Cloud Computing

Session 6

Infrastructure As A Service (IaaS)

BITS Pilani

Shwetha Vittal

IaaS

Really, what is IaaS???

Agenda

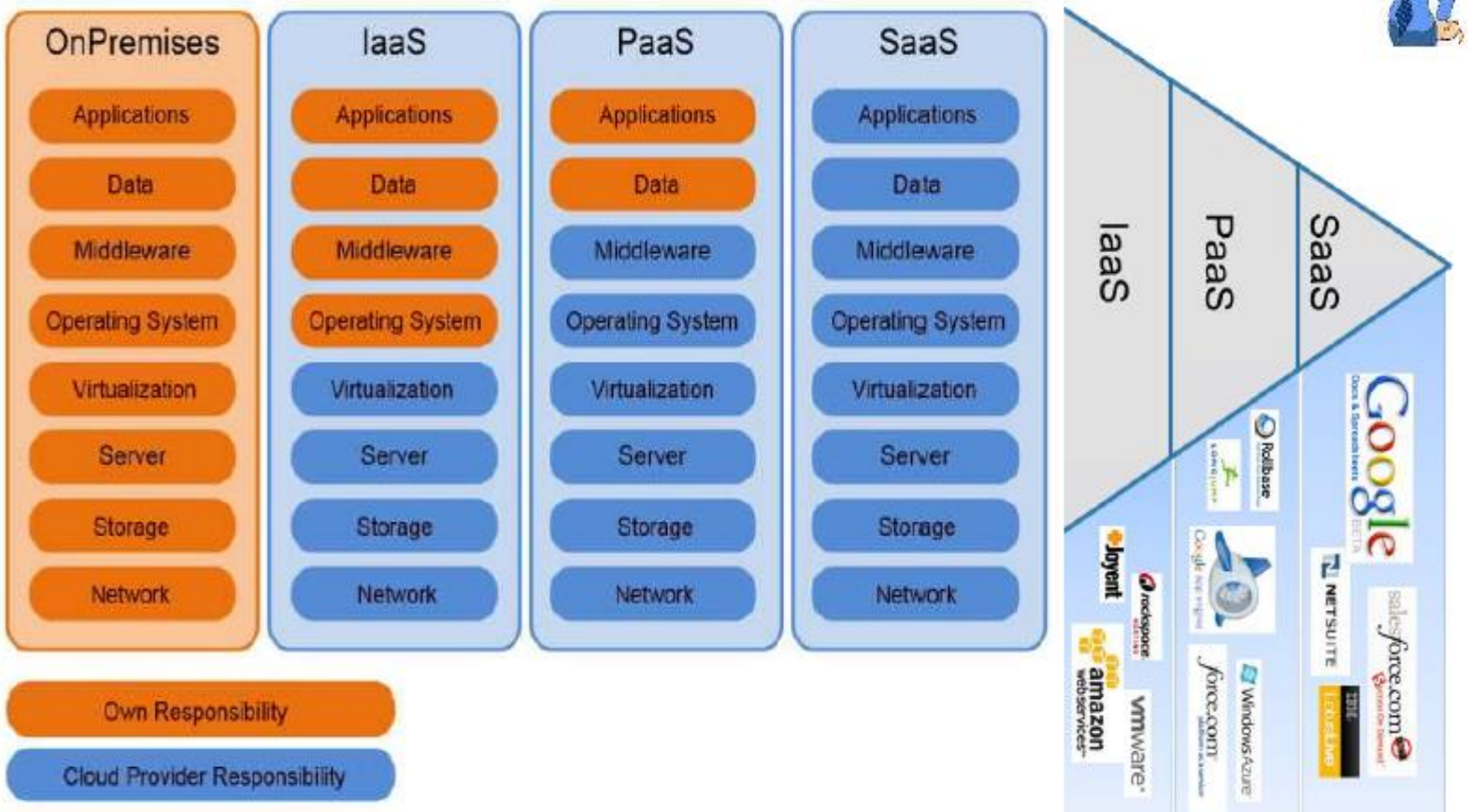
- Key Concepts & Functions of IaaS
- Introduction to AWS
- AWS Reference Model, Services
- AWS Region Versus Availability Zones
- AWS Shared Responsibility Model
- AWS IAM

Revisit 3-4-5 Rule of Cloud Computing

- Cloud computing is the **on-demand** delivery of compute power, database, storage, applications, and other IT resources via the **internet** with **pay-as you-go** pricing.
- 3-4-5 Rule of Cloud Computing
  - 3 Service Models
    - IaaS
    - PaaS
    - SaaS

Heard of 3 models of Cloud Computing?

