The Ultimate Guide
To AWS Storage Optimization

Introduction

Data storage is usually considered as an ancillary service and organizations often fail to optimize data storage once they have made the shift to cloud. It is also common to eliminate unused data storage, letting the services run for weeks and months as it keeps adding to the bill. According to **RightScale**, **around 7%** of the **overall cloud expenditure goes waste over unused storage volumes** and snapshots/copies of storage volumes.

AWS has a wide and agile range of data storage options that lets the customer switch among different tiers and types according to their requirements at a given time.

This whitepaper discusses about AWS storage optimization strategies and how to choose storage services that meet your needs at the lowest cost possible.

This document shall also act as a guide to navigate towards enhanced performance, availability, and durability for a budget.

Identify your data storage requirements

Understanding your **performance profile** for each of your workloads needs a detailed analysis to measure **input and output operations per second (IOPS)**, throughput, and other variables. This is your first step in understanding your data storage requirements.

AWS lets you optimize storage services for different scenarios. A data storage option ideal for some might not work for other workloads. Hence, it is important to consider storage options separately for every workload.

The following checklist can help you segment data inside each workload and help you figure out your storage needs.

How frequently and how fast do you need to access your data?

Based on your frequency, you can choose from a range of AWS storage options of frequently accessed, less frequently accessed, and rarely accessed.

Do you need high IOPS or throughput for your data?

There is also a wide range of storage options that are optimized for performance and throughput. Understanding the variables can help you estimate your requirements better to provision the right storage slab without overpaying.

How critical or durable is your data?

Critical or regulated data tend to cost a lot more than regular data, also because they are usually stored for a longer time.

How large is your data set?

You might be able to estimate the capacity and cost better if you know the total size of your data set.

How transient is the data?

Transient data is usually short lived and does not require a lot of durability.

What is your budget?

Having a set budget for the data storage will help you make informed decisions about your options.

Amazon offers its users with **3 umbrella categories of storage services: object, block, and file storage.** Each of these categories has been formulated to meet a different storage requirement to give you flexibility in finding solutions that work best for your storage scenarios.

Object storage

This facility includes **Amazon Simple Storage Service (Amazon S3)** and is very durable, all-purpose object storage. The service works well mainly for unstructured data sets like media content. Amazon S3 gives you the strongest level of data durability and availability on AWS cloud. This also gives you **3 stages** of storage for **hot, warm, and cold data**. If you see according to pricing, the colder data will also be cheaper to store but costlier to access.
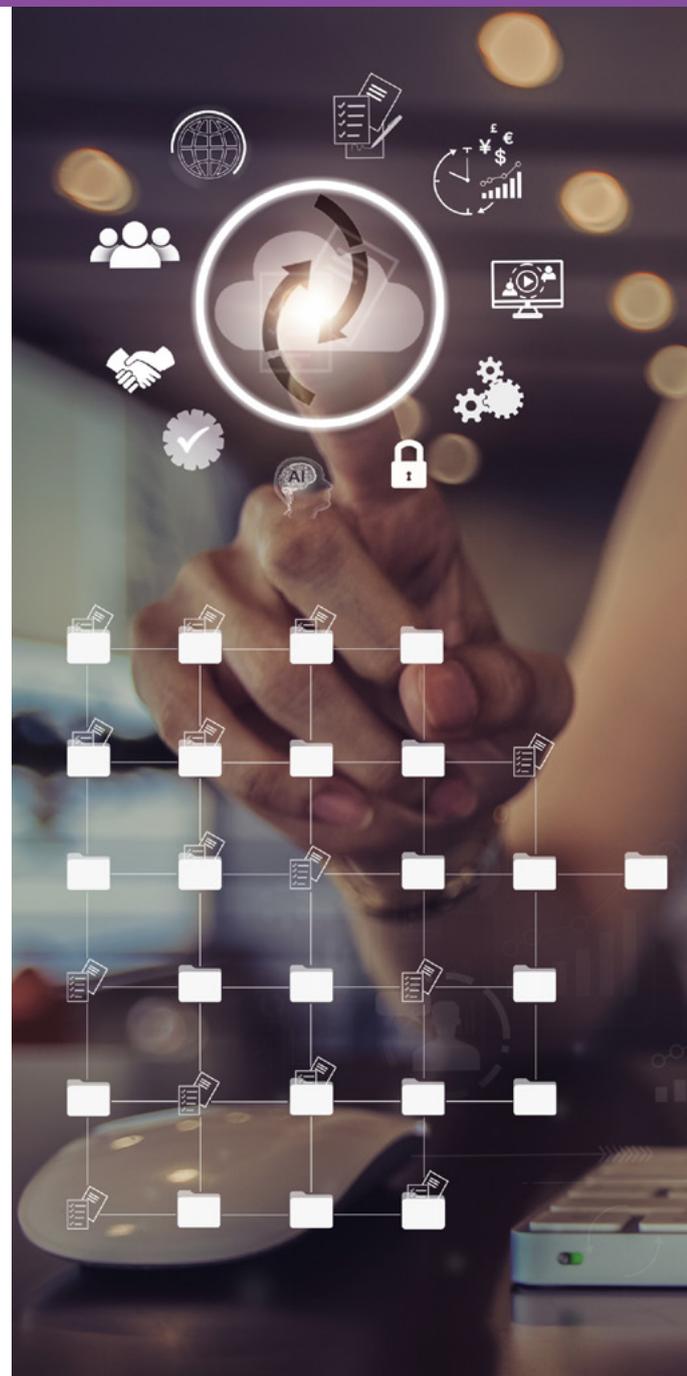
To optimize your storage costs, you can move your data easily between these options:

