УДК 004.8 **doi: 10.26906/SUNZ.2024.4.079**

Anton Kartashov, Larysa Globa

National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine

TOWARDS SEAMLEES MULTI-CLOUD INTEGRATION: STRATEGIC APPROACH

Abstract. **Background.** Cloud computing has transformed the IT landscape, offering scalable and cost-efficient solutions for data storage and access. The emergence of multi-cloud environments as a strategic approach to leverage various cloud service providers' strengths has introduced new challenges and opportunities. Existing multi-cloud frameworks and approaches lack versatility in addressing key concepts such as data security, scalability, cost optimization, and resource management. **Objective.** Designing and developing an ontological model and optimization function to enhance data management practices and decision-making in multi-cloud environments. **Methods.** The research employs an ontological approach to formalize domain concepts, relationships, and properties in multi-cloud environments. Additionally, an optimization function is proposed for selecting the best public cloud provider based on specific features. The study focuses on designing distributed storage techniques, optimizing data access latency, and developing security frameworks for multi-cloud settings. **Results.** The proposed ontological model successfully formalizes the domain's concepts, relationships, and properties in multi-cloud environments. The optimization function demonstrates effectiveness in selecting the most suitable public cloud provider based on the proposed features, enhancing data management practices automation and decision-making processes. **Conclusions.** This work addresses the critical challenge of improving data management and decision-making in multi-cloud environments through the development of an ontological model and optimization function. The research contributes to enhancing data security, scalability, cost optimization, and resource management in multi-cloud settings. Future work should focus on further refining the ontological model and optimization function, as well as exploring their application in various industry sectors.

Keywords: Cloud computing, multi-cloud environments, data storage, data access, ontological model, optimization function, data security, scalability, cost optimization, resource management.

Introduction

In recent times, cloud computing has transformed the landscape of the IT industry, providing scalable and cost-effective solutions for data storage and access. Notably, the emergence of multi-cloud environments has become a strategic approach to leverage the diverse strengths of various cloud service providers, thereby reducing the risks of vendor lock-in and optimizing overall system performance [1].

However, the process of migrating applications and/or data to the cloud has proven to be a complex endeavour. Numerous hurdles exist that impede the realization of cloud computing's full potential [2]. These difficulties often stem from the fact that existing applications have specific requirements and attributes that need fulfilment by cloud providers [3].

These challenges present an opening to establish a systematic approach for the assessment and quantification of cloud providers. Each provider extends analogous services with varying cost structures, performance metrics, security layers, and an array of features [4]. The large number of diverse service packages makes objective evaluation of different cloud providers' quality, performance, security, privacy, and dependability a daunting task [5].

This scientific paper delves into the domain of cloud computing, focusing specifically on data storage and access within multi-cloud setups. Our research explores the significant importance of this domain, encompassing critical aspects such as data security, scalability, cost optimization, and resource management.

The primary goal of this paper is to propose a comprehensive ontological model that formally represents key concepts, relationships, and properties within the domain of cloud computing and multi-cloud environments. This ontological model serves as a foundational framework to enhance our understanding of the complex interactions and interdependencies that exist in the cloud ecosystem. Moreover, it lays the groundwork for informed decision-making and innovative developments in cloud technologies.

To achieve this objective, the paper is structured as follows: Section 1 discusses the lessons learned from the literature and also defines cloud evaluation criteria conducted from the previous research. Section 2 introduces the ontological model designed to formalize the concepts, relationships, and properties in cloud computing and multi-cloud environments. It outlines the structure and key components of the model, elucidating how it helps capture the complexity and nuances of the cloud domain. Further Section 3 describes the Optimization Function for Cloud Provider Selection**.** Here, we present an optimization function that facilitates the selection of the most suitable public cloud provider based on a set of proposed features. This function leverages the ontological model to calculate a score for each cloud provider, enabling data-driven decisions and enhanced performance in multi-cloud setups. The conclusion provides a summary of the key contributions and insights gained from the research. It reiterates the significance of the proposed ontological model and optimization function and offers concluding remarks on their potential impact on the field of cloud computing.