Yes, Yes, IaaS, PaaS and SaaS



Cloud Service Models

More control over IT resources

Less control over IT resources

Key Components of IaaS

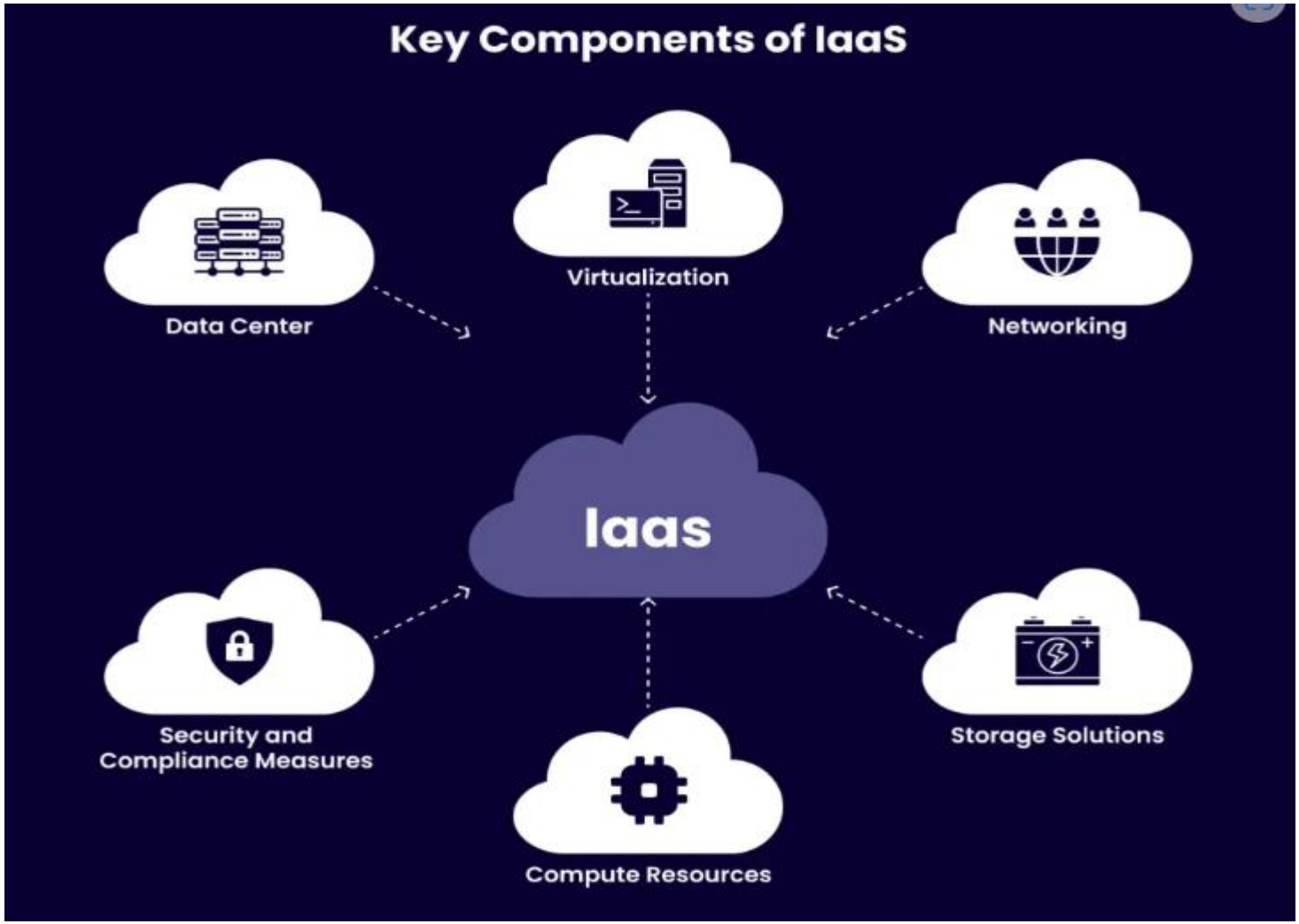


Image Courtesy: acecloud.ai

Key Functions of IaaS

1. Hypervisor - Virtualization
2. Resource pooling
3. Multi-tenant computing
4. Cloud bursting

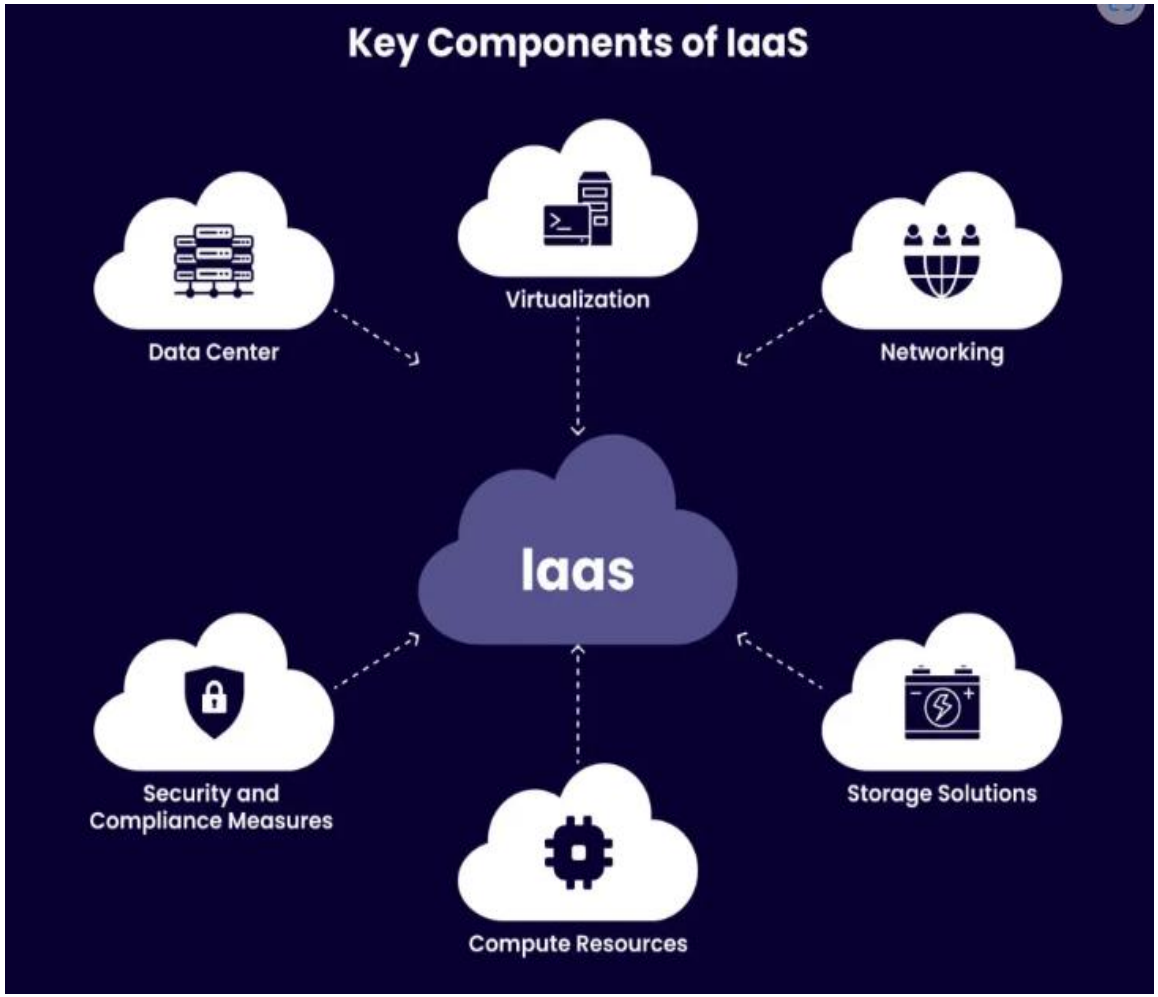
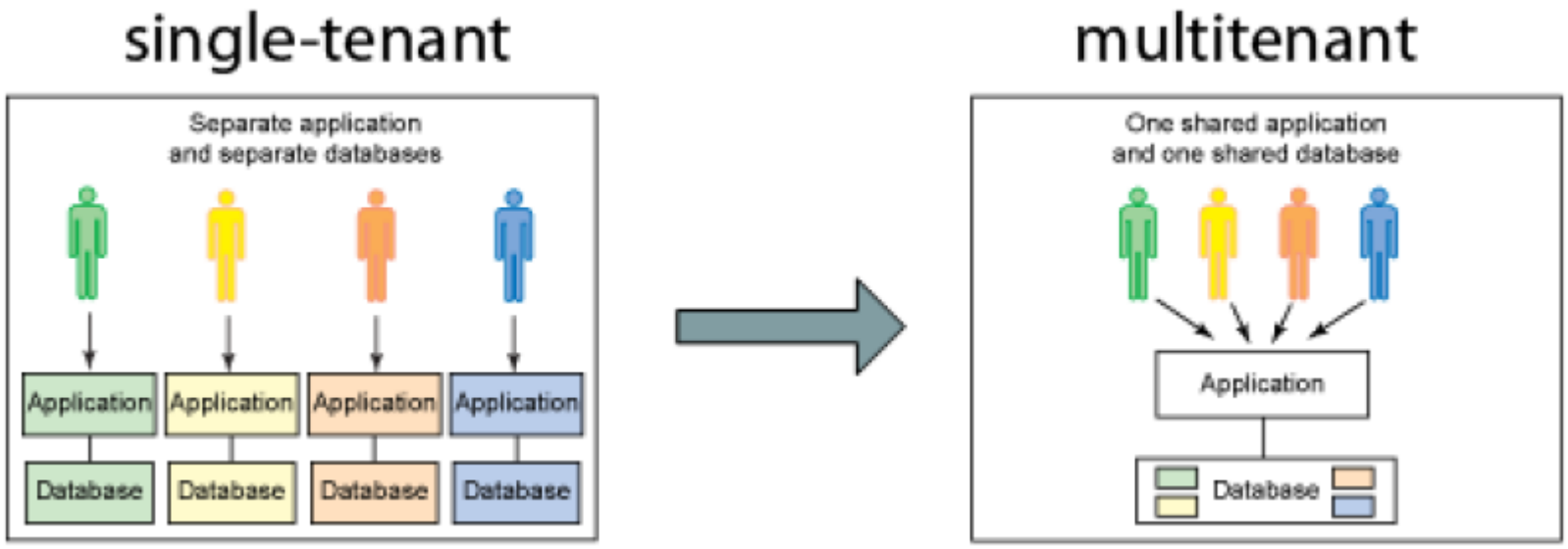


Image Courtesy: acecloud.ai

Multi-Tenant Computing



- **Single tenant Computing:** Separate application and databases per tenant.
- **Multi-tenant Computing:** Common application and database shared by multiple tenants.

Cloud Bursting

- The process of off-loading tasks to the cloud during times when the most compute resources are needed
- IT departments must be able to build and implement the software that handles the ability to re-allocate processes to an IaaS cloud.
- **Important considerations** to build and implement software that can manage such reallocation processes.

