 23 questions emphasizing TAM's primary constructs. The questions were formulated with Six questions focusing on Perceived Usefulness, six questions focusing on Perceived Ease of Use, four questions focusing on intention of Use, and finally, even open-ended questions to address the qualitative aspect of the survey as shown in Table 24: Question Categories & classifications.

Table 24: Question Categories & classifications

Question Category	Description & Focus
Perceived Usefulness (6 questions)	Designed to assess whether the respondents believe the AHP tool improves their decision-making process in selecting IaaS products.
Perceived Ease of Use (6 questions)	Focused on whether respondents found the AHP tool intuitive and easy to navigate.
Self-Predicted Future Usage (4 questions)	Intended to gauge the likelihood of future use and preference over traditional methods.
Suggestions & Recommendations (7 questions)	Aim to allow the respondents to add their feedback and offer suggestions for improvements, additional features, or any other aspects of the tool that they felt required attention.

Moreover, the questions were presented in a Likert-scale questionnaire. Each question was rated on a seven-point Likert scale, ranging from 'extremely unlikely' to 'extremely likely,' allowing for nuanced feedback on user experience from the respondents. Seven open-ended questions were added to add a qualitative dimension and exploratory methods rather than purely quantitative ones to the results collected to address the small sample size limitation. Table 34: Measurement Scales for Perceived Usefulness and Perceived Ease of Use in the appendix lists all the questions and how they were structured.

6.6.2 Survey Data Analysis and Feedback Integration

We analyzed the gathered data to understand how the AHP tool was perceived regarding the primary TAM elements evaluated in our questionnaire. This is a crucial step since the insights gathered from this analysis would be pivotal in refining the AHP tool and ensuring it meets industry standards and user expectations or formulating our future recommended work. The analysis focused on using both quantitative and qualitative methods. This ensured a comprehensive understanding of the expert feedback covering all 23 questions.

For each participant, we collected:

- Quantitative Data: Scores for each TAM item (PU, PEOU, Intention to Use).
- Qualitative Data: Verbal or written feedback on open-ended questions.

We organized the data in Part 1: Quantitative Scores and Responses to the TAM questions. Part 2: Qualitative Insights from the interviews or open-ended questions. Based on the analysis of the survey results, the key insights will be presented in the coming sections.

6.6.3 Key Insights from Survey Results

The key results are divided into two sections. The quantitative results are presented with the Perceived Usefulness AVG, Perceived Ease of Use (PEOU) AVG, and Behavioral Intention to Use AVG, with a very close AVG between all three aspects of the TAM-based validation scoring (5.7). This suggests that participants generally found the tool both useful and easy to use, with a moderate to high likelihood of adopting it in the future. Table Table 25: TAM Questionnaire Results shows the gathered average scores for each participant with an average score for all participants together.

Table 25: TAM Questionnaire Results

Participant	PU (average score)	PEOU (average score)	Intention to Use (average score)
P1	5.83	6.17	5.75
P2	6.00	6.17	6.00
P3	6.17	5.33	5.25
P4	5.83	5.67	6.00
P5	4.67	5.17	4.75
AVG	**5.70**	**5.70**	**5.55**

The second section of the results presents the qualitative results of the questionnaire based on the seven open-ended questions and the short interview conducted with the expert. The feedback gathered was structured in a theme-like structure to formulate a thematic analysis of feedback from the five participants. This thematic analysis revealed four key themes that we could conclude regarding the AHP tool's usability and potential improvements, which focused on the use and implementation of the tool, improvement suggestions, critical evaluation criteria that could be added to the framework, and finally, the technical aspects tackling the user experience and the tool's interface.

1. **Consultation & Business Utility**: Several participants suggested that the tool could be valuable for consultants who must present solutions to clients rather than being integrated within the organization. The decision-making process is more complex in real-world scenarios than in practical theory. A notable suggestion also was the inclusion of red flag criteria, which would help streamline the selection process for products that must meet specific compliance or specific organization standards. This would thereby reduce the complexity of the decision-making process and eliminate many choices that do not fit the criteria. The tool value for SME with limited resources has been discussed and emphasized on as well as one way of utilizing the tool more effectively.
2. **Improvement Suggestions**: Experts identified areas where the tool could be enhanced. Notably, the product scorecard could be made more intuitive by highlighting the top three products more straightforwardly. While extra information in the sheet was understandable for academic purposes, modifications for practical business use could be implemented on the tool. Additionally, almost all discussed or addressed AI's role in improving such a tool; integrating AI-based recommendations could improve the decision-making process, especially for organizations that rely on automated analysis for product interoperability and solution alignment. Another potential improvement was automating data collection through reliable repositories or web scraping.
3. **Critical Evaluation Criteria**: Participants emphasized the importance of including criteria other than those selected in the evaluation framework, such as the financial stability and reputation of IaaS providers. Other suggested evaluation factors included learning curves, DevOps readiness, and the availability of talent capable of working with the provider's platform.
4. **User Experience & Interface**: The user experience was generally positive, particularly regarding the scorecard and customizable scenarios. However, participants noted that the tool's interface could be confusing, especially without a demo, suggesting that improvements in UI design could help users navigate the tool more easily. The choice of Excel might not be the best choice, although it is unstable since there are a lot of calculations done in the tool.

6.6.4 Conclusion of 3rd party Validation Process

To conclude, the third-party validation of the AHP tool was conducted using a combination of qualitative feedback gathered from the interviews and quantitative analysis with the questions filled out afterward, both based on the Technology Acceptance Model (TAM). The validation results indicate that the tool is well-received by participants, with high ratings for **Perceived Usefulness (PU)** and **Perceived Ease of Use (PEOU)**. These findings are complemented by valuable insights from user feedback, highlighting areas of strength and potential for improvement.

