The Ultimate Guide to AWS EC2 Reserved Instances

1 UNDERSTANDING RIS

IT'S ALL VERY SIMPLE, REALLY.

EC2 Instances are priced in three ways: On Demand, where you pay full price; Spot, where your price is determined by market availability; and Reserved, where you receive a discount based on a commitment to pay for a certain amount of EC2 usage, whether you use the hours or not. When you purchase an EC2 reservation, you're essentially purchasing a coupon.

AWS Reserved Instances allow you to make a commitment to utilize specific instance types in return for a discount on your compute costs. In addition, some RIs have a capacity reservation that guarantees your ability to run an instance of this type in the future.

The benefits? Savings and guaranteed capacity (for some RI types). A reservation consists of 6 components:

- Platform (e.g. Linux)
- Instance type (e.g. m3.large)
- Scope (e.g. Regional or Availability Zone)
- Tenancy (e.g. the default, virtual private cloud, or classic)
- Term (typically 1 or 3 years, although variable terms can be found on the AWS Reserved Instance Marketplace)
- Type of reservation offering (e.g. No Upfront)
- Class of reservation (e.g. Convertible)

1 *UNDERSTANDING RIS*

Since reservations are a pricing discount applied to any instance usage of a specific type (e.g. m3.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. In other words, by giving AWS a "head's up," they give you a discount.

Amazon evaluates the available reservations and running instances on an hourly basis, and then randomly 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.

Since multiple reservation types (upfront amount and reservation term) and instance usage can match, the selection of a reservation gives preference toward applying the lowest hourly rate first. It's also worth noting that reservations have an affinity toward the account in which they were purchased.

The randomized approach of RIs is both a powerful feature and a source of constant confusion for customers. The confusion is often driven by customers purchasing RIs for a specific purpose (e.g. the marketing department), only to find its cost benefit is applied elsewhere.

CLOUDOLOGIST TIP: RI BASICS

- *Make a commitment to utilize specific instance types in return for a discount on your compute costs.*
- *Calculate your payback period exact number of months at 100% usage before there is a price benefit.*
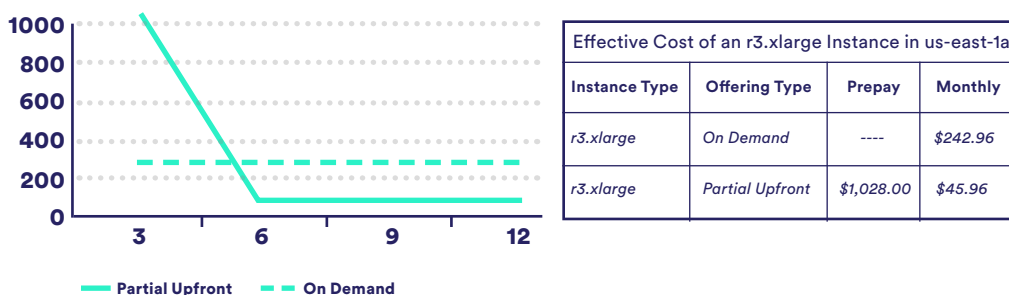- *Scale your cloud without scaling your IT workforce.*

2

COMMITTING

DON'T BE AFRAID TO JUST GO FOR IT.

Here's the reality: a 1-year term reservation will almost always break even after 6 months. This is when you can shut down an instance and still benefit from the reservation's pricing discount. For a 3-year reservation, the break-even point usually occurs around 9 months. All organizations who are concerned with their spend in the cloud need to be looking at RIs.

We see that an r3.xlarge in the us-east-1a region will cost 81% less per month with a 1-year partial upfront reservation.

Effective Cost of an r3.xlarge Instance in us-east-1a			
Instance Type	Offering Type	Prepay	Monthly
r3.xlarge	On Demand	----	$242.96
r3.xlarge	Partial Upfront	$1,028.00	$45.96

Partial Upfront   **On Demand**

We use what's called the "payback period" to calculate the exact number of months it would take before you see a price benefit, assuming 100% 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, since they are less expensive than On Demand immediately.

2

COMMITTING

(CONTINUED)

You can calculate your break-even point, by calculating the savings compared to On Demand pricing. For example, if you purchase a Partial Upfront Reservation for a r3.xlarge running Linux in us-east-1, the annualized cost of running it 24×7×365 On Demand is $2,917.08 ($0.333 per hour x 8760 hours in a year).

If the instance runs continuously throughout the year, and if you had purchased a reservation for that machine, the cost would be $1576.80 ($0.180 per hour * 8760 hours in a year). The difference is a 46% savings. That means you will break-even after 168 days or about 5.5 months, approximately 46% of the year.

