



# Rightsizing

## Technical Brief

## Summary

As organizations adopt the cloud, a typical approach is to overprovision resources. This enables operational teams to ensure that application performance and customer experience doesn't suffer, as well as accelerate migration of resources from the datacenter. Overprovisioning follows the datacenter mindset where additional resources were not readily available and procurement incurred capital expenditure (CapEx). However, the pay-as-you-go on-demand cloud model results in operational costs and requires a different mindset to resource utilization. Whether they're undergoing a lift-and-shift or a cloud native development, once organizations are comfortable running applications in the cloud, optimization becomes a critical enabler of continued success in their cloud deployment.

## Understanding the Challenge

Cloud resources are elastic, scalable and provisioned on-demand. Cloud providers offer a vast array of infrastructure options and services to choose from. Organizations can deploy infrastructure and services globally across regions and zones with the click of a button or a simple API call. Such elasticity and flexibility can also be daunting, however. According to Gartner, deploying an AWS EC2 instance can involve over 1.7 million combinations or decision points – an intimidating task.

Since there are so many variables for deploying instances, operators must be able to resize resources according to the use for each. Let's take the case of compute. Operators have the ability to choose between bare metal, general purpose, compute, memory, storage and GPU optimized compute instances. Rightsizing is not limited to compute only – it also extends to block storage, object storage, databases and in-memory datastores. As you can see, rightsizing is a multidimensional problem and each attribute adds to the complexity and decisions required to be made. The dimensions for consideration include:

- Type
- Size
- Cost

As a result, it is virtually impossible for humans to make effective decisions without extensive data collection and analysis. Since an underprovisioned instance can result in degraded application performance and user experience, which in turn results in negative business

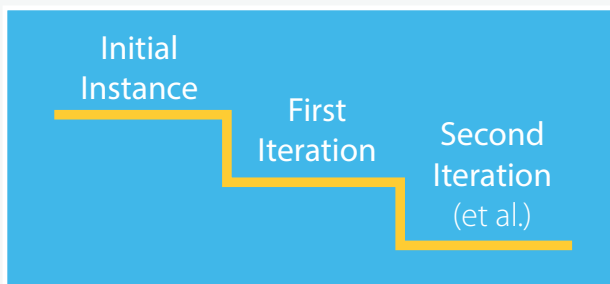
impact, it is important to make the best rightsizing decision. The possibility of making a wrong decision can handicap organizations' rightsizing decisions and their ability to optimize costs and deliver significant savings.

In the case of manual intervention, rightsizing is an iterative process where the right instance is determined by trial and error. Cloudability leverages your organization's historical data, nearly eight years of cohort data and best practices, performance metrics and usage data to deliver optimal rightsizing recommendations. The effort for rightsizing is opportunity cost – the cost and time for manual intervention. Cloudability is the solution that does the heavy lifting and delivers cost savings opportunities for your environment.

## Cloudability's Unique Perspective

### Without Cloudability (Trial and Error)

Typically in the same family with possibly no RI consideration

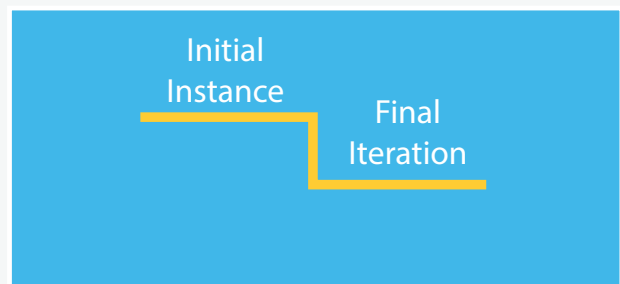


Possible reasons for iteration:

- Performance was hit
- Didn't save enough money
- Got the wrong shape

### With Cloudability (Data Driven)

Best option across families with RI consideration and known risk/savings



Let's dig deeper into how Cloudability's unique rightsizing capabilities help achieve your operational and business goals with confidence. Once set up, Cloudability starts collecting utilization metrics for supported infrastructure and services running across multiple cloud providers.

The first challenge in a rightsizing operation is to figure out the relevant utilization metrics for each instance type, the characteristics of that instance and the duration for analysis. Cloudability leverages nearly eight years of cloud operations experience optimizing the biggest cloud users across the globe. As a result of this experience, we've built rich machine learning algorithms that leverage the right input metrics for providing these recommendations. Our best practices and data suggest that utilization metrics collected over a 10-day period are the right duration for determining optimal rightsizing options. This duration ensures that Cloudability has enough data available before providing recommendations.

