Introduction to Cloud

Introduction to Cloud Computing

- Transitioning to the cloud is one of the most important things happening in IT today.
- It is a typical organization with a headquarters and some number of branch offices with users in them. Users rely on applications that run directly on the hardware that belongs to their organization.
- Cloud computing means that instead of all the computer hardware and software you're using sitting on your desktop, or somewhere inside your company's network, it's provided for you as a service by another company and accessed over the Internet, usually in a completely seamless way. Exactly where the hardware and software is located and how it all works doesn't matter to you, the user—it's just somewhere up in the nebulous "cloud" that the Internet represents.

Cloud from beginning

If you ever used

- Gmail (or yahoo email or something else in cloud)
- Google Drive or Apple drive
- YouTube
- Instagram

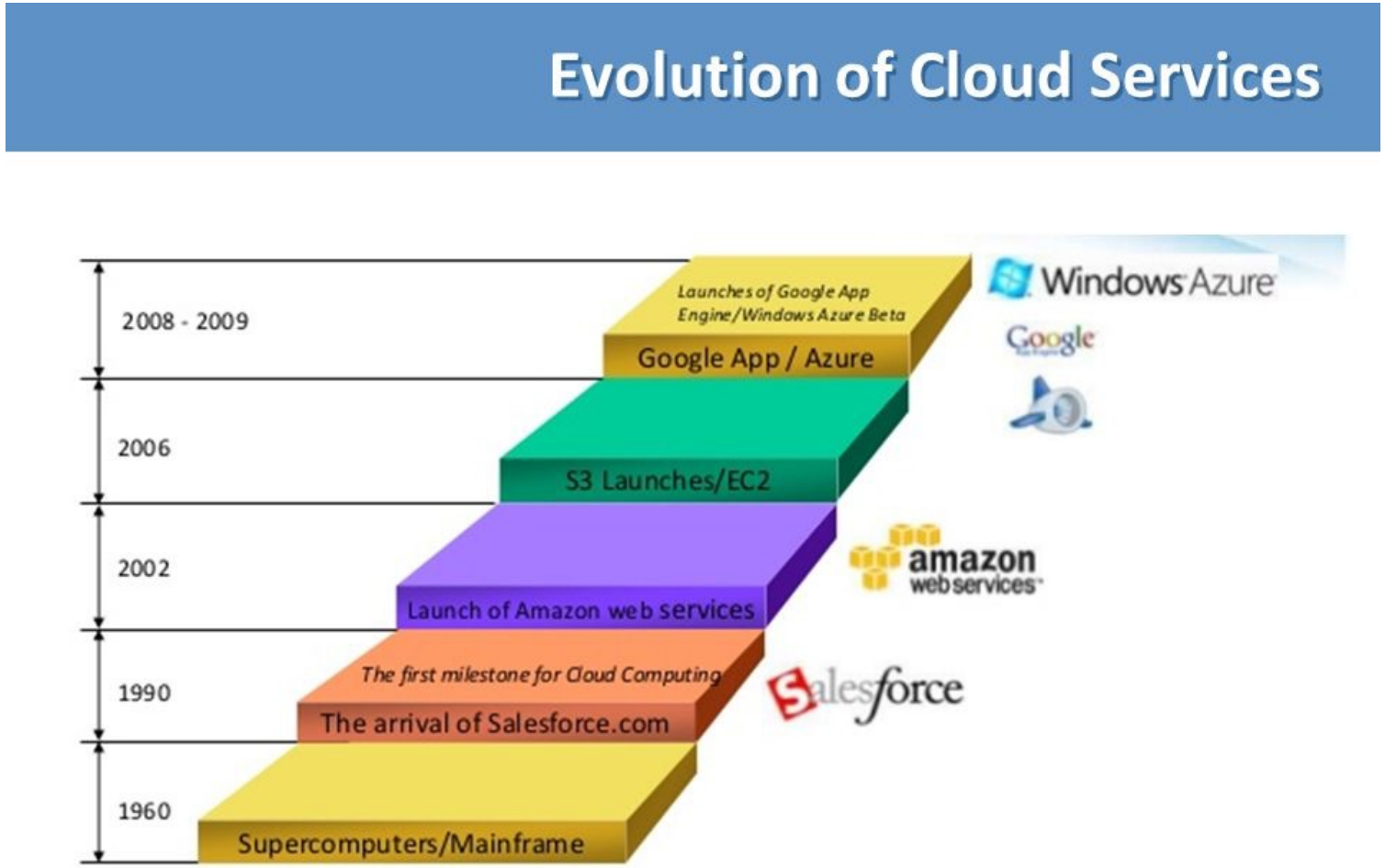
=> You already used cloud technology.