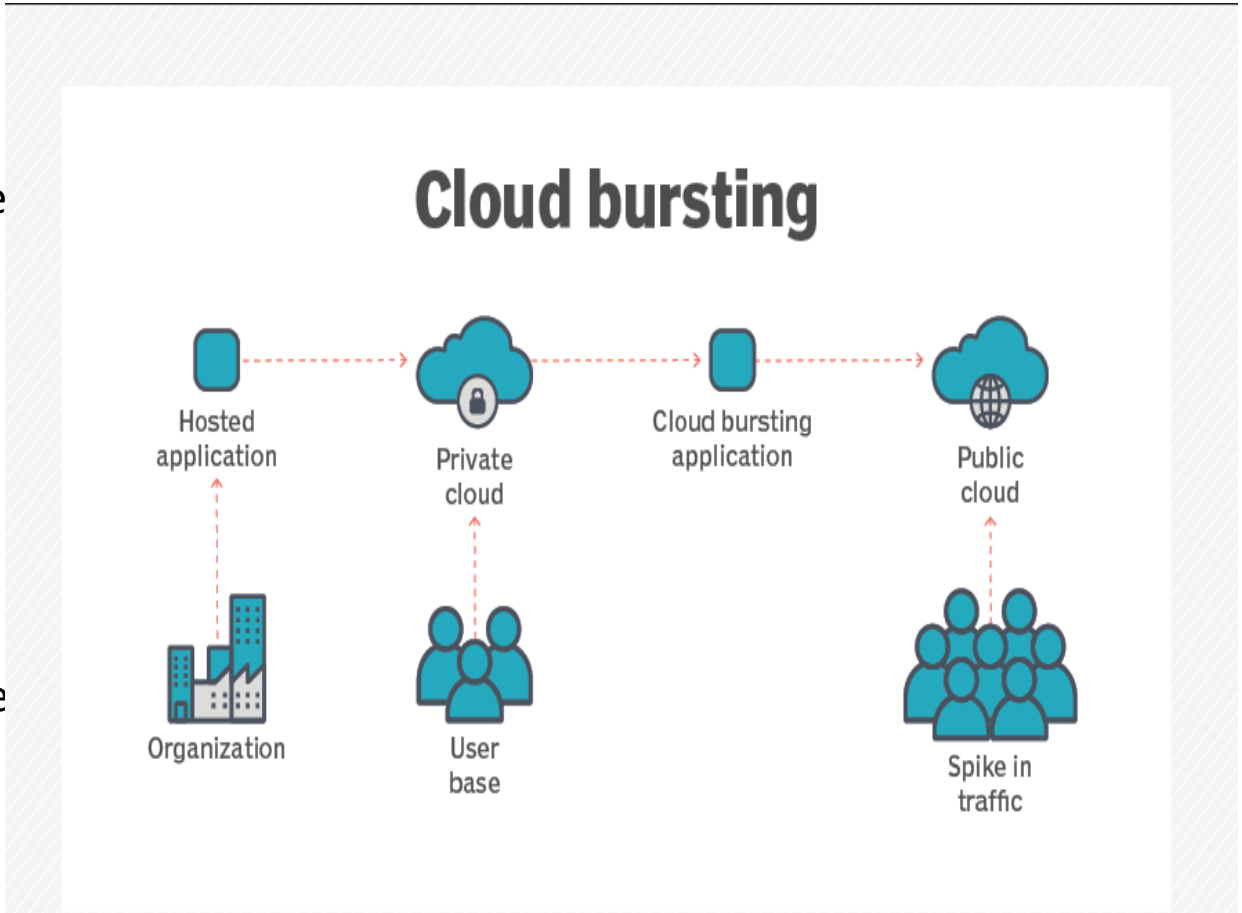
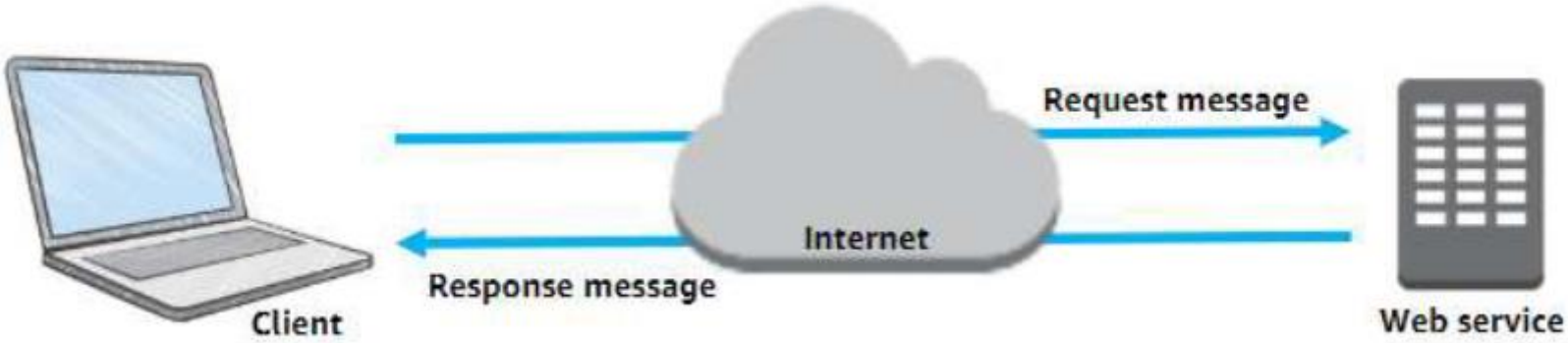


Image Courtesy: TechTarget

Key Considerations for Cloud Bursting

- Developing for a specific vendor's proprietary IaaS could prove to be a costly mistake
- The complexity of well-written resource allocation software is significant and do not come cheap
- What will you be sending off to be processed in the cloud?
  - Sending data such as personal identities, financial information, and health care data put an organization’s compliance at risk
  - Understand the dangers of shipping off processes that are critical to the day-to-day operation of the business

What are Web Services?

