



Benchmarking the Object Storage Services for Amazon and Azure

Wedad Ahmed^(✉), Hassan Hajjdiab, and Farid Ibrahim

College of Engineering and Computer Science,
Abu Dhabi University, Abu Dhabi, United Arab Emirates
1023649@students.adu.ac.ae, {hassan.hajjdiab,
farid.ibrahim}@adu.ac.ae

Abstract. Cloud computing is increasingly being used as a new computing model that provide users rapid on-demand access to computing resources with reduced cost and minimal management overhead. Data storage is one of the most famous cloud services that have attracted great attention in the research field. In this paper, we focus on the object storage of Microsoft Azure and Amazon cloud computing providers. This paper reviews object storage performance of both Microsoft Azure blob storage and Amazon simple storage service. Security and cost models for both cloud providers have been discussed as well.

Keywords: Cloud object storage · Microsoft azure blob storage
Amazon S3 storage

1 Introduction

Cloud computing is becoming the mainstream for application development. The cloud, which is a metaphor for the internet, provides high capability computing resources and storage services based on demand. Cloud support could be represented in terms of software support, platform support, and developmental tools support. Cloud computing comes in many forms: platform as a service (PaaS), where developers build and deploy their applications using the APIs provided by the cloud. Others offer infrastructure as a service (IaaS), where a customer runs applications inside virtual machines (VMs), using the APIs provided by their chosen host operating systems. SaaS uses the web to deliver applications that are managed by a third-party vendor and whose interface is accessed on the clients' side. IaaS cloud providers are responsible of providing the data center and their infrastructure software at a reduced cost and high availability. However, as a trade-off, cloud storage services do not provide strong guaranteed consistency. Due to replication, data could be inconsistent when a read immediately follows a write process. Across the web, many vendors offer data storage that resides in the cloud. There are several vendors offering cloud services in the market today such as Amazon, Google AppEngine and Microsoft Azure. Depending on access needs, one can access data stored in the cloud in three different ways: (1) using a web browser interface that enables moving files to and from the storage area; (2) through a mounted desk drive that look like a local desk drive letter or mounted file system in the

computer; (3) for application developers, the storage services could be handled using a set of application program interface (API) calls.

Cloud providers offers variety of storage services such as object storage, Block storage, File storage and VM disk storage. Object storage is designed for unstructured data, such as binary objects, user generated data, and application inputs and outputs. It can also be used to import existing data stores for analytics, backup, or archive. Object storage processes data as objects and can grow indefinitely by adding nodes and that what makes this kind of storage highly scalable and flexible. However, having very high scalability sometimes affects the performance requirements. A single object could be used to store Virtual Machine (VM) images or even an entire file system or database. Object storage requires less metadata to store and access files which reduce the overhead of managing metadata. The HTTPS/REST API is used as the interface to object storage system.

This paper will benchmark the Microsoft Azure Blob Storage (MABS) and Amazon simple storage service (Amazon S3). Amazon Simple Storage Services (S3) is a distributed data storage used to store and retrieve data as objects. Bucket is used as a container that holds unlimited number of objects. Objects are stored in different sizes from a minimum of 1 Byte to a maximum of 5 TB. Each bucket belongs to one geographical location including US, Europe and Asia. There are three storage classes in AWS: (1) Standard: used for frequently accessed data; (2) Standard - Infrequent Access (IA) used with less frequently accessed data but need fast access; (3) Amazon Glacier used for archiving and long-term backup. Microsoft Azure blob storage is used to store unstructured data in the form of objects. Blob storage is scalable and persistent. Microsoft replicates their data within the same data center or within different data center in multiple world locations for maximum availability and to ensure durability and recovery. Each object has attribute-value pairs. The Blob size in Microsoft ranges from 4 MB to 1 TB. Amazon however, uses a “bucket” to hold their objects, Microsoft use containers (such as Azure Table) to hold their blobs (objects). Azure tables facilitate query-based data management. Azure Blobs are in three types Block blob, Append blob and Page blob. Block blobs are suitable for large volume of blobs and for storing cloud objects. The Append blobs consist of blocks but they are enhanced for appending operations such as random read and write and Page blobs are best used for random accessible data. Azure implements two access tiers aspects: (1) Hot access tier which is dedicated for objects that is more frequently accessed at lower access cost. (2) Cool access tier which is dedicated for less frequently accessed objects at lower storage cost. For example, one can put Long-term back-up in cool storage and on-demand live video in the hot storage tier [2, 3, 5–10].