In the above examples you could use cloud application as a service: you don’t know where physically your email and documents are saved, but your expectation is they are available from any location.

Cloud infrastructure

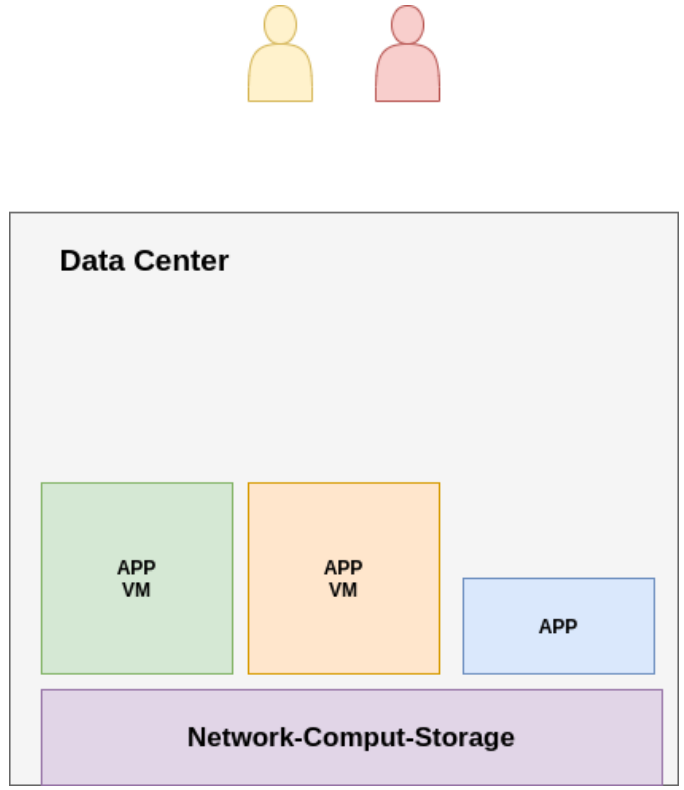
Why Cloud?
++++
Eliminate “heavy lifting”
Allows to “fail fast”
Pay-as-you-use model
Anywhere and anytime access
Affordable
Predictable costs
Worry-free IT
High levels of security
Quick deployment
Scalability