To present quantitative results, the average score for **PU** was **5.70**, suggesting that users recognize the tool's value in improving decision-making processes, particularly if placed in consultancy settings. Similarly, the average **PEOU** score of **5.70** indicates that participants found the tool easy to navigate, though some minor issues were raised regarding the tool's initial complexity for first-time users. While the **Intention to Use** score averaged at **5.55**, lower than others, it still shows a substantial likelihood of future use, with participants expressing satisfaction with the tool's capabilities and what it is trying to achieve. It is essential to mention that the participants considered the whole evaluation framework rather than the specific technical features of the tool itself.

On the other hand, the qualitative feedback revealed several recurring themes discussed earlier in section 6.1 insights that further support the tool's validation. Experts had positive feedback on the tool's scorecard feature and even more on the ability to accommodate customizable decision-making scenarios. However, suggestions for improvement were also noted, such as enhancing the scorecard visualization, integrating AI-based recommendations, and simplifying data collection processes. These suggestions reflect a broader interest in making the tool even more user-friendly and efficient, especially in a business context where time and accuracy are critical, and the role of AI is increasingly influencing the decision-making process.

The validation process has provided valuable insights into how the AHP tool can be further refined to meet users' needs. While the overall feedback was positive, with valuable suggestions to be incorporated into the tool, incorporating the suggested improvements—such as AI integration, enhanced data sourcing, and streamlined user navigation—is required in the future to ensure the tool's competitiveness and alignment with industry expectations.

Finally, while limited by sample size, this initial validation phase provides valuable insight and a solid foundation for future developments and broader testing. Addressing the areas identified for improvement will be critical in enhancing the tool's applicability across various sectors, ultimately increasing its adoption by industry professionals.

Chapter 7: Conclusion & Recommendations

This chapter presents the conclusion and recommendations by recapping the research objective and purpose, considering our main and sub-research questions proposed in the thesis. Then, we will illuminate critical findings, challenges, limitations, and recommendations for future work.

7.1 Recap of Research Objectives and Purpose

We have established that the current IaaS offering landscape requires evaluation across a multi-dimensional space, including advanced performance metrics, security standards, regulation compliance, service scalability, reliability, and integration capabilities with emerging technologies [3]. Organizations face the challenge of comparing different IaaS offerings on aspects beyond cost and basic performance metrics, as existing frameworks for comparing cloud services are often too complex, fall short of capturing the full spectrum of IaaS features, or are not updated frequently enough to reflect the latest industry developments [4]. The lack of a standardized set of objective comparison metrics hinders organizations' ability to make informed decisions that align with their strategic objectives. This gap in the literature and practice signifies a critical need for a robust framework that can assess IaaS offerings systematically, providing clear and actionable insights for businesses navigating this complex marketplace. This thesis proposes addressing this gap by developing an evaluation framework based on the AHP method to conduct a comprehensive and balanced evaluation of IaaS providers' products, facilitating decision-making that optimizes cost, performance, security, and alignment with business goals. To achieve our aim, we formulated the following research question.

M.Q. How can we develop objective comparison metrics encompassing IaaS, including availability, performance, reliability, scalability, and cost-effectiveness, to facilitate informed decision-making in cloud adoption?

This research question was instrumental in guiding the investigative process toward developing a comprehensive and standardized framework for IaaS assessment. It acknowledges the necessity for a multifaceted evaluation that does not solely rely on one or two dimensions but also provides a comprehensive approach to understanding the value proposition of IaaS offerings.

During our journey to answer our main research question, we established five sub-research questions that were formulated to dissect the central inquiry into manageable segments, each focusing on IaaS in relation to other cloud models such PaaS and SaaS, specific aspect of the IaaS evaluation, and pricing models used which we all answered in our literature review establishing the foundation base for designing the objective comparison metrics framework, and eventually facilitating the formalization of the final AHP evaluation framework we have created. These sub-questions were:

S.Q.1: Cloud Service Models: What are the most common cloud service models, and what are their architectures and differences? (SaaS, PaaS, and IaaS)

S.Q.2: Cloud Computing Service Metrics: What CCSM /frameworks currently exist, and how do they differ?

S.Q.3: Performance, Reliability, Scalability: How can the existing metrics be linked to measure performance, reliability, and scalability?

S.Q.4: Decision-making Framework: What frameworks can be utilized to establish the objective metrics in informed decision-making in cloud adoption?

S.Q.5: Cost-effectiveness: Which financial models and cost structures can be applied to compare the cost-effectiveness of IaaS providers transparently?

7.2 Summary of Key Findings

The key findings were divided into three main categories representing the AHP Framework and Tool Development, the Scenario Study Case Results, and finally the Third-Party Validations Outcomes.

7.2.1 AHP Framework and Tool Development

As presented in the chapter Designing an Objective Comparison Metrics Framework, we have created an evaluation framework that utilized the Analytical hierarchy method (AHP) for objective decision-making assessment of IaaS products. The Framework established six critical aspects to evaluate when assessing IaaS products. These six criteria were (Availability, Performance, Scalability, Reliability, Security, and Cost-effectiveness). The evaluation framework was designed with the end user perspective, focusing on flexibility, ease of use, implementation, and simplicity. It offers predefined scenarios focusing on performance, cost, and security while leaving the opportunity to create a custom one.

7.2.2 Scenario Study Case Results:

To demonstrate the tool's implementation in a practical, real-world scenario, we conducted a scenario study case where Ten IaaS products were evaluated across distinctive scenarios. The tool successfully and effectively demonstrated its ability to assess the performance of these products across different selected criteria, showing how they ranked, thus potentially assisting in the decision-making and selection process in case the tool is implemented by an organization in the future.

