White Paper: **March 2023**

The Ultimate Guide to AWS Reserved Instances and Savings Plans

Table of contents

Introduction

This guide has been compiled to help customers better understand Amazon Web Services (AWS) Reserved Instances and Savings Plans and how they might benefit from them. The objectives of this guide are to explain what AWS Reserved Instances and Savings Plans are, how they differ, and how they can best be used to minimize AWS cloud costs.

This guide includes example scenarios in which Reserved Instances and Savings Plans should be used, considerations when committing to Reserved Instances and Savings Plans, and an FAQ section. For specific advice about whether AWS Reserved Instances and Savings Plans are appropriate for your business, contact one of our AWS experts.

A brief history of AWS Reserved Instances

When the Amazon Elastic Compute Cloud (EC2) service was launched in August 2006, there was only one size, one region and one pricing option: pay as you go, or "On Demand" as it later became known. Three years later, there were five EC2 instance sizes and three regions (two in the U.S. and one in Europe). Today, there are more than 500 different Amazon EC2 instances across five different families in 24 regions and 77 Availability Zones, with choices across storage, networking, operating systems, and even type of processor.

Businesses took advantage of the low barrier to entry with pay-as-you-go pricing and quickly adopted cloud infrastructure, foregoing the upfront costs of building their own data center. But organizations soon realized that the promise of reduced costs in the cloud is only achieved with adequate governance. As newly formed DevOps teams spun up resources unchecked, cloud bills soared.

AWS Reserved Instances (RIs) for EC2 were first introduced in 2009 to help AWS customers reduce their compute costs. A Reserved Instance is a billing discount on the rate paid for a resource or service. RIs offer significant savings, but you pay for the entire RI regardless of usage, so it's critical to plan for and monitor your usage to get the most out of the investment. To accomplish both allocation/chargeback and ensure your investments are well utilized, you will need a robust cloud financial management strategy and the right tools to support visibility, optimization and governance.

A Reserved Instance consists of eight components:

- Platform (e.g., Linux)
- Instance type (e.g., m4.large)
- Scope (e.g., regional or Availability Zone)
- Network (e.g., virtual private cloud or classic)
- Tenancy (e.g., dedicated)
- Term (typically one or three years)
- Type of reservation offering (e.g., no upfront)
- Class of reservation (e.g., Convertible or Standard)

AWS evaluates the available reservations and running instances on an hourly basis, and then applies reservations to usage. Each usage of an instance for the hour gets evaluated to determine if there is an applicable reservation to cover it. Because reservations are a pricing discount applied to any instance usage of a specific type (e.g., m4.large in us-east-1a running Linux), if you launch an instance that matches the example instance type, family, location and operating system, you will be billed at a discounted percentage rather than the base On-Demand amount.

Standard and Convertible Reserved Instances

There are two classes of Reserved Instances: Standard and Convertible. Both are available for one-year or three-year terms, with greater discounts on three-year commitments. Standard RIs can offer up to 72 percent off On-Demand rates, but the discount can only be applied to instances that match certain attributes of the Reserved Instance you purchased.

In September 2016, AWS introduced Convertible RIs, which provide customers with the flexibility to change an RI's instance type, OS, tenancy or payment option during its term. This gives users greater flexibility when workloads change, allowing them to make longer-term commitment, and offer up to 66 percent off On-Demand prices.

Standard RIs enable you to modify the Availability Zone, scope, networking type, and instance size (within the same instance type) of your Reserved Instance. Convertible RIs enable you to exchange one or more Convertible Reserved Instance for another Convertible Reserved Instance with a different configuration, including instance family, operating system, and tenancy.

Regional and zonal Reserved Instances

When you purchase a Reserved Instance, you determine the scope of the Reserved Instance. The scope is either regional or zonal. Regional RIs apply to instances in the specified region within the same family, no matter the size of the instance. Zonal RIs apply only to the specified Availability Zone (AZ), but they have the benefit of reserving capacity. When you purchase a Reserved Instance in a specific Availability Zone, it guarantees capacity in that AZ for your workload.