This paper is organized as follows: Sect. 2 provide literature review of both cloud providers. Sections 3 and 4 discuss security and the pricing model implemented in Amazon and Azur storage. Section 5 evaluates which cloud provider has a better performance. In addition, it provides a quick comparison between the object storage services for Amazon and Microsoft. Finally, Sect. 6 concludes the paper and provides future scope.

2 Literature Review

Li et al. [6] have compared the performance of blobs (object) in Amazon AWS (referred to as C1 in the figure) and Microsoft Azure (C2). They used two metrics for performance comparison: operation response time and time to consistency. Operation response time metric measures how long it takes for a storage request to finish. The response time for an operation is the time that extends from the instance the client begins the operation to the instance when the last byte reaches the client. Time to consistency metric measures the time between the instance when a datum is written to the storage service and when all reads for the datum return consistent and valid results. Such information is useful to cloud customers, because their applications may require data to be immediately available with a strong consistency guaranteed. They first write an object to a storage service. The authors then repeatedly read the object and measure how long it takes before the read returns correct result.

Figure 1 shows the response time distributions for uploading and downloading one blob measured by their Java-based clients. They consider two blob sizes, 1 KB and 10 MB, to measure both latency and throughput of the blob store. The performance for the blob services for our selected providers depends on the type of operation. Microsoft has better performance than Amazon when it comes to downloading operations but for uploading operations, amazon performs way better than Microsoft. Figure 2 illustrates the time to download a 10 MB blob measured by non-Java clients. Compared to Fig. 1c, the non-Java clients perform much better for both providers, because it turns out that the Java implementation of their API is particularly inefficient.

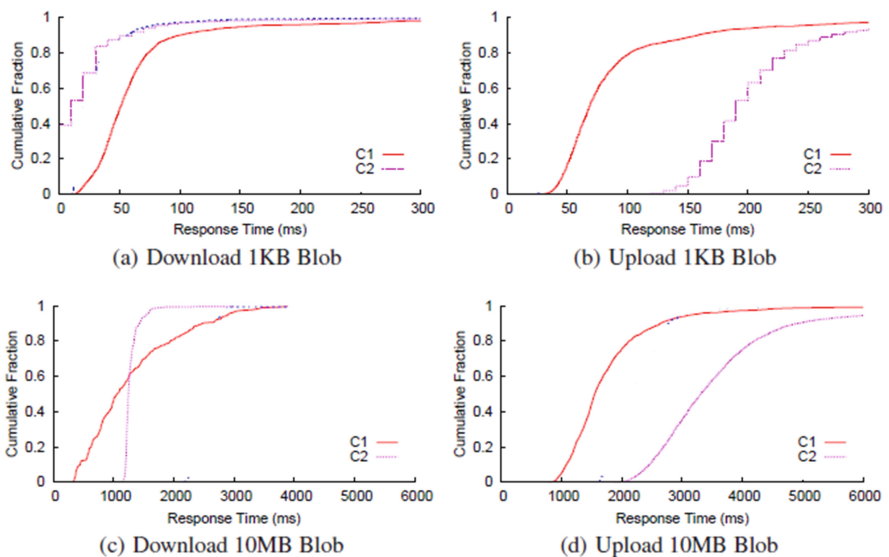


Fig. 1. The cumulative distribution of the response time to download or upload a blob using Java-based clients.

