



Persistent

Top Considerations for Commitment-based Cloud Pricing Models



Viraj Nandkarni
Chief Architect, Cloud Practice
Persistent

Commitment-based pricing models can offer strategic workarounds to reduce cloud spending. With FinOps gaining traction, commitment-based pricing models help enterprises align cloud expenditures more closely to their requirements and long-term business objectives.

However, taking a plunge toward a commitment-based pricing model with a less-than-optimal commitment strategy can fail to deliver the intended cost savings. Many organizations continue to be in the analysis-paralysis phase, unwilling to commit even when it makes more sense to do so and end up incurring higher cloud infrastructure costs.

It is important to compare the commitment-based model with the four most popular cloud pricing models. They all come with riders that must be understood and contextualized to cloud needs for real cost savings. We summarize these common cloud pricing models and their riders here.

Free Tier

- \ Free access
- \ Limited duration & usage
- \ Free credit limit possible services / features of a service

Pay-as-you-go

- \ Available for all services
- \ No upfront commitments
- \ Highest flexibility
- \ Most expensive model

Spot Pricing

- \ Discount spare capacity
- \ Usage or duration limits
- \ Not SLA-governed
- \ Not suitable for all use cases

Commitment-Based

- \ Upfront payments
- \ Commitment of one or three years to be made
- \ Discounts offered vary based on:
 - Commitment duration
 - Payment terms
 - Commitment criteria
 - Instance family

Two caveats before proceeding:

- 1 All options discussed below may only be available with some public cloud providers.
- 2 Cloud pricing is a moving target. An option available today may not be available a year down the line. Similarly, a better pricing model can come up in the future.

Key Considerations for Commitment-based Pricing

Do not over-assume savings percentages.

While planning to shift some pricing from pay-as-you-go to commitment-based models, calculating the real savings could be confusing. Since this is a move from a full OpEx to a part CapEx-based pricing,

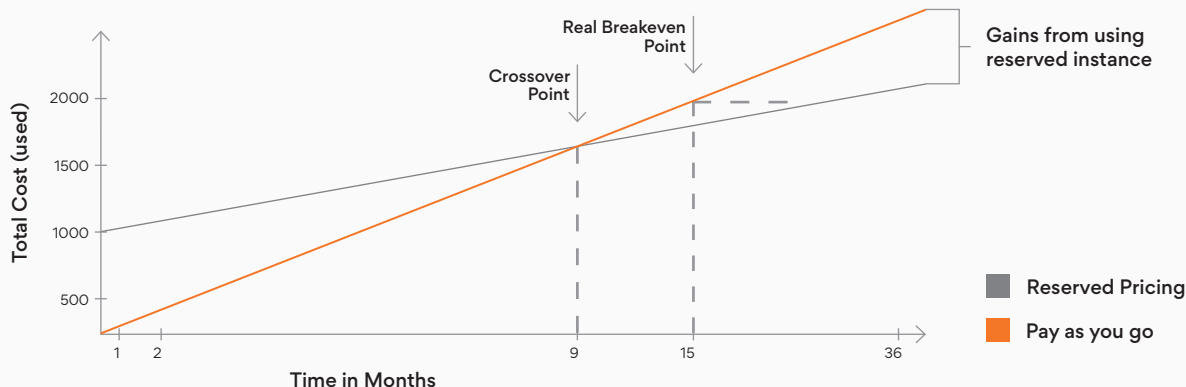
with a one- or three-year commitment term, it is important to understand what the stakeholders are willing to commit to and calculate savings within those constraints.

We vouch for a pragmatic approach that does not over-assume the savings percentage. Here's why:

- 1 Savings above 50% are normally for larger commitment terms and with larger upfront payments. Organizations have different risk appetites; some may be willing to commit to a large payment upfront and for a longer duration, whereas others may not.
- 2 Organizations may be evaluating cloud providers and unwilling to commit to one.
- 3 Since larger savings come with inflexible options, an organization with goals such as legacy modernization using containers or serverless architectures may be constrained to use more flexible models such as the savings plan compute option, which provides lesser discounts than the restricted commitment option.

