Cloud Optimization Through Discounts and Consolidated Billing

January, 2023

DRAFT

Federal Technology Investment Management (FTIM)
Community of Practice (CoP)
Cloud & Infrastructure (C&I) CoP

General Services Administration

Office of Government-wide Policy

Contents

Summary

This document looks primarily at cloud cost optimizations through vendor discounts that are applied through, savings plans, reserved instances and consolidated billing. This document focus on AWS and Azure. The Google Cloud Platform (GCP) was not part of the scope at the time of this writing.

The items discussed in the document are billing constructs that impact the way in which costs or savings are applied and are transparent to the performance and operation of the running virtual instances and other applicable resources

Definitions

Savings Plans vs Reserved Instances

Reserved Instances are based on the commitment to use an instance at a particular price over a specific period, while Savings Plans are based on the commitment to spend a particular dollar amount per hour over a specific period. With reservations, you commit to a specific virtual machine type in a particular region. For example, an Azure D2v4 VM in Japan East or an AWS M4 for one year. With a savings plan, you commit to spend a fixed hourly amount collectively on compute services. For example, $5.00/hour on compute services for one year. Reservations only apply to the identified compute service and region combination. Savings plan benefits are applicable to all usage from participating compute services across the globe, up to the hourly commitment.

Choose a Reservation

For highly stable workloads that run continuously and where you have no expected changes to the machine series or region, consider a reservation. Reservations provide the greatest savings.

Choose a Savings Plan

For dynamic workloads where you need to run different sized virtual machines or that frequently change datacenter regions, consider a compute savings plan. Savings plans provide flexible benefits and automatic optimization.

Savings Plans

Savings Plans offer significant savings over On-Demand or Pay-As-You-Go pricing in exchange for a commitment to use a specific dollar amount of computing power. This rate is measured by dollars-per-hour over a 1-year or 3-year commitment period.

Savings Plans can provide savings of up to 72% on your compute usage–regardless of instance family, size, OS type, tenancy, or Region. Savings plans are much simpler to manage and provide additional flexibility over Reserved Instances. They can apply across instance types and regions. They generally do not apply to database usage.

Reserved instances (RIs)

Reserved Instances are the original discount structure for the major cloud providers. RIs require specifying an instance type and size as well as a region. Amazon EC2 Reserved Instances (RI) or Azure reserved instances provide a significant discount (up to 72%) compared to On-Demand or Pay-As-You-Go pricing and provide a capacity reservation. Discount amounts are dependent on instance type, size, OS, etc.

The RI discount model provides a discounted cost for compute and does require you to coordinate your RI purchases and exchanges to ensure that you have an optimal mix that covers usage, which might change over time.

RIs are complex to manage and maintain in large cloud deployments. They lack flexibility in instance type, region, tenancy, etc. Reserve instances benefit when there is a known and steady demand for a particular resource type and size for an extended period of time.

Consolidate Billing - AWS

In an organization, the management account is responsible for paying all charges that the member accounts incur. If you're an administrator of a management account and you have the appropriate permissions, you can view aggregated usage costs for Reserved Instance discounts and volume tiering for all member accounts. You can use consolidated billing to view aggregated usage costs for accounts in the organization. Consolidated billing can also help you reduce those costs. For example, to ensure that you pay the lowest available prices for AWS products and services, AWS offers pricing tiers that reward higher usage with lower prices and discounted rates for purchasing instances in advance (known as *reservations* or *Reserved Instances*). Using consolidated billing, you can combine usage from multiple accounts into a single invoice, allowing you to reach the tiers with lower prices faster. You can also apply unused reservations from one account to another account's instance usage.

