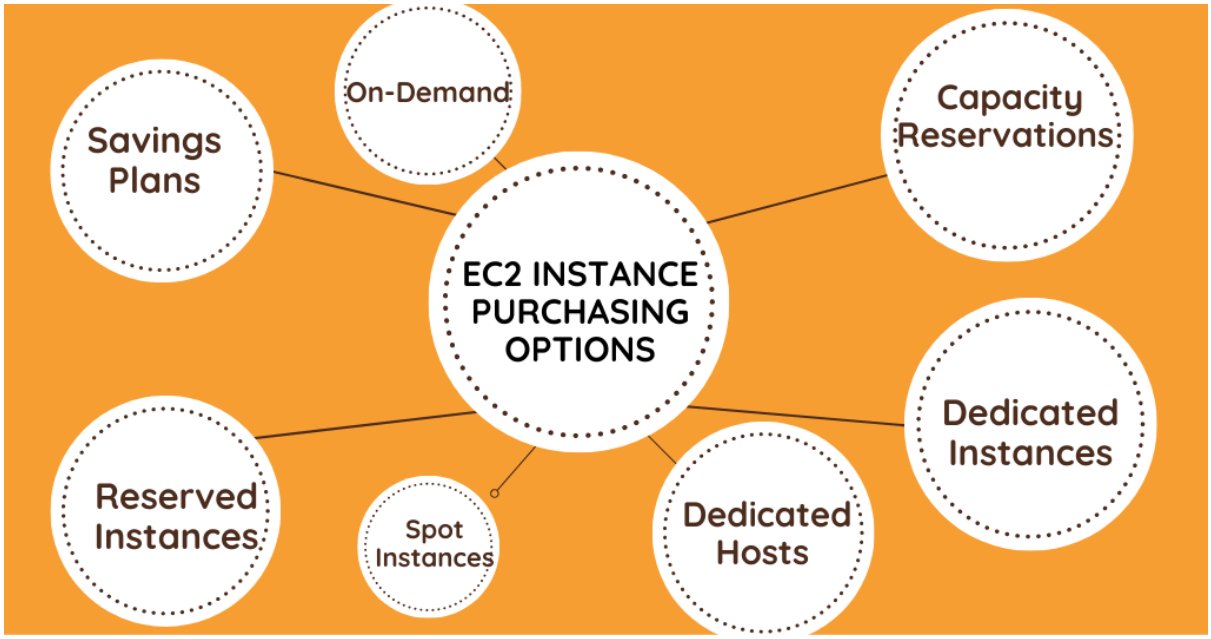
AWS EC2 Instance Purchasing Options

Introduction

Amazon EC2 (Elastic Compute Cloud) offers a variety of purchasing and configuration options from a cost and capacity management perspective, each tailored to meet specific needs and use cases. From cost-efficient solutions for flexible workloads to dedicated resources for compliance-sensitive applications, EC2 has it all. Let’s break down these options to help you decide the best fit for your needs.

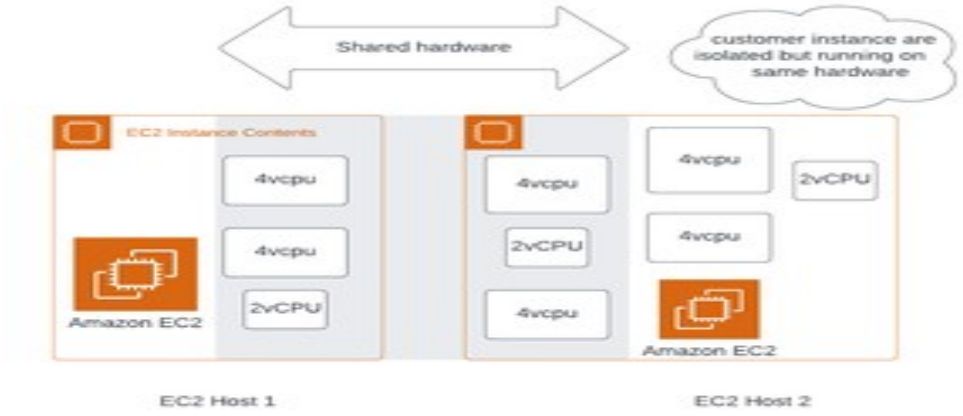


1. On-Demand Instances

Definition: On-Demand Instances let you pay for compute capacity by the hour or second (minimum of 60 seconds) with no long-term commitments or upfront payments.

Use Case: Ideal for short-term, irregular workloads that cannot be interrupted. For example, a startup testing a new application or a company running high-performance computing applications.

Real-World Example: An e-commerce website using On-Demand Instances during a flash sale to handle the sudden spike in traffic.



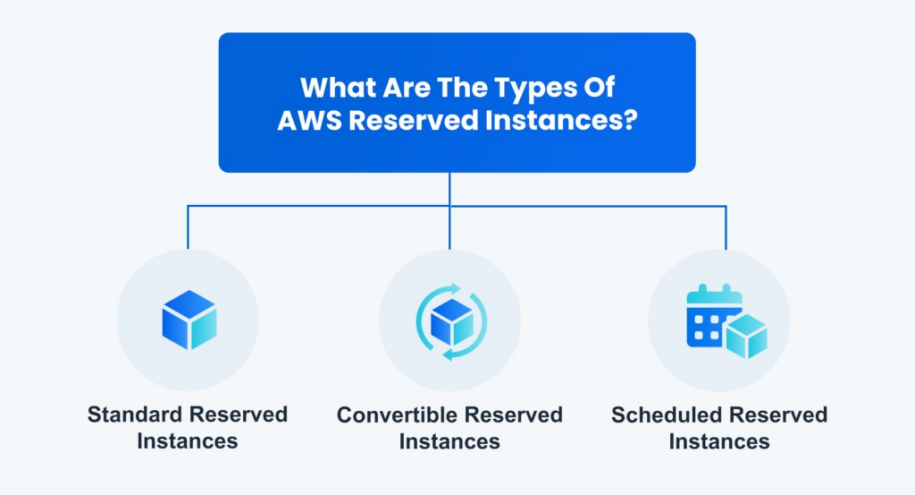
2. Reserved Instances (Standard and Convertible)

