



# A Taxonomy for Cloud Storage Cost

Akif Quddus Khan<sup>1</sup>, Nikolay Nikolov<sup>2</sup>, Mihhail Matskin<sup>3</sup>, Radu Prodan<sup>4</sup>,  
Christoph Bussler<sup>5</sup>, Dumitru Roman<sup>2</sup>, and Ahmet Soylu<sup>6</sup>(✉)

<sup>1</sup> Norwegian University of Science and Technology – NTNU, Gjøvik, Norway

`akif.q.khan@ntnu.no`

<sup>2</sup> SINTEF AS, Oslo, Norway

<sup>3</sup> KTH Royal Institute of Technology, Stockholm, Sweden

<sup>4</sup> University of Klagenfurt, Klagefurt, Austria

<sup>5</sup> Robert Bosch LLC, Ontario, CA, USA

<sup>6</sup> OsloMet – Oslo Metropolitan University, Oslo, Norway

`ahmet.soylu@oslomet.no`

**Abstract.** The cost of using cloud storage services is complex and often an unclear structure, while it is one of the important factors for organisations adopting cloud storage. Furthermore, organisations take advantage of multi-cloud or hybrid solutions to combine multiple public and/or private cloud service providers to avoid vendor lock-in, achieve high availability and performance, optimise cost, etc. This complicated ecosystem makes it even harder to understand and manage cost. Therefore, in this paper, we provide a taxonomy of cloud storage cost in order to provide a better understanding and insights on this complex problem domain.

**Keywords:** Cloud · Storage cost · StaaS · Taxonomy

## 1 Introduction

Cost is one of the important factors for organisations while adopting cloud storage; however, cloud storage providers offer complex pricing policies, including the actual storage cost and the cost related to additional services (e.g., network usage cost) [19]. Given the increasing use of storage as a service (StaaS) and its rapidly growing economic value [18], cost optimisation for StaaS has become a challenging endeavour for industry and also for research. Furthermore, while it is rare, deploying an application in a multi-cloud environment, which involves utilising multiple public cloud service providers (CSPs), can add further complexity to the cost structure. The goal is to minimise cost of data storage under complex and diverse pricing policies coupled with varying storage and network resources and services offered by CSPs [23]. Organisations take advantage of multi-cloud or hybrid solutions [40] to combine multiple public and/or private cloud storage providers to avoid vendor lock-in, to achieve high availability and performance, optimising cost, etc. [37]. An application deployed using multiple public and/or private cloud providers distributed over several regions can enhance the application's performance while reducing the cost. Nevertheless, the cost of using

cloud storage services is complex and often unclear structure, particularly in a multi-cloud or hybrid ecosystem.

The cloud storage providers tout ostensibly simple use-based pricing plans when it comes to pricing; however, a practical cost analysis of cloud storage is not straightforward [12], and there are a limited number of studies that focus on cost optimisation across multiple CSPs with varying price policies [16]. Comprehensive models and mechanisms are required to optimise the cost of using cloud storage services and storage service selection for data placement, for which it is essential to understand this complex cost structure. In this context, we collected and analysed data from the documentation of three major cloud service providers to find commonalities and differences, to provide a comprehensive taxonomy of cloud storage cost, and to provide a systematic and comprehensive framework for analysing and comparing the cost of different cloud storage solutions. It fills this gap by providing a structured approach, which can be used to develop a software tool for cost optimisation. It also provides a basis for more meaningful cost comparisons between cloud storage providers, which can help organisations to make more informed decisions about their cloud storage strategy. We aim that the work presented in this paper will provide researchers and practitioners working on cost optimisation, cost modelling, cloud provider selection in a multi-cloud or hybrid setting, etc., with a better understanding and insights regarding this complex problem domain.

The rest of the paper is structured as follows. Section 2 provides an overview of the key concepts, while Sect. 3 presents a taxonomy along with the related work. Finally, Sect. 4 concludes the paper and presents the future work.

## 2 Overview

Data intensive applications processing large amounts of data are ideal candidates for cloud deployment due to the need for higher storage and computing resources [26]. A single cloud storage provider with multiple regions or, as discussed earlier, due to concerns about cost, scalability, availability, performance, vendor lock-in, etc., a (geo-)distributed approach through a multi-cloud or hybrid solution could be opted for. In this paper, we will focus on a few of the major cloud service providers worldwide, such as Amazon Web Services (AWS), Microsoft Azure (Azure), and Google Cloud, among others like Alibaba Cloud and IBM Cloud. However, there is no guarantee that one of these multinational CSPs alone is optimal for an organisation's needs.