Amazon S3 standard- This is arguably one of the best storage options for data that you need to frequently access. S3 delivers low latency as well as throughput and is ideal for use cases like cloud applications, dynamic websites, content distribution, gaming, and data analytics.

Amazon S3 standard IA- S3 standard Infrequent Access is a storage option suitable for data that you use less often. These include items like long-term backups and disaster recovery. This offers cheap storage over time but the charges to retrieve/transfer data are high.

Amazon glacier- This service is especially designed for long term storage of rarely used data, including end-of-lifecycle, compliance, or regulatory backups. Amazon glacier offers you different data retrieval methods at different speeds and cost.
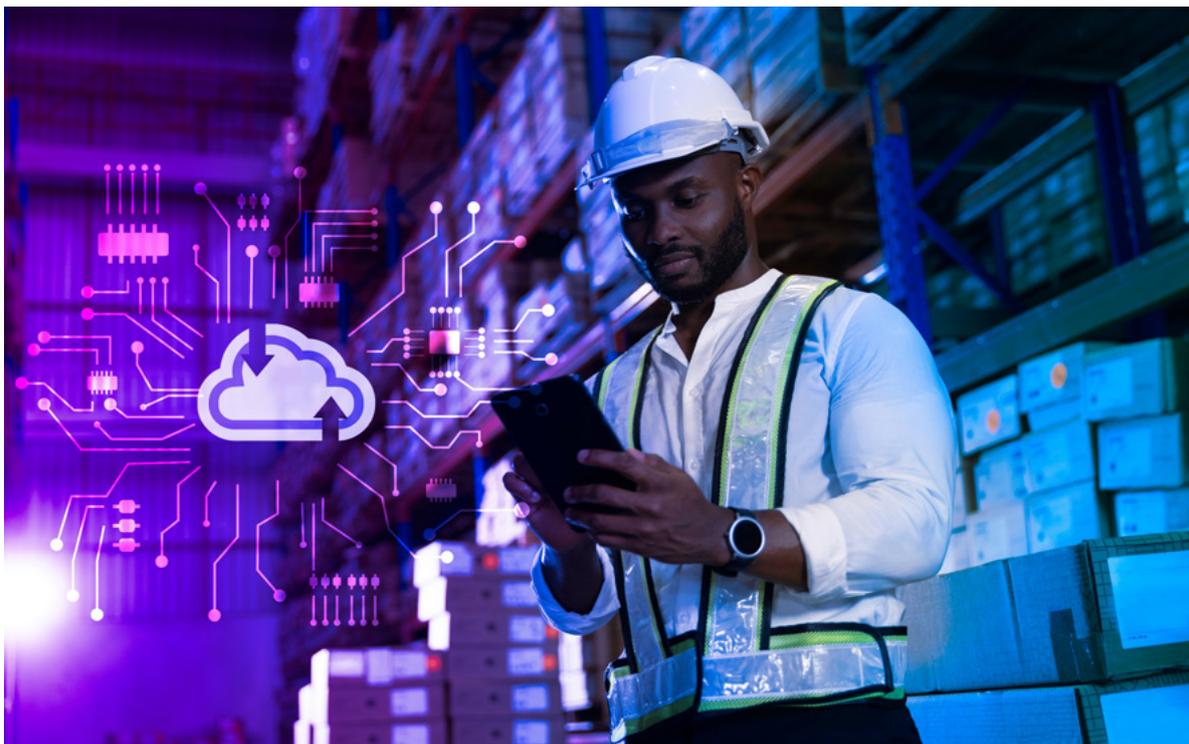
Block storage

