

# Compute Resources

After completing this episode, you should be able to:

- Discuss the various compute resource options available in the cloud today

**Description:** In this episode, you will learn about the various compute options that are available in the cloud today. This episode includes a discussion of VMs, containers, and serverless compute options.

## Compute Resources

Virtual machines (VMs) - cloud-based virtual machines (VMs) are virtualized computing resources provided by cloud service providers that mimic the functionality of physical servers, allowing users to run operating systems and applications as if they were on dedicated hardware. These VMs offer flexibility in terms of configuration, scalability, and resource allocation, enabling users to easily scale up or down based on demand. They are typically billed on a pay-as-you-go basis, providing cost-efficiency by charging only for the resources consumed. Cloud-based VMs can be quickly deployed, managed, and integrated with other cloud services, supporting a wide range of use cases such as application hosting, development and testing environments, and disaster recovery solutions. Examples of cloud-based VM services include Amazon EC2, Google Compute Engine, and Microsoft Azure Virtual Machines.

Containers - cloud-based containers are lightweight, portable computing environments that package applications and their dependencies into a single unit, ensuring consistency across various computing environments. Managed by container orchestration platforms like Kubernetes, these containers run on cloud infrastructure, enabling efficient scaling, deployment, and management of applications. Unlike traditional virtual machines, containers share the host system's OS kernel, making them more resource-efficient and faster to start. They are ideal for microservices architectures, facilitating continuous integration and deployment (CI/CD) by allowing developers to build, test, and deploy applications in isolated environments. Examples of cloud-based container services include AWS Elastic Kubernetes Service (EKS), Google Kubernetes Engine (GKE), and Azure Kubernetes Service (AKS).

Serverless - serverless compute in cloud technology allows developers to build and run applications without managing the underlying infrastructure, as the cloud provider automatically handles the provisioning, scaling, and maintenance of servers. In this model, code is executed in response to events or requests, with resources allocated dynamically, ensuring efficient utilization and billing based only on the actual compute time consumed. This abstraction enables developers to focus solely on writing and deploying code, enhancing productivity and speeding up development cycles. Serverless compute is highly scalable, automatically adjusting to meet demand, and supports various use cases, including web applications, APIs, data processing, and real-time file processing. Prominent examples of serverless compute services include AWS Lambda, Azure Functions, and Google Cloud Functions.

## Additional resources

- Choosing the Right Compute Option in GCP: <https://cloud.google.com/blog/products/compute/choosing-the-right-compute-option-in-gcp-a-decision-tree> (<https://cloud.google.com/blog/products/compute/choosing-the-right-compute-option-in-gcp-a-decision-tree>)