Optimization Strategy

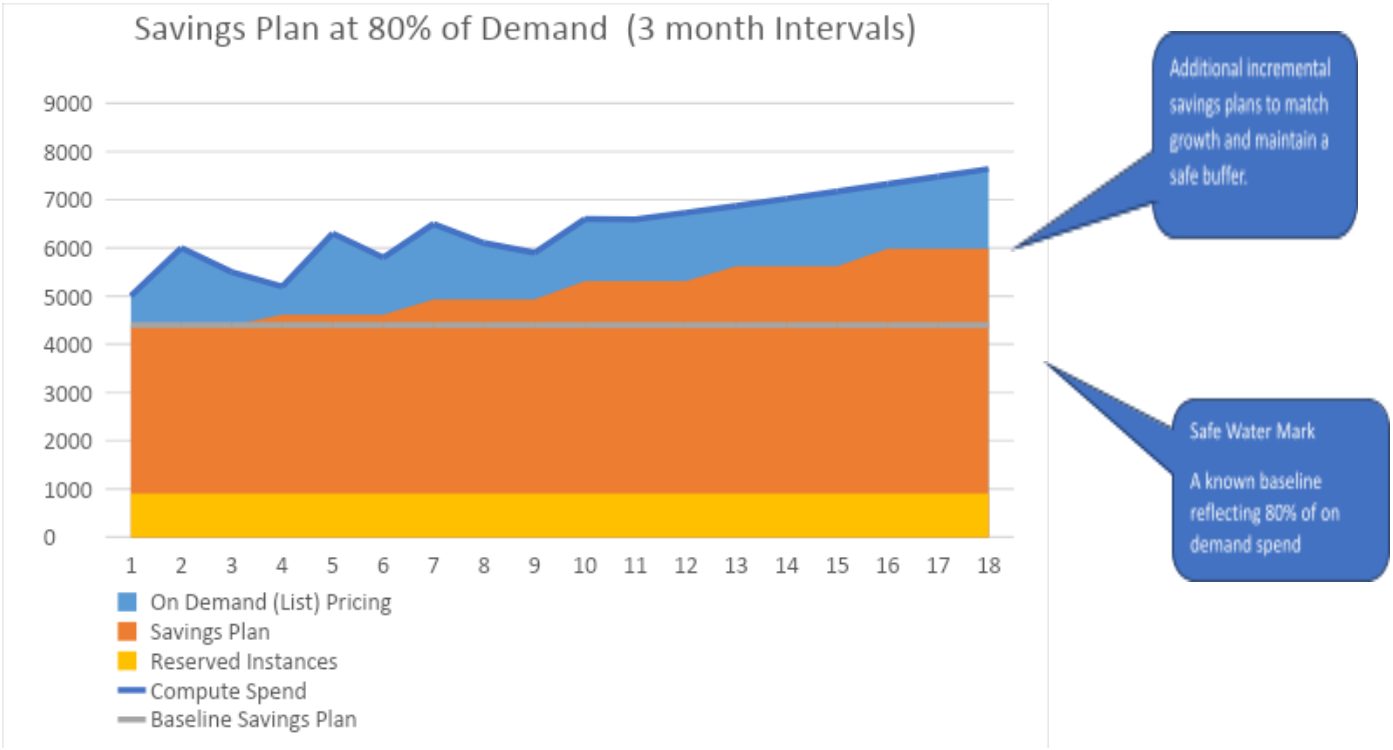
An optimization strategy is intended to maximize the discounts offered by the major cloud providers. It will likely consist of a combination of Reserved Instances and Savings Plans. Discounts can be purchased at any time during the calendar year and do not have to align to fiscal boundaries or provisioning of resources.

First look at known compute and look for consistency in instance type and size over time. Purchase Reserved Instances that match that consistent baseline of compute. After that, savings plans can be

purchased to obtain discounts on more variable cloud resources. They can be purchased in intervals and in varying amounts to meet your demand. Some recommendations state that a combination of RI and savings plans should be at a level equal to your minimum on demand level. This would be recommended if your cloud costs consistently grow but may put some of your discounts at risk if the variation in your demand dips during any given month. Judgement should be placed on plan coverage but 80% is generally a safe level because it allows for downward fluctuation without risk of losing on discounted commitments. Using this level as a baseline, continue to monitor your costs monthly and continue to purchase incremental savings plans to get closer to your optimal demand curve.

AWS Compute Savings Plans support Amazon EC2, AWS Fargate, and AWS Lambda. Savings Plans do not cover Amazon RDS, just as RIs do not cover Fargate for serverless applications.

Azure savings plans cover compute. They do not cover other software, Windows, networking, or storage charges.



Understanding AWS Savings Plan Recommendations & Cost

You can view the recommendations for your AWS account through the console. There are three primary metrics to be aware of:

- **Monthly On-Demand Spend:** An estimated monthly cost of total usage without discounts over a selected period (including all active Savings Plans). This metric is used to illustrate effectiveness of your existing Savings Plans versus On-Demand usage.
- **Estimated Monthly Spend:** A monthly cost projection based on current commitment recommendations. This metric includes any forecasted On-Demand usage due to hour-to-hour variations.

