CLOUD COMPUTING

cloud computing?

➢ cloud computing is the delivery of computing services like servers, storages and more over the Internet. The companies that offer these computing services are called cloud providers. They charge for **cloud computing services** based on usage.

➢ Cloud computing is usually classified on the basis of location, or on the service that the cloud is offering.

➢ Based on a cloud location, we can classify cloud as:

    1. Public Cloud

    2. Private Cloud

    3. Hybrid Cloud

1. **Public Cloud** – Whole computing infrastructure is located on the premises of a cloud computing company that offers the **cloud service.**

2. **Private Cloud** – Hosting all your computing infrastructure yourself and is not shared. The security and control level is highest while using a private network.

3. **Hybrid Cloud** – using both private and public clouds, depending on their purpose. You host your most important applications on your own servers to keep them more secure and secondary applications elsewhere.

Based on a service that the cloud is offering, we classify as:

1. IaaS (Infrastructure-as-a-Service):
IaaS is the most basic category of cloud computing services that allows you rent IT infrastructure (servers or VM's) from a cloud provider on a pay-as-you-go basis

2. PaaS(Platform-as-a-Service):

Platform-as-a-service (PaaS) refers to the supply an on-demand environment for developing, testing, delivering and managing software applications. It is designed to quickly create web or mobile apps, without worrying about setting up or managing the underlying infrastructure of servers, storage, network and databases needed for development.

3. SaaS(Software-as-a-Service):

Software-as-a-service (SaaS) is a method for delivering software applications over the Internet as per the demand and on a subscription basis. SaaS helps you host and manage the software application and underlying infrastructure and handle any maintenance (software upgrades and security patching).
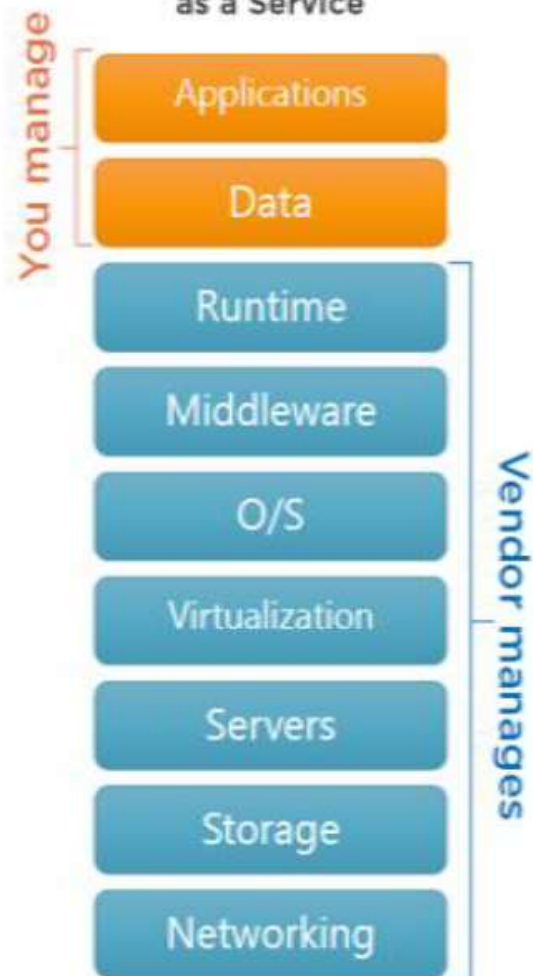
On-premises

- Applications
- Data
- Runtime
- Middleware
- O/S
- Virtualization
- Servers
- Storage
- Networking

You manage

Infrastructure
as a Service

- Applications
- Data
- Runtime
- Middleware
- O/S

You manage

- Virtualization
- Servers
- Storage
- Networking

Vendor manages

Platform
as a Service

- Applications
- Data

You manage

- Runtime
- Middleware
- O/S
- Virtualization
- Servers
- Storage
- Networking

Vendor manages

Software
as a Service

- Applications
- Data
- Runtime
- Middleware
- O/S
- Virtualization
- Servers
- Storage
- Networking