In cloud storage, data is stored in the form of objects (files, blobs, entities, items, records), which are pieces of data that form a dataset (collection, set, grouping, repository). Every object in cloud storage resides in a bucket. The term “bucket” is used by AWS and Google Cloud, whereas Azure refers to it as a “container”. Data could be stored and accessed in various structures, abstractions, and formats [26, 27]; users can choose the location where the storage bucket will be placed. Data could be distributed over multiple data stores to exploit the advantages of a multi-cloud and multi-region environment. It also plays an essential role in data compliance issues, where data must be stored in

particular geographical locations, e.g., GDPR [38], but it can also increase the cost. Yet, realising distributed data intensive applications on the cloud is not straightforward. Sharding and data replication [7] are the key concepts for data distribution. Sharding refers to splitting and distributing data across different nodes in a system, where each node holds a single copy of the data; it provides scalability in terms of load balancing and storage capacity and high availability. Data replication refers to continuous synchronisation of data or parts of it by copying it to multiple nodes in a system; it provides high availability and durability. However, data replication increases the cost and introduces the issue of data consistency due to synchronisation issues between the nodes under network partitioning; therefore, a trade-off between availability and consistency emerges [36]. CSPs offer storage services from datacenters located all around the world; therefore, communication and coordination among nodes could be hindered due to network issues in both cases, causing increased latency [8]. Data replication and sharding with an adequate data distribution strategy could also provide data locality and hence low latency by placing data closer to the computation early-on rather than moving it as needed later [3].

The location of a cloud storage server is characterised by continent, region, and availability zones (it is termed as zone by Google Cloud, replication zones by Azure, and availability zone by AWS). A continent is a geographical region such as North America, South America, the Middle East, etc. Each continent can have one or more regions, and each region features availability zones deployed within a latency-defined perimeter. They are connected through a dedicated regional low-latency network. Availability zones are physically separate locations within each region that are tolerant of local failures. A high-performance network connects the availability zones with extremely low round-trip latency. Each region often has three or more availability zones. Availability zones are designed so that if one zone is affected, regional services, capacity, and high availability are supported by the remaining two zones. Network infrastructure constitutes a major and integral part of the cloud continuum. Users are charged for using network services, reading and writing data to and from cloud storage (for most CSPs, data transfer-in is free). These are linked with data egress and ingress, while the former refers to data leaving the data container, and the latter refers to data entering a data container. Reading information or metadata from a cloud storage bucket is an example of egress. Uploading files to a cloud storage bucket or streaming data into a cloud-based data processing service are examples of data ingress in the cloud. Especially when data is distributed over multiple geographical areas over a distributed infrastructure managed by multiple third parties and transferred over the network, security and privacy concerns also need to be addressed. This is particularly challenging in complex multi-cloud and hybrid settings, as approaches that work seamlessly over multiple providers are required, apart from the additional cost introduced. In multi-cloud and hybrid settings, therefore, several challenges need to be addressed [8], such as multi-cloud management, security, workload and workflow management and cost optimisation under different contexts and parameters.

Cloud storage services offer a simple pay-as-you-go pricing plan; however, they do offer various pricing models as well [43]. In the block-rate pricing model, data ranges are defined, and each range has a different per GB price for storing data. Some CSPs, such as Azure, also offer a reserved pricing plan that helps lower the data storage cost by committing to and reserving storage for one year or three years. In addition to all these, with almost all the CSPs, there is an opportunity to directly contact the sales team and get a custom offer according to the requirements. A cloud service provider offers several different services with more or less the same functionality, but they cost differently because there's a difference in performance. For example, Amazon S3 and Reduced Redundancy Storage (RRS) are online storage services, but the latter compromises redundancy for lower cost [26]. An even more relevant example of this scenario is the model of storage tiers or classes that are offered not only by AWS but also by Google Cloud and Azure, i.e., the division or categorisation of storage services within AWS S3. Another strategy that the CSPs use is the bundling of services. It is not a strategy adopted recently, and not just by CSPs; it is being used intensively by a wide variety of other economic sectors as well [6]. Although the ultimate purpose of bundling is cost-effectiveness and increased customer satisfaction [39], it is also a strategy that can discourage new competitors from entering a market [31]. Following this strategy, CSPs bundle storage services with other related services. For example, network services have lower costs if data transfer between storage and other services is within the cloud environment, which means computing resources must also be from the same CSP.

### 3 A Taxonomy of Storage Cost

Cloud computing cost can be broken down into three groups: 1) storage cost concerns the amount of data stored in the cloud and its duration; 2) data transport cost concerns the amount of data moved over the cloud network; and 3) compute cost concerns the use of computing resources from the cloud continuum (e.g., VMs rented and duration). In this paper, we focus on storage cost and data transfer cost. Figure 1 shows the proposed cloud cost structure taxonomy. Storage costs comprise data storage, data replication, transaction, and network usage costs, whereas data transfer costs comprise data replication, transaction and network usage costs. In addition to that, storage cost also incorporates optional data security cost. We discuss the elements of the cloud storage cost structure based on how CSPs charge their users, including data storage, data replication, transaction, network usage, and data encryption costs. Storage and data transfer costs vary by storage tier, as discussed below. The taxonomy presented in this section is extracted by analysing the official pricing information provided publicly by AWS<sup>1</sup>, Google Cloud<sup>2</sup>, and Azure<sup>3</sup> in November 2022.

<sup>1</sup> <https://aws.amazon.com/s3/pricing/>.

<sup>2</sup> <https://cloud.google.com/storage/pricing>.