- **Estimated Monthly Savings:** An estimated net savings for a selected period based on the current (unpurchased) commitment recommendations.

AWS Reserved Instances Vs. Savings Plans

AWS Savings Plans	
Advantages	**Limitations**
AWS Savings Plans don't lock you into certain instance types like RIs, which means you can make changes within an instance family to be more cost-efficient.	AWS Savings Plans can't be purchased for RDS, Redshift, and other services.
AWS Savings Plans can benefit from price changes that occur during your commitment, unlike RIs.	AWS Savings Plans don't offer reseller opportunities to offload underutilized commitments.
AWS Savings Plans can be applied to Fargate as well as EC2 (but not RDS).	AWS Savings Plans charge On-Demand prices for overages.
AWS Savings Plans avoid the complexity of convertible RI exchanges.	AWS Savings Plans don't provide capacity reservations.
AWS Savings Plans require less infrastructure planning.	AWS Savings Plans don't often provide better discounts compared to RIs
AWS Savings Plans allow you to flexibly transfer workloads between instance types, sizes, and generations to meet changing demand and architecture.	
AWS Savings Plans provide discounts without committing you to any specific instance type.	

AWS Reserved Instances	
Advantages	**Limitations**
RIs for shorter terms can be purchased on the marketplace.	RI restricts a user to a specific instance family, such as the M4 instance,
RIs can include discounts for RDS, as well as EC2 (but not Fargate).	RIs only allows modification of the instance size, such as sizing down from an m4.2xlarge to m4.large. *.

RIs can provide some of the largest discounts on the higher-end (60% or more) in the case of some 3-year upfront terms.	RIs are limited to a specific availability zone within a certain region or availability zone.
Used for applications that require constant, "always-on" power for the better part of one to three years commitment time.	Limited to technology generation.
You have a reasonably consistent usage pattern over an extended period	

AWS Reserved Instance Details

Reserved Instances are a financial commitment to reserved discounted capacity over a fixed term of one or three years. Whenever you launch a machine that matches the specifications of your Reserved Instance, Amazon will apply the credit to your running instance in place of the on-demand rate. Each Reserved Instance credit can only be applied to a single running machine at any one time.

However, during any period where you have no active matching instances, your Reserved Instance credit will remain unused and go to waste. This underlines why it's so important to utilize your reserved capacity as much as possible — in order to achieve the maximum possible discount and return on investment.

Reserved Instances apply across Amazon EC2, Elasticsearch, Relational Database Service (RDS), and RedShift;

The level of discount you can expect to achieve also depends on the attributes you specify at the time you make your purchase.

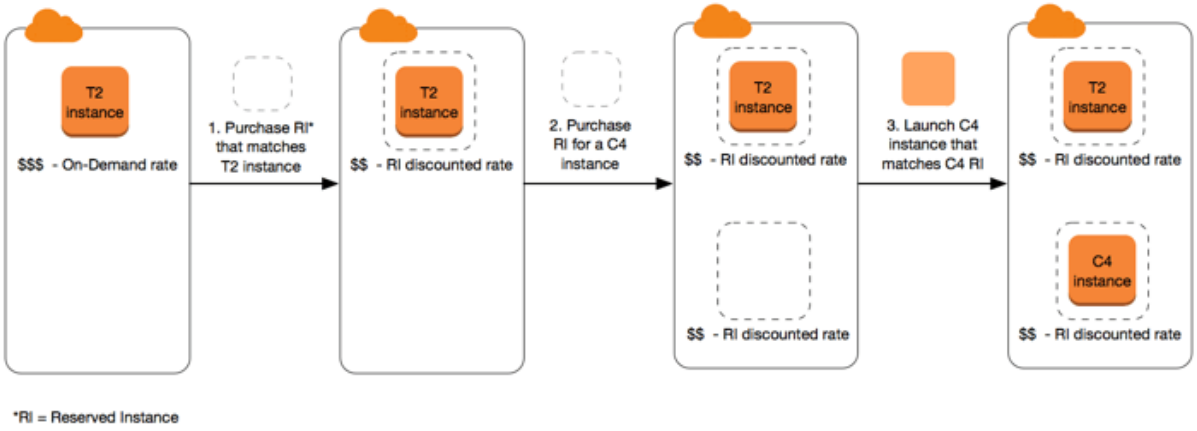
These include:

- **Instance type:** The compute resource you want to reserve
- **Length of term:** Either one or three years
- **Platform:** The operating system (OS)
- **Region:** The AWS region in which your Reserved Instance will apply
- **Tenancy:** A standard EC2 instance or single-tenant hardware
- **Amount of upfront payment:** All upfront, partial upfront or no upfront
- **Reserved Instance type:** Standard or Convertible

If you purchase a Reserved Instance for a specific account, Amazon can still apply your reservation to a matching instance in any of your company's linked accounts. Although it gives first priority to the account that made the purchase, other accounts may still consume your Reserved Instances even when you have a matching running machine.

This process is ideal for enterprises with many accounts because those instances can "float" between accounts. However, that flexibility may present cost allocation challenges for some organizations.

AWS then repeats the above process every hour. As Amazon only repeats the allocation process once an hour, you could launch a machine that matches one of your Reserved Instances, but still pay the on-demand rate for up to an hour before the reservation gets reallocated to it. In the meantime, another machine in another account could be benefitting from the Reserved Instance credit that you have paid for. This will be reflected in your AWS bill at the end of the month.



AWS savings Plans

When Is It Best To Use AWS Savings Plans?

Savings Plans offer flexibility, making them ideal for organizations and systems prone to usage changes. They are ideal when you:

- Want to save on compute services — not on database usage.
- Expect that your usage requirements will fluctuate significantly during the contract period. Compute Savings Plans allow exchanges between instance families, operating systems, sizes, tenancies, and regions.
- Need a setup that automatically adapts to changes in your infrastructure without constant, manual monitoring.
- Have serverless use cases with AWS Fargate and SageMaker applications
- Use Amazon EC2, Fargate, and Lambda that fluctuates over short periods, such as from season to season.

Azure Reserved Instances vs Savings Plans

Both Azure reservation and savings plan apply to the infrastructure part of the resource and not the license. Purchase models are similar in allowing monthly, partial upfront or all upfront billing arrangements. Savings Plans and RIs are both available for 1 or 3 year commitments and they are billed on a per hour basis.

Key difference between Azure reservation instances and savings plan are as below.

1. Azure reservations are purchased for a specific region and a specific SKU type (there is some flexibility around SKU family) but the only way to switch the SKU type is through RI exchange
2. In savings plans for compute, there is a greater level of flexibility where the only information that's required is choosing the scope (discussed later), Azure plan amount and the term. There is no need to choose a region or SKU type.

Azure savings plan will apply to the maximum discounted compute resource and you can switch it to any resource based on the scope, irrespective of the region or compute resource type.

Azure Savings Plans	
Advantages	**Limitations**
Azure Savings Plans don't lock you into certain instance types scope or region, series or OS	AWS Savings Plans can't be purchased for RDS, Redshift, and other services.
Azure Savings Plans can benefit from price changes, unlike RIs since they are based on hourly commitment of compute.	Savings plans cannot be traded/exchanged, returned, or cancelled
Savings plans apply to: ● Virtual Machines (only the compute costs; ie; it does not include storage, network, licensing, etc) ● Azure Dedicated Host (as with VMs, only the compute costs are covered by the savings plan) ● Container Instances ● Azure Premium Functions ● Azure App Services (not all are included; Pv3 or isolated v2 is required)	Azure Savings plans do not cover: ● Software ● OS ● Networking ● Storage charges. ● BareMetal Infrastructure ● A, G, and GS series VMs.
Savings Plans avoid the complexity of fixed or convertible RI and exchanges, refunds or cancelations.	Savings Plans don't often provide better discounts compared to RIs
Savings Plans require less infrastructure planning.	Azure Savings Plans charge On-Demand prices for overages.

Azure Reserved Instances	
Advantages	**Limitations**
RIs for shorter terms can be purchased on the exchange.	RI restricts a user to a SKU, region and scope.