1. Set of criteria for multi-cloud storage

The field of cloud computing has witnessed extensive research efforts aimed at understanding the intricacies of multi-cloud environments and optimizing data storage and access strategies. Several studies have highlighted the advantages of adopting multi-cloud architectures, such as increased redundancy, improved performance, and enhanced fault tolerance [6]. Multi-cloud solutions also provide organizations with the flexibility to choose cloud services based on specific requirements, ensuring cost-effectiveness and mitigating the risk of relying solely on one cloud provider [7].

Based on our literature review, conducted in the previous paper, and current cloud computing standards for storage and access, we define a complex set of criteria for storing data in multi-clouds that includes consideration of a wide range of factors that may affect the location, management, and retrieval of data across different cloud service providers [8]. Table 1 shows a comprehensive set of criteria that should be considered in the research.

Table 1 – **Comprehensive set of Criteria**

#	***Criteria Category***	***Specific Criteria***	***Measurement Metric (possible)***
1	Data Accessibility Criteria	Latency Requirements	Milliseconds (ms)
2		Redundancy and Availability	Availability Percentage (%)
3		Data Consistency	Data Consistency Index
4		Data Encryption	Encryption Strength (e.g., AES-256)
5	Cost and Resource Utilization Criteria	Cost Efficiency	Cost per GB/month ($)
6		Resource Allocation	Resource Utilization (%)
7		Data Lifecycle Management	Percentage of Archived Data (%)
8	Data Type and Format Criteria	Data Classification	Data Classification Score
9		Data Format	Data Format Compatibility
10	Compliance and Security Criteria	Regulatory Compliance	Compliance Audit Score
11		Data Ownership	Data Ownership Policy Adherence
12		Security Protocols	Security Protocol Strength
13	Scalability and Performance	Scalability	Scalability Factor
14		Performance Metrics	Throughput (requests/second)
15	Data Migration and Interoperability Criteria	Data Portability	Data Portability Index
16		Interoperability	Interoperability Score
17	Vendor Lock-In and Vendor Criteria	Vendor Lock-In Mitigation	Lock-In Reduction Score
18		Vendor Reputation	Vendor Reputation Rating
19	Disaster Recovery and Backup	Disaster Recovery Plan	Recovery Time Objective (RTO, hours)
20		Recovery Point Objective (RPO) and RTO	Recovery Point Objective (RPO, hours)
21		Data Backup Frequency	Frequency (e.g., per day, per week)
22		Backup Storage Redundancy	Redundancy Level (e.g., dual-site)
23	Monitoring and Reporting	Monitoring Tools	Tool Effectiveness (e.g., Score)
24		Reporting	Reporting Accuracy (e.g., Percentage)
25	Sustainability	Environmental Impact	Carbon Emission Reduction (%)
26		Energy Efficiency	Energy Usage (kWh)
27		Resource Sustainability	Resource Conservation Index

2. Defining Ontological Model

To formalize the concepts and relationships in the domain of cloud computing with a focus on data storage and access in multi-cloud environments, we propose an ontological model based on the defined set of criteria. This model represents the essential components, interconnections, and properties of cloud providers, cloud services, storage systems, access control mechanisms, data encryption algorithms, and other key entities. The ontological model enables a systematic and structured representation of the domain, facilitating better understanding, knowledge sharing, and future research:

A. Ontological Model

1. CloudProvider: Represents a cloud service provider, and it is characterized by the "hasName" property.

2. CloudService: Represents a cloud service offered by a provider, and it is characterized by the "hasName" property, and the "providesService" relationship, which relates a cloud provider to the services it offers.

3. StorageSystem: Represents a storage system used by a cloud service, and it is characterized by the "hasName" property, and the "usesStorageSystem" relationship, which relates a cloud service to the storage system it uses.

4. AccessControlMechanism: Represents an access control mechanism implemented by a cloud service, and it is characterized by the "hasName" property, and the "implementsAccessControl" relationship, which relates a cloud service to the access control mechanism it implements.

5. DataEncryptionAlgorithm: Represents a data encryption algorithm used by a cloud service, and it is characterized by the "hasName" property, and the "usesEncryptionAlgorithm" relationship, which relates a cloud service to the encryption algorithm it uses.