The tool successfully applied user-assigned pairwise comparisons to calculate weights to the evaluation criteria and executed the necessary calculations to score the products accurately. However, the process of collecting relevant metric data proved to be time-consuming. This was primarily due to challenges related to the data's availability, validity, and objectivity. Despite these hurdles, the data collection was manageable, though it became evident that the reliability of the data plays a critical role in determining accurate scores.

In summary, while the tool functions as expected and performs calculations flawlessly, its effectiveness depends on the quality and accuracy of the data the user provides. In our case, evaluating the IaaS products underscored the need for reliable data to ensure meaningful and objective results.

7.2.3 Third-Party Validation Using TAM

The AHP tool underwent third-party validation through qualitative feedback from interviews and quantitative analysis based on the Technology Acceptance Model (TAM), as mentioned in section 6.2.2. The validation results indicate strong acceptance of the tool, with high scores for Perceived Usefulness (PU) and Perceived Ease of Use (PEOU), averaging 5.70. These findings suggest that users find the tool valuable in improving decision-making, especially in consultancy settings, and easy to navigate despite minor complexities for first-time users. Meanwhile, the Intention to Use score was lower, with an average of 5.55, but it still reflected a positive inclination toward future use, though slightly lower than other metrics. It is essential to mention that participants considered the overall evaluation framework while conducting the evaluation rather than just technical features, appreciating the tool's ability to handle customizable decision-making scenarios reflected in the final score presented.

As a result, qualitative feedback echoed these findings, praising the scorecard feature and the tool's flexibility while recommending improvements such as enhanced scorecard visualization, AI-based recommendations, and simplified data collection processes. These suggestions align with AI's growing importance in decision-making and the need for efficient, user-friendly tools in business settings. Additionally, the feedback emphasized the point on the tool's usability and how it can be integrated within an organization. It underscored the tool's value in consultancy aspects and the value it adds for SMEs with limited resources for evaluating IaaS products. The third-party validation suggests that the tool will have a higher impact if used in the discussed context rather than within the organization's processes. Final thoughts to conclude while the validation results are promising, further refinement is necessary to incorporate the feedback and ensure the tool remains competitive and aligned with industry expectations, and an expanded validation with a larger pool is needed to fully validate the tool comprehensively.

7.3 Challenges and Limitations

Several aspects must be discussed and presented objectively to address our challenges and limitations while creating the evaluation framework. The main challenges were the critical aspects we needed to evaluate in the Framework, data availability related to metrics utilized in creating the Framework, comparison complexity as there is varying documentation and limited standardization across providers, and finally, subjective and weight assignments, as they may influence the evaluation's outcomes.

7.4 Recommendations for Future Work

Based on the case study and the third-party validation, the AHP-based evaluation tool developed in this study has shown significant potential in helping organizations evaluate IaaS products in a clear and objective structure. Nonetheless, several opportunities exist for improvement and expansion in future iterations, in which key areas where future work can focus on enhancing the tool's functionality and broader applicability.

While the current choice of an Excel-based tool serves its purpose as a functional prototype and was utilized due to its leverage in conducting complex calculations, its interface lacks comprehensive navigation assistance, which may appeal to only some users. Therefore, transitioning it into a web-based application would significantly enhance its usability and accessibility and is considered a necessity. Experts also support this conclusion and believe a web application would provide a more intuitive and user-friendly interface, improving the user experience. The development of web-based would have much-added value and benefits, allowing for greater scalability and facilitating integration with real-time data and sources. It will enhance the tool's ability to handle larger datasets efficiently and allow for multiple user inputs for weight assessment, especially if integrated directly with cloud service providers where they highlight their metric data related to their product offerings.

Interoperability and integration possibilities also positively impact data accessibility and accuracy, which is essential for improving the precision of the evaluation metrics, such as Total Cost of Ownership (TCO), recovery times, and network latency. Collaborating directly with IaaS providers or leveraging third-party portals that rate and evaluate cloud services could provide better access to critical data. These partnerships would enhance the accuracy of assessments and reduce reliance on assumptions, leading to more informed decision-making.

Additionally, future work should expand scenario customization options to increase the tool's relevance across various sectors, especially in sectors with unique requirements such as privacy regulations, compliance, and even sustainability criteria. Tailoring the tool to support industry-specific criteria would make it more versatile and applicable to a broader range of users, ensuring that the evaluations meet sector-specific needs. However, the focus at this stage should be on the validity and reliability of the tool itself.

We can only discuss future work by discussing the considerable potential for enhancing the tool by integrating AI-driven analytics and automation. Especially the IaaS product metric data collection, as it could automate the collection of real-time data from IaaS providers and third-party sources, significantly improving the accuracy and timeliness of evaluations. Additionally, AI could be used for predictive analysis, forecasting potential cost and performance trends, and recommending optimal cloud service configurations based on an organization's needs. This would add significant value to the evaluation framework by providing dynamic, data-driven insights.

While the current validation efforts through the scenario study case and third-party feedback have provided valuable insights and were sufficient for the current study, expanding testing across various industries is crucial for further framework improvements. Validating the tool's effectiveness in live environments would ensure its practical application and reveal potential improvements. Future work should focus on testing the tool within various organizational contexts, collecting real-world feedback, and iterating on the tool's design based on actual use cases. For example, testing the tool with a well-established consultant agency over a set period will provide valuable insight into its usability and the required future improvements.

Additionally, this version of the evaluation framework represents the first iteration, and as such, some areas require further refinement. Expanding the Framework to incorporate more detailed product specifications, such as hardware configurations and data center locations, were excluded due to time constraints; however, it would provide a more comprehensive assessment of IaaS products. Finally, as cloud technologies continue to evolve, the Framework must adapt to new advancements related to edge computing and AI-powered cloud services. Regular updates to the Framework will ensure its relevance in the fast-evolving cloud services landscape.

