**Cloud Provider Pricing Models**

**Spot VMs Definition:**

Unused compute capacity available at significantly lower prices. Spot VM instances are ideal for workloads that can be interrupted, providing scalability while reducing costs. Prices vary based on available capacity.

Spot VMs are ideal for the following types of workloads:

* Dev/test environments, including continuous integration and continuous delivery (CI/CD) workloads.
* Select high-performance computing scenarios, batch processing jobs, or visual rendering applications.
* Big data, analytics, container-based, and large-scale stateless applications.

**Savings Plan Definition:**

Get lower prices by agreeing to spend a fixed hourly amount on compute services for one or three years.

**Reserved Instances Definition:**

Reduce prices by committing to a specific instance configuration for 1 or 3 years.

**Azure**

Prices for single VMs:

* Linux VM: <https://azure.microsoft.com/de-de/pricing/details/virtual-machines/linux/>
* Windows VM: <https://azure.microsoft.com/de-de/pricing/details/virtual-machines/windows/>

Prices for VM scale sets:

* Linux: <https://azure.microsoft.com/en-us/pricing/details/virtual-machine-scale-sets/linux/>
* Windows: <https://azure.microsoft.com/en-us/pricing/details/virtual-machine-scale-sets/windows/>

Prices for Spot VMs:

* <https://azure.microsoft.com/en-us/pricing/spot-advisor/>

Price Models:

* **Pay-as-you-go**
  + Pay only for what you use
* **Spot VMs**
  + Up to 90% discount compared to pay-as-you-go prices
  + Pay only the maximum price you set in advance
  + Workloads will be evicted (within 30-seconds notice to finish running jobs or perform shutdown tasks) when:
    - The price is higher than the maximum price that you agreed to pay.
    - Azure no longer has available compute capacity and needs to reallocate its resources.
  + Eviction Policy is either deallocate (default) or delete
    - With deallocate charged for storage costs on the disks (no guarantee) -> can redeploy later, if there is now an instance available (not automatically)
    - Delete will delete all so no additional costs
  + Limitations:
    - Not supported for VM: B-series and Promo versions of any size (like Dv2, NV, NC, H promo sizes)
  + The following [offer types](https://azure.microsoft.com/support/legal/offer-details/) are currently supported:
    - Enterprise Agreement
    - Pay-as-you-go offer code (003P)
    - Sponsored (0036P and 0136P)
    - For Cloud Service Provider (CSP), see the [Partner Center](https://learn.microsoft.com/en-us/partner-center/azure-plan-get-started) or contact your partner directly.
* **Savings Plans**
  + Azure savings plan for compute is a flexible pricing model that provides savings up to 65 percent off pay-as-you-go pricing when you commit to spend a fixed hourly amount on compute services for one or three years.
  + Up-front payment or monthly
  + If your usage is less than or equal to your hourly amount:
    - Get lower, savings plan prices.
    - Pay the full hourly amount, even if usage is less.
    - Incur no additional costs—usage is covered by the plan.
  + If your usage is greater than your hourly amount:
    - Usage up to your hourly amount is billed at lower prices and included in the cost of your plan.
    - Additional usage is billed at pay-as-you-go prices and invoiced separately.
* **Reserved Instances** -> no official marketplace
  + Significantly reduce costs—up to 72 percent compared to pay-as-you-go pricing—with one-year or three-year terms for Windows and Linux virtual machines (VMs)
    - higher discount than savings plans, but lower flexibility
    - Can cancel a reservation for a termination fee
* Further price savings:
  + **Azure Hybird Benefit:** Is a cost-savings and migration benefit. Apply your existing Windows Server and SQL Server licenses with active Software Assurance or subscriptions to Azure Hybrid Benefit to achieve cost savings. Azure Hybrid Benefit allows users with existing on-premises licenses to apply them towards Azure services, resulting in significant cost savings.
  + **Azure Dev/Test pricing:** Azure offers discounted pricing for users who want to use Azure for development and testing purposes.
  + **Azure Free Account:** Azure offers a free account that provides limited access to Azure services for a limited time. This is an excellent option for users who want to try out Azure before committing to a paid plan.

**AWS**

Amazon Elastic Compute Cloud (Amazon **EC2**) is a web service that provides resizable computing capacity—literally, servers in Amazon's data centers—that you use to build and host your software systems. -> VMs in AWS

On-demand prices: <https://aws.amazon.com/ec2/pricing/on-demand/>

Spot prices: <https://aws.amazon.com/ec2/spot/pricing/>

Savings Plans prices: <https://aws.amazon.com/savingsplans/compute-pricing/>

Reserved Instances prices: <https://aws.amazon.com/ec2/pricing/reserved-instances/pricing/>

Dedicated Hosts prices: <https://aws.amazon.com/ec2/dedicated-hosts/pricing/>

Price models:

