# Reserved Instances and On-Demand Capacity Reservation

When we use EC2 instances, what we are essentially doing is renting computer infrastructure from AWS. The default renting agreement, i.e. On-Demand EC2 instances provides us with EC2 instances on-the-go, provisioning us with resources according to the load being put on their computers. However, this is not the only renting agreement available to us, and if we are already aware of the compute capacity that is required for our workload then AWS provides us with two other methods of renting EC2 instances: **Reserved Instances** and **On-Demand Capacity Reservation**.

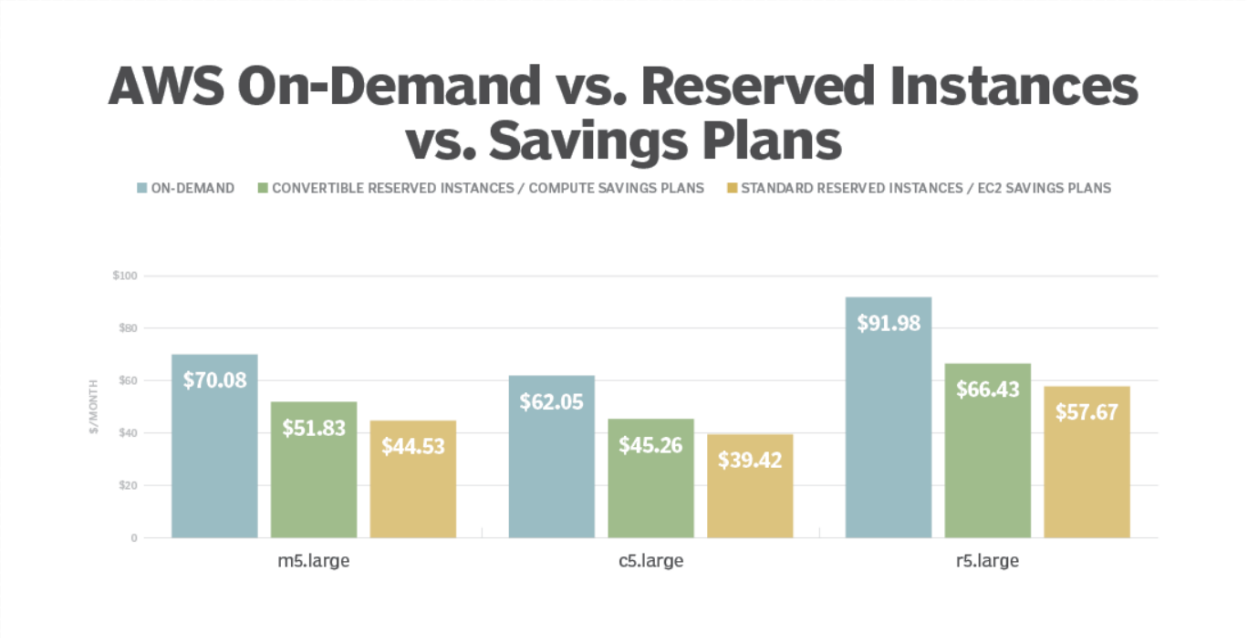
Leveraging these pricing models allow us to get a fixed number of EC2 instances of a specific hardware configuration potentially for a fraction of the price usually extracted when utilizing On-Demand EC2 instances.

**Reserved Instances** (RIs) in AWS are a cost-efficient way to secure compute capacity for long-term, predictable workloads. By committing to a one- or three-year term, RIs offer significant discounts (up to 75%) compared to On-Demand instance pricing. This pricing model is especially useful for workloads that have steady, consistent demand, such as databases, backend services, or enterprise applications. When you know your infrastructure needs in advance, RIs allow you to plan ahead and reduce overall cloud costs without sacrificing performance or availability.

RIs come in two types: **Standard** and **Convertible**. Standard RIs offer the deepest discounts but require a fixed commitment to a specific instance type, operating system, and tenancy. In contrast, Convertible RIs provide more flexibility by

allowing changes to the instance family, OS, or tenancy during the reservation period, albeit with slightly lower savings. Additionally, users can choose different payment options (All Upfront, Partial Upfront, or No Upfront) to balance between immediate cost savings and cash flow management.

While Reserved Instances can drastically reduce costs, it is crucial to remember that they do come with certain caveats which must be taken into consideration. If the workload pattern changes and the user does not fully utilize the reserved capacity for example, they are still obligated by contract to pay for the reservation. Therefore, careful planning is essential while utilizing Reserved Instances to avoid any unnecessary financial commitments.




Source: TechTarget and AWS Pricing Calculator

**On-Demand Capacity Reservations** offer AWS users the flexibility to reserve EC2 capacity for any fixed duration of time that is comfortable to them, without the need to specify exact instance types or Availability Zones (AZs). This approach provides much greater control over resource management, particularly for workloads requiring precise instance placement and configuration.

Unlike Reserved Instances (RIs), which necessitate a fixed one-year or three-year commitment, On-Demand Capacity Reservations are more suitable for tasks that have known capacity requirements but a shorter duration. If a workload demands reserved capacity for a temporary period, these reservations allow you to secure resources without long-term contracts, making them an efficient alternative to RIs. However, do note that On-Demand Capacity Reservation is simply a guarantee of compute capacity, and does not offer any reduction in cost when compared to normal On-Demand pricing, unlike reserved instances.

***TLDR;***

AWS provides us with two methods of reserving compute capacity: **Reserved Instances** and **On-Demand Capacity Reservation**.

**Reserved instances:** Provide the most savings but the contract must be for a 1 or 3 year duration.

**On-Demand Capacity Reservation:** Provide capacity without any long-term contracts, making them better for more temporary workloads though they are slightly pricier than Reserved instances.