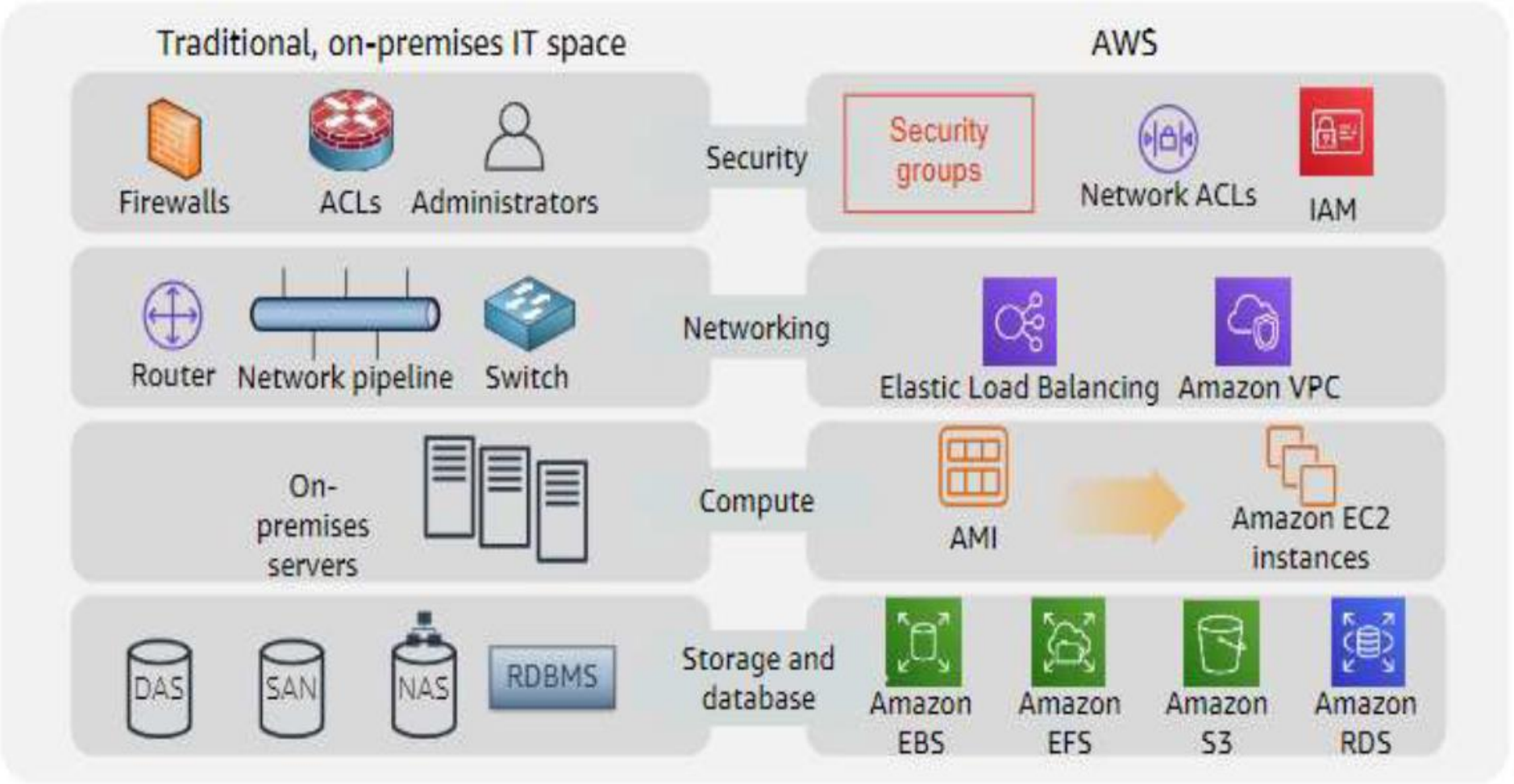


- A web service is any piece of software that makes itself available over the internet
- Uses a standardized format for the request and the response of an Application Program ming Interface(API) interaction
  - Extensible Markup Language (XML)
  - JavaScript Object Notation (JSON)

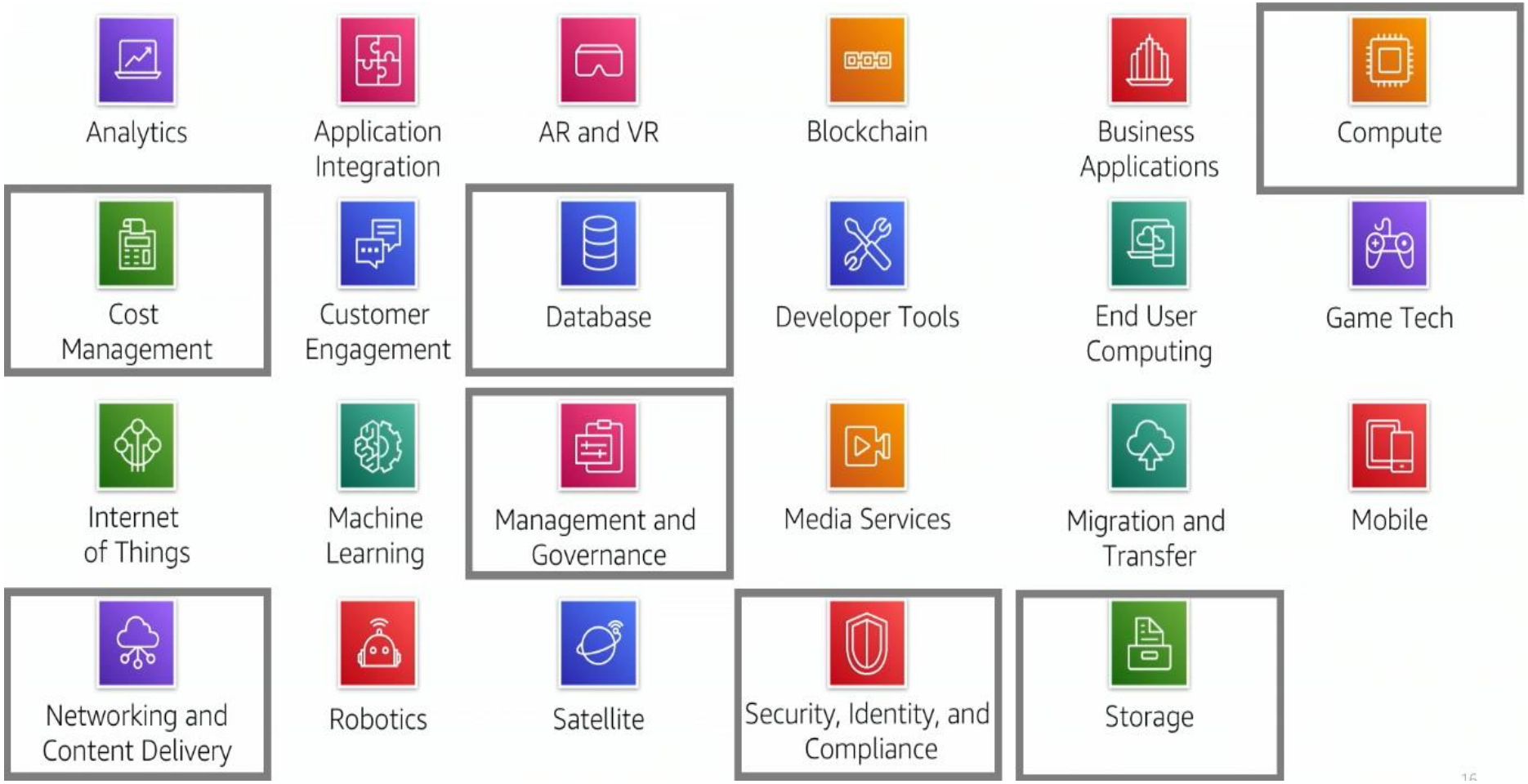
Amazon Web Services Cloud

- A secure cloud platform that offers a broad set of global cloud-based products
- Provides highly reliable and scalable infrastructure for deploying web-scale solutions
- More flexibility than own infrastructure, either on premise or at a data center facility
- AWS services work together like building blocks