7.5 Final Thoughts

The thesis provides a strong foundation for a comprehensive IaaS evaluation framework that utilizes the AHP method as its core for the decision-making process, as it addresses a critical need for objective cloud service assessments. While the validation results are promising and show potential practical utilization, especially within the consultancy domain, ongoing refinement and expanded validation efforts will be necessary to ensure the tool's continued relevance, practical implementation, and effectiveness in the rapidly evolving cloud services landscape.

References

[1] A. Modi, "IaaS vs. PaaS vs. SaaS," DEV Community. Accessed: Nov. 09, 2023. [Online]. Available: https://dev.to/cloudtech/iaas-vs-paas-vs-saas-41d2

[2] M. De Donno, K. Tange, and N. Dragoni, "Foundations and Evolution of Modern Computing Paradigms: Cloud, IoT, Edge, and Fog," *IEEE Access*, vol. 7, pp. 150936–150948, 2019, doi: 10.1109/ACCESS.2019.2947652.

[3] S. Shirur, "A Cloud Service Measure Index Framework to Evaluate Efficient Candidate with Ranked Technology," vol. 4, no. 3, 2013.

[4] J. Siegel and J. Perdue, "Cloud services measures for global use: The Service Measurement Index (SMI)," presented at the Annual SRII Global Conference, SRII, 2012, pp. 411–415. doi: 10.1109/SRII.2012.51.

[5] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, Dec. 2007, doi: 10.2753/MIS0742-1222240302.

[6] B. Aydin, "THE STATE-OF-THE-ART MATRIX ANALYSIS FOR USABILITY OF LEARNING MANAGEMENT SYSTEMS," *Computers in Education Journal of ASEE*, Jan. 2016.

[7] F. Liu *et al.*, "NIST Cloud Computing Reference Architecture," Sep. 2011.

[8] "Service Measurement Index Framework Version PDF Free Download.pdf."

[9] F. De Vaulx, E. Simmon, and R. Bohn, "Cloud computing service metrics description," National Institute of Standards and Technology, Gaithersburg, MD, NIST SP 500-307, Apr. 2017. doi: 10.6028/NIST.SP.500-307.

[10] G. S. Sureshchandar and R. Leisten, "A framework for evaluating the criticality of software metrics: an analytic hierarchy process (AHP) approach," *Measuring Business Excellence*, vol. 10, no. 4, pp. 22–33, Jan. 2006, doi: 10.1108/13683040610719254.

[11] "service-measurement-index-framework-version-2.pdf."

[12] S. J. Bigelow and K. Marko, "The history of cloud computing explained," WhatIs.com. Accessed: Nov. 20, 2023. [Online]. Available: https://www.techtarget.com/whatis/feature/The-history-of-cloud-computing-explained

[13] W. Chai and K. Casey, "What is SaaS (Software as a Service)? Everything You Need to Know," Cloud Computing. Accessed: Nov. 20, 2023. [Online]. Available: https://www.techtarget.com/searchcloudcomputing/definition/Software-as-a-Service

[14] W. Chai, K. Brush, and S. J. Bigelow, "What is PaaS? Platform as a Service Definition and Guide," Cloud Computing. Accessed: Nov. 20, 2023. [Online]. Available: https://www.techtarget.com/searchcloudcomputing/definition/Platform-as-a-Service-PaaS

[15] M. Boisvert, S. J. Bigelow, and W. Chai, "What is IaaS? Infrastructure as a Service Definition | TechTarget," Cloud Computing. Accessed: Nov. 09, 2023. [Online]. Available: https://www.techtarget.com/searchcloudcomputing/definition/Infrastructure-as-a-Service-IaaS

[16] B. Bhagat, S. Khobragade, and A. Arudkar, "The Evolution of Cloud Computing".

[17] N. Dimitri, "Pricing cloud IaaS computing services," *J Cloud Comp*, vol. 9, no. 1, p. 14, Dec. 2020, doi: 10.1186/s13677-020-00161-2.

[18] C. M. Mohammed and S. R. M. Zeebaree, "Sufficient Comparison Among Cloud Computing Services: IaaS, PaaS, and SaaS: A Review," Jan. 2021, doi: 10.5281/ZENODO.4450129.

[19] "Amazon EC2 – Secure and resizable compute capacity – AWS," Amazon Web Services, Inc. Accessed: Nov. 21, 2023. [Online]. Available: https://aws.amazon.com/ec2/pricing/