Azure Reservations can include discounts for VMs (windows and Linux), SQL DB, Cosmos DB, as well as storage reserved capacity.	RIs only allows exchange of reservations of the same type.
RIs can provide some of the largest discounts on the higher-end (80%) in the case of some 3-year windows commitments.	Reservations are limited to a specific availability zone, series and size.
Used for applications that require constant, "always-on" CPU utilization for the better part of one to three years commitment time.	Limited to technology generation.
You have a reasonably consistent usage pattern over an extended period	Scope must be managed. It can be limited to a single resource group, a single subscription, or shared scope across multiple subscriptions
B-Series instances can save 15% for burstable loads	Reservations don't apply A-series, Av2-series, or G-series VMs.

Azure Reserved Instance Details

Azure Reservations are pre-purchase commitments that reduce cloud consumption costs by reserving resources in advance. However, this pricing option does not apply to every resource on Azure. You can only leverage Azure Reservation discounts on a subset of virtual machines, app services, storage platforms, databases, or analytics services.

Microsoft splits Azure Reservations into two broad categories: Reserved Instances and Reserved Capacity.

- **Reserved Instances**: a pricing option for virtual machines
- **Reserved Capacity**: a set of discounts for Azure app, storage, and data services

- Once you purchase an Azure Reservation, Microsoft automatically applies the discount to resources matching the reservation's options and quantity. It implements the discounted pricing to the scope you set during the purchase process; this could be a subscription, resource group, or single resource.

- For example, reserving an instance and applying it to a resource group will use the Reserved Instance discount for any compute elements in that collection. Similarly, purchasing Reserved Capacity and setting the scope to the subscription will apply the discount to eligible Azure data services at that level.

- Determining the scope and reservation to purchase requires analysis. Microsoft recommends you only buy a reservation after examining the consistent base usage of the identified resource.

- In addition to analyzing usage data, Microsoft also provides resources that offer recommendations based on consumption patterns. The Azure Portal provides recommendations during the reservation purchase process. The VM BI Coverage report that forms part of the **Cost Management Power BI app** is another valuable resource that Enterprise customers can use

B-Series VMs

Many cloud workloads such as web servers, small databases, and development and test environments, often experience varying CPU utilization. These workloads will run for a long time using a small fraction of available CPU and then when traffic or processing increases, will quickly begin using all the CPU power in the VM. While most of the time these VMs are greatly underutilized (and hence not cost-effective), they cannot be easily downsized as there are times when full CPU capacity is required. To address this, Azure offers their B-Series burstable VMs. When not being used at a defined baseline of usage, these VMs build up credit so when full capacity is required, the cost is discounted, anywhere between 15-55% in comparison to equivalent VMs. However, if the VM becomes too CPU-intensive, it'll be throttled down to the defined baseline performance until enough credits are available. Therefore, some analysis is required to determine if your workloads are sufficiently low-use to benefit from Azure's B-Series.

Azure Savings Plan Details

Azure native utilities can provide savings plan purchase recommendations that are calculated by analyzing your hourly usage data over the last 7, 30, and 60 days. Azure calculates what your costs would have been if you had a savings plan and compares it with your actual pay-as-you-go costs incurred over the time duration. The calculation is performed for every quantity that you used during the time frame. The commitment amount that maximizes your savings is what is recommended.

For example, you might use 500 VMs most of the time, but sometimes usage spikes to 700 VMs. In this example, Azure calculates your savings for both the 500 and 700 VM quantities. Since the 700 VM usage is sporadic, the recommendation calculation determines that savings are maximized for a savings plan commitment that is sufficient to cover 500 VMs and the recommendation is provided for that commitment.

Note the following points:

- Savings plan recommendations are calculated using the on-demand usage rates that apply to you.
- Recommendations are calculated using individual sizes, not for the instance size family.
- The recommended commitment for a scope is reduced on the same day that you purchase a commitment for the scope.
    - However, an update for the commitment amount recommendation across scopes can take up to 25 days. For example, if you purchase based on shared scope recommendations, the single subscription scope recommendations can take up to 25 days to adjust down.
- Currently, Azure doesn't generate recommendations for the management group scope.

Savings Plan Scope

You can scope a savings plan to a shared scope, management group, subscription, or resource group scopes. Setting the scope for a savings plan selects where the savings plan savings apply. When you scope the savings plan to a resource group, savings plan discounts apply only to the resource group—not the entire subscription.