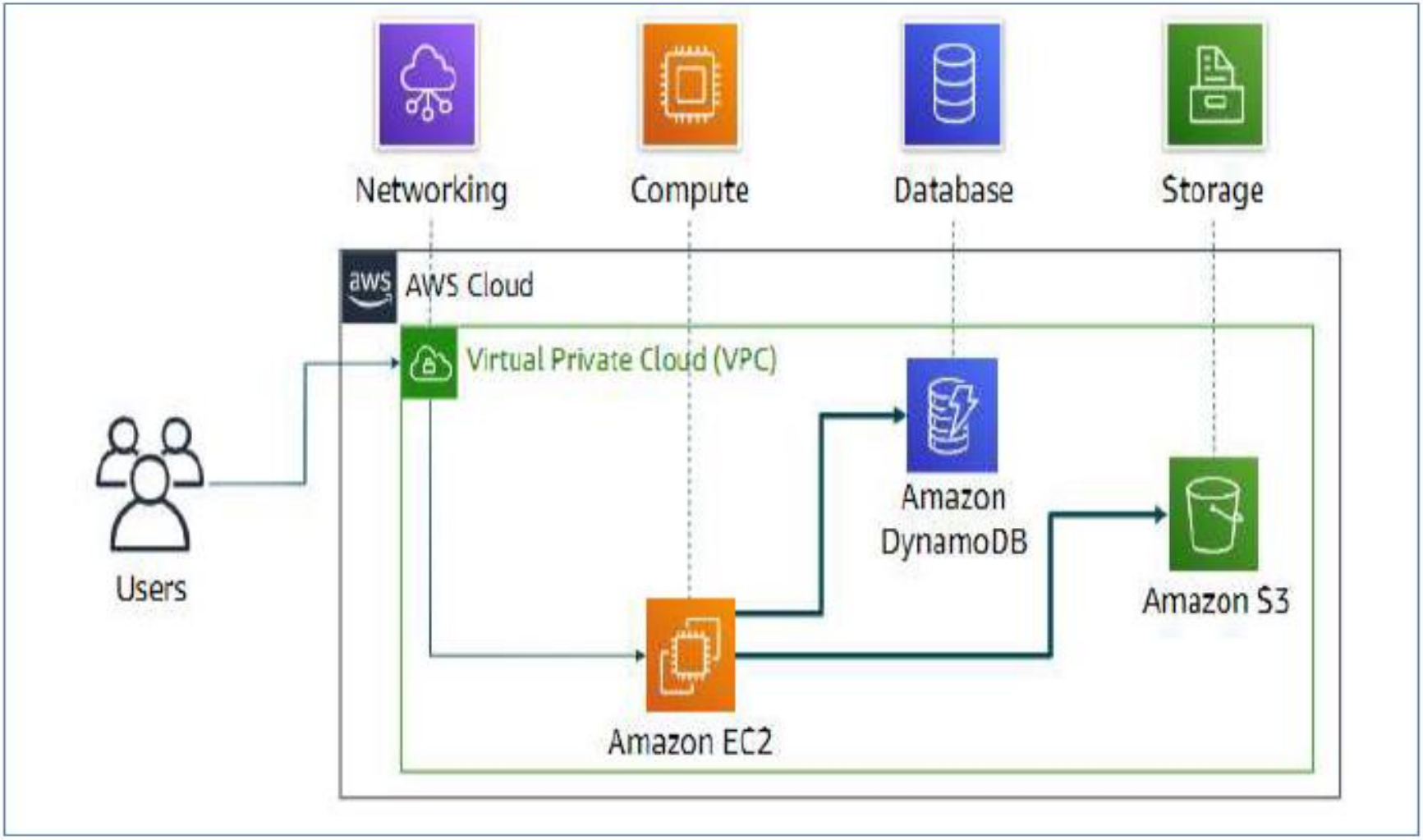
Similarities between AWS and Traditional IT