<sup>3</sup> <https://azure.microsoft.com/en-us/pricing/details/storage/blobs/>.

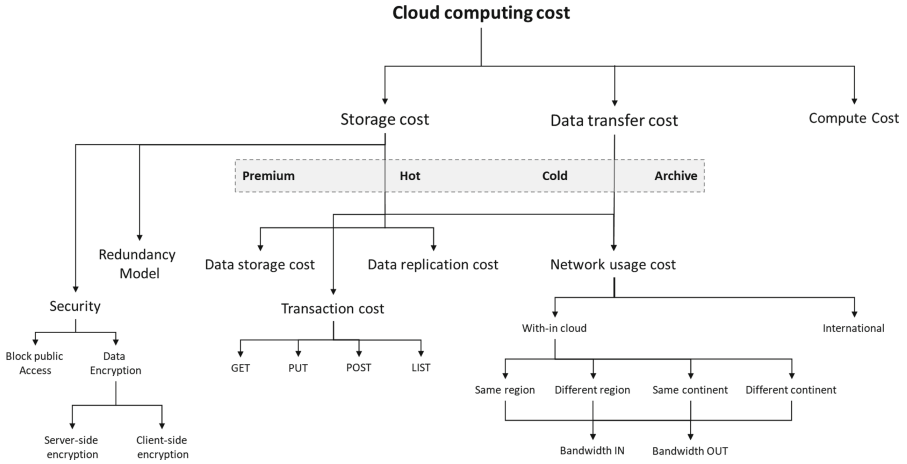


Fig. 1. Cloud computing cost taxonomy.

### 3.1 Storage Tiers

Every chunk of data stored in cloud storage uses a piece of information known as a “storage tier”, specifying its availability level and pricing structure. Storage servers on tiers don’t need to be connected to a virtual machine to store and read data. For example, AWS Elastic Block Storage can only be used with AWS EC2 instances; however, data stored on AWS S3 (tiered storage) can be accessed using a standard data transfer protocol. We collect storage tiers under four categories: premium, hot, cold, and archive. A summarised comparison of tiers offered by three different providers is shown in Table 1, whereas definitions and key characteristics of each storage tier are explained in the followings.

*Premium tier* is better suited for data that is frequently accessed and/or is only stored for short periods of time. This tier is called “Premium” in Azure, “Standard” in Google cloud, and “S3 Standard” in AWS. The premium tier costs more than the other tiers to store data but less to access the data. *Hot tier* is suggested for storing the data that is frequently accessed and modified. In Azure it is known as “Hot”, in Google cloud as “Nearline”, and in AWS as “S3 Standard – Infrequent Access”. This tier also has a higher cost as compared to the cold and archive storage tiers, but the associated network usage costs are comparatively lower. *Cold tier* is designed for storing data that is accessed and modified occasionally. For this tier, all cloud storage providers recommend that data must be stored for a specific minimum amount of time. Storage costs are less than premium and hot tiers, but network usage costs are higher. This tier is referred to as “Cool” in Azure, “Coldline” in Google cloud, and “S3 Glacier - Instant Retrieval” in AWS. *Archive tier* is designed for storing data that is rarely accessed and is stored for a longer period of time – basically an offline tier. Mostly, the data that is stored cannot be accessed immediately, but it varies from CSP to CSP. That is why it is recommended to store data that has

**Table 1.** Storage tiers comparison.

Tier	Azure	AWS	Google Cloud	Characteristics
Premium	Premium	S3 Standard, S3 Intelligent Tiering	Standard	Frequent data access; Data stored briefly; Expensive storage; Less network usage cost; No minimum storage time
Hot	Hot	S3 Standard Infrequent, S3 One Zone-IA	Nearline	Infrequent data access; Lower cost than Premium tier; Storage duration requirement may vary
Cold	Cool	S3 Glacier Instant Retrieval	Coldline	Less frequent data access; Cheaper than Premium and Hot tier storage; Higher network usage cost; Storage duration requirement may vary
Archive	Archive	S3 Glacier (Flexible or Deep Archive)	Archive	Rare data access; Long-term storage; Flexible latency required; Minimum storage time

flexibility in terms of latency requirements, i.e., on the order of hours. Unlike the cold tier, the minimum storage time is not just recommended but required; e.g., for Azure, it is 180 days. Azure and Google cloud term this tier as “Archive”, whereas AWS term the similar tiers as “S3 Glacier Flexible” and “S3 Glacier Deep Archive”.