Understand the crossover point versus the “real” break-even point

An important calculation to identify savings using a commitment-based model versus pay-as-you-go pricing, also known as the break-even point. Consider reserved pricing for a m5. xlarge instance **running under shared tenancy with Amazon Linux installed and runs in US East (Ohio)** with a three-year commitment with partial upfront payment.



In this case, the pay-as-you-go pricing per month (100% usage) is ~ \$140, and it catches up with the reserved pricing in about nine months. ($140 \times 9 = 1260$ for pay-as-you-go, $1009 + 27.4 \times 9 = 1252$ for reserved instance).

This is the crossover point where the total pay-as-you-go pricing at that time matches the total reserved instance pricing. After this, the pay-as-you-go pricing will cost more.

The key thing to remember is that reserved instance pricing is still levied as per commitment terms,

so if the instance is not used from this point onwards, organizations incur a loss in the case of reserved pricing versus pay-as-you-go.

A better measure is to determine when the pay-as-you-go pricing exceeds the total reserved instance pricing for the commitment period, which is $1009 + 27.4 \times 36 = \1995 . This figure will be reached for pay-as-you-go pricing just after the 14th month.

This is the “real” breakeven point because after this, the 3-year cost is already accounted for, and any usage post the 14th month will always be a saving.

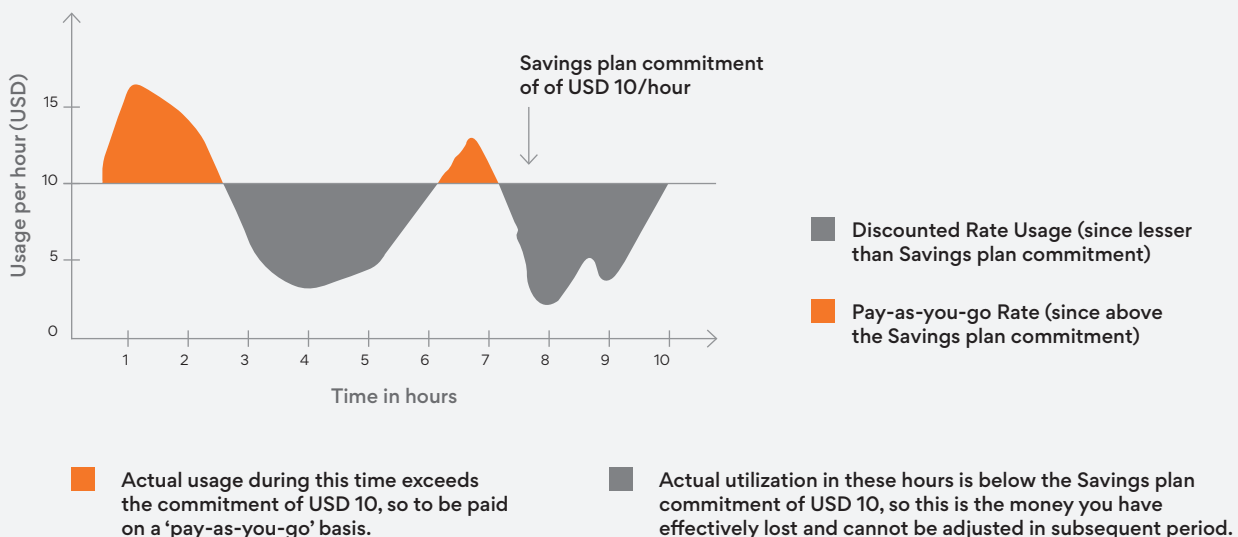
Hourly commitment and monthly pricing

A significant difference between savings plan pricing and reserved instance pricing is that discounts in savings plan are calculated on hourly usage whereas in the latter, it is averaged out on monthly usage. While this may not seem a big deal, it could very well be in case of large fluctuations in instance usage, especially in a savings plan.

With a savings plan, organizations commit to a fixed hourly usage, say, \$10 per hour. If the usage amounted to \$17 in hour X, the excess of \$7 would be

billed at pay-as-you-go cost. Similarly, if the usage in an hour Y amounts to \$2, then this lesser usage will not be adjusted against excess usage in the future.