[20] V. V. Rajendran and S. Swamynathan, “Parameters for comparing cloud service providers: A comprehensive analysis,” in *2016 International Conference on Communication and Electronics Systems (ICCES)*, Oct. 2016, pp. 1–5. doi: 10.1109/CESYS.2016.7889826.
[21] A. K. Bardsiri and S. M. Hashemi, “QoS Metrics for Cloud Computing Services Evaluation,” *IJISA*, vol. 6, no. 12, pp. 27–33, Nov. 2014, doi: 10.5815/ijisa.2014.12.04.
[22] T. D. Dabade, “INFORMATION TECHNOLOGY INFRASTRUCTURE LIBRARY (ITIL),” 2010.
[23] “ITIL - IT Infrastructure Library | IBM.” Accessed: Jan. 05, 2024. [Online]. Available: https://www.ibm.com/topics/it-infrastructure-library
[24] martinekuan, “Microsoft Cloud Adoption Framework for Azure - Cloud Adoption Framework.” Accessed: Jan. 06, 2024. [Online]. Available: https://learn.microsoft.com/en-us/azure/cloud-adoption-framework/
[25] 14:00-17:00, “ISO/IEC 20000-1:2018,” ISO. Accessed: Jan. 06, 2024. [Online]. Available: https://www.iso.org/standard/70636.html
[26] W. Pauley, “Cloud Provider Transparency: An Empirical Evaluation,” *IEEE Secur. Privacy Mag.*, vol. 8, no. 6, pp. 32–39, Nov. 2010, doi: 10.1109/MSP.2010.140.
[27] X. Zheng, P. Martin, K. Brohman, and L. D. Xu, “CLOUDQUAL: A Quality Model for Cloud Services,” *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1527–1536, May 2014, doi: 10.1109/TII.2014.2306329.
[28] S. S. Wagle, M. Guzek, P. Bouvry, and R. Bisdorff, “An Evaluation Model for Selecting Cloud Services from Commercially Available Cloud Providers,” in *2015 IEEE 7th International Conference on Cloud Computing Technology and Science (CloudCom)*, Nov. 2015, pp. 107–114. doi: 10.1109/CloudCom.2015.94.
[29] P. Kaushik, A. M. Rao, D. P. Singh, S. Vashisht, and S. Gupta, “Cloud Computing and Comparison based on Service and Performance between Amazon AWS, Microsoft Azure, and Google Cloud,” in *2021 International Conference on Technological Advancements and Innovations (ICTAI)*, Nov. 2021, pp. 268–273. doi: 10.1109/ICTAI53825.2021.9673425.
[30] Z. Li, L. O’Brien, H. Zhang, and R. Cai, “On a Catalogue of Metrics for Evaluating Commercial Cloud Services,” in *2012 ACM/IEEE 13th International Conference on Grid Computing*, Beijing, China: IEEE, Sep. 2012, pp. 164–173. doi: 10.1109/Grid.2012.15.
[31] A. Chauhan, “A Comparative Study of Cloud Computing Platforms,” *TURCOMAT*, vol. 11, no. 1, pp. 821–826, Apr. 2020, doi: 10.17762/turcomat.v11i1.13563.
[32] J. Ding, L. Sha, and X. Chen, “Modeling and evaluating IaaS cloud using performance evaluation process algebra,” in *2016 22nd Asia-Pacific Conference on Communications (APCC)*, Yogyakarta, Indonesia: IEEE, Aug. 2016, pp. 243–247. doi: 10.1109/APCC.2016.7581504.
[33] H. M. Alabool and A. K. Mahmood, “Review on cloud service evaluation and selection methods,” in *2013 International Conference on Research and Innovation in Information Systems (ICRIIS)*, Kuala Lumpur, Malaysia: IEEE, Nov. 2013, pp. 61–66. doi: 10.1109/ICRIIS.2013.6716686.
[34] N. Upadhyay, “Managing Cloud Service Evaluation and Selection,” *Procedia Computer Science*, vol. 122, pp. 1061–1068, 2017, doi: 10.1016/j.procs.2017.11.474.
[35] A. Guptha, H. Murali, and S. T, “A Comparative Analysis of Security Services in Major Cloud Service Providers,” in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, May 2021, pp. 129–136. doi: 10.1109/ICICCS51141.2021.9432189.
[36] X. Guerron, S. Abrahão, E. Insfran, M. Fernández-Diego, and F. González-Ladrón-De-Guevara, “A Taxonomy of Quality Metrics for Cloud Services,” *IEEE Access*, vol. 8, pp. 131461–131498, 2020, doi: 10.1109/ACCESS.2020.3009079.

[37] P. Costa, J. P. Santos, and M. M. Da Silva, “Evaluation Criteria for Cloud Services,” in *2013 IEEE Sixth International Conference on Cloud Computing*, Santa Clara, CA: IEEE, Jun. 2013, pp. 598–605. doi: 10.1109/CLOUD.2013.70.

[38] F. De Vaulx, E. Simmon, and R. Bohn, “Cloud computing service metrics description,” National Institute of Standards and Technology, Gaithersburg, MD, NIST SP 500-307, Apr. 2017. doi: 10.6028/NIST.SP.500-307.

[39] Atlassian, “ITIL in Modern ITSM: A Complete Guide,” Atlassian. Accessed: Jan. 05, 2024. [Online]. Available: https://www.atlassian.com/itsm/itil

[40] D. Wright and D. Smith, “Gartner Reprint.” Accessed: Dec. 20, 2023. [Online]. Available: https://www.gartner.com/doc/reprints?id=1-2ES4ML14&ct=230823&st=sb&trk=44f67619-4f3b-42e8-93b9-32ad8a123845&sc_channel=el

[41] D. S. Kim, J. B. Hong, T. A. Nguyen, F. Machida, J. S. Park, and K. S. Trivedi, “Availability Modeling and Analysis of a Virtualized System Using Stochastic Reward Nets,” *2016 IEEE International Conference on Computer and Information Technology (CIT)*, pp. 210–218, Dec. 2016, doi: 10.1109/CIT.2016.97.

[42] D. Ford *et al.*, “Availability in Globally Distributed Storage Systems,” presented at the USENIX Symposium on Operating Systems Design and Implementation, Oct. 2010. Accessed: Jul. 22, 2024. [Online]. Available: https://www.semanticscholar.org/paper/Availability-in-Globally-Distributed-Storage-Ford-Labelle/a090e4f09ab702c21b3e71eeda265550f59f997a

[43] M. Nabi, M. Toeroe, and F. Khendek, “Availability in the cloud: State of the art,” *Journal of Network and Computer Applications*, vol. 60, pp. 54–67, Jan. 2016, doi: 10.1016/j.jnca.2015.11.014.

[44] R. D. S. Matos, P. R. M. Maciel, F. Machida, Dong Seong Kim, and K. S. Trivedi, “Sensitivity Analysis of Server Virtualized System Availability,” *IEEE Trans. Rel.*, vol. 61, no. 4, pp. 994–1006, Dec. 2012, doi: 10.1109/TR.2012.2220711.