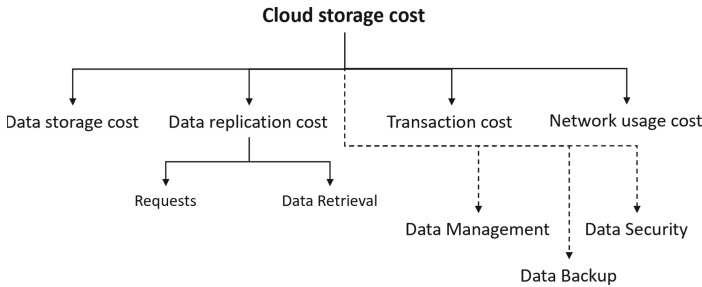
Storing a data object in only one tier at all times can be costly and inefficient. Mansouri and Erradi [24] present an example where storing 30 GB of data (with a large number of objects) and having 10K reads and 10K writes incur 1 GB of data retrieval in the US-South central region of Azure blob storage. Based on the pricing in January 2018, the cost in the cool tier is 79.55% more than that in the hot tier. However, as the data size increases to 60 GB, while the number of read and write requests approaches to zero, the cost of storing the blob in the cool tier becomes 84% less than the cost in the hot tier. Other studies also provide cost optimisation by moving data between different tiers during the data life cycle [10]. Krumm and Hoffman [12] developed a tool designed specifically for cost and usage estimation for laboratories performing clinical testing. It provides a way to explore storage options from different CSPs, cost forecasts, data compression, and mechanisms for rapid transfer to the cold tier. Jin et al. [11] developed a framework for cost-effective video storage in cloud storage, while Mansouri and Erradi [5] developed a cost optimisation algorithm for data storage and migration between different storage tiers. Nguyen et al. [33] proposed a cost-optimal two-tier fog storage technique for streaming services. Liu et al. [18–20] developed multiple algorithms presented in various studies for cost optimisation using multi-tier cloud storage.

Storage tiers can be effectively used to achieve high data durability and availability. For example, Liu et al. [17] developed an algorithm (PMCR) and did extensive numerical analysis and real-world experiments on Amazon S3.

Extending the work on high availability, Wiera et al. [35] presents a unified approach named Wier to achieve high availability and data consistency. Wier makes it possible to set data management policies for a single data centre or for a group of data centres. These policies let the user choose between different options and get the best performance, reliability, durability, and consistency. In this kind of situation, it finds the best place to store user data so that one can quickly find the best way to balance different priorities, such as cost, availability, durability, and consistency. Similarly, Zhang et al. [44] presents a bidding algorithm for tiered cloud storage to achieve low latency.

### 3.2 Cost Structure

The cost structure for cloud storage can be broken down into four main groups: 1) data storage, 2) data replication, 3) transaction, and 4) network usage. Figure 2 shows the cost taxonomy. The four elements mentioned above and those shown in Fig. 2 with solid lines are mandatory cost elements that a user can optimise but cannot altogether avoid. The other three elements, which are data management, data backup, and data security, are optional. CSPs do not provide these for free, but they are not mandatory. A user might have to pay for third-party data management services as well in the context of a multi-cloud environment.



**Fig. 2.** Cloud storage cost taxonomy.

*Data storage cost* refers to storing data in the cloud. It is charged on a per-GB-per-month basis. Each storage tier has different pricing. It also depends on the amount of data that is being stored. Some CSPs offer block-rate pricing, i.e. the larger the amount of data; the lower the unit costs are [30]. When it comes to big data, data storage costs could be huge. However, data compression techniques can reduce the size of the data by efficiently compressing the data, hence reducing the storage cost. Hossain and Roy [9] developed a data compression framework for IoT sensor data in cloud storage. In their two-layered compression framework, they compressed the data up to 90% while maintaining an error rate of less than 1.5% and no bandwidth wastage. On the other hand, distributed data storage comes with its own challenges. One of the challenges of storing data in a

distributed environment is the efficient repair of a failed node, i.e., minimising the data required to recover the failed node. Coding theory has evolved to overcome these challenges. Erasure encoding is used for reliable and efficient storage of big data in a distributed environment [14]. Several erasure coding techniques developed over time; Balaji et al. [2] present an overview of such methodologies (Table 2).

**Table 2.** Storage cost terminology comparison for the three providers.

Cost element	Definition	Azure	AWS	Google Cloud
Storage cost	It refers to data storage costs and is charged per GB per month.	Data storage	Storage	Data storage
Data replication cost	It refers to the process of replicating data from on-premises storage to the cloud or from one cloud zone to another. It is charged on a per GB basis	Data replication	Replication	Operation charges
Transaction cost	Cost for requests made against storage buckets and objects. They are charged on the quantity of requests. DELETE and CANCEL requests are free. Types of requests include PUT, COPY, POST, LISTS, GET, SELECT. Data retrieval is charged on a per GB basis	Operations & data transfer	Requests & data retrieval	Operation charges and retrieval fees
Network usage cost	Cost of all bandwidth, into and out of the cloud storage server. It is also charged on a per GB basis	Data transfer	Data transfer	Network egress

*Data replication cost* refers to replicating data from on-premises storage to the cloud or from one cloud zone to another. Data storage systems adopt a 3-replicas data replication strategy by default, i.e., for each chunk of data that is uploaded, three copies are stored, to achieve high data reliability and ensure better disaster recovery (AWS S3<sup>4</sup>, Azure Blob Storage<sup>5</sup>, Google Cloud Storage<sup>6</sup>). This means that for users to store one gigabyte of data, they have to pay for the cost of three gigabytes as well as the cost of making data copies, known as “data replication”. This significantly affects the cost-effectiveness of cloud storage [15]. The cost of data replication is charged on a per-GB basis. Several data replication strategies are available to achieve various objectives. For example, Mansouri et al. [25], Liu et al. [17], and Edwin et al. [4] developed data replication strategies to achieve optimal cost. Mansouri and Javidi [22] and Nannai and Mirnalinee [32] focused on achieving low access latency by developing dynamic data replication strategies. Ali et al. [1] presented a framework

<sup>4</sup> <https://aws.amazon.com/s3/faqs/>.