Vendor manages
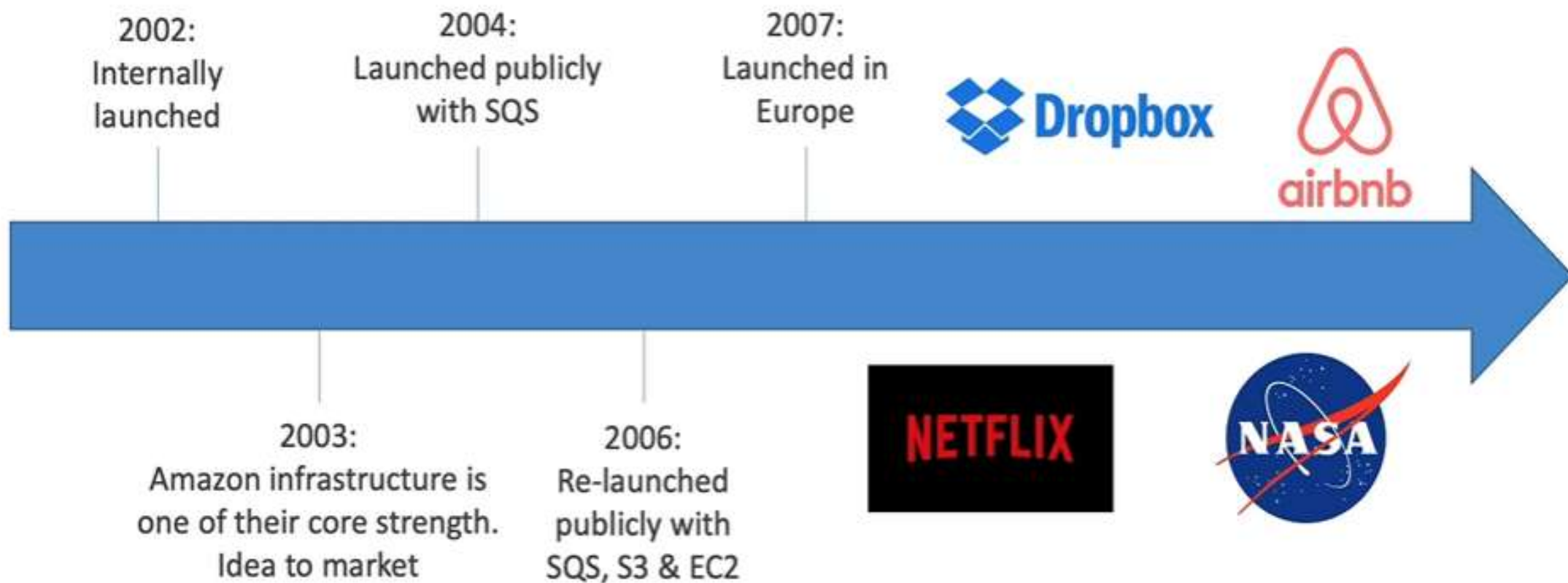
List of top 10 cloud service providers:

1. Microsoft Azure
2. Amazon Web Services (AWS)
3. Google Cloud
4. Alibaba Cloud
5. IBM Cloud
6. VMWare
7. Oracle
8. Salesforce
9. SAP
10. Rackspace Cloud

Gartner Survey:

	AWS	Azure	Google Cloud
Company	AWS Inc.	Microsoft	Google
Launch year	2006	2010	2008
Geographical Regions	25	54	21
Availability Zones	78	140 (countries)	61
Key offerings	Compute, storage, database, analytics, networking, machine learning, and AI, mobile, developer tools, IoT, security, enterprise applications, blockchain.	Compute, storage, mobile, data management, messaging, media services, CDN, machine learning and AI, developer tools, security, blockchain, functions, IoT.	Compute, storage, databases, networking, big data, cloud AI, management tools, Identity and security, IoT, API platform
Compliance Certificates	46	90	
Annual Revenue	$33 billion	$35 billion	$8 billion