6. ServiceLevelAgreement: Represents a service level agreement associated with a cloud service, and it is characterized by the "hasDescription" property, and the "hasSLA" relationship, which relates a cloud service to the service level agreement.

7. DataReplication: Represents a data replication mechanism performed by a storage system, and it is characterized by the "hasName" property, and the "performsReplication" relationship, which relates a storage system to the data replication mechanism it performs.

8. Application: Represents an application hosted by a cloud service, and it is characterized by the "hasName" property, and the "hostedBy" relationship, which relates an application to the cloud service hosting it.

9. DataAccessMethod: Represents a data access method used by an application, and it is characterized by the "hasName" property, and the "usesAccessMethod" relationship, which relates an application to the data access method it uses.

10. DataStorageTier: Represents a storage tier available in a storage system, and it is characterized by the "hasName"

property, and the "hasStorageTier" relationship, which relates a storage system to the available storage tiers.

11. DataConsistencyModel: Represents a data consistency model followed by a cloud service, and it is characterized by the "hasName" property, and the "followsConsistencyModel" relationship, which relates a cloud service to the data consistency model it follows.

12. DataBackupStrategy: Represents a data backup strategy used by a storage system, and it is characterized by the "hasName" property, and the "hasBackupStrategy" relationship, which relates a storage system to the data backup strategy.

13. DataTransferProtocol: Represents a data transfer protocol supported by a cloud service, and it is characterized by the "hasName" property, and the "supportsTransferProtocol" relationship, which relates a cloud service to the supported data transfer protocol.

14. DataCompressionAlgorithm: Represents a data compression algorithm used by a cloud service, and it is characterized by the "hasName" property, and the "usesCompressionAlgorithm" relationship, which relates a cloud service to the compression algorithm it uses.

15. DataIndexingMethod: Represents a data indexing method used by a cloud service, and it is characterized by the "hasName" property, and the "usesIndexingMethod" relationship, which relates a cloud service to the indexing method it uses.

16. DataVersioningMethod: Represents a data versioning method used by a cloud service, and it is characterized by the "hasName" property, and the "usesVersioningMethod" relationship, which relates a cloud service to the versioning method it uses.

17. DataLifecycleManagement: Represents a data lifecycle management mechanism used by a cloud service, and it is characterized by the "hasName" property, and the "hasLifecycleManagement" relationship, which relates a cloud service to the data lifecycle management mechanism.

18. DataMigrationStrategy: Represents a data migration strategy followed by a cloud service, and it is characterized by the "hasName" property, and the "followsMigrationStrategy" relationship, which relates a cloud service to the data migration strategy.

19. DataSharingMechanism: Represents a data sharing mechanism supported by a cloud service, and it is characterized by the "hasName" property, and the "supportsSharing Mechanism" relationship, which relates a cloud service to the supported data sharing mechanism.

The ontological model presented above represents a comprehensive domain of cloud computing, with a specific focus on data storage and access in multi-cloud environments.

Each concept is enriched with properties and relationships that allow for detailed descriptions and associations.

The axioms are logical statements that define specific relationships, constraints, and properties within the ontological model.

The logical representation of the axioms further strengthens the ontological model. These axioms provide a formal foundation for the representation and reasoning of cloud-related concepts, enabling the development of optimized algorithms and decision-making processes in selecting the best cloud provider based on the proposed features.

By combining the axioms with the ontological model, we can gain deeper insights into the domain.