Table 1: Comparison of regional and zonal Reserved Instances[1]		
	Regional RIs	Zonal RIs
Availability Zone flexibility	Applies in any Availability Zone in the specified region	No Availability Zone flexibility; the Reserved Instance discount applies to instance usage in the specified Availability Zone only
Capacity reservation	No capacity reservation; a regional Reserved Instance does not provide a capacity reservation	Provides a capacity reservation in the specified Availability Zone
Instance size flexibility	Applies within the instance family, regardless of size; only supported on Amazon Linux/ Unix Reserved Instances with default tenancy	No instance size flexibility; the Reserved Instance discount applies to instance usage for the specified instance type and size only

Introduction of AWS Savings Plans

As RIs grew in popularity and the requirements of customers changed, the options—and the management overhead—multiplied. The management overhead was further complicated by the launches of new regions and new instance families while Reservations were active. AWS acknowledged that managing commitments at scale was becoming more complicated, particularly for customers operating in dynamic environments, and AWS Savings Plans were created.

While Savings Plans provide greater degrees of flexibility, they don't alleviate all the management burden of Reservations, and come with greater challenges in some areas. This white paper will explain what these challenges are, explore the consequences of failing to address them, and help you and your business get prepared to take advantage of and effectively manage AWS Reserved Instances and Savings Plans.

1. Amazon Web Services. "Amazon EC2 Reserved Instances and Other AWS Reservation Models." March 2021.

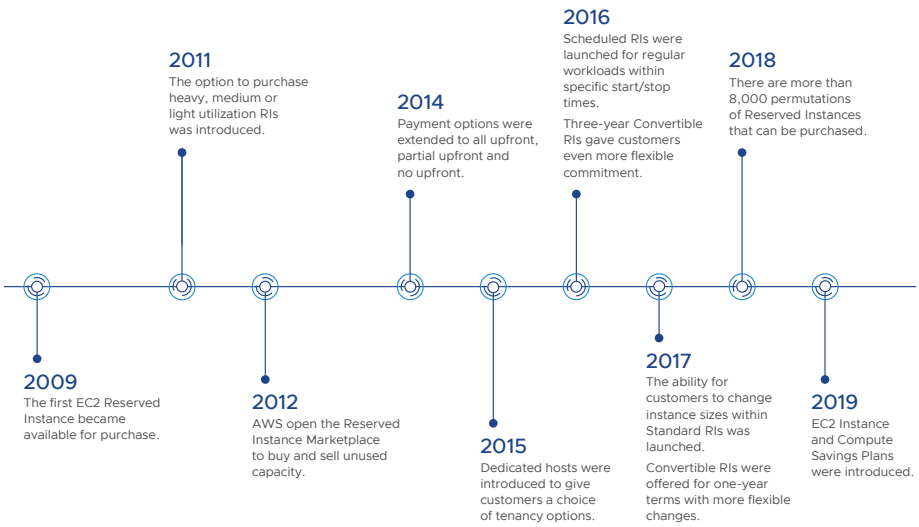

Figure 1: The history of Reserved Instances and Savings Plans.

Types of Savings Plans

In 2019 when Savings Plans were first introduced, there were two types: EC2 Instance Savings Plans and Compute Savings Plans. These types give customers a choice between maximizing the financial benefits of the discount program (by sacrificing flexibility) or maximizing flexibility while benefiting from a smaller discount. In this respect they are similar to Standard RIs and Convertible RIs, and it's no coincidence that the maximum Savings Plans discounts match those of the RI discount program.

EC2 Instance Savings Plans

EC2 Instance Savings Plans offer savings of up to 72 percent compared against On-Demand rates depending on the term of the commitment, payment option used, and instance family. This plan can only be applied to EC2 instances in a specific instance family in a specific region (e.g., M5 EC2 instances in North Virginia) but will continue to apply if instance sizes, operating systems, or tenancies are changed.