[45] A. Whitepaper, “Availability and Beyond: Understanding and Improving the Resilience of Distributed Systems on AWS - AWS Whitepaper”.

[46] A. Lenk, M. Menzel, J. Lipsky, S. Tai, and P. Offermann, “What Are You Paying For? Performance Benchmarking for Infrastructure-as-a-Service Offerings,” in *2011 IEEE 4th International Conference on Cloud Computing*, Washington, DC, USA: IEEE, Jul. 2011, pp. 484–491. doi: 10.1109/CLOUD.2011.80.

[47] S. P. Ahuja and N. Deval, “On the Performance Evaluation of IaaS Cloud Services with System-Level Benchmarks:,” *International Journal of Cloud Applications and Computing*, vol. 8, no. 1, pp. 80–96, Jan. 2018, doi: 10.4018/IJCAC.2018010104.

[48] M. Sajjad, A. Ali, and A. Salman, “Performance Evaluation of Cloud Computing Resources,” *ijacsa*, vol. 9, no. 8, 2018, doi: 10.14569/IJACSA.2018.090824.

[49] A. Stephen, S. Benedict, and R. P. A. Kumar, “Monitoring IaaS using various cloud monitors,” *Cluster Comput*, vol. 22, no. S5, pp. 12459–12471, Sep. 2019, doi: 10.1007/s10586-017-1657-y.

[50] R. Ramanathan, “AHP Approach - an overview | ScienceDirect Topics.” Accessed: Dec. 05, 2023. [Online]. Available: https://www-sciencedirect-com.ezproxy2.utwente.nl/topics/economics-econometrics-and-finance/ahp-approach

[51] J. Wątróbski, P. Ziemba, J. Jankowski, and M. Zioło, “Green Energy for a Green City—A Multi-Perspective Model Approach,” *Sustainability*, vol. 8, no. 8, p. 702, Jul. 2016, doi: 10.3390/su8080702.

[52] T. L. Saaty and M. Sodenkamp, “The Analytic Hierarchy and Analytic Network Measurement Processes: The Measurement of Intangibles”.

[53] J. A. Alonso and M. T. Lamata, “CONSISTENCY IN THE ANALYTIC HIERARCHY PROCESS: A NEW APPROACH,” *Int. J. Unc. Fuzz. Knowl. Based Syst.*, vol. 14, no. 04, pp. 445–459, Aug. 2006, doi: 10.1142/S0218488506004114.

[54] I. Tomashevskii, “Eigenvector ranking method as a measuring tool: Formulas for errors,” *European Journal of Operational Research*, vol. 240, pp. 774–780, Feb. 2015, doi: 10.1016/j.ejor.2014.07.050.

[55] K. D. Goepel, “Implementing the Analytic Hierarchy Process as a Standard Method for Multi-Criteria Decision Making in Corporate Enterprises – a New AHP Excel Template with Multiple Inputs,” presented at the International Symposium on the Analytic Hierarchy Process, Jun. 2013. doi: 10.13033/isahp.y2013.047.

[56] VPSBenchmarks, “Compare Web Runs Performance,” VPSBenchmarks. Accessed: Sep. 09, 2024. [Online]. Available: https://www.vpsbenchmarks.com/compare/performances/web_runs?selected_plan_ids=87%2C417%2C994%2C1059%2C1380%2C1491

[57] “TechRadar Pro | Expert buying advice and guides for small business | TechRadar Pro,” TechRadar. Accessed: Sep. 16, 2024. [Online]. Available: https://www.techradar.com/pro

[58] “Home - Geekbench.” Accessed: Sep. 16, 2024. [Online]. Available: https://browser.geekbench.com/

Appendix:

Table 26:Use Case Description

Use Case	Definition	Rationale
Express a Description for an Existing Metric (UC1)	*To express existing metric description in plain English using the base concepts – i.e., metric, expression, rule, parameter – (e.g., starting from an existing cloud SLA metric).*	often the metric(s) information is scattered over a document text (measurement rules, exceptions, underlying quantities, and metrics, etc.) and is mixed with related info that is not part of the metric definition per say (performance objectives, remediation measures and penalties, etc.). Distinguishing metric definition in a specific structure has proven to be of excellent value to understand the metric in use.
Create a Description for New Metric (UC2)	*To develop metric descriptions from scratch (e.g., elasticity metric)* *Figure 8 CSM Model Scenario Examples*	Engineers as well as SLA writers and auditors need some framework to describe and design metrics. A metric model or structure helps define a sharable representation and detect missing components. The metric can be defined using existing components. This is important to ensure new metrics are created using a structure that can match existing metrics, so they can be used or exchanged harmoniously. This use case provides the template and process for creating a metric
Formalize a Metric Description (UC3)	*To formalize metric expressions and rule statements from plain English to formal languages thus creating a path to metric description maturation (e.g., go from CSLA metric English expression language to ISO80000 expression language).*	it is convenient to use the same overall metric structure, when translating a plain text description of a metric into a more formal representation closer to its execution. These are two steps (plain text, formal) in the design process of a metric. Converting a metrics expression from plain language to formal language is necessary to dissect a metric into its variable components to relate the metric to the CSM so that similar metrics can be minimized to reduce duplication.
Generalize a Metric Description (UC4)	*To generalize metric description and comparison based on a category (e.g., generic availability metric) to develop a blueprint.*	Defining the foundational elements of the metric gives users a reusable starting point for the creation of new metrics and helps identify when metrics are uniquely different or just variations on the same general blueprint. In many cases it is desirable to share the same metric foundation if not the same metric. For example, there are many variants of a "service availability percentage" metric across providers. But can we say these providers share a similar general meaning of what is measured (availability percentage)? And how can we identify this common base? Extracting a metric blueprint from a set of metrics is a way to identify this base. A "service availability metric blueprint" captures what is common to several flavors of "service availability" while making the differences easier to spot across providers (often important "details," e.g., exception rules, etc.). This blueprint in turn makes it easier for a Cloud user to compare these metric variants. A metric description blueprint can be seen as partially defined metric.
Reuse Metrics Elements Across Metrics (UC5)	*To define standalone metric elements like rules, parameters or expressions*	Reuse of certain standard elements comprising cloud service metrics can help to ensure consistency across metrics and ease the process of creating them. A catalog of reusable elements could include such things as standard expressions for unit conversion (time, temperature, etc.) and standard parameters for the

	that can be reused in different descriptions.	number of days in a month, which one would expect to be the same across many or all metrics.