B. Logical representation of the axioms:

1. Each CloudProvider provides at least one CloudService.
   - ∀xCloudProvider(x)→ ∃y CloudService(y) ∧ providesService(x, y)
2. Each CloudService uses exactly one StorageSystem.
   - ∀xCloudService(x)→∃y StorageSystem(y) ∧ usesStorageSystem(x, y)
   - ∀x,y,z(CloudService(x)∧ usesStorageSystem(x,y)∧ usesStorageSystem(x, z) ∧ y ≠ z) → False
3. Each CloudService implements at least one AccessControlMechanism.
   - ∀xCloudService(x)→∃yAccessControlMechanism(y)∧ implementsAccessControl(x,y)
4. Each CloudService uses at least one DataEncryptionAlgorithm.
   - ∀xCloudService(x)→∃yDataEncryptionAlgorithm(y) ∧ usesEncryptionAlgorithm(x, y)
5. Each CloudService has exactly one ServiceLevelAgreement.
   - ∀xCloudService(x)→∃!y ServiceLevelAgreement(y) ∧ hasSLA(x, y)
6. Each StorageSystem performs at least one DataReplication mechanism.
   - ∀xStorageSystem(x)→∃yDataReplication (y) ∧ performsReplication(x, y)
7. Each Application is hosted by exactly one CloudService.
   - ∀x Application(x) → ∃!y CloudService(y) ∧ hostedBy(x, y)
8. Each Application uses at least one DataAccessMethod.
   - ∀xApplication(x)→∃y DataAccessMethod (y) ∧ usesAccessMethod(x, y)
9. Each StorageSystem has at least one DataStorageTier.
   - ∀xStorageSystem(x)→∃y DataStorageTier(y) ∧ hasStorageTier(x, y)
10. Each CloudService follows at least one DataConsistencyModel.
   - ∀xCloudService(x)→∃y DataConsistencyModel(y)∧ followsConsistencyModel(x, y)

These axioms express logical statements that define the relationships and constraints within the ontology, specifying the required conditions and properties for the concepts involved.

In the current context of the ontological model, relationships play a vital role in defining the connections and interactions between various concepts. Each relationship is expressed through a logical statement that establishes a link between two entities in the domain.

C. Descriptions of the relationships

1. **providesService** (CloudProvider, CloudService) Indicates that a specific CloudProvider delivers a particular CloudService to users.

2. **usesStorageSystem** (CloudService, StorageSystem) Specifies that a particular CloudService relies on a specific StorageSystem for data storage and management.

3. **implementsAccessControl** (CloudService, AccessControlMechanism) Indicates that a specific CloudService employs a particular AccessControlMechanism to regulate data access and user permissions.

4. **usesEncryptionAlgorithm** (CloudService, DataEncryptionAlgorithm) Specifies that a particular

CloudService employs a specific DataEncryptionAlgorithm to protect data confidentiality.

5. **hasSLA** (CloudService, ServiceLevelAgreement) Indicates that a specific CloudService is bound by a particular ServiceLevelAgreement that defines the quality of service and performance guarantees.

6. **performsReplication** (StorageSystem, DataReplication) States that a particular StorageSystem implements a specific DataReplication mechanism to duplicate data for fault tolerance and availability.

7. **hostedBy** (Application, CloudService) Indicates that a specific Application is hosted and executed by a particular CloudService.

8. **usesAccessMethod** (Application, DataAccessMethod) Specifies that a particular Application utilizes a specific DataAccessMethod for data interactions.

9. **hasStorageTier** (StorageSystem, DataStorageTier) Specifies that a particular StorageSystem offers a specific DataStorageTier with distinct performance characteristics.

10. **followsConsistencyModel** (CloudService, DataConsistencyModel) Indicates that a particular CloudService adheres to a specific DataConsistencyModel to maintain data integrity.

In summary, the logical statements describing these relationships provide essential insights into the associations and interactions between different components in the ontology of cloud computing with a focus on data storage and access in a multi-cloud environment.

These relationships form the backbone of the ontological model, which is visually shown on the Fig. 1, enabling a comprehensive understanding of the domain and facilitating the optimization of cloud provider selection based on proposed features.