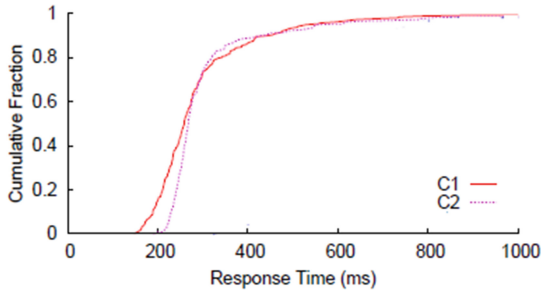
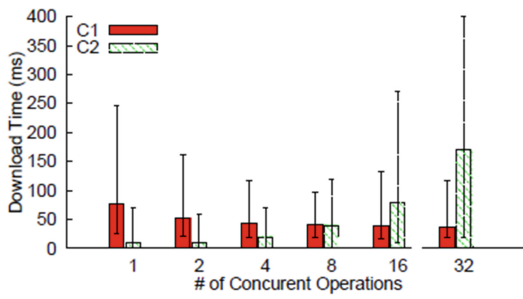
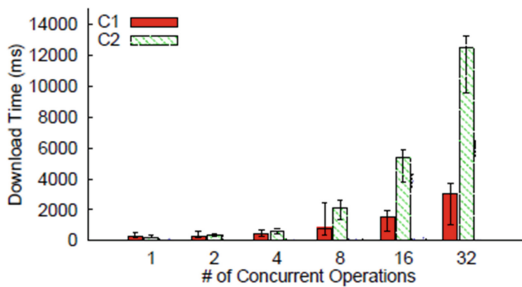


Fig. 2. The cumulative distribution of the time to download a 10 MB blob using non-Java clients.

Figure 3 compares the scalability of the blob services. They send concurrent operations. They use non- java clients since the implementation of Java is inefficient - as mentioned above. The figure only demonstrates the requests for downloading since the uploading results are similar in trend. Amazon shows good scaling performance for small size blobs, but it cannot handle large number of simultaneous operations very well when it comes to large blob size.



(a) 1KB Blob



(b) 10MB Blob

Fig. 3. The blob downloading time from each blob service under multiple concurrent operations. The number of concurrent requests ranges from 1 to 32. Note that the x-axes are on a logarithmic scale.

Wada et al. [1] compared the consistency between Amazon S3 and blob storage in Microsoft. First, they measured the consistency for AWS S3 during 11-day period of time. They updated an object in a bucket with the current timestamp as its new value. In this experiment, they measure five configurations: a write and a read run in a single thread, different threads, different processes, different VMs or different regions. S3 provides two kinds of write operations: standard and reduced redundancy. A standard write operation is for durability of at least 99.999999999%, while a reduced redundancy write durability goal is to reach at least 99.99% probability. Amazon S3 buckets offer eventual consistency for overwrite set operations. However, stale data was never found in their study regardless of write redundancy options. It seems that staleness and inconsistency might be visible to a consumer of Amazon S3 only, while carrying out the operations such that there is a failure in the nodes where the data is stored. The experiment was also conducted on Windows Azure blob storages for eight days. They did measurements for four configurations: a write and a read run in a single thread, different threads, different VMs or different regions. On Windows blob storage a write updates a blob and a reader reads it. No stale data found at all. It is known that all types of Windows Azure storages support strong data consistency and the study in [4] confirms that.

Persico et al. [9], assess Amazon S3 performance of remote data delivery using the standard HTTP GET method of cloud-to-user network. They study standard AWS S3 storage class in different four cloud regions distributed in four continents, namely, United States, Europe, Asia Pacific, South Africa. In each region they created a bucket that contains files of various sizes, from 1B to 100 MB. Bismark platform is used to simulate clients worldwide. Bismark nodes (vantage points (VPs)/source regions) are distributed in 77 locations distributed such that the United States (US, 36 VPs), Europe (EU, 16), Central-South America (CSA, 4), Asia-Pacific Region (AP, 12), and South Africa (ZA, 9). The study confirms that the size of the downloaded object heavily affects the measured performance of the network, independently from the VP. Considering the goodput average values from all the source regions, the cloud regions reported 3562, 2791, 1445, and 2018 KiB/s, respectively for objects of 100 MB size. So, United States and Europe represent the best available choices for cloud customer. Often the best performance is obtained when a bucket and VP are in the same geographic zone. However, [9] found that the US and EU cloud regions give better performance in terms of goodput (+45.5%, on average), though sometimes this choice leads to suboptimal performance.