<sup>5</sup> <https://docs.microsoft.com/en-us/azure/storage/blobs/storage-redundancy>.

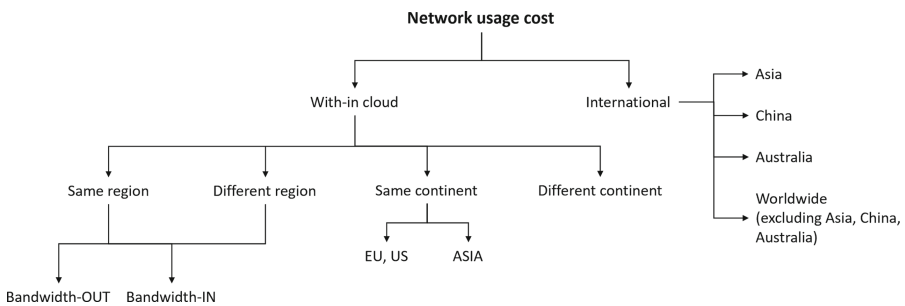
<sup>6</sup> <https://cloud.google.com/storage/docs/redundancy>.



(DROP) to pull off maximum performance and security, whereas Tos et al. [41] and Mokadem et al. [29] developed approaches to attain high performance and increase providers’ profit.

*Transaction cost* refers to the costs for managing, monitoring, and controlling a transaction when reading or writing data to cloud storage [34]. Cloud storage providers charge not only for the amount of data that is transferred over the network but also for the number of operations it takes. Both READ and WRITE operations have different costs. They are charged based on the number of requests. DELETE and CANCEL requests are free. Other requests include PUT, COPY, POST, LISTS, GET, and SELECT. On the other hand, data retrieval is charged per GB basis. Google Cloud has a different term for transaction costs, which is “operation charges”, defined as the cost of all requests made to Google cloud storage.

*Network usage cost* refers to network consumption or usage based on the quantity of data read from or sent between the buckets. Data transmitted by cloud storage through egress is reflected in the HTTP response headers. Hence, the term network usage cost is defined as the cost of bandwidth into and out of the cloud storage server. It is charged on a per-GB basis. Google Cloud has two tiers of network infrastructure: premium and standard. These differ from Azure and AWS, as they only offer a single network tier. Although network performance varies by storage tier, meaning CSPs have multiple network tiers, users cannot explicitly choose between them. For Google Cloud, the cost to use the premium network tier is more than the standard network tier, but it offers better performance. The network usage cost is a complex combination involving several factors, such as the route of data transfer, whether within the same cloud or outside. In the case of the same cloud, the cost varies depending on whether data is moved in regions within and across continents. Figure 3 shows the taxonomy for the network usage cost.



**Fig. 3.** Network usage cost taxonomy.

*Data encryption cost* is an essential element of the security costs. Cloud storage providers encrypt data using a key managed by the provider or the client, with no extra cost for the server-managed key. However, customer-managed keys

incur charges as they are stored on the provider’s infrastructure. The cost of key management is categorised into monthly billed key cost, number of operations using the key, and per hour billed HSM (hardware security module, a physical device that provides extra security for sensitive data). Key rotation is an additional cost. Though optional, data encryption affects the total cost of cloud storage. The cost of encryption and encryption/decryption keys is pretty much similar for all providers, while HSM costs vary.

### 3.3 Redundancy Model

Redundancy implies the service provider replicates valuable and important data in multiple locations. A client should ideally have several backups so that large server failures won’t impair their ability to access information [21]. Cloud storage providers offer to store data with three different redundancy options. In single-region, data is stored in a single geo-graphic location such as *eu-west*. In a dual-region mode, a user can store data in two geo-graphical locations of his choice. For example, this mode can be a suitable option if the data is frequently accessed in two different regions, such as Europe and the US. A multiple-region mode can be selected if the data is frequently accessed from different regions. The redundancy model not only improves the durability of the data, but also the availability [28]. A summary of cloud storage redundancy models for three providers is given in Table 3.