Compute Savings Plans

Compute Savings Plans offer more flexibility but a smaller discount—up to 66 percent compared against On-Demand rates. The plans can be applied to EC2 instances, AWS Fargate services, and AWS Lambda services in any region, across any instance family, and between services (for example, if container workloads running on EC2 are moved to the Fargate service).

There are advantages to committing to Compute Savings Plans as opposed to EC2 Instance Savings Plans inasmuch as they are the more flexible of the two options. Compute Savings Plans enable customers to purchase with greater confidence they will be able to take full advantage of the discount if their workload needs change over the term of the commitment.

Amazon SageMaker Savings Plans

In 2021, AWS announced Amazon SageMaker Savings Plans, offering up to 64 percent off On-Demand rates. Amazon SageMaker is a fully managed service for building, training and deploying machine learning (ML) models. With similar flexibility to Compute Savings Plans, Amazon SageMaker Savings Plans help maximize savings regardless of the ML instance family, size, or AWS region.

Table 2: Types of Savings Plans[2]

Plan	Maximum discount	Applies to	Flexible across
Compute Savings Plans	66 percent off On-Demand rates (as with Convertible RIs)	EC2, Fargate, Lambda	• Instance family (e.g., move from C5 to M5) • Region (e.g., change from EU Ireland to EU London) • Size (e.g., change from c5.xlarge to c5.large) • OS (e.g., change from Windows to Linux) • Tenancy (e.g., switch dedicated to default)
EC2 Instance Savings Plans	72 percent off On-Demand rates (as with Standard RIs)	EC2 instances in the selected family (e.g., C5 or M5) and region	• Size (e.g., move from c5.xlarge to c5.large) • OS (e.g., change from m5.xl Windows to m5.xl Linux) • Tenancy (e.g., modify m5.xl Dedicated to m5.xl Default)
Amazon SageMaker Savings Plans	64 percent off On-Demand rates	SageMaker	• Instance family (e.g., move from ml.m5 to ml.c5) • Region (e.g., change from EU Ireland to EU London) • Size (e.g., change from ml.c5.xlarge to ml.c5.large) • Component (e.g., migrate from Notebook to Training)

2. Amazon Web Services. "Savings Plans User Guide." Version 1.0. August 2020.

Similarities between Reserved Instances and Savings Plans

At a high level, there are a few similarities between RIs and Savings Plans. Both are commitment discount plans that give you a better rate on AWS resources in exchange for a commitment.

Three ways to pay

With all iterations of AWS Reserved Instances and Savings Plans, you have three payment options: all upfront, partial upfront, and no upfront. The more money paid upfront, the greater the discount.

Two commitment term options

All AWS Reserved Instances and Savings Plans can be purchased for either a one-year or a three-year commitment. The three-year commitment has higher discounts for On-Demand prices than the one-year commitment. Although variable terms can be found on the Amazon Reserved Instance Marketplace, new RIs from AWS come as a one-year or three-year commitment.

Use it or lose it

Unused RIs and Savings Plans do not accumulate. For example, if you commit to $10 per hour but only consume $8 of services in any hour, you will lose the remaining $2. Therefore, it's critical that you plan for and track usage of commitment discounts to minimize waste. However, if you're using linked or consolidated accounts, it's likely that the unused Savings Plan commitment would have floated to benefit another account.

Float

One major similarity between Reserved Instances and Savings Plans is their ability to float between linked accounts, with an affinity for the purchasing account. This means that, each hour, RIs and Savings Plans will first look in the purchasing account for instances that match the reservation. If no instances in the purchasing account are running that match the reservation characteristics, the RI or Savings Plan will then be applied to instances in linked accounts. Because multiple reservation types and instances can match, the selection of a reservation gives preference toward applying the lowest hourly rate first.

The concept of float is both a powerful feature and a source of confusion for customers. The confusion is often driven by customers purchasing RIs and Savings Plans for a specific purpose, only to find the cost benefit is applied elsewhere.

