BEST PRACTICES FOR REDUCING SPEND ON AWS

INTRODUCTION

Most of the companies rely on using Public Cloud services as they are Easy to Use, Flexible, Cost-Effective, Reliable, Scalable and Secure. These are the reasons why companies opt for AWS so as to spend more time on business strategies and application development by reducing time and effort on maintaining the infrastructure and investment of money thereby boosting their profits.

In fact, many of the organizations have seen their monthly cloud bills increasing three times or higher than expectations. This happens mostly because of the improper planning of infrastructure set up where most of the resources fall into a state of unfulfilled excess demand and not implementing the best practices of using AWS services.

Use Reserved Instances

Reserved Instances are a great way to can save you up to 75% discount when you do not expect to change the EC2 instance type or its availability zone for a longer duration. AWS offers reserved instances for 1-year or 3-year terms and gives you the option to choose three payment options: All Upfront, Partial Upfront, and No Upfront. If you choose the Partial or No Upfront payment option, the remaining balance will be due in monthly increments over the term follows different payment options for providing huge discounts.

Delete Unattached/Unused EBS Volumes

When an EC2 instance is launched, EBS volume is usually attached as local storage for the instance and with ‘Delete on Termination' by default option checked using the console to avoid deletion of EBS volumes. Certain times this option is unchecked and when an EC2 instance is terminated, EBS volume remains and gets continuously charged by AWS despite the fact that it is not in use.

It is always ideal to delete unattached EBS volumes or any EBS volume which has very low I/O activity and rarely used. (Ensure that EBS volume is always backed up before terminating the instance and volume to ensure important data is not lost).This will help you to avoid unnecessary charges on your AWS bill and also your data is not compromised or available to others.

Delete Older EBS Volume Snapshots

An EBS snapshot is a point-in-time copy of your Amazon EBS volume. Organizations schedule automated EBS volume snapshots daily, weekly or monthly based on the business requirement. Since EBS volume snapshots are incremental copies of data, deleting older snapshots will lower your monthly costs and also it does not affect the ability to restore the data from later snapshots available.

When an EBS volume is deleted, its snapshots are never deleted automatically and they remain in your environment always ensure when any EBS volume is deleted its snapshots are also deleted and verified.

BEST PRACTICES FOR REDUCING SPEND ON AWS

Rightsize Resources /EC2 Instances, EBS Volumes & RDS databases

Most organizations incur more than 50% of the cost on EC2 instances, EBS Volumes and RDS services as AWS provides wide variety of instance types and sizes. Developers or employees often tend to spin up new resources that are larger than needed and end up over-provisioning resources leading to higher costs also by not doing the right assessment of the resources based on the business needs.

Most of the times resources are underutilized either by provisioning resources more than required or either workload is no longer resource intensive.

To avoid this, we need to do a proper readiness assessment of the existing applications before migrating to cloud, enable auto-scaling for EC2 instances and then enable CloudWatch monitoring to get alerts based on metrics i.e CPU utilization, Memory utilization, Disk utilization, Network utilization. Based on the metric details, we can rightsize the resources.

Use Spot Instances

Amazon EC2 Spot Instances let you take advantage of unused EC2 capacity in the AWS cloud which is available at a 90% discount compared to On-Demand Instances. Spot Instances can be utilized for CI/CD development, Big Data Processing, Scaling of EC2 instances by adding Spot Instances when needed and any less important workloads which are needed up to 6-10 hours usage without interruption. Most organizations are getting benefited on using Spot Instances for most of their scenarios and also we have the advantage of quoting the price of our own rather an AWS provided a price for longer usage. This is the most important factor to keep in mind as they save around 40% of our monthly EC2 usage if used properly.

Optimize Amazon S3 Storage by Using Appropriate Storage Tier and Enable Life Cycle Management to Reduce Cost.