If you've purchased RIs in either the consolidated account or one linked to the consolidated billing account, and there is no instance usage in a given hour in this purchase account to utilize the reservation, the reservation can be applied to matching instance usage in any other linked account within the consolidated bill.
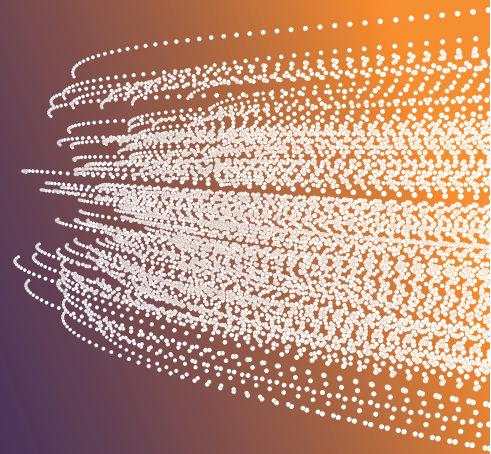
Now this is where things can get tricky. Many organizations link one or more accounts together into a consolidated bill. Doing this unleashes one of the more powerful and, in some cases, confusing behaviors of RIs - their ability to "float" across accounts.

By default, reservations have an affinity for the account in which they were purchased, so the "float" will only occur if there is no instance usage within its account that can take advantage of the reservation. While the price reduction benefit of RIs "float" across accounts within a consolidated bill, it's important to note that the capacity reservation does not. So if you have an available reservation in Account A, Amazon will not guarantee that you can launch an equivalent instance in Account B, even if these accounts are linked into the same consolidated billing account.

Regardless, a cursory view of your existing usage should give you a pretty good idea of your potential savings. Now you just have to go for it.

KEY QUESTIONS TO ASK

1. What percentage of this group do I expect will be running one year from now? Three years from now?

2. How likely are the instances in this group to stay within their current region? Ignore availability zone for now.

3. How likely am I to change to instance type family for the instances in this group (e.g. switching from m1 to m3)?

3

GET INVOLVED

RIS PUT YOU BACK IN CONTROL, ENJOY IT.

Amazon has different costs for different types of images (e.g. Linux, Windows) that can be launched and each type has different pricing. At any reasonable scale, purchasing RIs on a per instance basis will be almost certainly unmanageable, so group your instances based on one or more topics (e.g. environments, function, application) so that you can evaluate the cost by group.

Once you have grouped your infrastructure, focus on the most expensive group first (e.g. MongoDB servers). Since the RIs are really targeted at "always-on" infrastructure, you can choose to not evaluate groups whose infrastructure is only on < 65% of the time.

Before making a purchase you will also need to identify where your reservations will live within each group you want to purchase. Your decision really comes down to one of simplifying the purchase versus maximizing the cost and capacity benefit of the purchase. The general best practice is to purchase reservations where you have specific usage. AWS makes it easy to modify your entire reservation or just a subset of them. Standard Reservations can be modified in the following ways:

- Switching Regional and Availability Zone (AZ) Scope.
- Switching between Availability Zones for reservations scoped to a specific zone within the same region.
- Switching between Classic EC2 and Virtual Private Cloud.
- Switching the instance size within the same family.

Convertible Reservations are much more flexible than Standard Reservations. You can exchange them in all the ways you can for a Standard Reservation, but in addition you may:

- Exchange for a new instance type
- Exchange for a new operating system
- Exchange for a different tenancy
- Exchange for a different instance size

When you exchange a Convertible Reservation, 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.

Regional RIs are EC2 Reserved Instances (either Standard or Convertible) which automatically apply across all Availability Zones in a region. As of March 2017, Regional RIs can also be 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 and is available at no additional cost.

If it's not already, continuously modifying your AZ scoped reservations should be a part of your overall RI management strategy in order 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 2 c3.4xlarge standard reservations with no matching on demand usage but 4 c3.2xlarge instances being used 100% of the time, all in the in eu-central-1 Region. The On Demand monthly charges for each of the c2.2xlarge instances will cost you $752.95 per month. If you modify and split your 2 c3.4xlarge reservations into 4 c3.4xlarge instances, you will benefit immediately from the cost savings, all while ensuring that your prepaid reservation is being used. Modifications, like purchases, can be submitted through the AWS console, directly through the API, or automatically with CloudHealth.

4 *AUTOMATION*

SAVE MONEY WHILE YOU SLEEP

The goal of any scaling organization is to identify patterns in usage that translate into time- and money-saving efficiencies. In the RI world, those efficiencies are the product of automations. But when you're dealing with billions of data points across multiple systems, where do you begin?

Using CloudHealth, you can easily model a purchase for specific accounts, regions (e.g. us-east), instance types (e.g. m3 instance types), functions (e.g. Elasticsearch Clusters) or even business groups (e.g. marketing department). All those billions of data points are tracked, organized, and analyzed instantly. You can view your potential savings and upfront price, new reservation rate, and, perhaps most importantly, the payback period.

While the recommendations provided from CloudHealth will be the optimum for the settings you have configured, you're still in complete control. For example, if CloudHealth recommends purchasing 10 Partial Upfront reservations for m3.large usage in us-east-1a for Linux, you can choose to purchase only 7 based on your organizational knowledge of future instance usage (e.g. you expect a reduction in load that affects this instance type).

Whether you're actively pursuing Reserved Instances as a cost-saving measure, or just beginning to explore it as a possibility, CloudHealth can help.