Source: Gartner (July 2021)

- AWS enables you to build sophisticated, scalable applications

- Applicable to a diverse set of industries

- Use cases include
    - Enterprise IT, Backup & Storage, Big Data analytics
    - Website hosting, Mobile & Social Apps
    - Gaming

AWS Global Infrastructure

- AWS Regions

- AWS Availability Zones

- AWS Data Centers

- AWS Edge Locations / Points of Presence

- https://infrastructure.aws/
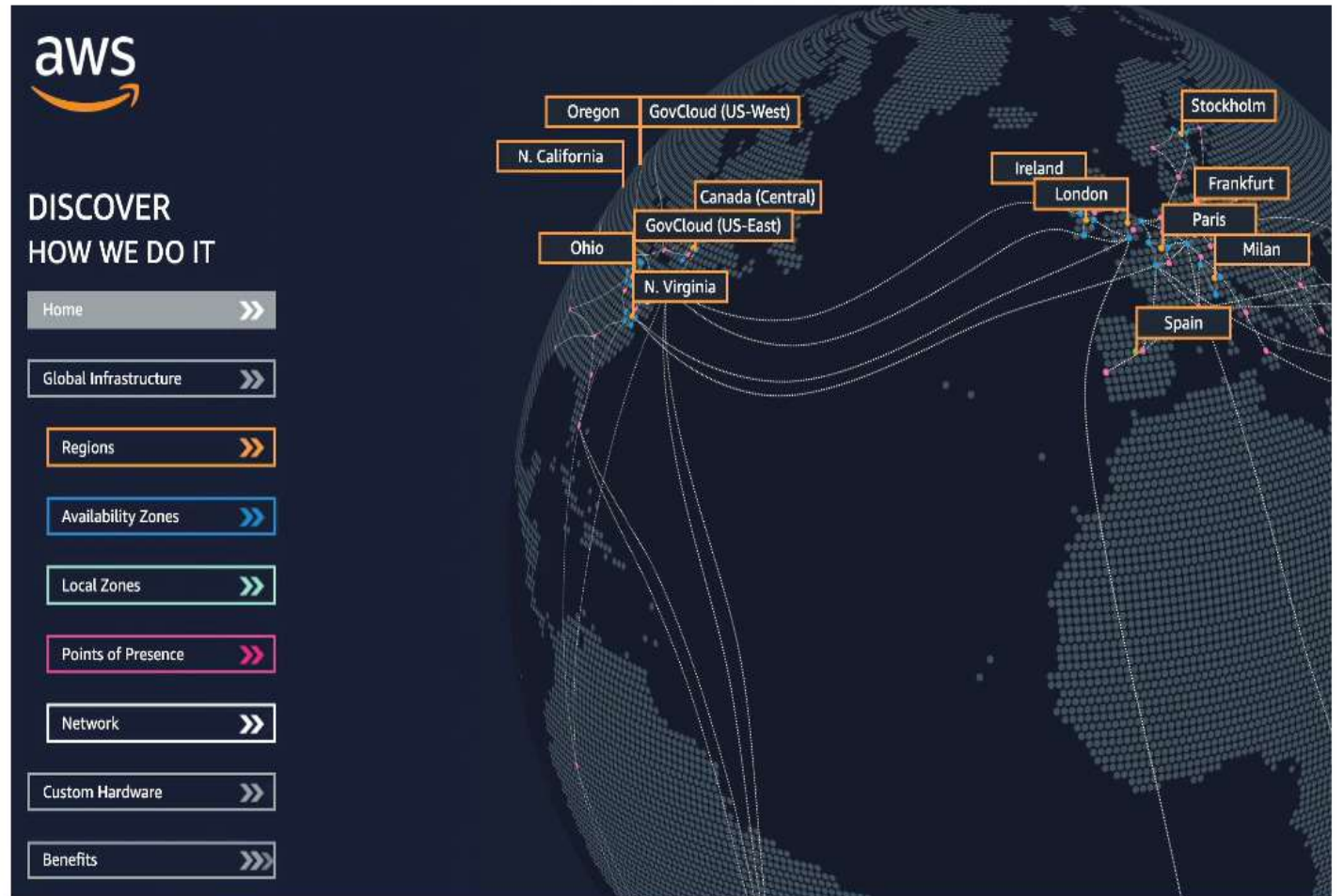
Regions and Availability Zones:

- 27 Regions

- Each Region has multiple availability zones.