This includes Amazon **Elastic Block Store (EBS)**, which provides a durable block storage option for EC2 instances. Amazon EBS is usually good for data that requires long term persistence and quick access at guaranteed levels of performance. The **two types** of block storage include **solid state drive (SSD)** storage **and hard disk drive (HDD)** storage.

SSD storage helps you optimize transnational workloads with performances closely tied to the IOPS. The two volume options that come under SSD include, EBS provisioned IOPS SSD (io1) and EBS general purpose SSD (gp2). The first one works best for latency-sensitive workloads requiring specific minimum-guaranteed IOPS. With io1 volume the user can pay separately for provisioned IOPS. Therefore, making them a better match for lower cost slabs unless you need high levels of provisioned IOPS. However, the later one is designed more for a general use, offering a balance between cost and performance.

HDD storage is mainly formulated for throughput-intensive workloads like data warehouses and log processing. The two types of HDD volumes comprise of Throughput optimized HDD (st1) and Cold HDD (sc1). While the st1 is best used for frequently accessed throughput-intensive workloads, sc1 is designed for less frequently accessed workloads.

File storage

Amazon elastic file system (Amazon EFS) provides the user with simple and scalable file storage options for EC2 instances. Amazon EFS supports any number of instances simultaneously with a storage capacity scaling from gigabytes to petabyte of data without provision storage. This facility is designed for workloads and applications such as big data, media processing workflows, content management, and web serving. It also supports file synchronization capabilities to efficiently synchronize files from on-premise or cloud file system at a speed at least 5 times faster than the standard copy tools.

Although S3 and Amazon EFS allocate storage based on the usage and also set the pricing accordingly, EBS volumes charge you for provisioned allocated storage whether or not you use it. You can optimize the cost without affecting the functionality by maximizing the use of Amazon S3 wherever possible.

Optimizing Amazon S3 storage

Amazon S3 allows the user to analyze data access patterns, make inventory lists, and configure lifecycle policies. The user can also set up rules that automatically shift data to cheaper S3 storage tiers if the data is less frequently accessed. This also lets you delete unused data after a set expiration date. You can also tag/categorize your S3 objects and put a filter on the tags in your data lifecycle policies.

Amazon S3 storage class analysis comes handy when you need to determine when to transition data to another storage class and to analyze storage access patterns. Infrequent data access patterns of a filtered data set over a period of time can indicate if you need to choose a more appropriate storage class, improve lifecycle policies, and make predictions about future growth and usage.

S3 inventory is yet another management tool which audits and reports about the duplication or encryption status of S3 objects weekly/monthly. This service also gives you CSV output files listing objects with their corresponding metadata. This allows you to

configure a number of inventory lists for a single bucket that is organized by different s# metadata tags. The user can also query Amazon S3 inventory with the help of standard SQL using Amazon Athena, Amazon Redshift Spectrum, etc.

