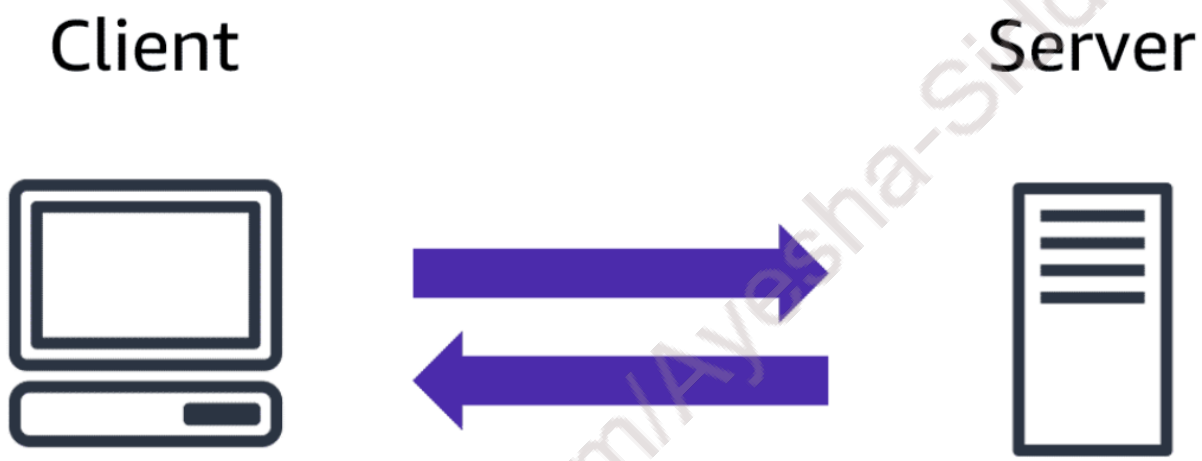
AWS Module 1 - Intro to AWS

What is a client-server model?

- In computing, a **client** can be a web browser or desktop application that a person interacts with to make requests to computer servers. A **server** can be services, such as Amazon Elastic Compute Cloud (Amazon EC2) – a type of virtual server.
- For example, suppose that a client makes a request for a news article, the score in an online game, or a funny video. The server evaluates the details of this request and fulfills it by returning the information to the client.



The main key principle is : **Pay for what you need**

What is Cloud computing?

- Cloud computing is the on-demand delivery of IT resources over the internet with pay-as-you-go pricing.
- Cloud computing provides developers with the ability to focus on what matters most and avoid infrastructure procurement, maintenance, and capacity planning, or undifferentiated heavy lifting.
- Cloud computing provides a simple way to access servers, storage, databases and a broad set of application services over the Internet. A cloud services platform such as Amazon Web Services owns and maintains the network-connected hardware required for these application services, while you provision and use what you need via a web application.

What is undifferentiated heavy lifting?

- In short, undifferentiated heavy lifting is everything that an application needs to do but doesn't increase its competitive advantage in the eyes of its customers.

Extra Info:

Companies That Might Prefer Lower-Level Infrastructure (IaaS):

1. **Large Enterprises:** Some large organizations have well-established IT departments with extensive technical expertise. They may choose lower-level infrastructure to have full control over their cloud resources and tailor their infrastructure to meet unique requirements.
2. **Highly Regulated Industries:** Companies operating in highly regulated industries (e.g., finance, healthcare) often have strict compliance requirements. They may opt for lower-level infrastructure to have greater control over security configurations and data governance.
3. **Customized Solutions:** Businesses with complex or highly customized applications that cannot be easily accommodated by off-the-shelf solutions might choose lower-level infrastructure to build and manage their specialized environments.
4. **Cost Efficiency:** In some cases, larger companies may choose lower-level infrastructure to optimize costs by fine-tuning resource allocation and utilizing their existing IT expertise efficiently.

Companies That Might Prefer Higher-Level Services (PaaS, SaaS):

1. **Startups and Small Businesses:** Smaller companies with limited IT resources and technical expertise may choose higher-level services to accelerate development, reduce operational overhead, and focus on their core product or service.
2. **Fast-Growing Startups:** Startups experiencing rapid growth often prioritize speed and scalability. They may opt for higher-level services to quickly deploy and scale applications without worrying about infrastructure management.
3. **Cost-Efficient Solutions:** Companies looking to minimize capital expenditures and reduce operational costs may prefer higher-level services because they typically follow a pay-as-you-go model, eliminating the need for large upfront investments.
4. **Short Development Cycles:** Businesses aiming for quick time-to-market might choose higher-level services to streamline the development process, leveraging pre-built components and services.
5. **Non-Technical Companies:** Companies in non-technical industries that don't have in-house IT expertise may find it more practical to rely on managed services for their technology needs.
6. **Scalability Needs:** Organizations expecting fluctuating workloads or rapid growth may prefer higher-level services that can automatically scale resources up or down based on demand It's important to note that the choice between lower-level and higher-level services is not binary. Many companies use a combination of both, often referred to as a hybrid approach, to meet various application requirements efficiently. The specific choice will depend on factors like the complexity of the application, budget constraints, technical capabilities, and the company's overall strategic objectives.

Cloud Computing Deployment Models

What is cloud based deployment?

- Run all parts of the application in the cloud.
- Migrate existing applications to the cloud.
- Design and build new applications in the cloud.

What is on-premises deployment?

- **On-premises deployment** is also known as a *private cloud* deployment. In this model, resources are deployed on premises by using virtualization and resource management tools.
- For example, you might have applications that run on technology that is fully kept in your on-premises data center. Though this model is much like legacy IT infrastructure, its incorporation of application management and virtualization technologies helps to increase resource utilization.