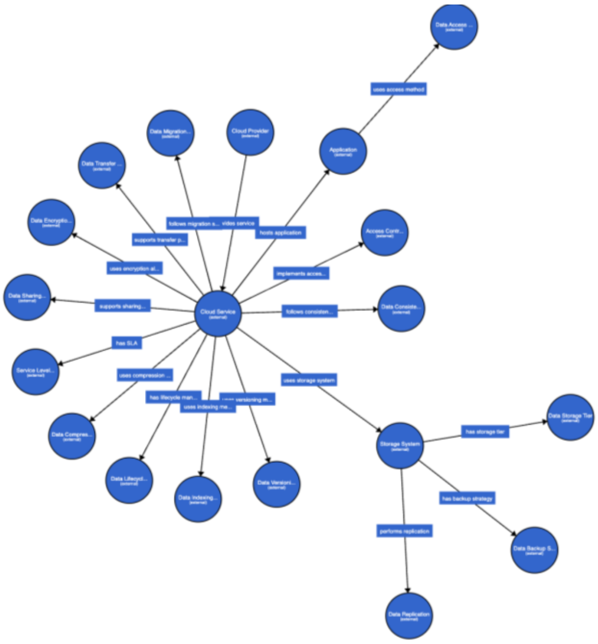

Fig.1. Visual representation of the Ontological model

3. Setting Optimisation Function

In order to make informed decisions when selecting the best public cloud provider in a multi-cloud environment, we introduce an optimisation function.

This function incorporates a weighted scoring mechanism based on relevant criteria such as data security, performance, cost-effectiveness, compliance, and scalability. By assigning appropriate weights to these criteria, the optimisation function calculates a score for each cloud provider, aiding decision-makers in choosing the most suitable option that aligns with their specific requirements and objectives.

To define the optimization function in algebraic form, we can express it as a weighted sum of the desired features:

- V = Set of cloud vendors (AWS, Azure, GCP);
- F = Set of desired features;
- W = Set of weights corresponding to each desired feature.

The optimization function can be defined as follows:

$$Score = \\ = \sum [\, feature \in F \cap vendor \,] (W[\, feature \,]), \quad (1)$$

where:

- $Score(vendor)$ represents the score of a specific cloud vendor based on the presence of desired features and their corresponding weights;
- $feature \in F \cap vendor$ denotes that the feature is both desired and offered by the vendor;
- $W[feature]$ represents the weight assigned to each desired feature.

The function calculates the score for each vendor by summing the weights of the desired features that are present in the vendor's offerings. The higher the score, the more suitable the vendor is considered for data storage based on the desired features and their assigned weights.

More complex optimization function could involve additional factors or constraints, such as cost, performance, and reliability. An example of an extended optimization function that considers cost and performance along with the presence of desired features:

- V = Set of cloud vendors (AWS, Azure, GCP);
- F = Set of desired features;
- W = Set of weights corresponding to each desired feature;
- $C(vendor)$ = Cost factor for a specific vendor;
- $P(vendor)$ = Performance factor for a specific vendor.

The optimization function can be defined as follows:

$$Score = \\ = \sum [\, feature \in F \cap vendor \,] (W[\, feature \,]) + \\ + \alpha \cdot C(vendor) + \beta \cdot P(vendor), \quad (2)$$

where α and β are coefficients that determine the relative importance of cost and performance in the optimization function.

The function calculates the score for each vendor by summing the weights of the desired features present in the vendor's offerings and adjusting it based on the cost and performance factors.

The coefficients α and β control the balance between cost and performance considerations.

For example, if a particular feature's contribution to the score is dependent on a continuous variable (e.g., time, data volume, latency), the function could involve integrals to evaluate the accumulated effect over a range of values. Similarly, derivatives can be utilized to capture the rate of change or sensitivity of certain factors.

Future research could involve expanding the ontological model to include additional criteria and to account for changes in technological infrastructure for cloud storage.

It is also crucial to continually refine the ontological model by adapting it to new requirements and emerging technologies.

Furthermore, the development of an optimization function within the ontological model could enhance its efficacy in dynamically adjusting data distribution strategies based on real-time performance metrics and cost considerations.

Conclusion

In conclusion, the research domain of cloud computing with a focus on data storage and access in multi-cloud environments represents a vital area of study with profound implications for data management practices. This scientific paper highlights the importance and scientific goals, emphasizing the need for advanced security measures, scalable storage solutions, and efficient resource management strategies.

By proposing an ontological model and an optimisation function, we aim to enhance data management practices, drive innovation in cloud technologies, and facilitate informed decision-making in multi-cloud environments. They both serve as a foundation for further research and the enhancement of strategies for multi-cloud environments.