How Reservations and Savings Plans differ

The primary difference between RIs and Savings Plans is that, with Saving Plans, customers commit to a minimum dollar per hour spend rather than committing to a number of instances level of utilization.

Choose amount to commit with Partial Upfront Savings Plans

Another key difference between Savings Plans and Reservations is that Savings Plans allow you to decide how much money you want to put down upfront for a Partial Upfront Savings Plan, anywhere from 50 to 99 percent. While the savings rate doesn't change, it gives customers the opportunity to stretch a capital expense budget or potentially spend a use-it-or-lose-it budget at the end of a fiscal period.

Reservable resources and services

Although AWS recommends Savings Plans[3] over Reserved Instances[4] because they offer the same savings with additional flexibility, fewer resources and services can be covered with Savings Plans.

Currently, Savings Plans can only be purchased for the following compute resources[5]:

- Amazon EC2
- AWS Lambda
- AWS Fargate, including Amazon Elastic Container Service (ECS) and Amazon Elastic Kubernetes Service (EKS)
- Amazon SageMaker

Reservation models are available for[6]:

- Amazon EC2
- Amazon Relational Database Service (RDS)
- Amazon ElastiCache
- Amazon OpenSearch Service
- Amazon Redshift
- Amazon DynamoDB

If your business uses Reserved Instances to save money on the cost of these services, you will have to continue using Reserved Instances until Savings Plans are available.

Reselling on the AWS Marketplace

One limitation that exists with Savings Plans compared to Reserved Instances is that, at present, Savings Plans cannot be sold in the Reserved Instance Marketplace. If you overcommit to a Savings Plan, it's a financial obligation

3. Amazon Web Services. "Savings Plans FAQ."
4. Amazon Web Services. "Amazon EC2 Reserved Instances Pricing."
5. Amazon Web Services. "Savings Plans User Guide." Version 1.0. August 2020.
6. Amazon Web Services. "Amazon EC2 Reserved Instances and Other AWS Reservation Models." March 2021.

Although Savings Plans do not provide a capacity reservation, you can reserve capacity with On-Demand Reservations and apply Savings Plans discounts to them.

you are going to be stuck with for up to three years. It will be essential to have effective tools, such as VMware Tanzu CloudHealth, at your disposal to accurately forecast demand and make intelligent commitments.

Note that only Amazon EC2 Standard Reserved Instances can be sold in the Reserved Instance Marketplace. Amazon EC2 Convertible Reserved Instances cannot be sold. Reserved Instances for other AWS services, such as Amazon RDS and Amazon ElastiCache, cannot be sold either.

Further differences exist between RIs and Saving Plans, but due to the similarities between Standard RIs and EC2 Instance Savings Plans, and Convertible RIs and Compute Savings Plans, it's easier to explain the differences by comparing each pairing separately.

Standard RIs and EC2 Instance Saving Plans

Standard RIs and EC2 Instance Savings Plans are both family-specific and region-specific (e.g., M5 EC2 instances in North Virginia), but EC2 Instance Savings Plans can also be applied to different operating systems and tenancies, whereas customers currently need separate Standard RIs for each operating system or tenancy.

The situation regarding capacity reservations is slightly more complicated. With Standard RIs, customers have the option of assigning the RI to a specific Availability Zone to reserve capacity. This option doesn't exist in the Savings Plans discount program. However, it's possible to apply Savings Plans discounts to On-Demand Capacity Reservations[7] and indirectly get both the benefits of reserved capacity and a commitment discount.