Bocchi et al. [11] compared the performance of Microsoft azure and AWS S3 storage for a customer located in Europe. When downloading a single file of 1 MB, 10 MB, 100 MB and 1 GB size. The download rate was faster for large files; this is due to TCP protocol. Microsoft Azure blob storage performs better than Amazon in case of 1 MB and 10 MB file sizes. However, for 100 MB and 1 GB file size Amazon S3 download rate had higher throughput than Microsoft Azure. In case of downloading multiple objects of different sizes at the same time, the throughput was minimized for both cloud service providers. The same test was also repeated over time to understand whether there are other factors affecting the performance of both cloud providers such as traffic peaks during a day. They downloaded and uploaded 10 MB file making 3500 transactions in a week. The performance of uploading file was the same for both

providers but the download was faster than the upload. For both types of operations, there was no correlation found between the time of the day and the performance.

Hill et al. [12] evaluated Windows Azure blob storage performance. They measured the maximum throughput in operations/sec or MB/sec for 1 to 192 concurrent clients. They analyzed the performance of downloading a single blob of 1 GB and uploading a blob of the same size to the same container by multiple concurrent users. This test was run three times each day at different times. There was no significant performance variation across the different days. The maximum throughput found was 393.4 MB/s for 128 clients, downloading the same Blob, and 124.25 MB/sec for 192 clients uploading a blob to the same container. Obviously, as the clients increase, the performance decreases. Figure 4 demonstrates the average client bandwidth as a function of the number of concurrent users downloading the same blob. Both operations are sensitive to the number of concurrent clients accessing the object. For uploading an object into the same container, the data transfer rate is decreased to 50% with 32 clients as compared to one client. The performance of downloading an object is limited by the client's bandwidth; there was a drop of 1.5 MB per client bandwidth when the number of concurrent clients doubled. The overall blob upload speed is much slower than the download speed. For example, the average upload speed is only ~ 0.65 MB/s for 192 VMs, and ~ 1.25 MB/s for 64 VMs.

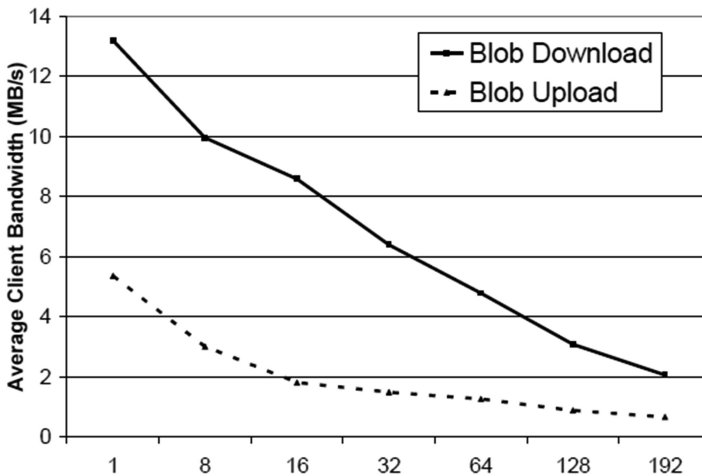


Fig. 4. Average client blob download bandwidth as a function of the number of concurrent clients.

3 Cost

Both cloud providers have no fixed costs and apply ‘pay-as-you-go’ model. There are three factors in which expenses are charged: (1) Raw storage: Amazon S3 and Windows Azure start charging \$ 0.085 per GB per month. The price increases as the stored objects increase and vice versa. (2) Requests: Microsoft and Amazon charge

users according to the number of requests they make. Amazon S3 charges \$0.005 per 1,000 PUT, COPY, LIST requests and \$0.004 every 10,000 GET requests. Windows Azure charges \$0.005 per 100,000 requests. (3) Outgoing bandwidth: uploading to the cloud is free, but download is charged by size. Both cloud providers charges 0.12 \$ per GB and declining with the amount of capacity consumed. All in all, the cost is calculated based on three factors: stored object size, the number of download requests, and traffic transferred volume. In addition, there are other aspects that determine the pricing model such as the location of the object and the storage tiers of cloud providers- Hot or Cool access tiers in Microsoft Azure or standard, infrequent access, or Glacier in Amazon [9, 11].