Savings plan scoping options

You have the following options to scope a savings plan, depending on your needs:

- **Shared scope** - Applies the savings plan discounts to matching resources in eligible subscriptions that are in the billing scope. If a subscription was moved to a different billing scope, the benefit no longer applies to the subscription. It does continue to apply to other subscriptions in the billing scope.
    - For Enterprise Agreement customers, the billing scope is the enrollment. The savings plan shared scope would include multiple Active Directory tenants in an enrollment.
    - For Microsoft Customer Agreement customers, the billing scope is the billing profile.
    - For Microsoft Partner Agreement, the billing scope is a customer.
- **Single subscription scope** - Applies the savings plan discounts to the matching resources in the selected subscription.
- **Management group** - Applies the savings plan discounts to the matching resource in the list of subscriptions that are a part of both the management group and billing scope. To scope a savings plan to a management group, you must have at least read permission on the management group.
- **Single resource group scope** - Applies the savings plan discounts to the matching resources in the selected resource group only.

Appendix

AWS Comparison of Offerings

AWS Offering Types

Unit	Reserved Instances	AWS Savings Plans
Types	Standard RIs and Convertible RIs	EC2 Instance Savings Plans and Compute Savings Plans
Commitment	One year or three years	One year or three years
Payment	All upfront, partial upfront, and no upfront	All upfront, partial upfront, and no upfront

AWS Discount & Coverage Comparisons

Pricing Category	Standard RIs	Convertible RIs	EC2 Instance Savings Plans	Compute Savings Plans
Savings over On-Demand	Up to 72%	Up to 66%	Up to 72%	Up to 66%

Lower price for monetary commitment	No	No	Yes	Yes
Applies to any family	No	No	No	Yes
Applies to any size	No	No	Yes	Yes
Applies to any tenancy or OS	No	No	Yes	Yes
Applies to Fargate & Lambda usage	No	No	No	Yes
Applies across any AWS region	No	No	No	Yes
Offers 1- and 3-year terms	Yes	Yes	Yes	Yes

AWS Feature Comparisons

Type of Feature	Standard RI	Convertible RI	EC2 Instance Savings Plan	Compute Savings Plan
Scope	Zonal (for Zonal RI) and Regional (for Regional RI)	Zonal (for Zonal RI) and Regional (for Regional RI)	Global	Global
Estimated cost reduction	Up to 72%	Up to 66%	Up to 72%	Up to 66%
Average discount for 1 year	38%	29%	38%	29%
Average discount for 3 years	58%	51%	58%	51%
Availability of 1 year tenure	Yes	Yes	Yes	Yes

Availability of 3 years tenure	Yes	Yes	Yes	Yes
Instance family flexibility	No	Yes (conditions apply)	No	Yes
Operating system flexibility	No	Yes	Yes	Yes
Regional flexibility	No	No	No	Yes (except China)
Tenancy flexibility	No	No	Yes	Yes
Payment options	All, Partial, No Upfront	All, Partial, No Upfront	All, Partial, No Upfront	All, Partial, No Upfront
Availability of capacity reservation	Yes	Yes	No	No
Applicability of account limits	Yes	Yes	No	No
Entitled to sell in the marketplace	Yes	No	No	No
Instance size flexibility	Yes, for Linux OS	Yes, for Linux OS using Exchange Reserved Instances API and console	Yes	Yes
Availability Zone flexibility	Yes, for Regional RI Linux OS using Modify Reserved Instances API and console	Yes, for Regional RI Linux OS using Exchange Reserved Instances API and console	Yes	Yes
List of Supported AWS services	EC2, RDS, Elastic Cache, DynamoDB, Redshift	EC2	EC2	EC2, Fargate, Lambda
Offload/Exchange Commitment	Sell under AWS Marketplace	Exchange for another instance size, family, OS, tenancy, term, or payment option	None	None

AWS Term Comparisons

Unit	Reserved Instance	EC2 Instance Savings Plan	Compute Savings Plan
Average 1 Year Discount	38%	29%	29%
Average 3 Year Discount	58%	58%	51%
Instance Family	Fixed	Fixed	Flexible
Instance Size	Fixed (except Linux)	Flexible	Flexible
Geography	1 Region	1 Region	Flexible
OS	Fixed	Flexible	Flexible
Service	EC2 / RDS	EC2	EC2 / Fargate