Long-term costs
Less customizable

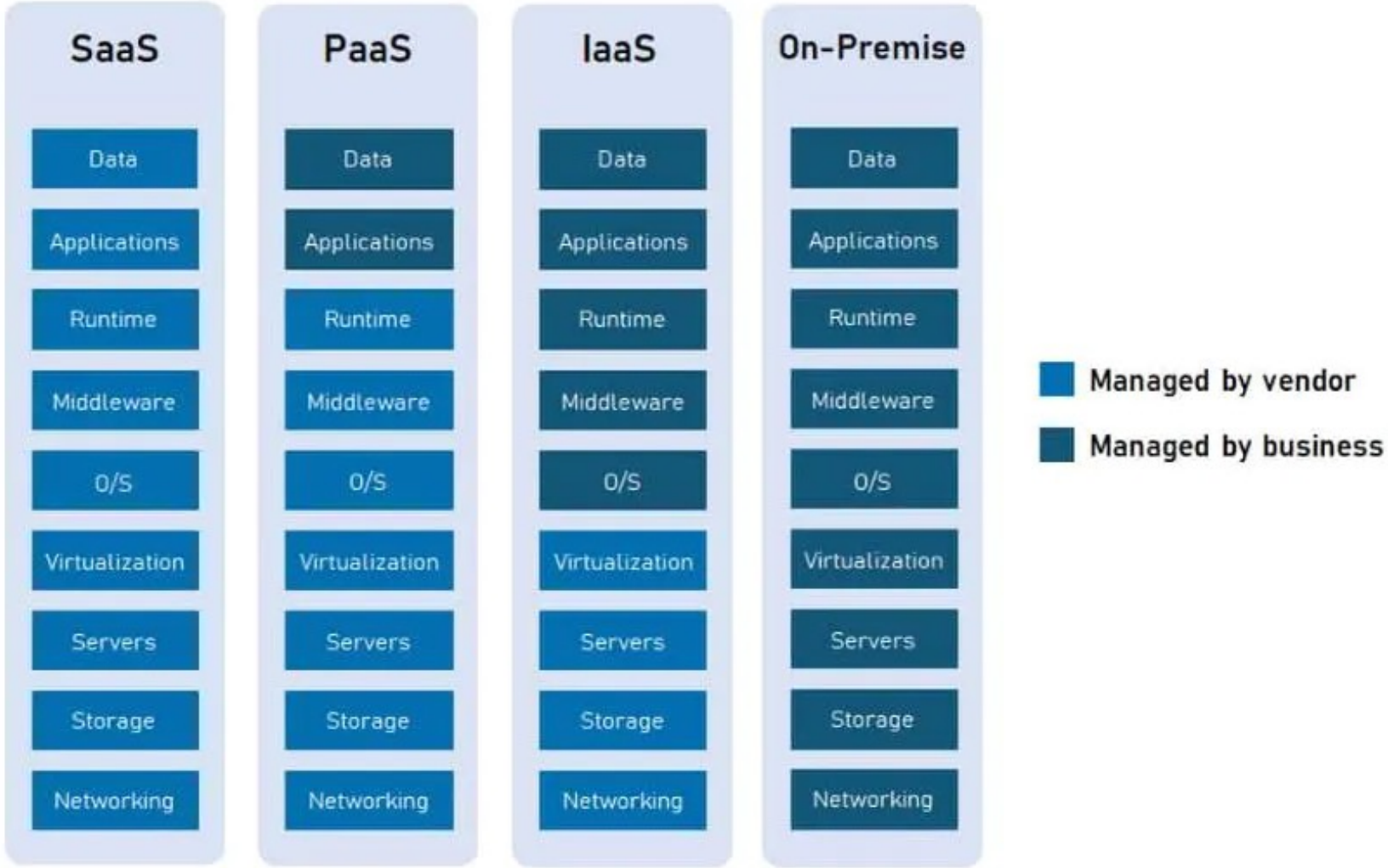


Data center

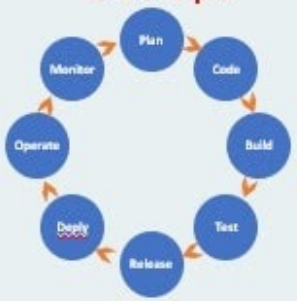
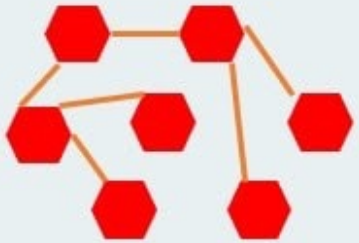
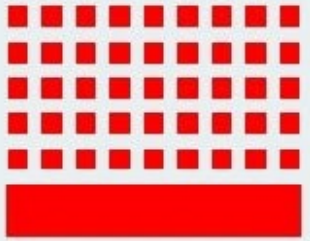
- The organization needs to provide and manage the required power cooling and real estate resources for maintaining the data center.
- The organization could be developing some apps in-house for this to happen, the developers need some development platforms to develop test, and run their software. This needs to be managed by the IT organization meaning new hardware and software assets to procure setup and manage.
- This familiar picture has been so common to us and in use around the world today. With the rising popularity of cloud computing in recent years, this picture is changing in multiple ways.