References

1. Hong, Jiangshui & Dreibholz, Thomas & Schenkel, Joseph & Hu, Jiaxi. (2019). An Overview of Multi-cloud Computing. 10.1007/978-3-030-15035-8_103.
2. Alonso, J., Orue-Echevarria, L., Casola, V. et al. Understanding the challenges and novel architectural models of multi-cloud native applications – a systematic literature review. J Cloud Comp 12, 6 (2023). https://doi.org/10.1186/s13677-022-00367-6
3. Tomarchio, O., Calcaterra, D. & Modica, G.D. Cloud resource orchestration in the multi-cloud landscape: a systematic review of existing frameworks. J Cloud Comp 9, 49 (2020). https://doi.org/10.1186/s13677-020-00194-7
4. [18] T. G. Papaioannou, N. Bonvin, and K. Aberer, "Scalia: An adaptive scheme for efficient multi-cloud storage," in Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, ser. SC '12. Los Alamitos, CA, USA: IEEE Computer Society Press, 2012, pp. 20:1–20:10.
5. Celesti, A., Galletta, A., Fazio, M. and Villari, M., 2019. Towards hybrid multi-cloud storage systems: Understanding how to perform data transfer. Big Data Research, 16, pp.1-17.
6. Li J, Lin D, Squicciarini AC, Li J, Jia C (2017) Towards privacy preserving storage and retrieval in multiple clouds. IEEE Trans Cloud Comput 5(3):499–509. https://doi.org/10.1109/TCC.2015. 2485214
7. Tchernykh, A., Babenko, M., Miranda-López, V., Drozdov, A.Y. and Avetisyan, A., 2018, May. WA-RRNS: Reliable data storage system based on multi-cloud. In 2018 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW) (pp. 666-673). IEEE.
8. Anton Kartashov and Larysa Globa Overview of the Approaches to Managing Distributed Storage and Access to Cloud Data//Proceedings of International Conference on Applied Innovation in IT. Volume 11, Issue 2, pp. 19-29. (DOI:10.25673/112990)

Received (Надійшла) 12.09.2024
Accepted for publication (Прийнята до друку) 06.11.2024

Шлях до безшовної мультихмарної інтеграції: стратегічний підхід

А. Д. Карташов, Л. С. Глоба

Анотація. Проблематика. Хмарні обчислення трансформували ІТ-індустрію, пропонуючи масштабовані та економічно ефективні рішення для зберігання та доступу до даних. Поява мультихмарних середовищ як стратегічного підходу до використання переваг різних постачальників хмарних послуг створила нові виклики та можливості. Існуючим мультихмарним фреймворкам та підходам бракує універсальності у вирішенні ключових концепцій, таких як безпека даних, масштабованість, оптимізація витрат та управління ресурсами. **Мета досліджень**. Розробка та впровадження онтологічної моделі та функції оптимізації для вдосконалення практик управління даними та прийняття рішень у мультихмарних середовищах. **Методика реалізації**. Дослідження використовує онтологічний підхід для формалізації концепцій, зв'язків та властивостей онтологічної моделі у мультихмарних середовищах. Крім того, запропоновано функцію оптимізації для вибору найкращого публічного хмарного провайдера на основі конкретних характеристик. Дослідження зосереджується на розробці методів розподіленого зберігання даних з урахуванням запропонованого набору вимог для мультихмарних середовищ. **Результати досліджень**. Запропонована онтологічна модель успішно формалізує концепції, зв'язки та властивості онтологічної моделі в мультихмарних середовищах. Функція оптимізації демонструє ефективність у виборі найбільш підходящого публічного хмарного провайдера на основі запропонованих характеристик, покращуючи автоматизацію практик управління даними та процесів прийняття рішень. **Висновки**. Ця робота пропонує оптимізоване рішення критичної проблеми вдосконалення управління даними та прийняття рішень у мультихмарних середовищах шляхом розробки онтологічної моделі та функції оптимізації. Дослідження сприяє підвищенню безпеки даних, масштабованості, оптимізації витрат та управління ресурсами в мультихмарних середовищах. Майбутні дослідження повинні зосередитися на подальшому вдосконаленні онтологічної моделі та функції оптимізації, а також на вивченні їх застосування в різних галузях промисловості.

Ключові слова: хмарні обчислення, мультихмарні середовища, зберігання даних, доступ до даних, онтологічна модель, функція оптимізації, безпека даних, масштабованість, оптимізація витрат, управління ресурсами.