AWS Compute Services: Building Blocks of the Cloud

Compute services form the essential foundation of AWS cloud infrastructure, providing the processing power that drives modern applications. These services offer secure, scalable compute resources that can be dynamically adjusted to meet changing workload requirements.

From traditional virtual machines to cutting-edge serverless architectures, AWS compute options power everything from simple websites to complex enterprise applications, machine learning workloads, and high-performance computing scenarios.

Virtual Servers: Amazon EC2

Elastic Compute Cloud (EC2)

EC2 provides resizable virtual machines in the cloud, giving you complete control over your computing resources. Choose from a wide range of instance types optimized for different use cases:

- General Purpose: Balanced compute, memory, and networking
- Compute Optimized: High-performance processors
- Memory Optimized: Fast performance for memory-intensive workloads
- Storage Optimized: High, sequential read/write access to large datasets

Cost Optimization Features

- EC2 Spot Instances: Up to 90% savings for interruptible workloads
- Auto Scaling: Automatically adjust capacity based on demand
- Reserved Instances: Discounted pricing with capacity reservation

Platform as a Service Solutions

AWS Elastic Beanstalk

A fully managed service that handles infrastructure deployment and management for web applications. Simply upload your code, and Beanstalk automatically handles capacity provisioning, load balancing, auto-scaling, and application health monitoring.

Supports multiple platforms including Java, .NET, PHP, Node.js, Python, Ruby, and Docker.

Amazon Lightsail

The simplest way to launch and manage a virtual private server with AWS. Provides pre-configured development stacks like WordPress, LAMP, and MEAN with predictable, low monthly pricing.

Ideal for small websites, simple web applications, and dev/test environments with straightforward networking and storage options.

Both services significantly reduce the operational complexity of deploying and managing applications, allowing developers to focus more on writing code than managing infrastructure.

Container Services

Amazon Elastic Container Service (ECS)

A fully managed container orchestration service that supports Docker containers. ECS makes it easy to deploy, manage, and scale containerized applications without the complexity of managing container infrastructure.

AWS Fargate

A serverless compute engine for containers that works with both ECS and EKS. Eliminates the need to provision and manage servers, letting you specify and pay for resources per application.

Amazon Elastic Kubernetes Service (EKS)

A managed Kubernetes service that simplifies running Kubernetes on AWS without needing to install and maintain your own Kubernetes control plane. Certified Kubernetes conformant for compatibility with existing applications.

AWS App Runner

A fully managed service that makes it easy to deploy web applications and APIs at scale without any infrastructure experience. Automatically builds and deploys web applications from source code or container images.