Cloud services vs On-Premise

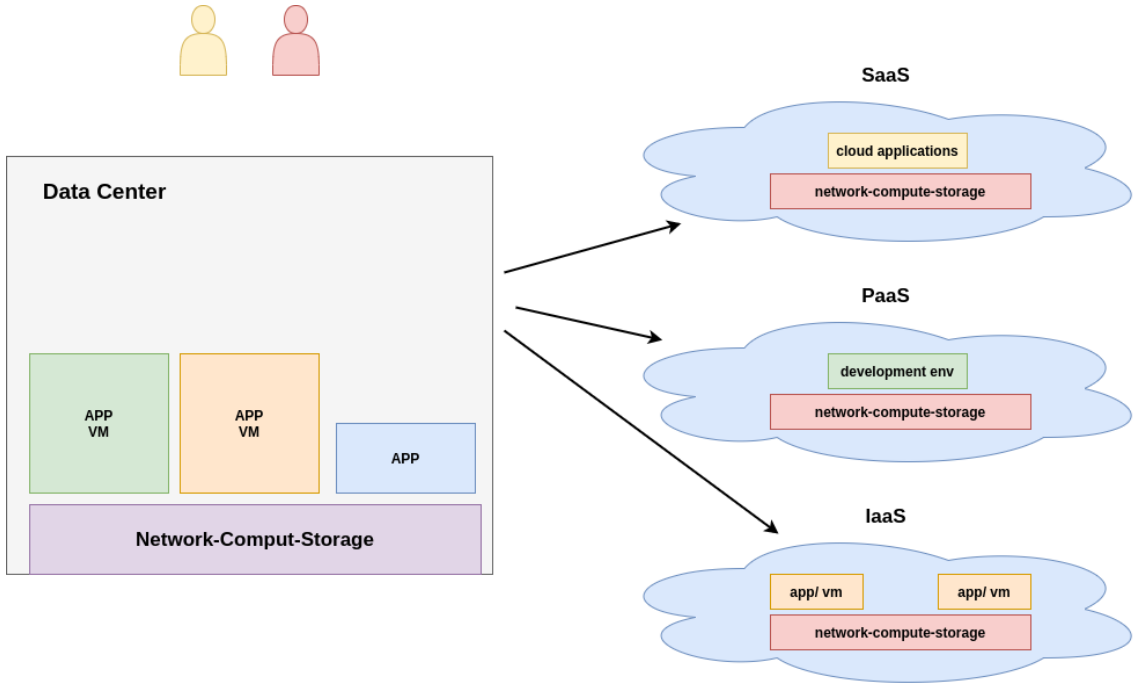


Evolution of software development

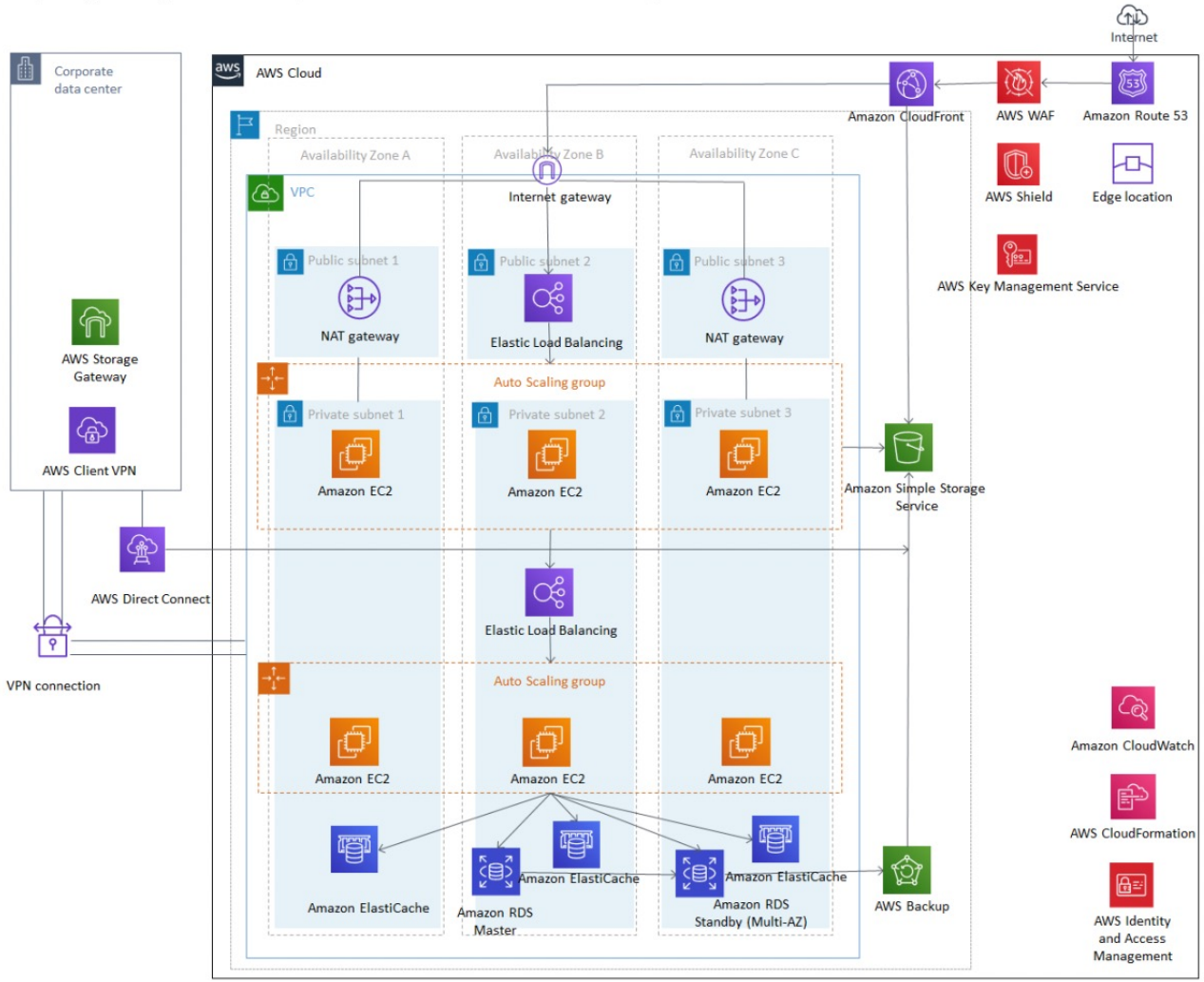
	Development Process	Application Architecture	Deployment and Packaging	Application Infrastructure
~ 1980 ~ 1990	Waterfall 	Monolithic 	Physical Server 	Datacenter 
~ 2000	Agile 	N-Tier 	Virtual Servers 	Hosted 
~ 2010 Now	DevOps 	Microservices 	Containers 	Cloud 