Table 3: Comparison of Standard RIs and EC2 Instance Savings Plans

Type	Standard Regional RIs	Standard AZ Scoped RIs	EC2 Instance Savings Plans
Commitment unit	Per Instance	Per Instance	Dollars per hour
Geography	Region specific	Zone specific	Any region
Instance family	Fixed	Exchangeable	Any family
Instance size	Any size	Must modify	Any size
Tenancy	Fixed	Exchangeable	Any tenancy
Operating system	Fixed	Exchangeable	Any OS

7. Amazon Web Services. "Amazon EC2 FAQs."

Compute Savings Plans automatically apply to EC2, Lambda and Fargate (and ECS and EKS using Fargate). The automatic application of discounts significantly reduces the time needed to manage the commitment discount, compared to Convertible Reserved Instances.

Convertible RIs and Compute Saving Plans

Customers currently using Convertible RIs will likely benefit the most by switching to Compute Savings Plans because of the significant reduction in management overhead. Customers can exchange Convertible RIs for a different size, region, instance family, tenancy or operating system, but the process is manual.

With Compute Savings Plans, discounts are applied automatically across all components. The key word is "automatically." The automatic application of discounts to EC2, Fargate, Lambda and SageMaker services significantly reduces the management overhead. This factor alone should influence customers to choose AWS Savings Plans instead of RIs wherever they are available.

Table 4: Comparison of Convertible RIs and Compute Savings Plans

Type	Convertible Regional RIs	Convertible AZ Scoped RIs	Compute Savings Plans
Commitment unit	Per Instance	Per Instance	Dollars per hour
Geography	Region specific	Zone specific	Any region
Instance family	Exchangeable	Exchangeable	Any family
Instance size	Any size	Must modify	Any size
Tenancy	Exchangeable	Exchangeable	Any tenancy
Operating system	Exchangeable	Exchangeable	Any OS

Table 5: Comparing RIs and Savings Plans[8]				
	Savings Plans		Reserved Instances	
	Compute Savings Plans	EC2 Instance Savings Plans	Convertible RIs	Standard RIs
Savings over On Demand	Up to 66%	Up to 72%	Up to 66%	Up to 72%
Low price in exchange for monetary commitment	•	•		
Automatically applies pricing to any instance family	•			
Automatically applies pricing to any instance size	•	•	Regional convertible RIs and regional Standard RIs provide instance size flexibility	Regional convertible RIs and regional Standard RIs provide instance size flexibility
Automatically applies pricing to any tenancy or OS	•	•		
Automatically applies to Amazon ECS and Amazon EKS using AWS Fargate	•			
Automatically applies to AWS Lambda	•			
Automatically applies pricing across AWS regions	•			
1- and 3-year term length options	•	•	•	•
Sellable on the AWS Reserved Instance Marketplace				EC2 only

8. Amazon Web Services. "Savings Plans User Guide." Version 1.0. August 2020.

Making a commitment

One of the biggest mistakes that companies make when getting started with RI and Savings Plans management is failing to commit fast enough. Often, businesses think they need to rightsize all their existing infrastructure before committing to usage or spending on those resources.[9]

Just go for it

The reality is that a one-year term commitment will almost always break even after six months. This is when you can shut down an instance and still benefit from the reservation's pricing discount. For a three-year reservation, the break-even point usually occurs around nine months. All organizations that are concerned with their spend in the cloud need to be looking at RIs and Savings Plans as a first step toward reducing spend.

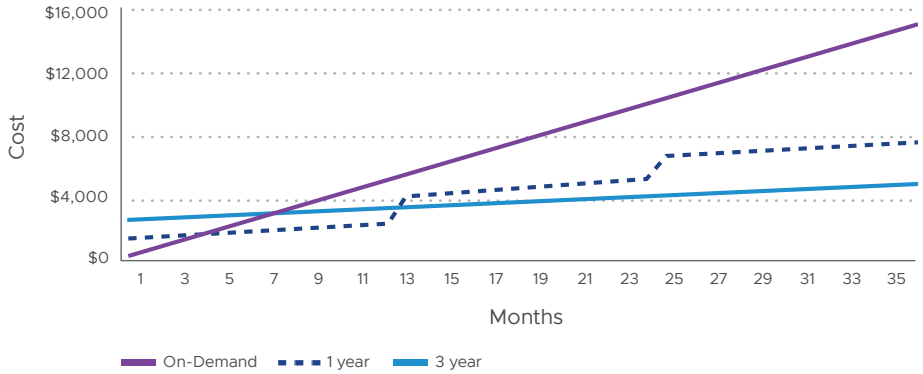

Figure 2: Costs for a partial upfront reservation of an r4.2xlarge instance in us-east-1 running Linux/Unix.