Table 27: Saaty's CIr Values for Matrices

Size of Matrix	Random Consistency (CIr)	# of Criteria
1	0	1
2	0	2
3	0.58	3
4	0.90	4
5	1.12	5
6	1.24	6
7	1.32	7
8	1.41	8
9	1.45	9
10	1.49	10

Table 28: Uptime Percentage Scale

Uptime %	Score (1 to 9)	Description
100.00%	9	Products that achieve a perfect availability of 100% receive the highest score of 9, indicating superior reliability with no downtime.
99.9999%	7	Products with an availability of 99.9999% are highly reliable, with minimal downtime, and are awarded a score of 7. This level of availability represents a very high standard, with only a few seconds of downtime per year.
99.999%	5	Availability of 99.999% is also strong, but slightly below the top tier. Products achieving this level receive a score of 5, reflecting their high uptime but with slightly more potential for downtime compared to the top two categories.
99.99%	3	Products with 99.99% availability are still reliable but may experience occasional downtime. These products are scored three, as they offer good performance, though not at the exceptional level of higher categories.
≤99.99%	1	Products with an availability of 99.99% or lower receive the lowest score of 1, indicating that they have significant downtime and are less reliable compared to higher-scoring products.

Table 29: Recovery Feature Scale

Recovery Method	Score (1 to 9)	Description
Backup	***1***	*The least efficient recovery method, requiring manual processes and slow recovery.*
Backup & Restore	***3***	*Slightly better than Backup but still requires manual intervention.*
Auto Backup	***5***	*Automatic backups improve ease of use and reduce downtime, making it more efficient.*

Auto Backup & Restore	***7***	*A highly automated process, with better recovery speed and lower downtime.*
More (Advanced Features)	***9***	*The most advanced method with full automation, fastest recovery speed, and minimal downtime*

Table 30: Latency Performance Scale

Latency (ms)	Score (1 to 9)	Explanation/Use Case
Under twenty ms	***9***	Excellent for real-time applications like gaming, financial trading, and low-latency video streams.
20 - 40 ms	8	Very good for gaming, VoIP, and interactive video calls. Slightly higher but still low enough for most real-time uses.
40 - 60 ms	7	Good for web browsing, video streaming, and VoIP. May cause slight delays in interactive applications.
60 - 80 ms	5	Acceptable for general web use, video streaming, and moderate video conferencing. Noticeable delays in gaming.
80 - 100 ms	2	Borderline acceptable for video conferencing and VoIP. Noticeable delays, especially in interactive content.
Above one hundred **ms**	1	Poor for real-time applications, VoIP, and gaming. Streaming and general web browsing may still be tolerable.

Table 31: Storage Durability Scale

Durability %	Score (1 to 9)	Explanation/Use Case
99.999999999%	**9**	Exceptional durability, almost zero data loss risk, suitable for mission-critical storage and archives.
99.99999999%	**8**	Extremely high durability, excellent for long-term storage of sensitive or high-value data.
99.9999999%	**7**	Very high durability, typically used in enterprise-class systems where minimal data loss is essential.
99.999999%	**6**	High durability, ideal for important but not necessarily mission-critical data storage.
99.99999%	**5**	Reliable durability for general enterprise storage, sufficient for most business applications.
99.9999%	**4**	Good durability, suitable for backup storage where some minimal risk of data loss is tolerable.
99.999%	**3**	Average durability, acceptable for non-critical data or short-term storage needs.
99.99%	**2**	Lower durability, higher risk of data loss, appropriate for non-essential data storage.
99.9%	**1**	Minimal durability, significant data loss risk, suitable only for temporary or disposable data

Table 32: Data Transfer Fees Scale

Fees after 1st TB	Score (1 to 9)	
0/GB	10	Scoring scale was created based on charges for data transfer fees out after the first TB since several providers offer first TB to be fee
$0.01/GB or lower	9	
$0.02 to $0.05/GB	7	
$0.06 to $0.09/GB	5	
$0.10 to $0.12/GB	3	
Higher than $0.12/GB	1	