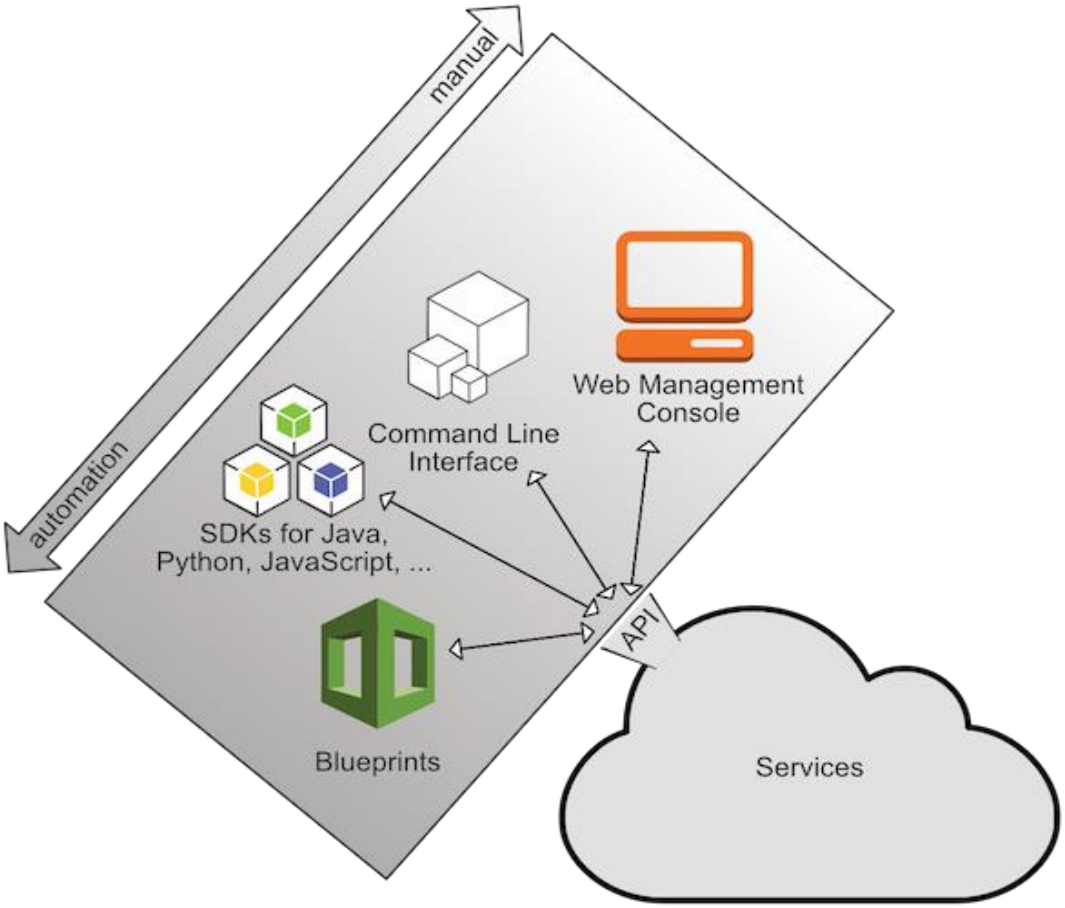
Categories of AWS services



Simple Solution Example



Three Ways to Interact with AWS



1. AWS Management Console - Easy-to-use graphical interface
2. Command Line Interface (AWS CLI) - Access to services by discrete commands or scripts
3. Software Development Kits (SDKs) Access services directly from your code (such as Java, Python, and others)

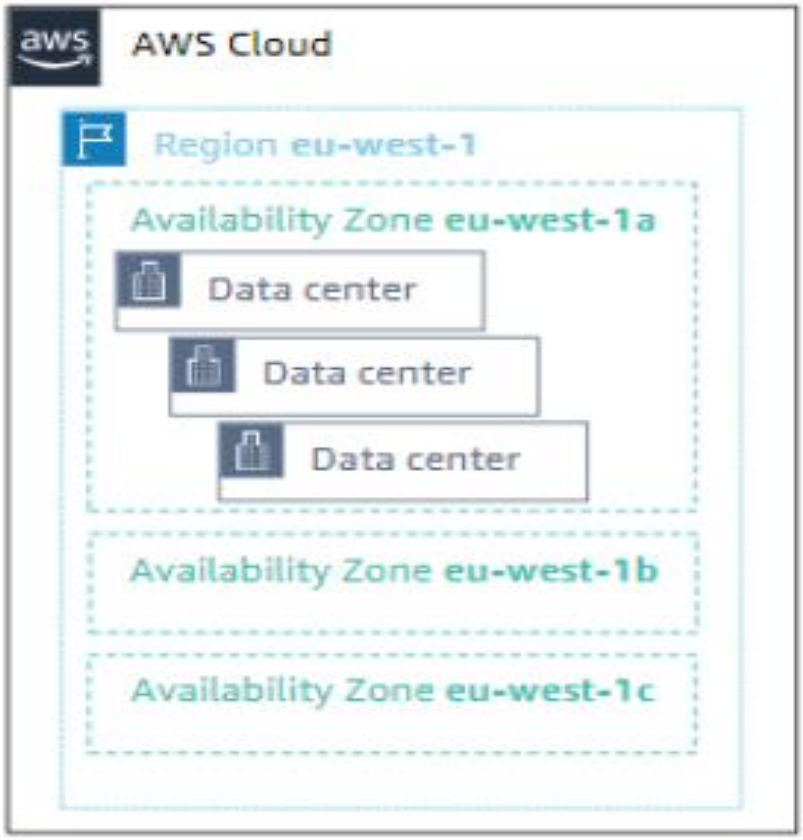
AWS Global Infrastructure

https://aws.amazon.com/about-aws/globalinfrastructure/#AWS_Global_Infrastructure_Map