Calculating the break-even point

The payback period is used to calculate the exact number of months it would take before you see a price benefit, assuming 100 percent usage. This metric is invaluable for mitigating the risks of reservations by identifying how long you must actually use them before they break even.

The payback period is applied to partial and all upfront reservations. It's calculated by comparing the cash outlay for On-Demand usage and the proposed offering over each month in a term, and then identifying the month at which the cost for the On-Demand instance usage exceeds the cost for the reserved offering. There is no payback period for a no upfront reservation because they are less expensive than On-Demand immediately.

You can figure out your break-even point by calculating the savings compared to On-Demand pricing. For example, if you purchase a partial upfront reservation for an r4.2xlarge running Linux in us-east-1, the annualized cost of running it 24x7x365 On-Demand is $2,917.08 ($0.532 per hour x 8,760 hours in a year).

9. O'Reilly Media. "Cloud FinOps." J.R. Storment and Mike Fuller. 2020.

If the instance runs continuously throughout the year, and if you purchased a reservation for that machine, the cost would be $2,796.12. The difference is a 40 percent savings. That means you will break even after six months.

Modify and exchange reservations as needs change

When you exchange a Convertible RI, you effectively purchase a new reservation that has the same or greater cost than what you started with. The value of the Convertible RI is based on the total sum of all the payments you will make over the lifetime of the original RI.

RIs with a regional scope (either Standard or Convertible) automatically apply across all Availability Zones in a region. Regional RIs can also be instance size flexible, meaning they are automatically applied to any instance within the same family and region, regardless of size. Size flexibility is offered on Linux/Unix regional RIs at no additional cost.

If it's not already, continuously modifying your Availability Zone scoped reservations should be part of your overall RI management strategy to reap their cost and capacity reservation benefits. This is, after all, why you bought them.

The following example highlights a compelling reason for modifying your reservations. Let's assume you have two c5.4xlarge standard reservations, but four c5.2xlarge instances are being used 100 percent of the time, all in another Availability Zone in the same region. The On-Demand charges for each of the c5.2xlarge instances will cost you about $250 per month. If you modify and split your two c5.4xlarge reservations into four c5.2xlarge reservations, you will benefit immediately from the cost savings, all while ensuring that your prepaid reservation is being used.

The flexibility of Convertible RIs and Compute Savings Plans makes it less risky to commit to these discounts because they apply to a greater number of resources and services without any intervention, making them more likely to be fully utilized.

Purchases and modifications can be submitted through the AWS console, directly through the API, or automatically with Tanzu CloudHealth.

Layer RIs and Savings Plans in the same account

Most customers will have both reservations and Savings Plans in many of their accounts. Every hour, AWS will assess the possible discount programs that usage qualifies for and will apply compute discounts in the following order, from the least flexible to the most flexible purchases:

1. Standard Reserved Instances
2. Convertible Reserved Instances
3. EC2 Savings Plans
4. Compute Savings Plans

To best balance discount programs, it's necessary for customers to have visibility into and understand the current RI inventory, utilization and burn rate, and then layer Savings Plans onto RIs incrementally.