What is hybrid deployment?

- Connect cloud-based resources to on-premises infrastructure.
- Integrate cloud-based resources with legacy IT applications.
- For example, you have legacy applications that are better maintained on premises, or government regulations require your business to keep certain records on premises.
- For example, suppose that a company wants to use cloud services that can automate batch data processing and analytics. However, the company has several legacy applications that are more suitable on premises and will not be migrated to the cloud. With a hybrid deployment, the company would be able to keep the legacy applications on premises while benefiting from the data and analytics services that run in the cloud.

Benefits of cloud computing

Consider why a company might choose to take a particular cloud computing approach when addressing business needs?

1.Upfront Expense Vs Variable Expense?

- Upfront expense refers to data centers, physical servers, and other resources that you would need to invest in before using them. Variable expense means you only pay for computing resources you consume instead of investing heavily in data centers and servers before you know how you're going to use them.

2.Spending money to maintain data centers

- Computing in data centers often requires you to spend more money and time managing infrastructure and servers.
- A benefit of cloud computing is the ability to focus less on these tasks and more on your applications and customers.

3.Stop guessing capacity.

- With cloud computing, you don't have to predict how much infrastructure capacity you will need before deploying an application.
- For example, you can launch Amazon EC2 instances when needed, and pay only for the compute time you use. Instead of paying for unused resources or having to deal with limited capacity, you can access only the capacity that you need. You can also scale in or scale out in response to demand.

4.Benefit from massive economies of scale

- When organizations use cloud services, they essentially share the costs of the underlying infrastructure with other customers. As a result, they can access computing, storage, and networking resources at a lower cost compared to building and managing their own data centers.
- Customers benefit from lower prices because cloud providers can offer competitive rates due to their cost efficiencies.

5.Increased speed and agility

- The flexibility of cloud computing makes it easier for you to develop and deploy applications
- This flexibility provides you with more time to experiment and innovate. When computing in data centers, it may take weeks to obtain new resources that you need. By comparison, cloud computing enables you to access new resources within minutes.

6.Go Global in mins

- The global footprint of the AWS Cloud enables you to deploy applications to customers around the world quickly, while providing them with low latency. This means that even if you are located in a different part of the world than your customers, customers are able to access your applications with minimal delays.

Cloud Computing Models

What is IaaS?

- Infrastructure as a Service (IaaS) contains the basic building blocks for cloud IT
- typically provides access to networking features, computers (virtual or on dedicated hardware), and data storage space.
- IaaS provides you with the highest level of flexibility and management control over your IT resources and is mostly like the existing IT resources that many developers are familiar with today.

What is PaaS?

- Platform as a Service (PaaS) removes the need for you to manage the underlying infrastructure (usually hardware and operating systems) and allows you to focus on the deployment and management of your applications.
- This helps you be more efficient because you don’t need to worry about resource procurement, capacity planning, software maintenance, patching, or any of the other undifferentiated heavy lifting involved in running your application.

What is SaaS?

- Software as a Service (SaaS) provides you with a completed product that is run and managed by the service provider.
- With a SaaS offering you do not have to think about how the service is maintained or how the underlying infrastructure is managed
- you only need to think about how you will use that piece of software

- A common example of a SaaS application is web-based email which you can use to send and receive email without having to manage feature additions to the email product or maintain the servers and operating systems that the email program is running on.
- In most cases, people referring to SaaS are referring to end-user applications which are cloud-based software applications that are readily accessible to end-users via the internet. These applications offer the convenience of not requiring local installation or maintenance, making them a popular choice for businesses and individuals looking for accessible and user-friendly software solutions.

Global Infrastructure

- AWS has infrastructure all over the world, so developers can deploy applications in multiple physical locations with just a few clicks. By putting your applications in closer proximity to your end users, you can reduce latency and improve the user experience.
- AWS Cloud infrastructure is built around AWS Regions and Availability Zones. A Region is a physical location in the world where we have multiple Availability Zones. Availability Zones consist of one or more discrete data centers, each with redundant power, networking, and connectivity, housed in separate facilities.

How does AWS design its Availability Zones and Regions for fault tolerance and stability?

- Each Amazon Region is isolated from other Regions, providing maximum fault tolerance and stability.
- Availability Zones (AZs) within a Region are interconnected through low-latency links.
- AWS allows for the placement of instances and data storage across multiple geographic regions and Availability Zones within a Region.
- Each Availability Zone is an independent failure zone, physically separated and often located in low-risk flood plains.
- Data centers in different Availability Zones have discrete uninterruptible power supply (UPS) and onsite backup generation facilities.
- Independent substations supply data centers in different Availability Zones, reducing the risk of power grid events affecting multiple AZs.
- All Availability Zones are redundantly connected to multiple tier-1 transit providers for robust network connectivity.

What is cloud computing?

- ( ) Backing up files that are stored on desktop and mobile devices to prevent data loss
- ( ) Deploying applications connected to on-premises infrastructure
- ( ) Running code without needing to manage or provision servers
- (•) On-demand delivery of IT resources and applications through the internet with pay-as-you-go pricing

SUBMIT

What is another name for on-premises deployment?

- (•) Private cloud deployment
- ( ) Cloud-based application
- ( ) Hybrid deployment
- ( ) AWS Cloud

SUBMIT

How does the scale of cloud computing help you to save costs?

- ✗ You do not have to invest in technology resources before using them.
- ✓ The aggregated cloud usage from a large number of customers results in lower pay-as-you-go prices.
- ✗ Accessing services on-demand helps to prevent excess or limited capacity.
- ✗ You can quickly deploy applications to customers and provide them with low latency.