For publishing storage, request, and data transfer metrics Amazon S3 can also use Amazon CloudWatch. Storage metrics which go in as a daily report are available at 1-minute intervals for granular visibility. These can be collected as well as reported as a whole for the entire bucket or for smaller subsets of objects.

AWS has also announced the launch of two new tools Amazon S3 select and Amazon Glacier select to help users further optimize the costs relevant to storage and data retrieval. Although traditionally data in object storage had to be retrieved as a whole, S3 select lets you access your data in smaller parts irrespective of the size of the object using simple SQL expressions. You can also efficiently retrieve data using Amazon glacier without having to restore archived objects. Both S3 select and Glacier select can uncover more insights from the data, regardless of the storage tier and thus, help you cut costs.

Optimizing Amazon EBS storage

You need to be mindful while using Amazon EBS as you would be paying for the provisioned capacity and performance whether the volume is unattached with low write activity or not. You will have to regularly monitor volumes to identify the ones that are unattached or underutilized/overutilized and adjust the provisioning to match the actual usage.

Delete unattached Amazon EBS Volumes

Finding and deleting unattached volumes is a fool-proof way to reduce wasted expenses. You must consider the attached EBS volumes with EC2 instances that don't expire automatically once the instance is stopped or terminated, continuing to incur charges.

Resize or change the EBS volume type

You can also identify and downsize underutilized volumes or change their volume types. Monitoring the read-write access of EBS volumes should help you determine if the throughput is low. The following pointers can be useful in optimizing your EBS volumes.

Optimize capacity for general-purpose SSD gp2 volumes to only pay for what you use Pay close attention to IOPS utilization instead of throughput since you are directly paying for it and you can always provision **10-20%** more than the maximum IOPS utilization By reducing provisioned IOPS or switching to a general-purpose SSD gp2 volume from IOPS SSD io1 volume, you can cut down your maintenance costs

1. You can convert to a cold HDD sc1 volume if the volume is 500 gigabytes or larger to save big bucks on your storage rate

2. You should also remember that you can go back to the original settings if and when needed.

3. Delete stale Amazon snapshots

4. Considering that you have a backup policy that captures daily/weekly EBS snapshots, you might want to look at snapshots that are older than 30 days and delete them to reduce storage costs. Deleting a snapshot will have no effect on the volume and you can use third-party tools such as Command Line Interface or Skeddly for the same.

Storage Optimization is an ongoing process

Maintaining the right-sized and right-priced architecture is a continuous process that requires some diligence. You should consider optimizing the storage every month to make the most efficient use of the storage spend. A few key points that can help you streamline the process are as follows:

1. A continuous mechanism for optimizing the storage and setting up storage policies

2. Monitoring cost as closely as possible using AWS cost and reporting tools

3. Enforcing S3 object tagging and establishing S3 lifecycle policies to continuously optimize data storage throughout the lifecycle

Conclusion

The process of storage optimization includes the evaluation of changes in storage usage, needs, and choosing the most cost-efficient/appropriate AWS storage option. You can streamline the workflow of storage optimization by designing a monthly schedule for the task. You must take advantage of the powerful tools of AWS third-party vendors to regularly monitor storage costs and evaluate volume usage for steady optimization.

Aspire Systems is a global technology services firm serving as a trusted technology partner for our customers. We work with some of the world's most innovative enterprises and independent software vendors, helping them leverage technology and outsourcing in our specific areas of expertise. Our core philosophy of "Attention. Always." communicates our belief in lavishing care and attention on our customer and employees.

For more info contact: info@aspiresys.com or visit www.aspiresys.com

USA	SINGAPORE	INDIA	BELGIUM
+ 1 630 368 0970	+65 3163 3050	+91 44 6740 4000	+ 32 3 204 1942

NETHERLANDS	POLAND	MEXICO
+ 31 (0)30 800 92 16	+48 58 732 77 71	+52 222 980 0115