* **On-demand**
* Pay for the instances that you use by the second or the hour, with a minimum of 60 seconds, with no long-term commitments or upfront payments
* On-Demand instances are recommended for:
  + Users that prefer the low cost and flexibility of Amazon EC2 without any upfront payment or long-term commitment
  + Applications with short-term, spiky, or unpredictable workloads that cannot be interrupted
  + Applications being developed or tested on Amazon EC2 for the first time
* **Spot instances**
  + Amazon EC2 Spot instances allow you to request spare Amazon EC2 computing capacity for up to 90% off the On-Demand prices
  + Spot instances are recommended for:
* Applications that have flexible start and end times
* Applications that are feasible only at very low compute prices
* Users with urgent computing needs for large amounts of additional capacity
* **Savings Plans**
  + You can reduce your Amazon EC2 costs by making a commitment to a consistent amount of usage, in USD per hour, for a term of 1 or 3 years.
  + AWS recommends Savings Plans over RIs due to the higher flexibility
  + Applicable for EC2, Fargate and Lambda
* **Reserved Instances**
  + You can reduce your Amazon EC2 costs by making a commitment to a specific instance configuration, including instance type and Region, for a term of 1 or 3 years.
  + AWS provides further discounts based on volume (500k – 4mio = 5%, 4mio – 10 mio = 10%, over = even more)
  + RI instance types:
* Standard RIs: These provide the most significant discount (up to 72% off On-Demand) and are best suited for steady-state usage.
* Convertible RIs: These provide a discount (up to 54% off On-Demand) and the capability to change the attributes of the RI as long as the exchange results in the creation of Reserved Instances of equal or greater value. Like Standard RIs, Convertible RIs are best suited for steady-state usage.
* Scheduled RIs: These are available to launch within the time windows you reserve. This option allows you to match your capacity reservation to a predictable recurring schedule that only requires a fraction of a day, a week, or a month.
  + RI attributes: When the attributes of the RI match the attributes of a used instance, the discount of the RI is applied and the capacity reservation is utilized. These attributes are:
* [Instance type](https://aws.amazon.com/ec2/instance-types/): Instance types comprise varying combinations of CPU, memory, storage, and networking capacity. For example, m4.2xlarge.
* Platform description: RIs can be purchased for Amazon EC2 instances running Linux/UNIX, SUSE Linux, Red Hat Enterprise Linux, Microsoft Windows Server, and Microsoft SQL Server platforms.
* Tenancy: If you are using instances with a default tenancy, purchase default tenancy Reserved Instances. If you are using dedicated instances, purchase Reserved Instances with a tenancy of dedicated.
* Availability Zone (optional): By selecting an Availability Zone, the RI you purchase will provide a capacity reservation and a discount for the use of instances in the selected Availability Zone. When an Availability Zone isn’t specified, the RI discount will apply to a running instance of any size (within the same family) in the region. As an example, let’s say you own an m4.2xlarge Linux/UNIX RI with default tenancy in US East (N.Virginia). Then this RI’s discounted rate can automatically apply to two m4.xlarge instances in us-east-1a or four m4.large instances in us-east-1b.
  + Reserved Instances Marketplace
    - Reserved Instance Marketplace allows other AWS customers to list their Standard RIs for sale. Third-party Standard RIs are no different from the Standard Reserved Instances purchased directly from AWS except they are often listed at lower prices and shorter terms.
* **Dedicated Hosts**
  + Amazon EC2 Dedicated Hosts allow you to use your eligible software licenses from vendors such as Microsoft and Oracle on Amazon EC2, so that you get the flexibility and cost effectiveness of using your own licenses, but with the resiliency, simplicity and elasticity of AWS.
  + An Amazon EC2 Dedicated Host is a physical server fully dedicated for your use, so you can help address corporate compliance requirements.
* Further price savings:
  + Free Tier: As part of the AWS Free Tier, you can get started with Amazon EC2 for free. This includes 750 hours of Linux and Windows t2.micro instances (t3.micro for the regions in which t2.micro is unavailable), each month for one year. To stay within the Free Tier, use only EC2 Micro instances.

**GCP**

Google Compute Engine is a cloud-based infrastructure as a service (IaaS) offering from Google Cloud Platform (GCP). It allows users to run virtual machines (VMs) on Google's global infrastructure.

Compute Engine Prices (On demand, Spot, commitment): <https://cloud.google.com/compute/all-pricing>

SUD Prices: <https://cloud.google.com/compute/docs/sustained-use-discounts>

* **Pay-as-you-go**
  + Charged for a minimum of one minute, then in 1 second increments
* **Spot VMs**
  + Automatic discount between 60% and 91%
  + Spot VMs provide significant discounts for VMs. Additionally, you receive discounts for [GPUs](https://cloud.google.com/compute/gpus-pricing), [local SSDs](https://cloud.google.com/compute/disks-image-pricing#localssdpricing), and [external IP addresses](https://cloud.google.com/vpc/network-pricing#ipaddress) that are attached to Spot VMs.
  + If a preempted VM was stopped, you are not charged for the VM hours while it remains in a TERMINATED state. You can access and recover data from any persistent disks that are attached to the VM, but those disks still incur storage charges until you delete them.
  + Spot prices, the prices for Spot VMs, change over time, up to once every 30 days
  + Spot VMs cant apply CUDs or SUDs
* **Commited Use Discounts (CUDs)**
  + up to a 70% discount for memory-optimized machine types and up to a 57% discount for all other machine types.
  + When you purchase a commitment, you commit either to a minimum amount of resource usage or to a minimum spend amount for a specified term of one or three years.
  + **Resource-based committed use discounts:** You receive these CUDs when you purchase a resource-based commitment and commit to use a minimum level of Compute Engine resources in a particular region.
    - vCPU for 1 year gets discount of 37%
    - 3 year commitment gets 70% percent discount on memory optimized machined and 57% on all other machine types
  + **Compute Engine flexible committed use discounts:** You avail flexible CUDs when you purchase a spend-based (or flexible) commitment for Compute Engine and commit to a minimum amount of hourly spend.
    - A 28% discount over your committed hourly spend amount for a 1-year commitment
    - A 46% discount over your committed hourly spend amount for a 3-year commitment
* **Sustained Use Discounts**
  + Compute Engine offers automatic discounts for sustained use of vCPUs and memory that are not receiving any other discounts.
  + Whenever you use an applicable vCPU or an applicable GB of memory for more than 25% of a month, Compute Engine automatically gives you a sustained use discount for every incremental hour that you continue to use that resource.
  + The discount increases with usage and you can get up to a 30% net discount off of the vCPU and memory cost for instances that run the entire month.