Benefit from the cloud without doing any upfront investment

- First, organizations can use cloud applications commonly called **Software As A Service (SaaS)** which really means applications running at data centers owned by a cloud operator and accessed via the internet.
- Organizations can also use cloud development platforms which are commonly called **Platform as a Service(PaaS)** for their software developers.
- Developers could consume the tools and the development environment they need from the cloud.
- The third way organizations can leverage the cloud is by running some of their applications on computing resources at data centers across the Internet. This is commonly called the **Infrastructure as a Service(IaaS)**consumption model.



AWS for classic n-tier Web application



AWS Terms

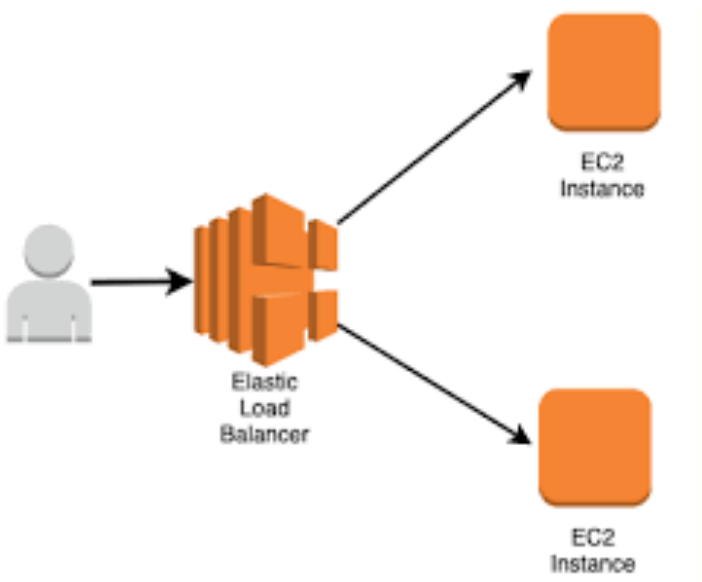
AWS has the concept of a **Region**, which is a physical location around the world where we cluster data centers.

We call each group of logical data centers an **Availability Zone**. Each AWS Region consists of a minimum of three, isolated, and physically separate AZs within a geographic area.

A **VPN** or Virtual Private Network creates a private network connection between devices through the internet. VPNs are used to safely and anonymously transmit data over public networks. They work by masking user IP addresses and encrypting data so it's unreadable by anyone not authorized to receive it.