4 Security

Cloud computing faces new and challenging security threats. Microsoft and Amazon are public cloud providers and their infrastructure and computational resources are shared by several users or organizations across the world. Security concerns go around data Confidentiality, Integrity and Availability (CIA) [9, 10].

Amazon AWS confidentiality: Amazon uses AWS Identity and Access Management (IAM) to ensure confidentiality. To access AWS resources, a user should first be granted a permission that consists of an identity and a unique security credentials. IAM applies the least privilege principal. AWS Multi-Factor Authentication (MFA) and Key rotation are used to ensure confidentiality as well. MFA is used to boost the control over AWS resources and account settings for registered users. MFA requires a user to provide a code along with username and password. This code is sent to user by an authenticating device or a special application on a mobile phone. For Key rotation, AWS recommends that access keys and certificates be rotated regularly, so that users can alter security settings and maintain their data availability [9, 10].

Microsoft Azure confidentiality: Microsoft uses Identity and Access Management (IAM) so that only authenticated users can access their resources. A user needs to use his/her credit card credentials to subscribe to Windows Azure. Then a user can access resources using Windows Live ID and password. In addition, Windows use isolation to provide protection. This keeps data and containers logically or physically segregated. Moreover, Windows azure encrypt its internal communication using SSL encryption. It provides a user with a choice of encrypting their data during storage or transmission. In addition, Microsoft Azure gives the choice to integrate the .NET “Cryptographic Service Providers (CSPs)” by extending its SDK with .NET libraries. Thus, a user can add encryption, hashing and key management to their data [9, 10].

AWS integrity: Amazon user can download/upload data to S3 using HTTPS protocol through SSL encrypted end points. In general, there are server-side encryption and client-side encryption in Amazon S3. In the server side, amazon manages the encryption key for its users. In the client side, a user can encrypt its data before uploading them to Amazon S3 using client encryption library. For integrity, Amazon uses a Hash-based Message Authentication Code (HMAC). A user has a private key that is used when a user makes a service request. This private key is used to create

HMAC for the request. If the created HMAC with the user request matches the one stored in the server then the request is authenticated, and its integrity is maintained [9, 10].

Microsoft Azure integrity: Microsoft provides architecture to ensure integrity. It uses Cryptographic Cloud Storage Service with searchable encryption schema to ensure data integrity. This service ensures that only authorized access is permitted and enables encryption (SSE and ASE) and decryption of files [9, 10].

Availability: Microsoft Azure's have two storage tiers each with different availability; hot storage availability tier which is 99.9 and 99% is the percentage for cool storage tier. Both Amazon Standard and standard – infrequent access storage tiers have 99.9% availability. To ensure high availability, both Microsoft and Amazon provide geo-replication. Geo-replication is a type of data storage replication in which the same data is stored on servers in multiple distant physical locations [9, 10].

5 Results and Evaluation

In summary, there is no clear winner when it comes to evaluating the performance of object storage of Amazon and Microsoft. Many combinations of factors can affect the throughput of object storage such as:

- Object size
- The location of cloud data center
- The location of clients with respect to their cloud resources
- Type of operation (download, upload, etc.)
- Concurrency level of both client's and operations
- Files size
- Client's bandwidth
- Client's hardware
- Cloud provider policy

The following are some points that summarize some of the overall differences and similarities in the two selected cloud providers:

- There are different types of access classes in both Microsoft Azure Blob Storage (MABS) and AWS S3. Hot access tier in Microsoft corresponds to standard access tier in amazon. In addition, Cool tier in Microsoft correspond s to Amazon Glacier Access type.
- While Microsoft Azure stores data in a container, AWS stores data in a bucket.
- Microsoft offers storage objects in three types: block blob, page blob and append blob, while Amazon does not classify their object storage.
- For security and access management both Amazon and Microsoft allow only the account owner to access their data. However, account owners could make some objects public for sharing purposes.
- Both Microsoft and Amazon use Identity and Access Management to ensure security and encryption to ensure confidentiality. Both provide the facility of data encryption before downloading or while uploading data.