**Table 3.** Cloud storage redundancy options across the three providers.

Region	Azure	AWS	Google Cloud
Single	Locally redundant storage (LRS), Geo-redundant storage (GRS)	AWS Region	Single
Dual	Read-Access Geo Redundant Storage (RA-GRS)	Cross-region. Replication (CRR) (one way replication)	Dual-region
Multi	Zone-redundant storage (ZRS), Geo-zone-redundant storage (GZRS)	Cross-region, Replication (CRR) (two way replication)	Multi-region

Moving from single-region to dual or multiple regions can reduce access latency but comes at a cost. The higher the data redundancy, the higher the cost of data storage, both storage and replication costs. To determine which redundancy solution is ideal, it is advised to weigh the trade-offs between reduced costs and higher availability. Azure offers two types of replication in dual and multi-region replication. Using geo-replication, data is replicated to a secondary region remote from the primary region to protect against local disasters. The data in the secondary region can only be used for disaster recovery and has no read

access. Using geo-replication with read access, a secondary region also provides read access. Waibel et al. [42] formulated a system that incorporates multiple cloud services to determine redundant yet cost-efficient storage by considering factors such as storage and data transfer costs in different cloud providers. The system recommends the most cost-effective and redundant storage solution. To increase application performance, different parts of the dataset can be stored and loaded from different availability zones or regions to ensure that the application's performance is not compromised due to network throughput bottlenecks.

*Single-Region:* A single geographic area, like Sao Paulo, is referred to as a single-region. For data consumers, e.g., analytics pipelines [13], operated in the same region, a single-region is utilised to optimise latency and network capacity. Since there are no fees levied for data replication in regional locations, single-regions are a particularly advantageous alternative for short-lived data. In comparison to data kept in dual and multi-region, single-region has the lowest cost. *Dual-region:* A particular pair of areas, such as Tokyo and Osaka, is referred to as a dual-region. When single-region performance benefits are required but improved availability from geo-redundancy is also desired, a dual-region is employed. High-performance analytics workloads that can run active-active in both regions at once are very well suited for dual-regions. This indicates that users will enjoy good performance while reading and writing data in both regions to the same bucket or data container. Due to the high consistency of dual-regions, the view of the data remains constant regardless of where reads and writes are occurring. Dual-region data storage is more expensive than single-region, but less expensive than multi-region and provides better availability and low latency. *Multi-region:* A vast geographic area, like the United States, that encompasses two or more geographic locations is referred to as a multi-region. When a user has to provide content to data consumers dispersed across a wide geographic area and not connected to the cloud network, a multi-region approach is employed. Generally, the data is kept near where the majority of the users are. The multi-region model is the most expensive model of data storage; however, it also addresses a wide range of security, privacy, availability, and data durability issues.

## 4 Conclusions

In this paper, we presented a storage cost taxonomy for the cloud to guide practitioners and researchers. Our taxonomy confirms that storage cost for the cloud is a complex structure, especially in a multi-cloud setting, where a broad spectrum of differences may exist between CSPs. Furthermore, cost needs to be considered inline with other quality of service (QoS) attributes and service level agreements (SLAs), which may also affect the cost directly or indirectly (e.g., availability, consistency, etc.). Our future work will include analysis of cost in relation with other QoS attributes, trade-offs between different cost elements (e.g., computing vs. storage), as well as review of existing literature for cost optimisation. These will provide a deeper understanding of cloud storage cost and uncover the existing literature's limitations and weaknesses.

**Acknowledgments.** This research is partially funded by DataCloud project (EU H2020 101016835).

## References

1. Ali, M., Bilal, K., Khan, S.U., et al.: DROPS: division and replication of data in cloud for optimal performance and security. *IEEE Trans. Cloud Comput.* **6**(2), 303–315 (2018). <https://doi.org/10.1109/TCC.2015.2400460>
2. Balaaji, S., Krishnan, M.N., Vajha, M., Ramkumar, V., et al.: Erasure coding for distributed storage: an overview. *Sci. China Inf. Sci.* **61**(10), 1–45 (2018). <https://doi.org/10.1007/s11432-018-9482-6>
3. Barika, M., Garg, S., Zomaya, A.Y., et al.: Orchestrating big data analysis workflows in the cloud: research challenges, survey, and future directions. *ACM Comput. Surv.* **52**(5) (2019). <https://doi.org/10.1145/3332301>
4. Edwin, E.B., Umamaheswari, P., Thanka, M.R.: An efficient and improved multi-objective optimized replication management with dynamic and cost aware strategies in cloud computing data center. *Clust. Comput.* **22**(5), 11119–11128 (2019). <https://doi.org/10.1007/s10586-017-1313-6>
5. Erradi, A., Mansouri, Y.: Online cost optimization algorithms for tiered cloud storage services. *J. Syst. Softw.* **160**, 110457 (2020). <https://doi.org/10.1016/j.jss.2019.110457>
6. Georgios, C., Evangelia, F., Christos, M., Maria, N.: Exploring cost-efficient bundling in a multi-cloud environment. *Simul. Model. Pract. Theory* **111**, 102338 (2021). <https://doi.org/10.1016/j.simpat.2021.102338>
7. Gessert, F., Wingerath, W., Friedrich, S., Ritter, N.: NoSQL database systems: a survey and decision guidance. *Comput. Sci. Res. Dev.* **32**(3–4), 353–365 (2017). <https://doi.org/10.1007/s00450-016-0334-3>
8. Hong, J., Dreiholz, T., Schenkel, J.A., Hu, J.A.: An overview of multi-cloud computing. In: Barolli, L., Takizawa, M., Xhafa, F., Enokido, T. (eds.) *WAINA 2019. AISC*, vol. 927, pp. 1055–1068. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-15035-8\\_103](https://doi.org/10.1007/978-3-030-15035-8_103)
9. Hossain, K., Roy, S.: A data compression and storage optimization framework for IoT sensor data in cloud storage. In: *Proceedings of the 21st International Conference of Computer and Information Technology (ICCIT 2018)*, pp. 1–6. IEEE (2018). <https://doi.org/10.1109/ICCITECHN.2018.8631929>
10. Irie, R., Murata, S., Hsu, Y.F., Matsuoka, M.: A novel automated tiered storage architecture for achieving both cost saving and QoE. In: *Proceedings of the 8th International Symposium on Cloud and Service Computing (SC2 2018)*, pp. 32–40. IEEE (2018). <https://doi.org/10.1109/SC2.2018.00012>
11. Jin, H., Wu, C., Xie, X., Li, J., et al.: Approximate code: a cost-effective erasure coding framework for tiered video storage in cloud systems. In: *Proceedings of the 48th International Conference on Parallel Processing (ICPP 2019)*, pp. 1–10. ACM (2019). <https://doi.org/10.1145/3337821.3337869>
12. Krumm, N., Hoffman, N.: Practical estimation of cloud storage costs for clinical genomic data. *Pract. Lab. Med.* **21**, e00168 (2020). <https://doi.org/10.1016/j.plabm.2020.e00168>
13. Lee, C., Murata, S., Ishigaki, K., Date, S.: A data analytics pipeline for smart healthcare applications. In: Resch, M.M., Bez, W., Focht, E., Gienger, M., Kobayashi, H. (eds.) *Sustained Simulation Performance 2017*, pp. 181–192. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-66896-3\\_12](https://doi.org/10.1007/978-3-319-66896-3_12)