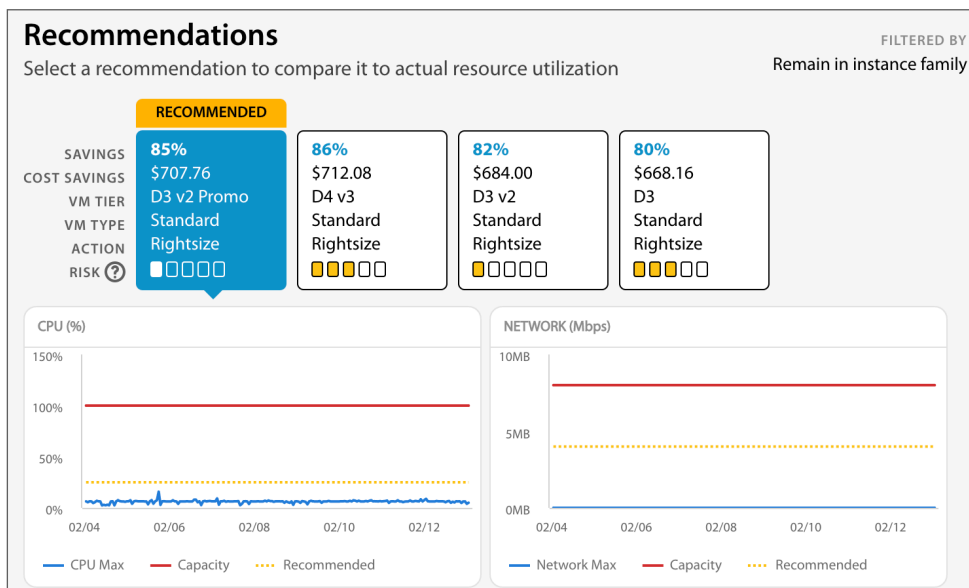
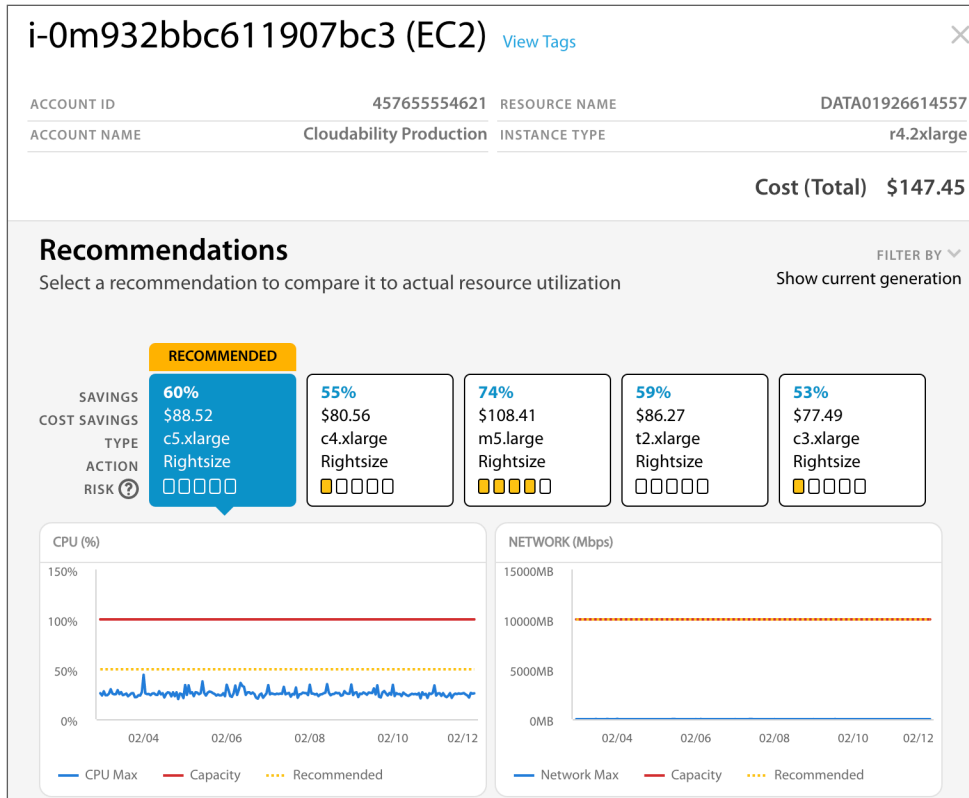
When determining rightsizing, two key attributes are at play: risk and savings. Generally these are inverse to each other - i.e., with a higher acceptance of risk, there are more options for significant savings. Calculating potential savings is fairly straightforward (even if there's some decent number crunching); it's being able to zone in on valid resource types and finding a way to represent/mitigate risk where the rubber hits the road.