This is effectively a loss of \$8. Hence, while a savings plan can provide more flexible options, calculations of how much to commit can be far trickier here than in the case of reserved instances where organizations pay a fixed amount for the entire month irrespective of usage hours. Therefore, savings plan is not always the best choice for all types of workloads.



Making short-term commitments

When clients are ready for a commitment, they could prefer shorter durations compared to the minimum commitment duration of one year offered by cloud providers. In this case, exploring the reserved instance marketplace first is advised instead of directly purchasing the reservations or savings plan commitments. Organizations should check if there is a seller willing to offload reservations with a limited commitment duration. Tools are available in the market that provide functionality to monitor

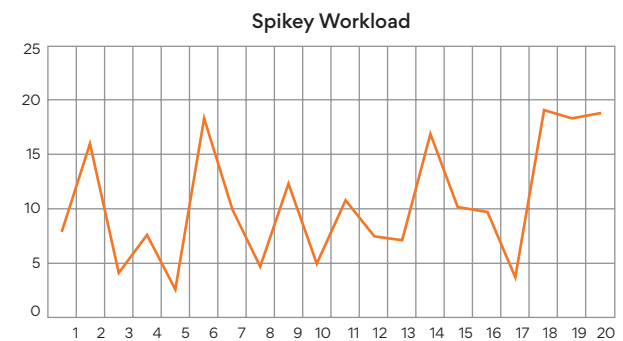
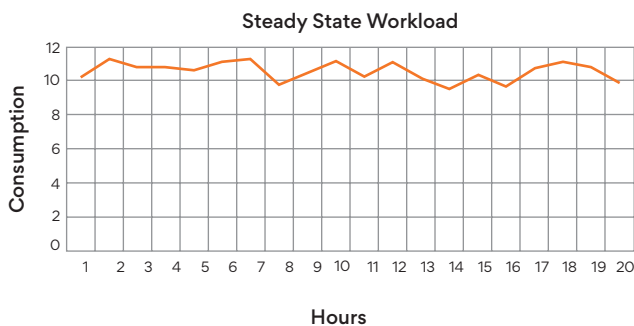
reservation opportunities actively. Feel free to explore these options, as doing this manually can be daunting. If this works, organizations receive discounted pricing for shorter-term commitments.

However, this option can only sometimes be relied upon. The reserved marketplace may not offer the right instance option when you need it, or for all commitment types, e.g., Amazon Web Services (AWS) allows selling only standard reserved instances.

A savings plan is not always the right option

A broad range of computing services can be utilized under the savings plan umbrella, with the flexibility of the different instance types being leveraged.

However, the savings plan has some limitations versus reserved instances, so it is **not always** appropriate. As highlighted above, some usage patterns could be more expensive under the savings plan. The figures below are sample examples of steady state and spikey workloads.



A savings plan, once committed to, can be canceled, and it cannot be sold on the marketplace, while reserved instances of certain types can be sold in the marketplace. Microsoft Azure allows canceling reserved instances. Azure also allows trading an existing reservation for a savings plan but not vice-versa.

Besides this, the Savings plan is majorly restricted to compute services as of now, whereas cloud providers such as AWS extend the reservation model to other services, including Relational Database Service, DynamoDB, and Redshift. This may change, and the savings plan could be extended to support other services.

Usage, Pricing Changes, and New Instance Offerings

To round off savings calculations, let us understand a few important things about the pay-as-you-go model.

Usage Assumption

One of the assumptions clients make to simplify savings between pay-as-you-go and commitment-based models is that instances are utilized 24 / 7 for 12 months. This may be the case; however, under the pay-as-you-go model, there is no compulsion to keep the instances running all the time. But with commitment-based pricing, organizations pay whether the instances are running or not, 24 / 7.

Hence, if the usage of instance(s) being moved over to commitment-based models was lesser in case of pay-as-you-go model, this should be accounted for in the savings calculations by considering pay-as-you-go usage at actuals but commitment-based usage at 100%. This will help organizations identify the actual cost savings versus what was calculated by assuming a 100% utilization.