Amazon **CloudFront** is a content delivery network operated by Amazon Web Services. Content delivery networks provide a globally-distributed network of proxy servers that cache content, such as web videos or other bulky media, more locally to consumers, thus improving access speed for downloading the content.

Load balancer



- Many businesses use a proxy server to route and secure traffic between the networks.
- A load balancer acts as the “traffic cop” sitting in front of your servers and routing client requests across all servers capable of fulfilling those requests in a manner that maximizes speed and capacity utilization and ensures that no one server is overworked, which could degrade performance.
- A load balancer is **a device that acts as a reverse proxy and distributes network or application traffic across a number of servers**. Load balancers are used to increase capacity (concurrent users) and reliability of applications.
- **Elastic Load Balancing (ELB)** is a load-balancing service for Amazon Web Services (AWS) deployments. ELB automatically distributes incoming application traffic and scales resources to meet traffic demands.

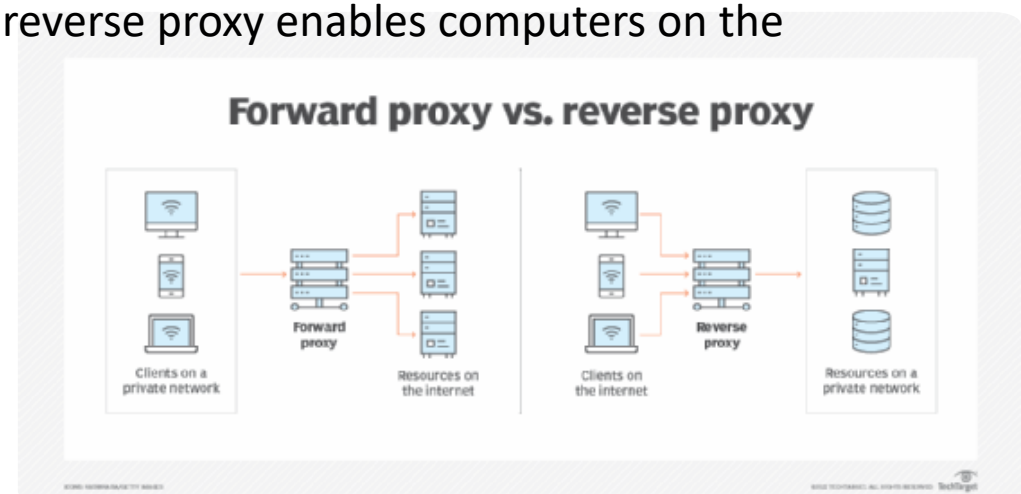
Forward vs. reverse proxies

Forward and reverse proxies secure and isolate resources that reside on a private network, but they play different roles in modern enterprise architectures.

The biggest similarity between a forward and reverse proxy is that they both protect devices connected to a private network against threats from the internet and other external networks.
Both forward and reverse proxies can limit the types and sizes of files that pass through them and disallow users who have not authenticated to send requests through them.

Both forward and reverse proxies can perform port and protocol switching, which can further disguise the access patterns used to access resources hidden behind them.

The key difference between a reverse proxy and a forward proxy is that a forward proxy enables computers isolated on a private network to connect to the public internet, while a reverse proxy enables computers on the internet to access a private subnet.



AWS Terms

EC2, Amazon Elastic Compute Cloud is a part of Amazon.com's cloud-computing platform, Amazon Web Services, that allows users to rent virtual computers on which to run their own computer application.

Amazon Elastic Block Store (Amazon EBS) is an easy-to-use, scalable, high-performance block-storage service designed for Amazon Elastic Compute Cloud (Amazon EC2).

Amazon S3 is object storage built to store and retrieve any amount of data from anywhere. It’s a simple storage service that offers industry leading durability, availability, performance, security, and virtually unlimited scalability at very low costs.

S3

Amazon S3 is a simple key-based object store. When you store data, you assign a unique object key that can later be used to retrieve the data. Keys can be any string, and they can be constructed to mimic hierarchical attributes. Alternatively, you can use S3 Object Tagging to organize your data across all of your S3 buckets and/or prefixes.

Amazon S3 gives any developer access to the same highly scalable, highly available, fast, inexpensive data storage infrastructure that Amazon uses to run its own global network of web sites. The S3 Standard storage class is designed for 99.99% availability, the S3 Standard-IA storage class, S3 Intelligent-Tiering storage class, and the S3 Glacier Instant Retrieval storage classes are designed for 99.9% availability, the S3 One Zone-IA storage class is designed for 99.5% availability, and the S3 Glacier Flexible Retrieval and S3 Glacier Deep Archive class are designed for 99.99% availability and an SLA of 99.9%.