14. Li, J., Li, B.: Erasure coding for cloud storage systems: a survey. *Tsinghua Sci. Technol.* **18**(3), 259–272 (2013)
15. Li, W., Yang, Y., Yuan, D.: A novel cost-effective dynamic data replication strategy for reliability in cloud data centres. In: *Proceedings of the 9th International Conference on Dependable, Autonomic and Secure Computing*, pp. 496–502. IEEE (2011). <https://doi.org/10.1109/DASC.2011.95>
16. Liu, G., Shen, H.: Minimum-cost cloud storage service across multiple cloud providers. *IEEE/ACM Trans. Netw.* **25**(4), 2498–2513 (2017). <https://doi.org/10.1109/ICDCS.2016.36>
17. Liu, J., Shen, H., Narman, H.S.: Popularity-aware multi-failure resilient and cost-effective replication for high data durability in cloud storage. *IEEE Trans. Parallel Distrib. Syst.* **30**(10), 2355–2369 (2018). <https://doi.org/10.1109/TPDS.2018.2873384>
18. Liu, M., Pan, L., Liu, S.: To transfer or not: an online cost optimization algorithm for using two-tier storage-as-a-service clouds. *IEEE Access* **7**, 94263–94275 (2019). <https://doi.org/10.1109/ACCESS.2019.2928844>
19. Liu, M., Pan, L., Liu, S.: Keep hot or go cold: a randomized online migration algorithm for cost optimization in STaaS clouds. *IEEE Trans. Netw. Serv. Manage.* **18**(4), 4563–4575 (2021). <https://doi.org/10.1109/TNSM.2021.3096533>
20. Liu, M., Pan, L., Liu, S.: Effeclouds: a cost-effective cloud-of-clouds framework for two-tier storage. *Futur. Gener. Comput. Syst.* **129**, 33–49 (2022). <https://doi.org/10.1016/j.future.2021.11.012>
21. Expedient LLC: System Redundancy in Cloud Computing. <https://www.expedienttechnology.com/blog/cloud/system-redundancy-in-cloud-computing/>
22. Mansouri, N., Javidi, M.: A new prefetching-aware data replication to decrease access latency in cloud environment. *J. Syst. Softw.* **144**, 197–215 (2018). <https://doi.org/10.1016/j.jss.2018.05.027>
23. Mansouri, Y., Buyya, R.: To move or not to move: cost optimization in a dual cloud-based storage architecture. *J. Netw. Comput. Appl.* **75**, 223–235 (2016). <https://doi.org/10.1016/j.jnca.2016.08.029>
24. Mansouri, Y., Erradi, A.: Cost optimization algorithms for hot and cool tiers cloud storage services. In: *Proceedings of the 11th International Conference on Cloud Computing (CLOUD 2018)*, pp. 622–629. IEEE (2018). <https://doi.org/10.1109/CLOUD.2018.00086>
25. Mansouri, Y., Toosi, A.N., Buyya, R.: Cost optimization for dynamic replication and migration of data in cloud data centers. *IEEE Trans. Cloud Comput.* **7**(3), 705–718 (2017). <https://doi.org/10.1109/TCC.2017.2659728>
26. Mansouri, Y., Toosi, A.N., Buyya, R.: Data storage management in cloud environments: taxonomy, survey, and future directions. *ACM Comput. Surv.* **50**(6), 1–51 (2017). <https://doi.org/10.1145/3136623>
27. Mazumdar, S., Seybold, D., Kritikos, K., Verginadis, Y.: A survey on data storage and placement methodologies for cloud-big data ecosystem. *J. Big Data* **6**, 15 (2019). <https://doi.org/10.1186/s40537-019-0178-3>
28. Melo, R., Sobrinho, V., Feliciano, F., Maciel, P., et al.: Redundancy mechanisms applied to improve the performance in cloud computing environments. *J. Adv. Theor. Appl. Inform.* **4**(1), 45–51 (2018)
29. Mokadem, R., Hameurlain, A.: A data replication strategy with tenant performance and provider economic profit guarantees in cloud data centers. *J. Syst. Softw.* **159**, 110447 (2020). <https://doi.org/10.1016/j.jss.2019.110447>