Serverless Compute: AWS Lambda

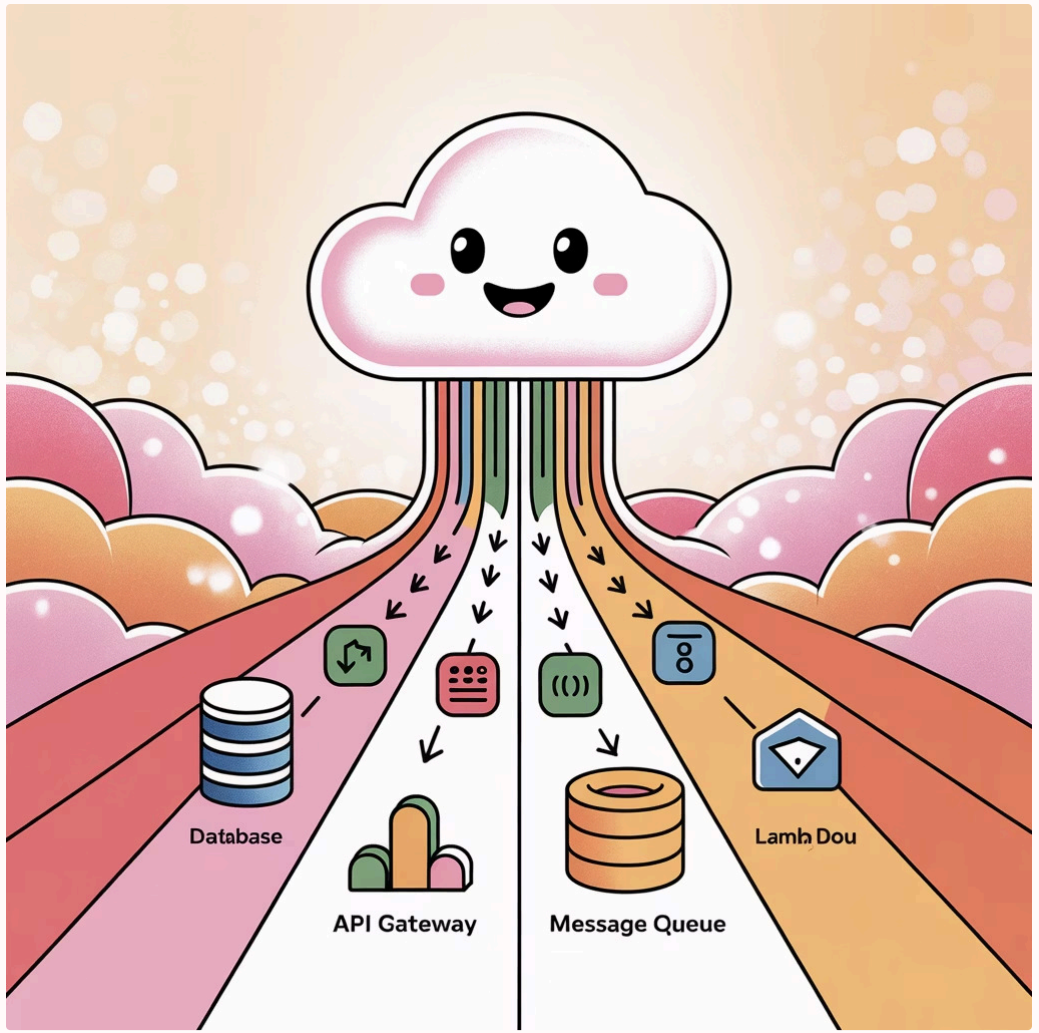
Run Code Without Thinking About Servers

AWS Lambda lets you run code without provisioning or managing servers. You pay only for the compute time you consume—there is no charge when your code is not running.

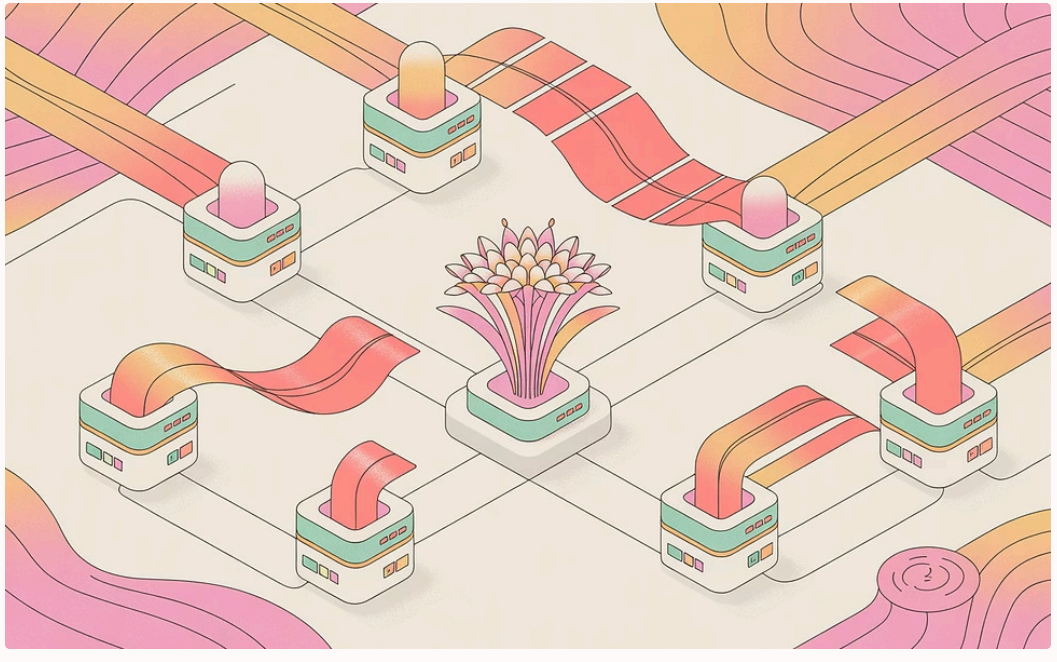
Lambda functions can be triggered by over 200 AWS services and SaaS applications, executing your code in response to events like changes to data in an S3 bucket, updates to a DynamoDB table, or HTTP requests via API Gateway.

Key Benefits

- Zero server management or capacity planning required
- Automatic scaling from a few requests per day to thousands per second
- Built-in fault tolerance and high availability
- Supports multiple languages: Node.js, Python, Java, Go, .NET, Ruby