Definition:

- **Standard Reserved Instances**: Offer up to 75% off the On-Demand price, with a commitment of 1 or 3 years.
  For example, an always-up, right-sized database server that will be running continuously for 1 year.
- **Convertible Reserved Instances**: Provide a discount (up to 54% off On-Demand) and the flexibility to change the attributes of the RI as long as the exchange results in the creation of Reserved Instances of equal or greater value.

Use Case: Best for steady-state workloads. Standard RIs are ideal for applications with predictable usage, while Convertible RIs are suitable for workloads with some variability but generally consistent operation.

Real-World Example: A financial services firm using Standard Reserved Instances for their consistent backend processes, and Convertible Reserved Instances for their customer-facing applications that may need scaling or changing over time.



3. EC2 Savings Plans

Definition: Savings Plans offer significant savings over On-Demand pricing, in exchange for a commitment to use a specific amount of compute power (measured in $/hour) for a 1 or 3-year period.

Use Case: Suitable for users with a consistent amount of compute usage over a long period. This can range from running micro services to data processing applications.

Real-World Example: A media company running continuous data processing workloads on EC2 instances, committing to a certain amount of compute usage with an EC2 Savings Plan to reduce costs

4. EC2 Spot Instances

Definition: Spot Instances provide access to unused EC2 capacity at up to 90% off the On-Demand price.

Use Case: Perfect for flexible, stateless, and fault-tolerant applications like big data analysis, containerized workloads, or CI/CD pipelines. Not suitable for critical jobs or those intolerant to interruptions.

Real-World Example: A research institution conducting large-scale genomic analysis using Spot Instances, taking advantage of the lower costs for its flexible and interruptible tasks.

5. Dedicated Hosts

Definition: Dedicated Hosts are physical servers with EC2 instance capacity fully dedicated to your use. They allow you to use your existing server-bound software licenses and can help you meet compliance requirements by providing physical isolation from other AWS customers.

Use Case: Particularly useful for businesses with strict regulatory compliance needs or those who need to use their existing software licenses in the cloud.

Real-World Example: A user intends to leverage their current software licenses based on per-socket, per-core, or per-virtual machine criteria for a Microsoft Windows server operating on Amazon Web Services (AWS),

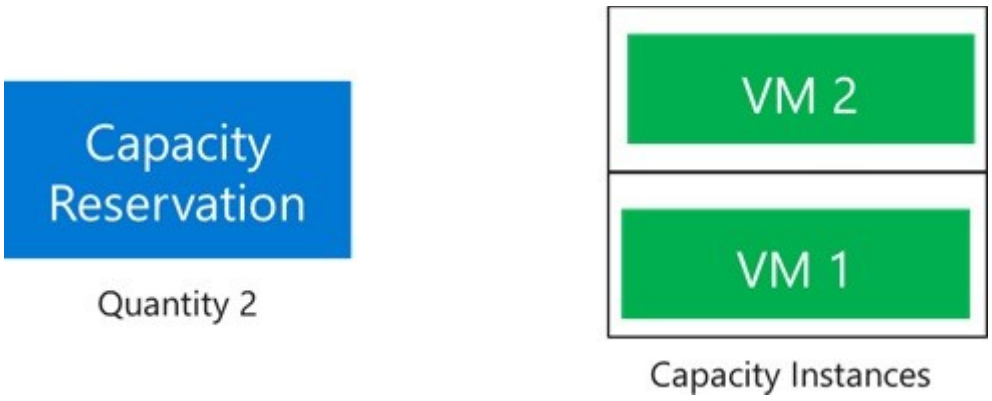
6. Dedicated Instances

Definition: Dedicated Instances run in a Virtual Private Cloud (VPC) on hardware that's dedicated to a single customer. Unlike Dedicated Hosts, Dedicated Instances do not offer the same level of control over the physical host, but they do ensure your instances are isolated at the hardware level from instances that belong to other AWS accounts.

Use Case: Ideal for applications that require isolation from other customers' instances for security or compliance reasons, but do not require the use of existing software licenses.

Real-World Example: A government agency using Dedicated Instances to process sensitive data, ensuring compliance with strict regulatory standards without the need for physical control over the server

7. Capacity Reservations



Definition: Capacity Reservations allow you to reserve capacity for your EC2 instances in a specific Availability Zone for any duration.

Use Case: Useful for applications that need a guaranteed availability in a specific AWS region, ensuring capacity for critical applications.

Real-World Example: An online gaming company might use Capacity Reservations to ensure they have enough capacity for a new game launch in a specific geographic region.

Conclusion

Choosing the right EC2 instance purchasing option depends on your specific use case, budget, and flexibility requirements. Whether it’s the cost-effective Spot Instances for interruptible workloads or Dedicated Hosts for compliance-heavy applications, AWS provides a diverse range of options to cater to various business needs. By understanding each option and aligning it with your operational requirements, you can optimize your AWS investment effectively