## How Cloudability Delivers Rightsizing

Cloudability is delivering game-changing capabilities for rightsizing by:

- Taking into account the entire time series for your resources. We create a statistical model around average and peak utilization periods.
- Having a massive data pool to work with. We have confidence in the performance characteristics across instance families/types (even between regions). We use this information to determine valid instance types for your workload.
- Using proprietary algorithms to rank risk between these different options. This includes evaluating the likelihood of clipping and how much headroom you are likely to have.
- Most importantly, giving you a graphical representation to assess this risk by enabling you to overlay your workload on top of the most applicable instance types in your region. Here we are talking about representing the entire time series, and being able to visually inspect your headroom and potential for clipping. This is across all four key metrics of CPU utilization, disk (for instance types with local disk), memory and bandwidth.

Being able to cycle through these recommendations and simulate resource utilization ahead of actioning is game changing for rightsizing initiatives. Operations teams can analyze the risk/savings profile for each rightsizing option suggested. Selecting a recommendation provides the anticipated utilization profile of that option if the resource was to be resized, as shown in the figures below.



As shown in the image below, each Cloudability rightsizing recommendation includes cost savings that can be achieved. We recommend starting with the instances that can deliver maximum savings.

The screenshot displays the Cloudability Rightsizing interface. At the top, there are navigation tabs for 'Analyze', 'Optimize', and 'Predict'. Below this, the 'Rightsizing' section is titled, followed by a subtitle: 'A ranked list of underutilized resources based on data for the last 10 days. Learn more about rightsizing in our Knowledge Base.' The interface includes filters for 'AWS' and 'AZURE', and a sub-filter for 'EC2'. The main table lists several EC2 instances with the following columns: Resource ID, Resource Name, Account Name, Idle, Cost (Total), Current, New, Action, and Cost Savings. Each row also has a 'Details' button.

Resource ID	Resource Name	Account Name	Idle	Cost (Total)	Current	New	Action	Cost Savings
i-0b234e...	production-hbase-phoe...	AWS Production	0	\$434.11	i3.8xlarge	x1e.xlarge	Rightsize	\$233.95
i-0d6db...	production-hbase-phoe...	AWS Production	0	\$434.11	i3.8xlarge	x1e.xlarge	Rightsize	\$233.95
i-0fb27...	production-hbase-phoe...	AWS Production	0	\$434.11	i3.8xlarge	x1e.xlarge	Rightsize	\$233.95
i-04b5e...	staging-ecs-cluster-inst...	AWS Production	0	\$575.39	r3.8xlarge	c3.8xlarge	Rightsize	\$172.19
i-0abf9c...	production-hbase-phoe...	AWS Production	0	\$450.56	i3.8xlarge	i3.4xlarge	Rightsize	\$151.04
i-0c355...	production-hbase-phoe...	AWS Production	0	\$436.82	i3.8xlarge	i3.4xlarge	Rightsize	\$137.30
i-06760...	production-hbase-phoe...	AWS Production	0	\$434.11	i3.8xlarge	i3.4xlarge	Rightsize	\$134.59

Another powerful capability in Cloudability is filtering the information by views. With proper tagging and billing account structures, this information can be meaningfully segmented for the responsible teams, making it easy for them to take action. Cloudability provides a prioritized list of savings opportunities for rightsizing, enabling organizations to undertake the effort with confidence. Cloudability rightsizing recommendations span across multiple products for each cloud vendor, thereby increasing your surface area for savings.

For additional information, please visit us at [www.cloudability.com](http://www.cloudability.com) or [sign up for a free trial](#).

Cloudability is a multi-cloud True Cost™ management platform that delivers precise, accurate and timely cloud financials to enable visibility, optimization and forecasting, empowering customers to operate their cloud with certainty, agility and cutting-edge strategy.