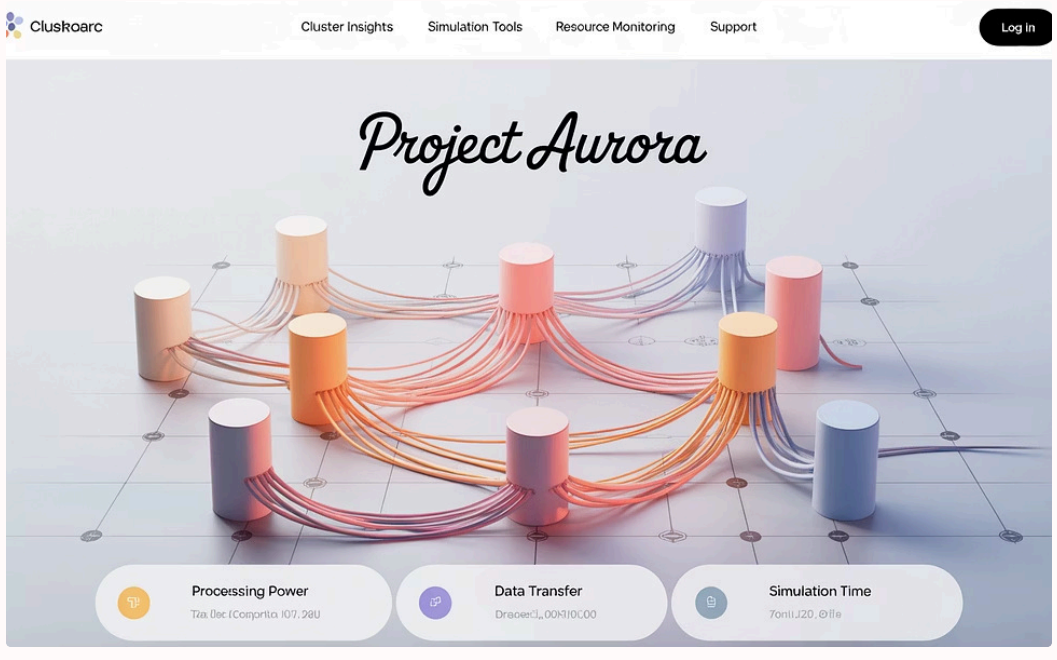


Specialized Compute Services



AWS Batch

Dynamically provisions the optimal quantity and type of compute resources based on the volume and specific requirements of batch jobs submitted. Ideal for processing large genomics datasets, financial risk models, or media transcoding.



AWS ParallelCluster

An open-source cluster management tool that makes it easy to deploy and manage High Performance Computing (HPC) clusters on AWS. Enables scientists and engineers to solve complex computational problems with applications that require high network performance.

These specialized services are designed for computation-intensive workloads in fields such as scientific research, AI/ML model training, engineering simulations, and big data analytics where traditional compute resources would be insufficient.

Hybrid and Edge Compute

AWS Outposts

Extends AWS infrastructure, services, and tools to virtually any on-premises facility. Outposts is a fully managed solution with AWS-designed hardware that provides consistent hybrid experience for applications requiring low latency access to on-premises systems.

AWS Local Zones

Places compute, storage, database, and other select AWS services closer to large population and industry centers. Enables applications that require single-digit millisecond latency for use cases like media & entertainment content creation and real-time gaming.

AWS Wavelength

Embeds AWS compute and storage services within 5G networks, providing mobile edge computing infrastructure for developing, deploying, and scaling ultra-low-latency applications for 5G devices.

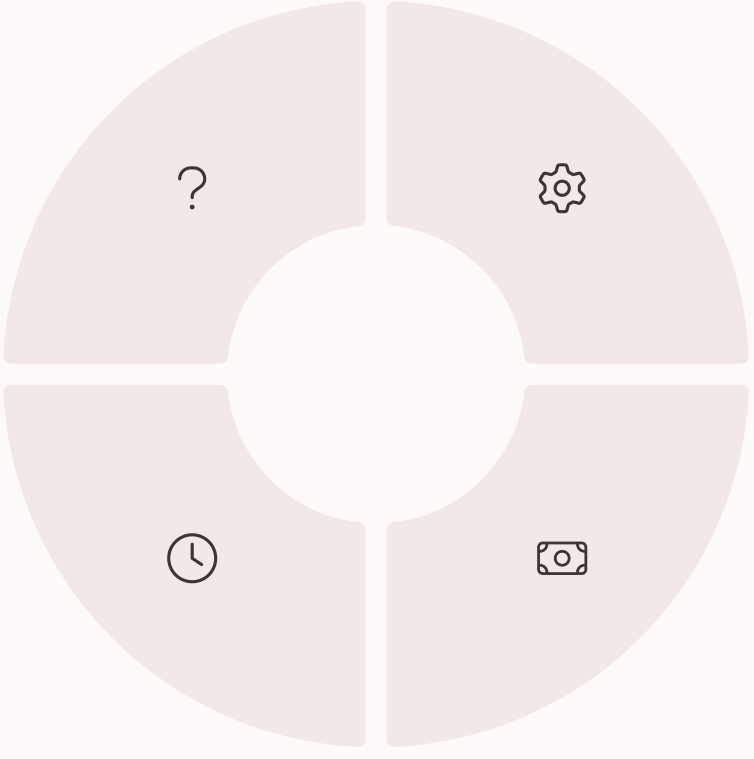
Selecting the Right AWS Compute Service

Workload Type

- Traditional applications → EC2
- Containerized apps → ECS/EKS
- Event-driven functions → Lambda
- Web applications → Elastic Beanstalk

Performance Requirements

- Low latency → Local Zones
- High performance → HPC clusters
- Local data processing → Outposts
- Edge computing → Wavelength



Management Preference

- Full control → EC2
- Minimal management → Fargate
- Zero infrastructure → Lambda
- Simplified deployment → Beanstalk

Cost Considerations

- Steady workloads → Reserved Instances
- Variable traffic → On-demand/Spot
- Pay-per-use → Serverless
- Predictable pricing → Lightsail

The flexibility of AWS compute services allows you to choose the right tools for your specific requirements, and even combine multiple services in a single architecture to optimize performance, cost, and operational efficiency.