Amazon Simple Storage Service (Amazon S3) is one of the most popular Amazon Web Services (AWS) offering with flexible pricing for storing data which is scalable, high-speed, low-cost, secure and durable. It is designed for storing hot data and cold data and has a simple web interface that you can use to store and retrieve any amount of data, at any time, from anywhere using the internet.

Amazon offers 6 different tiers for storing data i.e Amazon S3 Standard, Amazon S3 Standard IA, Amazon S3 One Zone IA, Amazon Glacier and Amazon Glacier Deep Archive.

Amazon S3 Standard: S3 Standard offers high durability, availability, and performance object storage for frequently accessed data. Because it delivers low latency and high throughput, S3 Standard is appropriate for a wide variety of use cases, including cloud applications, dynamic websites, content distribution, mobile and gaming applications, and big data analytics.

Amazon S3 Standard-Infrequent Access (S3 Standard-IA): S3 Standard-IA is for data that is accessed less frequently, but requires rapid access when needed. S3 Standard-IA offers the high durability, high throughput, and low latency of S3 Standard, with a low per GB storage price and per GB retrieval fee. This combination of low cost and high performance makes S3 Standard-IA ideal for long-term storage, backups, and as a data store for disaster recovery files.

BEST PRACTICES FOR REDUCING SPEND ON AWS

Amazon S3 One Zone-Infrequent Access (S3 One Zone-IA): S3 One Zone-IA is for data that is accessed less frequently, but requires rapid access when needed. Unlike other S3 Storage Classes which store data in a minimum of three Availability Zones (AZs), S3 One Zone-IA stores data in a single AZ and costs 20% less than S3 Standard-IA. S3 One Zone-IA is ideal for customers who want a lower-cost option for infrequently accessed data but do not require the availability and resilience of S3 Standard or S3 Standard-IA.

Amazon S3 Glacier (S3 Glacier): S3 Glacier is a secure, durable, and low-cost storage class for data archiving. You can reliably store any amount of data at costs that are competitive with or cheaper than on-premises solutions. To keep costs low yet suitable for varying needs, S3 Glacier provides three retrieval options that range from a few minutes to hours

Amazon S3 Glacier Deep Archive (S3 Glacier Deep Archive): S3 Glacier Deep Archive is Amazon S3's lowest-cost storage class and supports long-term retention and digital preservation for data that may be accessed once or twice in a year. It is designed for customers — particularly those in highly-regulated industries, such as the Financial Services, Healthcare, and Public Sectors — that retain data sets for 7-10 years or longer to meet regulatory compliance requirements. S3 Glacier Deep Archive can also be used for backup and disaster recovery use cases and is a cost-effective and easy-to-manage alternative to magnetic tape systems, whether they are on-premises libraries or off-premises services

To optimize the cost of your data storage, enable object lifecycle management that automatically transitions data between the storage classes. For instance, you can automatically move your data from S3 Standard to IA after 30 days, archive data to Glacier after 90 days, or set up a delete policy to expire specific objects after 180 days depending on your business needs.

Amazon Storage Classes Table

	S3 Standard	S3 Standard IA	S3 Z-IA	Glacier	Glacier Deep Archive
Redundancy	99.999999999% (11 9’s)	99.999999999% (11 9’s)	99.999999999% (11 9’s)	99.999999999% (11 9’s)	99.999999999% (11 9’s)
Availability	100%	100%	100%	100%	100%
Availability Zones	≥3	≥3	100%	≥3	≥3
The minimum period of storage	Unlimited	30 days	30 days	90 days	180 days
The minimum size of the object	Unlimited	128KB (for less you are charged as for 128KB)	128KB (for less you are charged as for 128KB)	40KB (for less you are charged as for 40KB)	40KB (for less you are charged as for 40KB)
Access to the object	Milliseconds	Milliseconds	Milliseconds	1 minute - 12 hours, depending on the retrieval options	12 hours-48 hours, depending on the retrieval options
Charge for data retrieval	None	Per GB of the data retrieved	Per GB of the data retrieved	Per GB of the data retrieved	Per GB of the data retrieved

Life Cycle Transitions	Yes	Yes	Yes	Yes	Yes

Reduce AWS costs with EC2 Instance Scheduling/ Stop and Start Instances on a Schedule

Most organizations can reduce Amazon Elastic Compute Cloud (Amazon EC2) usage by stopping and starting EC2 instances automatically at certain times (weekdays during nights (6PM-8AM) or low business hours, weekends or during holidays) or utilization thresholds.