- Total 84 availability zones
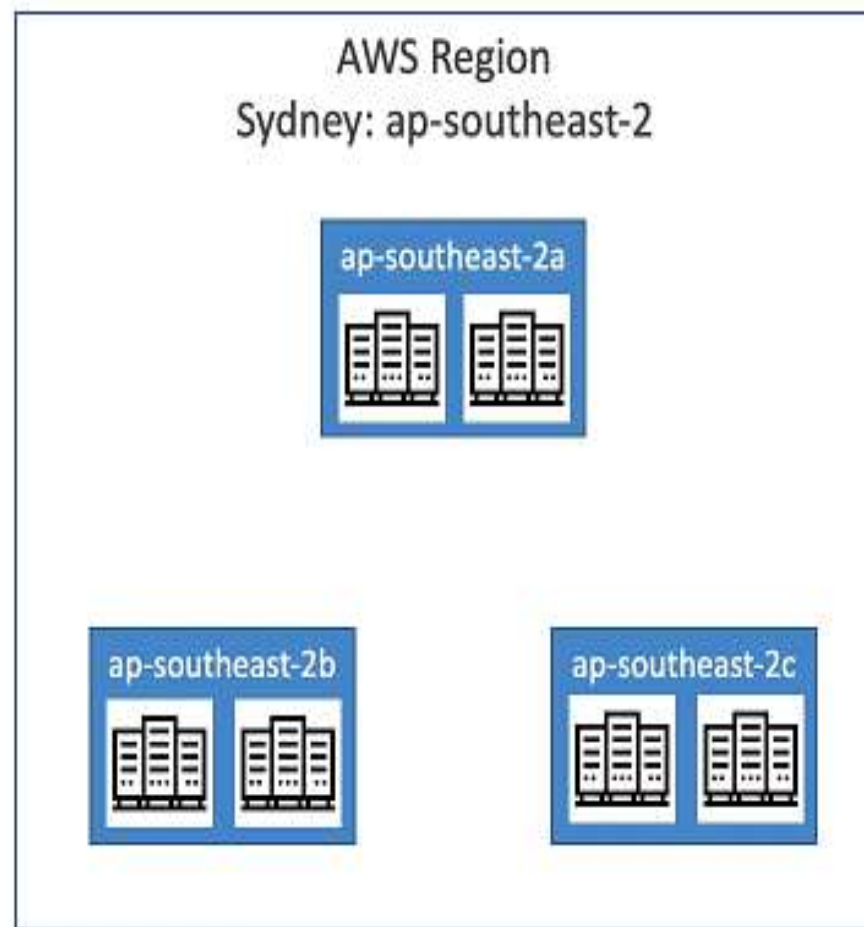
How to choose region?

If you need to launch a new application, where should you do it?

- Compliance with data governance and legal requirements: data never leaves a region without your explicit permission

- Proximity to customers: reduced latency

- Available services within a Region: new services and new features aren't available in every Region

- Pricing: pricing varies region to region and is transparent in the service pricing page

- Each region has many availability zones (usually 3, min is 2, max is 6). Example:
  - ap-southeast-2a
  - ap-southeast-2b
  - ap-southeast-2c

- Each availability zone (AZ) is one or more discrete data centers with redundant power, networking, and connectivity

- They're separate from each other, so that they're isolated from disasters

- They're connected with high bandwidth, ultra-low latency networking

AWS Region
Sydney: ap-southeast-2

THANK YOU