AWS Standard RI Flexibility Comparisons

Type	Standard Regional RIs	Standard Zonal RIs	EC2 Instances Savings plans
Commitment Unit	Per instance	Per instance	Per Dollar
Geography	Region specific	Zone specific	Region specific
Instance Family	Fixed	Fixed	Fixed
Tenancy	Fixed	Fixed	Any tenancy
Operating System	Fixed	Fixed	Any OS

AWS Convertible RI Flexibility Comparisons

Type	Convertible Regional RIs	Convertible Zonal RIs	AWS Compute Savings plans
Commitment Unit	Per instance	Per instance	Per Dollar
Geography	Region specific	Zone specific	Any region
Instance Family	Exchangeable	Exchangeable	Any family
Tenancy	Exchangeable	Exchangeable	Any tenancy
Operating System	Exchangeable	Exchangeable	Any OS

Azure Comparison of Offerings

Azure Reserved Instances

Azure Reserved Instances is a pricing option that allows you to reserve capacity on a subset of virtual machines for a period of one or three years. By committing and prepaying for the Azure virtual machine and compute component, you can reduce the cost by up to 72 percent. However, this discount only applies to your virtual machine cost—it does not apply to any pre-installed software, networking, or storage costs.

The table below details the inclusions and exclusions:

Resource Type	Included	Excluded
Reserved Virtual Machine Instance	Virtual machine and cloud services compute costs	Additional software, Windows, networking, or storage charges.
Azure Dedicated Host	Compute costs	Virtual machine, cloud services compute costs, additional software, Windows, networking, or storage charges.

Azure Reserved Instances Minimum Requirements

You can apply Reserved Instance pricing to both Windows and Linux virtual machines running on Azure. However, not every configuration is eligible for this discount. This pricing plan excludes virtual machines that form part of the A-series, Av2-series, and G-series. Any promotional virtual machines or images in preview are also ineligible.

Reserved Instance Options

Reserved Instance Type	Savings vs Pay-as-You-Go
Windows VMs	Up to 80%
Linux VMs	Up to 72%

Azure Reserved Capacity

Reserved Capacity, like Reserved Instances, offers pricing discounts for pre-committing to services. The difference between the two offerings is the selected resources. Reserved Instances refers to a pricing plan that applies to virtual machines. Reserved Capacity covers everything else eligible for an Azure Reservation. Depending on the service and length of commitment, Reserved Capacity savings can range up to 65 percent. Azure also excludes individual components. Depending on the consumed service, you may still pay the full price for software, networking, and storage.

The table below details the inclusions and exclusions:

Resource Type	Included	Excluded
Azure Disk Storage reservations	Premium SSDs of P30 size or greater.	Other disk types or sizes smaller than P30.
Azure Storage reserved capacity	Standard storage accounts for Blob storage or Azure Data Lake Gen2 storage.	Bandwidth and transaction rates.
Azure Cosmos DB reserved capacity	Throughput.	Storage and networking.
SQL Database reserved vCore	SQL Managed Instance and SQL Database Elastic Pool/single database.	SQL license billed separately.
Azure Synapse Analytics	DWU usage.	Storage and networking.
Azure Databricks	DBU usage.	Compute, storage, and networking.

App Service stamp fee	Stamp usage.	Workers or any other resources associated with the stamp.
Azure Database for MySQL	Compute costs	Software, networking, and storage.
Azure Database for PostgreSQL	Compute costs.	Software, networking, and storage.
Azure Database for MariaDB	Compute costs	Software, networking, and storage.
Azure Data Explorer	Markup charges.	Compute, networking, and storage.
Azure Cache for Redis	Compute costs.	Networking and storage.

Azure Reserved Capacity Minimum Requirements

Like Reserved Instances, Reserved Capacity is not a universal discount for all data services running on Azure. Microsoft applies Reserved Capacity pricing to data solutions running on a minimum of 8 vCores for SQL Databases or 20,000 Request Units (RUs) for Azure Cosmos DB. Reservation discounts on App Services are only available on the Premium V3 and Isolated tiers.