Amazon S3 was designed from the ground up to handle traffic for any internet application. Pay-as-you-go pricing and unlimited capacity ensures that your incremental costs don’t change and that your service is not interrupted. Amazon S3’s massive scale lets you spread the load evenly, so that no individual application is affected by traffic spikes.

Main Cloud players

	aws	Azure	Google Cloud
Compute:	EC2 Instance	Virtual Machine (VM)	Compute Engine VMs
Object Storage:	S3	Blob Storage	Cloud Storage
Logical Data Centre:	VPC	VNet	VPC
Private Connectivity (L2):	Direct Connect	ExpressRoute	Interconnect
Gateways:	TGW, VGW, DGW	VNet Gateway	Cloud Router

Cloud provides agility and speed with economies of scale benefits

No need to say organizations will still have some applications that run locally. Not every service is consumed from the cloud. There could be different motivations behind that like data security, performance cost, and so on.

But still, the cloud is a viable alternative for writing some of the services and it is here to stay. The biggest promise of cloud computing is the promise of agility and effectiveness coupled with the economies of scale provided by large cloud providers.

Instead of waiting for weeks or months to roll applications users can now access information and applications from a cloud provider.

The organization might build new applications themselves on a cloud platform. At the same time, they could run those in-house built applications or the applications they license from a third party.

IaaS, PaaS, and SaaS are the 3 delivery models of cloud computing.

Cloud computing

- It’s a type of internet-based computing that provides shared computer processing resources and data to any type of consumer device on demand.
- It is a model for enabling ubiquitous on-demand access to a shared pool of configurable computing resources. Computer resources could be computer networks, servers, storage, applications, and services that can be rapidly provisioned and released with minimal management efforts.

On-demand self-service

- The first one is having the On-Demand self-service capability, so a consumer can provision computing capabilities such as server network storage or an application by him or herself without requiring human interaction with the service provider.

Broad network access

- This emphasizes the fact that services should be available over the network and accessed through standard mechanisms. Any client platform available like mobile phones, tablets, laptops, or high-end servers could be used for that.

Resource pooling

- Resource pooling means the providers' computing resources are pooled to serve multiple consumers using a multi-tenant model. Different physical and virtual resources could be dynamically assigned and reassigned according to consumer demand.
- There is a sense of location independence in that the consumer generally has no control or knowledge over the exact location of the provided resources but may be able to specify a location at a higher level of abstraction.
- For example country state or data center. Examples of resources include storage, processing capacity, memory, and network bandwidth.

Rapid elasticity

- Meaning resources can be elastically provisions and released in some cases. This is automatic. This provides the ability to scale rapidly outward and inward along with the demand to the consumer.
- The capabilities available for provisioning often appear to be unlimited and can be changed in any quantity at any time.

Measured service

- Cloud systems automatically control and optimize resource usage by leveraging a metering capability at some level of abstraction appropriate to that type of service.
- Depending on the service metered, resources could be storage capacity, processing capacity, bandwidth and active user counts. Resource usage can be monitored controlled and reported, providing transparency for both the provider and consumer of the utility service in the end.

Cloud scaling strategy

Implementing and managing a cloud scaling strategy is:

- Convenient: You can easily increase or decrease storage capacity as needed.
- Flexible and fast: You can quickly respond to changing demands to keep customers up and running without delays in service.
- Cost-effective: You don’t have to pay for expensive hardware or provide the space to store it.
- Fault-tolerant: Resources can automatically be scaled to accommodate redundancies and to facilitate disaster recovery.
- Time-saving: Upgrading existing hardware and installing new hardware on-site can be very time-consuming. Cloud computing can take care of the scaling for you. This frees you up to focus on innovation and process improvement rather than troubleshooting errors and other issues.

Conclusion

- Cloud computing provides individual users and organizations with various capabilities to store and process their data in third-party data centers that may be located anywhere far from the user ranging in the distance from across the city to across the world.
- Cloud computing relies on sharing of resources to achieve coherence and economy of scale similar to a utility like the electricity grid over an electricity network.