Table 33: Other Metrics Scales

Metric	Explanation/Use Case
CPU Performance	Score is made of the Sysbench and Geekbench CPU test number of operations per second and of the Endurance test number of iterations per hour metric. Converted Score from Third party comparison Platform. Platform Score is a score out of twenty. Tool Score = Portal Score/2
Data Transfer Rates	Score is made of the Sysbench and Fio tests measuring random and sequential storage speeds. Converted Score from Third party comparison Platform. Platform Score is a score out of twenty. Tool Score = Portal Score/2
Network Performance	Score is made of Speed test and Iperf3 network upload and download transfer speeds. Converted Score from Third party comparison Platform. Platform Score is a score out of twenty. Tool Score = Portal Score/2
SLA	Similar Scoring to Uptime Percentage
R Recovery Time	
Recovery Feature	Similar Scoring to Availability Recovery Feature
Auto Scaling	The difference between the Yes and No scores should increase as the weight increases. In other words, the more important a criterion (higher weight), the larger the gap between Yes and No. The range should scale from 1 (small difference for low weight) to 9 (large difference for high weight). **Yes** starts at a base score of 5, and the score increases linearly with the weight (up to 9 at maximum weight, 25%). **No** starts at a lower score (e.g., 1) and the difference from Yes grows with the weight.
Manual Scaling (Vertical or Horizontal)	
Data Encryption Levels	Each product is evaluated based on five features, and the user will indicate whether each feature is available by selecting **Yes or No** (or using checkboxes that return **TRUE or FALSE**). The scoring system assigns a score of 2 points for each feature that is marked Yes (**or TRUE**) and zero points for each feature marked **No (or FALSE**). The total score is calculated by summing the points for each feature. Since there are five features, the maximum possible score is ten points (if all features are available), and the minimum score is 0 points (if none of the features are available).

Global Compliance Certifications	The formula is likely used in a compliance certification scoring system where: If the certification count is "10 or more" (indicating a high level of compliance or certification), a fixed score of 10 is assigned. Otherwise, the actual certification number (which could be less than 10) is used as the score.
Cost per Resource Unit	Instead of using a fixed multiplier (like 4 or 8), we can introduce a **dynamic scaling factor** that adjusts based on the difference between the highest and lowest prices. This way, the scoring system is **adaptive**: small price ranges result in smaller score differences, while large price ranges spread the scores more evenly. The formula now adjusts the scoring based on the price range relative to the average price of the products using Dynamic Scaling. This means that: If the price range is small, the score differences will be minimized. If the price range is large, the score differences will be maximized. Multiplying by the Range-to-Average Ratio
Total Cost of Ownership (TCO)	

Figure 26: Magic Quadrant for Strategic Cloud Platform Services [24]

Table 34: Measurement Scales for Perceived Usefulness and Perceived Ease of Use

Perceived Usefulness (PU):

Q1. Using the AHP tool helps me accomplish decision-making tasks more quickly.

UNLIKEY	Extremely	Quite	Slightly	Neither	Slightly	Quite	Extremely	LIKEY

Q2. The AHP tool improves my job performance in evaluating IaaS products.

UNLIKEY	Extremely	Quite	Slightly	Neither	Slightly	Quite	Extremely	LIKEY

Q3. Using the AHP tool increases my productivity when comparing multiple cloud providers.

UNLIKEY	Extremely	Quite	Slightly	Neither	Slightly	Quite	Extremely	LIKEY

Q4. The AHP tool enhances the overall effectiveness of my decision-making process.

UNLIKEY	Extremely	Quite	Slightly	Neither	Slightly	Quite	Extremely	LIKEY

Q5. The AHP tool makes it easier for me to assess different cloud service providers.

UNLIKEY	Extremely	Quite	Slightly	Neither	Slightly	Quite	Extremely	LIKEY

Q6. I find the AHP tool useful in conducting thorough evaluations of IaaS products.

UNLIKEY	Extremely	Quite	Slightly	Neither	Slightly	Quite	Extremely	LIKEY

Perceived Ease of Use (PEOU):

Q1. Learning to use the AHP tool was easy for me.

UNLIKEY	Extremely	Quite	Slightly	Neither	Slightly	Quite	Extremely	LIKEY

Q2. I find it easy to get the AHP tool to perform the tasks I need.

UNLIKEY | | | | | | | | LIKEY

Extremely Quite Slightly Neither Slightly Quite Extremely

Q3. My interaction with the AHP tool is clear and understandable.

UNLIKEY | | | | | | | | LIKEY

Extremely Quite Slightly Neither Slightly Quite Extremely

Q4. It was easy for me to become skillful at using the AHP tool.

UNLIKEY | | | | | | | | LIKEY

Extremely Quite Slightly Neither Slightly Quite Extremely

Q5. It is easy to remember how to use the AHP tool after learning it.

UNLIKEY | | | | | | | | LIKEY

Extremely Quite Slightly Neither Slightly Quite Extremely

Q6. I find the AHP tool intuitive and user-friendly.

UNLIKEY | | | | | | | | LIKEY

Extremely Quite Slightly Neither Slightly Quite Extremely

Behavioral Intention to Use (Self-predicted Future Usage):

Q1. Assuming the AHP tool is available in my future projects, I predict I will use it regularly.

UNLIKEY | | | | | | | | LIKEY

Extremely Quite Slightly Neither Slightly Quite Extremely

Q2. I would prefer using the AHP tool over other methods for evaluating cloud services.

UNLIKEY | | | | | | | | LIKEY

Extremely Quite Slightly Neither Slightly Quite Extremely

Q3. I am likely to recommend the AHP tool to other professionals in my field.

UNLIKEY								LIKEY
	Extremely	Quite	Slightly	Neither	Slightly	Quite	Extremely	

Q4. I am satisfied with the performance of the AHP tool and intend to use it in future.

UNLIKEY								LIKEY
	Extremely	Quite	Slightly	Neither	Slightly	Quite	Extremely	

evaluations.

Open-ended Questions

Q1. What improvements would you suggest for the AHP tool?

Q2. Which features of the AHP tool do you find most helpful?

Q3. What features or aspects of the AHP tool did you find particularly easy to use? Please explain why.

Q4. Were there any parts of the tool that you found confusing or difficult to navigate? If so, please describe them and suggest how they could be improved.

Q5. Do you think additional instructions or tooltips would make the tool easier to use? If so, what specific areas or features would benefit from more guidance?

Q6. How would you compare the ease of use of this AHP tool to other similar decision-making tools you have used in the past? Please provide examples.

Q7. Were there any tasks or functions that took longer than expected to complete using the AHP tool? If so, please explain.