As an example, in the “real” break-even point example provided earlier, if the pay-as-you-go utilization was considered at 75% and the same specs were used, then the real break-even point will be hit after 24 months (2 years) as compared to after 14 months when the pay-as-you-go utilization was considered as 100%.

Some cloud providers, such as AWS, provide options like scheduled reserved instances to counter this to a certain extent, but they have issues. These include minimum commitment hours per week, month, and year. Moreover, changing schedules is not allowed once committed. Finally, savings over pay-as-you-go achieved using scheduled reserved instances is in the range of only 5 – 10% based on whether they are used in peak or off-peak hours.

Price changes

The discounted rates obtained when you reserve an instance are on the instance price at the time of commitment. This discount then carries over throughout the commitment term. But price reductions of the Pay-as-you-go pricing model

are common, so it is often likely that the actual savings with reservation use may be less than initially projected. However, it is unlikely that cloud providers will reduce the prices to impact the savings obtained through a commitment-based model significantly.

Newer (and cheaper) instances

Newer, better-performant instances can be introduced at cheaper prices. Cloud providers do [this](#) routinely. For the past few years AWS has been providing more performant and cheaper new-generation instances in the same instance family

with Graviton processors. Whereas the pay-as-you-go model allows changes in instance types at short notice, not all flavors of commitment-based pricing models may allow the switch.

Commitments with other Optimization Strategies

Clients optimizing their workloads often need clarity about how and when to plan their commitments during the optimization process tend to significantly delay commitment based spends until rightsizing optimizations are “completed”, thereby losing benefits. The thing to understand is both these options are mutually exclusive and are best done simultaneously. Without steady-state workloads with extremely predictable usage patterns, rightsizing and reservations are anyway complex processes that evolve.

The key here is not to delay either but to take small steps and use the flexibility available in each option to course correct and improve. For example, in the case of commitment-based models, this may mean that initial commitments are made in options that provide flexibility for change. Also, organizations can use reserved instances and savings plan options

together or add a new savings plan commitment over an existing one, etc.

Centralizing the reservation strategy is another great option for multiple teams and accounts. Ideally, organizations should slightly under-commit rather than grossly overcommit, and this can be challenging to estimate for volatile workloads. The benefit of having a larger coverage area is that if one team has fluctuations in resource usage, they could be compensated for by usage by another team. Using this model also leads to better usage visibility across teams and greater organizational efficiency. Based on how resources are procured centrally but used across multiple teams (e.g., what constitutes the show back versus chargeback), the centralized structure can get a bit complicated from a billing perspective, but given its benefits, opting for this structure is worth it.

Conclusion

If utilized optimally, commitment-based models can be an excellent option, leading to considerable savings.

It is important to assess and plan out the commitment journey to determine the best pricing model or combination of pricing models. Understanding usage patterns and deciding when, how much, and how frequently to commit is crucial.

Over time, optimizing commitments adaptively and automating commitment-based spending is also important.

This can be overwhelming. As a digital transformation leader in cloud services, in cloud services, Persistent Systems can help identify the right commitment strategy to optimize cloud spending or fast-track the adoption of FinOps.

Get in touch with us.

[Learn More](#)

About Persistent

With over 23,000 employees located in 21 countries, Persistent Systems (BSE & NSE: PERSISTENT) is a global services and solutions company delivering Digital Engineering and Enterprise Modernization. We work with the industry leaders including 14 of the 30 most innovative companies as identified by BCG, 8 of the top 10 largest banks in the US and India, and numerous innovators across the healthcare and software ecosystems. As a participant of the United Nations Global Compact, Persistent is committed to aligning strategies and operations with universal principles on human rights, labour, environment, and anti-corruption, as well as take actions that advance societal goals.

USA

Persistent Systems, Inc.
2055 Laurelwood Road, Suite 210
Santa Clara, CA 95054
Tel: +1 (408) 216 7010
Fax: +1 (408) 451 9177
Email: info@persistent.com

India

Persistent Systems Limited
Bhageerath, 402
Senapati Bapat Road
Pune 411016
Tel: +91 (20) 6703 0000
Fax: +91 (20) 6703 0008



Persistent