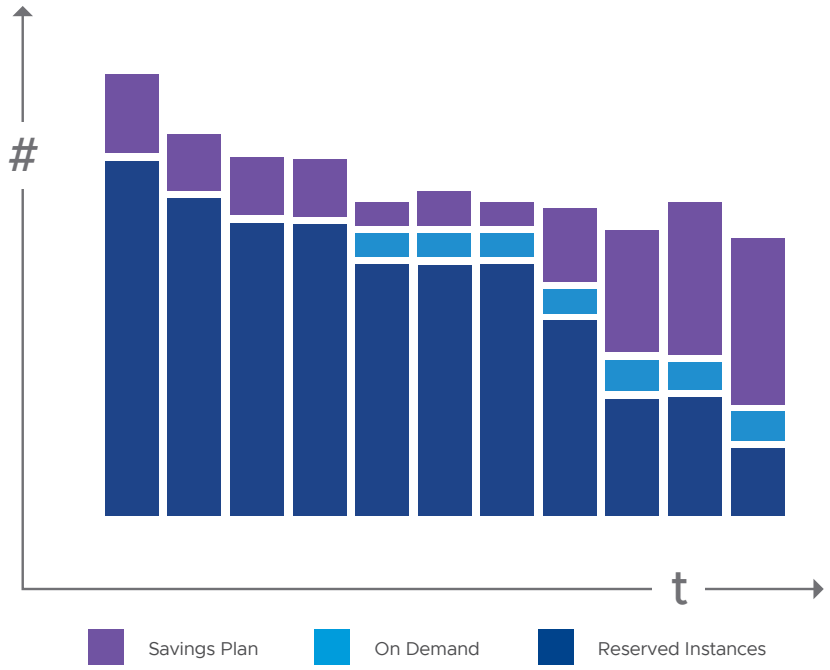

Figure 2: Laying an AWS Savings Plan on top of RI coverage can reduce AWS cloud costs.

Frequently asked questions

Q. Does AWS intend to replace RIs with Savings Plans?

A. AWS recommends purchasing Savings Plans as your RIs expire. Does this mean they will phase out RIs in favor of Savings Plans? No one knows, but both discount programs are expected to coexist for the foreseeable future as several high-cost services (e.g., Amazon RDS) do not yet have Savings Plans available.

Q. What's the biggest advantage of Savings Plans in comparison to Reserved Instances?

A. Different customers will find different things to like about AWS Savings Plans compared to RIs. For some, increased flexibility may be the biggest advantage. For example, the ability to apply Compute Savings Plans commits across multiple regions is a big advantage for large global enterprises. For other customers, the reduction in management overhead will be the biggest benefit.

Q. What's the biggest disadvantage of switching to Savings Plans?

A. One of the biggest disadvantages of Savings Plans is that they cannot be exchanged or sold on a secondary marketplace, such as the Reserved Instance Marketplace. This means you must take every step to ensure you're making the best commitment; otherwise, you run the risk of wasted Savings Plans.

Q. Can you reserve capacity when you commit to an AWS Savings Plan like you can with zonal RIs?

A. Unlike One-Zone Reserved Instances, there is no scope attribute with Savings Plans. However, customers that require reserved capacity can take advantage of On-Demand Capacity Reservations (ODCRs) and apply Savings Plans discounts to the ODCR deployments.

Q. Do the Savings Plans mean I don't have to rightsize EC2 instances anymore?

A. Optimization is just as important as it's always been. If a Savings Plan is applied to an overprovisioned EC2 instance, the discount is being wasted on services that aren't being used. Inasmuch as Savings Plans can help minimize AWS Cloud costs, you'll still have to take responsibility for optimization.

Conclusion

While Convertible RIs and Savings Plans provide greater flexibility, they don't eliminate the need to properly forecast your future requirements. Making commitments for future needs will always require visibility into your cloud infrastructure and workload requirements. Tanzu CloudHealth is here to help guide you toward the most optimal commitments for your business. Reach out to our Strategic Savings Desk—a complimentary service for Tanzu CloudHealth customers—at cht-savingsdesk@groups.vmware.com. Our team will be happy to guide you through the discount programs and explain the benefits of using Tanzu CloudHealth to manage your AWS environment.

Item No: 1703106aq-wp-ult-gde-aws-ris-sav-plans-uslet 3/23