- While Amazon verifies the integrity of data stored using checksums, Microsoft provides independent architecture to eliminate integrity issues using their Cryptographic Cloud Storage Service.

6 Conclusion

The time for computing-as-a-utility is now used by many organizations around the world. Cloud services, such as object storage, are the building blocks for cloud applications. Cloud storage allows objects of different nature to be shared or archived. Cloud object storage is mostly used for its high scalability and its ability to store different types of data. It is challenging to compare cloud providers, since they all provide almost the same features for an end user point of view. However, each cloud provider has its own implementation features. In this context, a comparison of performance, cost, and security has been done for Amazon S3 and Microsoft Azure Blob storage. This paper presents several studies that tested the performance of object storage for both cloud providers. We have summarized many factors that affect the performance of storage operation from end user point of view. In addition, we presented a review on cost and security models for both cloud providers. As a future work, we are going to explore additional services provided by AWS and Microsoft Azure and we will do our own practical experiment to test and compare the performance of Microsoft Azure Blob Storage and Amazon object storage.

References

1. Wada, H., Fekete, A., Zhao, L., Le, K., Liu, A.: Data consistency properties and the tradeoffs in commercial cloud storages: the consumers' perspective. In: CIDR 2011, Fifth Biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA (2011)
2. Ruiz-Alvarez, A., Humphrey, M.: An automated approach to cloud storage service selection. In: Proceedings of the 2nd international workshop on Scientific Cloud Computing, ScienceCloud 2011, pp. 39–48 (2011)
3. Agarwal, D., Prasad, S.K.: AzureBench: benchmarking the storage services of the Azure cloud platform. In: 2012 IEEE 26th International Parallel and Distributed Processing Symposium Workshops and PhD Forum, Shanghai, pp. 1048–1057 (2012)
4. Krishnan, S.: Programming Windows Azure: Programming the Microsoft Cloud. O'Reilly (2010)
5. Jamsa, K.: Jones & Bartlett Publishers, Mar 22, 2012, cloud computing
6. Li, A., Yang, X., Kandula, S., Zhang, M.: CloudCmp: comparing public cloud providers. In: Proceeding IMC 2010 Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, pp. 1–14 (2010)
7. Samundiswary, S., Dongre, N.M.: Object storage architecture in cloud for unstructured data. In: 2017 International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, pp. 1–6 (2017)
8. Tajadod, G., Batten, L., Govinda, K.: Microsoft and Amazon: a comparison of approaches to cloud security. In: 4th IEEE International Conference on Cloud Computing Technology and Science Proceedings, Taipei, pp. 539–544 (2012)

9. Persico, V., Montieri, A., Pescapè, A.: On the network performance of Amazon S3 cloud-storage service. In: 2016 5th IEEE International Conference on Cloud Networking (Cloudnet), Pisa, pp. 113–118 (2016)
10. Calder, B., Wang, J., Ogus, A., Nilakantan, N., Skjolsvold, A., McKelvie, S., Xu, Y., Srivastav, S., Wu, J., Simitci, H., Haridas, J., Uddaraju, C., Khatri, H., Edwards, A., Bedekar, V., Mainali, S., Abbasi, R., Agarwal, A., Haq, M., Haq, M., Bhardwaj, D., Dayanand, S., Adusumilli, A., McNett, M., Sankaran, S., Manivannan, K., Rigas, L.: Windows azure storage: a highly available cloud storage service with strong consistency. In: SOSp 2011 Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles, pp. 143–157 (2011)
11. Bocchi, E., Mellia, M., Sarni, S.: Cloud storage service benchmarking: methodologies and experimentations. In: 2014 IEEE 3rd International Conference on Cloud Networking (CloudNet), Luxembourg, pp. 395–400 (2014)
12. Hill, Z., Li, J., Mao, M., Ruiz-Alvarez, A., Humphrey, M.: Early observations on the performance of Windows Azure. In: Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing, pp. 367–376 (2010)