INTRODUCTION TO AWS RESERVED INSTANCES

Cloud computing gained traction in the SMB market segment with its promised value proposition to replace high CapEx with affordable ongoing OpEx. SMBs embraced the cloud movement to take advantage of the cost variability and were able to focus resource investments on building the business instead of building the IT infrastructure. For organizations that scale rapidly, the Total Cost of Ownership (TCO) cloud may be hardly more attractive than the traditional IT service delivery model for use cases that require predictable compute power for prolonged time period.

In such cases, the Pay-as-You-Go pricing model presents limited cost advantages and often prevents organizations from committing to long-term partnerships with cloud vendors. On the other hand, on-premise infrastructure deployments offer significant cost savings over the long term.

For cloud vendors, there is a provision as well as incentive to facilitate long-term partnership with customers by promising a lower TCO. AWS emerged as the leading cloud vendor by identifying this opportunity and rolling out a new pricing model that rewards long-term commitments to AWS infrastructure services. It is called AWS Reserved Instances.

(This tutorial is part of our AWS Guide. Use the right-hand menu to navigate.)

Standard Reserve Instances: Think of AWS Reserved Instances as a discount coupon. You purchase the discount coupon applicable to instances with specific pre-defined attributes. You purchase these instances and reserve them for future use – hence the term Reserved Instances. Whenever your IT workloads trigger operations on those instances, AWS applies the discount of up to 75 percent of the regular On-Demand Instance pricing. The only difference from On-Demand Instances is that customers need to purchase the Reserve Instances with a term commitment of 1 to 3 year by paying a percentage of the cost upfront. AWS offers the flexibility to change the Availability Zone,

size and network type of AWS Standard RIs after the purchase. Standard RIs offer the most practical viability for workloads with long-term steady-state usage. Once purchased, customers cannot terminate their contracts for RIs. They can however, modify, sell or exchange the RIs.

Convertible Reserved Instances: Since Standard RIs offer low flexibility in terms of changing attributes of the instances, customers can opt for Convertible RIs that allow the flexibility to change instance families, platforms, tenancy and payment options following a purchase. AWS offers a maximum of 54 percent discount on Convertible RIs based on the upfront payment and commitment terms ranging 1 to 3 years. Convertible RIs can be exchanged with other Convertible RIs as long as the exchange is of equal or a greater dollar value. More terms and conditions are available here. As of now, customers cannot sell their unused Convertible RIs, although this feature is in the works and will roll out soon according to the official source.

Scheduled Reserved Instances: For customers with recurring periodic workload processing requirements, AWS offers the ability to reserve instances during a specific period of the day, week or month on a regular basis. As a result, AWS can offer a 5-10 percent discount on the reserved instances if customers choose to run the infrastructure resources during off-peak hours. Unlike Spot Instances that are purchased to run at a specific time period and run automatically once the schedule has reached, customers can choose to take advantage of the Scheduled RI pricing when their instance usages matches the necessary criteria. AWS doesn't require a multi-year commitment for Scheduled RIs and the percentage savings is lowest among the three Reserve Instance offering classes. Read this AWS resource for a detailed description of the AWS Scheduled RI lifecycle.

For all three offering classes, the discount is primarily determined by the term commitment and percentage of the cost paid upfront. AWS offers three payment options: all upfront, partial upfront and no upfront. The more you pay upfront and the longer you commit to the resources, the greater is the discount as compared to on-demand instance pricing options.

Think Beyond Cost Savings

AWS Reserved Instances isn't all about cost savings. In fact, customers can use their ability to reserve AWS instances based on pre-defined criteria to maximize the value potential of their upfront or long-term investments into AWS resources. Consider the ability to reserve a specific AWS Region and Availability Zone for your workload requirements. You can choose the data center locations based on computing performance in terms of low latency, as well as other factors such as cost, regulatory compliance and geographic presence. Reserving instances in a specific Availability Zone means that your AWS investments for known and steady-state workload requirements are consistent with your decisions based on those factors.

Reservation of capacity and Availability Zones are also useful to accommodate usage spikes of mission-critical workloads to accommodate sudden increase in network traffic or app popularity without compromising end-user experience. Similarly, customers can choose multiple Availability Zones to dynamically distribute the workloads as a failsafe against location-specific outages or natural disasters. By reserving instances that are the most cost effective, technically feasible and compliant to stringent regulations, customers no longer need to reevaluate their AWS investments based on changing pricing patterns and unforeseen limitations in purchasing on-demand instances.

A Word of Caution

AWS Reserved Instances offer great value to customers with a specific set of resource and capacity requirements. However, AWS RIs are not suitable for all customers with steady-state usage needs. For instance, committing to RIs for a period of 1-3 years means that you will run the same hardware resources over the term commitment. As a result, you won't take advantage of improved compute power at a lower cost as AWS upgrades the hardware offerings for its on-demand instances. Secondly, the RI offering classes limit the instance attributes that are available at a discount. In the real-world, these attributes may not align with the evolving and unpredictable workload requirements that your organization will face over the long term. Finally, you need to plan for RIs well ahead of schedule. Once the RI period ends, you will need to immediately purchase the next set of RIs or risk investing in on-demand instances at a higher cost and with potential limitations on Regions and Availability Zones based on compliance, pricing and available capacity, among other key factors.

One strategy to address these concerns is to purchase RIs on a minimal viable and iterative basis: adding frequency to RI investments allow you to purchase RIs based on changing instance profiles as well as improving AWS pricing and hardware upgrades. Additionally, you can strategize RI purchases on an ongoing basis while aligning the investments with technical and business goals of the organization.