30. Naldi, M., Mastroeni, L.: Cloud storage pricing: a comparison of current practices. In: Proceedings of the 2013 International Workshop on Hot Topics in Cloud Services (HotTopiCS 2013), pp. 27–34. ACM (2013). <https://doi.org/10.1145/2462307.2462315>
31. Nalebuff, B.: Bundling as an entry barrier. *Q. J. Econ.* **119**(1), 159–187 (2004)
32. Nannai John, S., Mirnalinee, T.: A novel dynamic data replication strategy to improve access efficiency of cloud storage. *Inf. Syst. e-Bus. Manage.* **18**(3), 405–426 (2020). <https://doi.org/10.1007/s10257-019-00422-x>
33. Nguyen, S., Salcic, Z., Zhang, X., Bisht, A.: A low-cost two-tier fog computing testbed for streaming IoT-based applications. *IEEE Internet Things J.* **8**(8), 6928–6939 (2020). <https://doi.org/10.1109/JIOT.2020.3036352>
34. Nuseibeh, H.: Adoption of cloud computing in organizations. In: AMCIS 2011 Proceedings - All Submissions, p. 372 (2011)
35. Oh, K., Qin, N., Chandra, A., Weissman, J.: Wiera: policy-driven multi-tiered geo-distributed cloud storage system. *IEEE Trans. Parallel Distrib. Syst.* **31**(2), 294–305 (2019). <https://doi.org/10.1109/TPDS.2019.2935727>
36. Priya, N., Punithavathy, E.: A review on database and transaction models in different cloud application architectures. In: Shakya, S., Du, K.L., Haoxiang, W. (eds.) Proceedings of Second International Conference on Sustainable Expert Systems. Lecture Notes in Networks and Systems, vol. 351, pp. 809–822. Springer, Singapore (2022). [https://doi.org/10.1007/978-981-16-7657-4\\_65](https://doi.org/10.1007/978-981-16-7657-4_65)
37. Ramamurthy, A., Saurabh, S., Gharote, M., Lodha, S.: Selection of cloud service providers for hosting web applications in a multi-cloud environment. In: Proceedings of the International Conference on Services Computing (SCC 2020), pp. 202–209. IEEE (2020). <https://doi.org/10.1109/SCC49832.2020.00034>
38. Shah, A., Banakar, V., Shastri, S., Wasserman, M., et al.: Analyzing the impact of GDPR on storage systems. In: Proceedings of the 11th USENIX Conference on Hot Topics in Storage and File Systems. USENIX Association (2019)
39. Simon, H., Wuebker, G.: Bundling—a powerful method to better exploit profit potential. In: Fuerderer, R., Herrmann, A., Wuebker, G. (eds.) Optimal Bundling, pp. 7–28. Springer, Heidelberg (1999). [https://doi.org/10.1007/978-3-662-09119-7\\_2](https://doi.org/10.1007/978-3-662-09119-7_2)
40. Tomarchio, O., Calcaterra, D., Modica, G.D.: Cloud resource orchestration in the multi-cloud landscape: a systematic review of existing frameworks. *J. Cloud Comput.* **9**(1), 49 (2020). <https://doi.org/10.1186/s13677-020-00194-7>
41. Tos, U., Mokadem, R., Hameurlain, A., Ayav, T., et al.: Ensuring performance and provider profit through data replication in cloud systems. *Clust. Comput.* **21**(3), 1479–1492 (2018). <https://doi.org/10.1007/s10586-017-1507-y>
42. Waibel, P., Matt, J., Hochreiner, C., et al.: Cost-optimized redundant data storage in the cloud. *Serv. Orient. Comput. Appl.* **11**(4), 411–426 (2017). <https://doi.org/10.1007/s11761-017-0218-9>
43. Wu, C., Buyya, R., Ramamohanarao, K.: Cloud pricing models: taxonomy, survey, and interdisciplinary challenges. *ACM Comput. Surv.* **52**(6) (2019). <https://doi.org/10.1145/3342103>
44. Zhang, Y., Ghosh, A., Aggarwal, V., Lan, T.: Tiered cloud storage via two-stage, latency-aware bidding. *IEEE Trans. Netw. Serv. Manage.* **16**(1), 176–191 (2018). <https://doi.org/10.1109/TNSM.2018.2875475>