Reserved Capacity Options

Reserved Instance Type	Savings vs Pay-as-You-Go
Azure SQL Database	Up to 80%
Azure Cosmos DB	Up to 65%
Azure Synapse Analytics	Up to 65%
App Service	Up to 55%
Azure Storage Reserved Capacity	Up to 38%

Azure Terms Comparison

Depending on the service, Microsoft applies reservation pricing as detailed in the table below:

Resource Type	One Year Commitment	Three Year Commitment
Reserved Virtual Machine Instance	Up to 32% discount	Up to 72% discount
Azure Disk Storage reservations	Up to 5% discount	Not on offer
SQL Database reserved vCore	Up to 21% discount	Up to 33% discount
Azure Synapse Analytics	Up to 37% discount	Up to 65% discount
Azure Databricks	Up to 39% discount	Up to 61% discount
Azure Database for MySQL	Up to 42% discount	Up to 61% discount
Azure Database for PostgreSQL	Up to 39% discount	Up to 59% discount
Azure Database for MariaDB	Up to 42% discount	Up to 61% discount
Azure Cache for Redis	Up to 33% discount	Up to 65% discount
Azure App Service Premium V3	Up to 25% discount	Up to 40% discount
Azure App Service Isolated	Not on offer	Up to 40% discount

Conclusion

Most organizations will operate in a hybrid cloud and on-prem environment as well as a multi-cloud environment. This document provided some of the discount programs offered by two of the major cloud providers. Your needs will dictate your environment scope and constraints. This document is in no way exhaustive but takes a sampling of the most salient discount program best practices and puts them in one place.

Cloud optimization is not a once and done endavor. It is a continuous part of the ongoing operations and a critical part getting the best value out of your cloud spend.

Monitoring Savings Plans

Monitoring your usage is an essential part of managing your AWS Savings Plans. This helps you understand how discounts are applied and what usage types are covered under your Savings Plans. You can manage usage applicable to your Savings Plans in many ways, such as:

Using Reserved Instances and Saving Plans Together

Purchasing a Savings Plan doesn't free you from following cost optimization strategies. Many consumers who plan on using Savings Plan have already purchased Reserved Instances for lower prices. Trying to abandon your RIs by selling them all and buying a Savings Plan isn't an ideal solution for your budget; conversely, having a portfolio of just Reserved Instances often lacks the flexibility your business needs. Using both models balance more significant discounts with long-term flexibility.

If you have existing RIs as a base layer for known compute your savings plan will then be applied after RI commitments up to your savings plan commitment level. This will maximize savings on top of your Savings Plan, you have a backup discount program that can reduce cloud costs until the expiration of your reservations. If you have predictable resources that aren't covered by RIs, applying a Savings Plan on them can be a safe option.

AWS Consolidated billing - Calculating Consolidated Bills for Storage Savings

In an organization, the management account is responsible for paying all charges that the member accounts incur. If you're an administrator of a management account and you have the appropriate permissions, you can view aggregated usage costs for Reserved Instance discounts and volume tiering for all member accounts. You can also view the charges that individual member accounts incur, because AWS creates a separate bill for each member account based on that account's usage. AWS also includes invoice summaries for each account in the management account invoice. During each billing period, AWS calculates your estimated charges several times each day so that you can track your costs as your organization incurs them. Your bill is not finalized until the beginning of the next month.

Note

Like member accounts, a management account can incur usage charges. However, as a best practice you shouldn't use the management account to run AWS services. An exception is for services and resources that are required to manage the organization itself. For example, as part of managing your consolidated billing you might create an S3 bucket in the management account to store AWS Cost and Usage Reports.

Pricing Tiers

Some AWS services are priced in *tiers*, which specify unit costs for defined amounts of AWS usage. As your usage increases, your usage crosses thresholds into new pricing tiers that specify lower unit costs for additional usage in a month. Your AWS usage is measured every month. To measure usage, AWS treats all accounts in an organization as a single account. Member accounts don't reach tier thresholds individually. Instead, all usage in the organization is aggregated for each service, which ensures faster access to lower-priced tiers. As each month begins, your service usage is reset to zero.

Each AWS service publishes its pricing information independently. You can access all individual pricing pages from the AWS Pricing page.

Examples of consolidated billing can be found on the Calculating Costs for Amazon S3 Standard Storage page.

The example at the above linked page shows how using consolidated billing in AWS Organizations helps lower the overall monthly cost of storage. If you calculate the cost for each member account separately, the total cost is $7,660 rather than $6,720. By aggregating the usage of the three accounts, you reach the lower-priced tiers sooner.

Resources

AWS Compute savings plans

AWS Reservations

AWS Well Architected Framework

How AWS Reserved instances are applied

Azure Savings Plans vs RIs

Azure Savings Plans

Azure Well Architected Framework