Outside of production, there are many instances related to development or test environment that are not required to run 24/7/365. These instances can be turned off between 6PM-8AM on weekdays, weekends and during business holidays to avoid charges for EC2 instances and reduce billing per month. But remember though we stop EC2 instances outside of business hours we are still charged for EBS volumes as they are attached to EC2 instance and have very low I/O activity and can avoid data loss. These can be automated using various ways with the help of scripting or using AWS services.

Delete Unused/Unattached Elastic IP Addresses

An Elastic IP address known as Static IP address is a reserved public IP address that can be associated to any EC2 instance. As long as Elastic IP address is in use you are not charged, but the moment Elastic IP address becomes idle either because of not being associated to any EC2 instance or not deleting after the EC2 instance termination AWS charges for the unused Elastic IP addresses.

AWS can allocate 5 Elastic IP addresses for an AWS account per region, but if we want to increase the limit then we will have to reach out to the AWS team to increase the limit.

Use Auto Scaling for EC2 Instances

AWS EC2 Auto Scaling is one of the best practices that need to be implemented for EC2 instances as they help you ensure that the correct number of instances are always available to handle the load for the application.

Some of the benefits of using EC2 Auto Scaling:

- **Better fault tolerance**: Amazon EC2 Auto Scaling can detect when an instance is unhealthy, terminate it, and launch an instance to replace it. You can also configure Amazon EC2 Auto Scaling to use multiple Availability Zones. If one Availability Zone becomes unavailable, Amazon EC2 Auto Scaling can launch instances in another one to compensate.
- **Better availability**: Amazon EC2 Auto Scaling helps ensure that your application always has the right amount of capacity to handle the current traffic demand.
- **Better cost management:** Amazon EC2 Auto Scaling can dynamically increase and decrease capacity as needed. Because you pay for the EC2 instances you use, you save money by launching instances when they are needed and terminating them when they aren't.

BEST PRACTICES FOR REDUCING SPEND ON AWS

Use Tags in AWS to Improve Cost Allocation and Usage Optimization

AWS allows customers to assign metadata to their AWS resources in the form of tags. Each tag is a label that you assign to AWS resources that consists of a key and value. You can use tags to organize your resources and cost allocation tags to track your AWS costs. An effective tagging is a first step ensuring towards compliance, usage optimization, and cost control mechanism for organizations.

Minimize AWS Data Transfer Costs

AWS charges data transfer costs between AWS and the Internet which are highly dependent on the region you select and within AWS i.e Data transfer across regions and Data transfer across within regions.

Usually, AWS data transfer costs are generally higher between regions when compared to inter-region data transfers between availability zones, but again AWS data transfer charges between availability zones are higher compared to with-in an availability zone

Points to Remember:

- Inbound data from the Internet to AWS is free but outbound data from AWS to the Internet is charged as per the AWS pricing model.
- Between EC2 instances in the same Availability Zone Intra-region data in/out is free with a Private IP.
- Between EC2 instances across the Availability Zone's data in/out is not free and they are charged as per the AWS pricing model.
- Inter-Region data in is free and data out is charged as per the AWS pricing model.

The best way to minimize data transfer cost is too

- Architect or Design systems in such a way there are fewer data across AWS regions or availability zones.
- As different AWS regions have different associated data transfer costs, it is always better to decide the region first and then plan based on the pricing model.
- Always use private IP addresses instead of the public or elastic IP address for data transfers.
- Use Amazon CloudFront (AWS’s CDN) as much as you can when serving static files. The data transfer out rates are cheaper from CloudFront and free between CloudFront and EC2 or S3.