1. Regions
2. Availability Zones
3. Data Centers

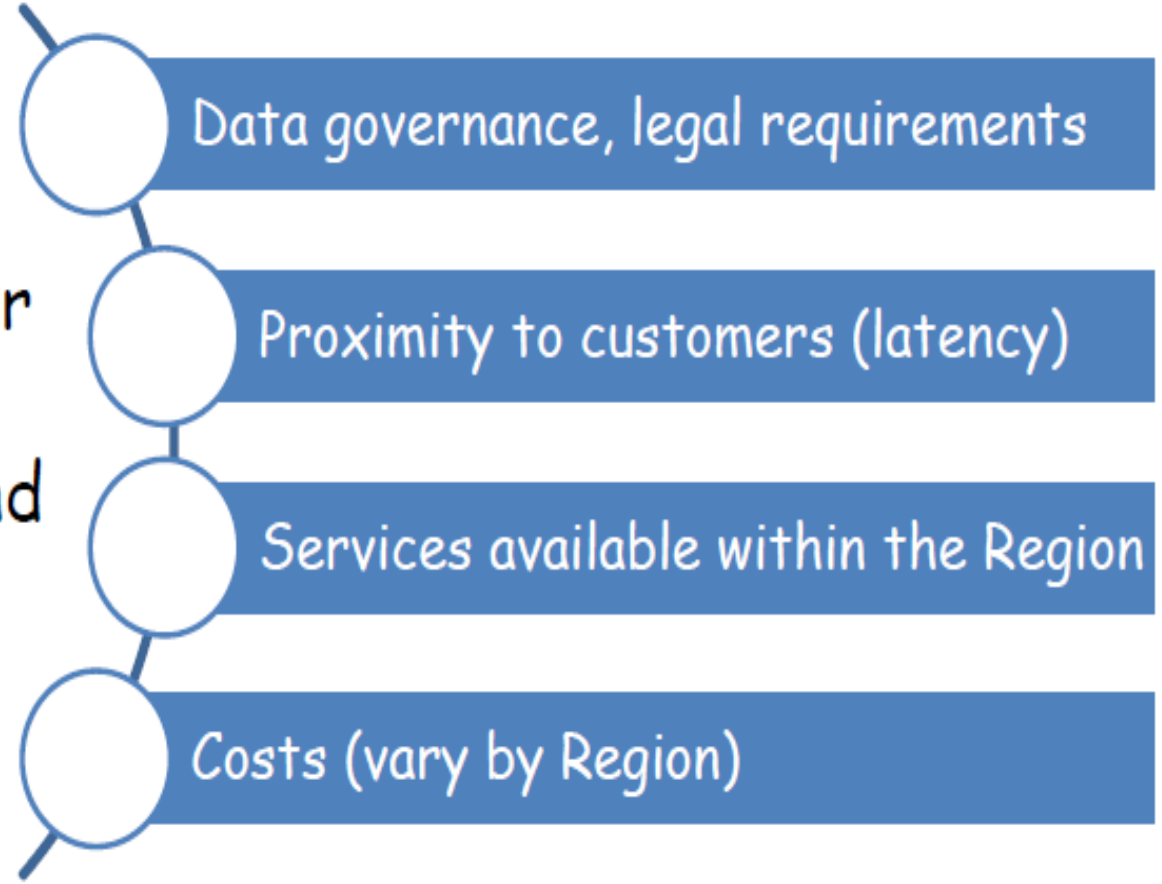
Region

- An AWS Region is a geographical area.
- Data replication across Regions is controlled by you.
- Communication between Regions uses AWS backbone network infrastructure.
- Each Region provides full redundancy and connectivity to the network.
- A Region typically consists of two or more Availability Zones.

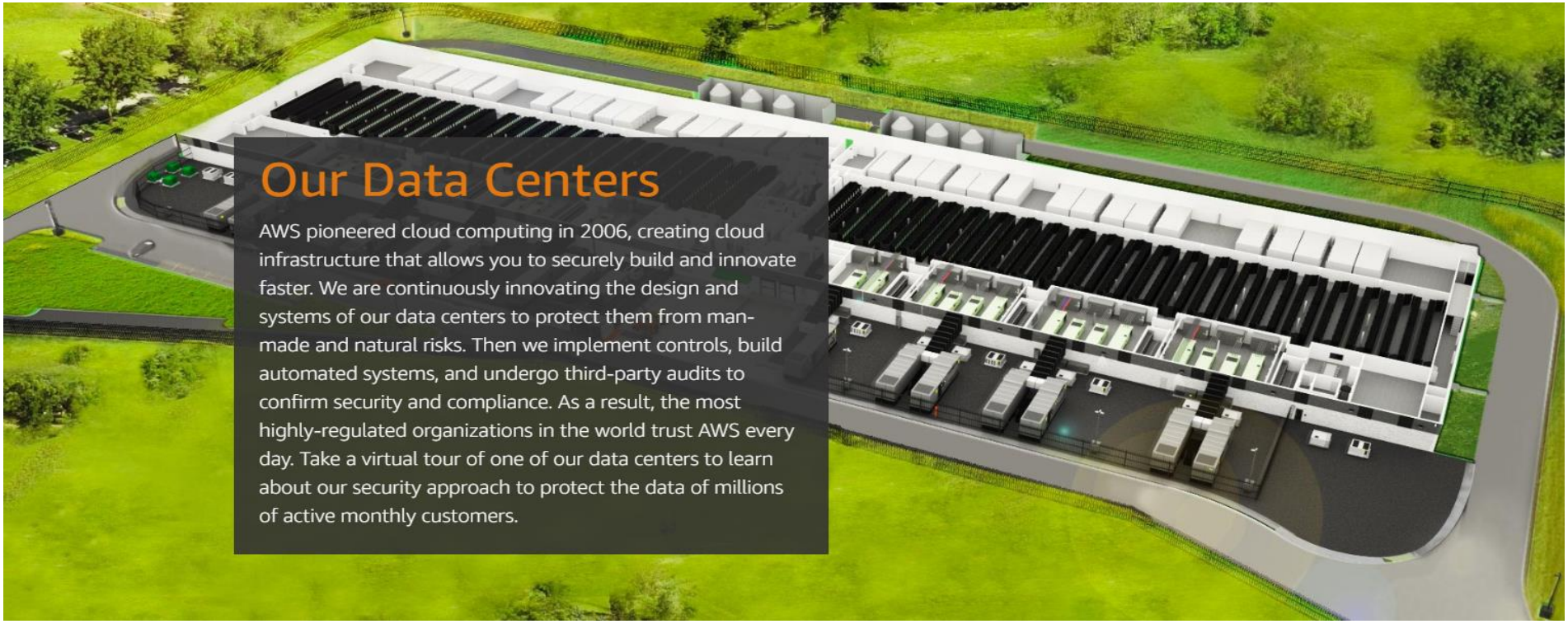


Region

Determine the right Region for your service, applications, and data based on these factors



AWS Data Centers



AWS Data Centers - Our Data Centers

- Data centers are where the data resides and data processing occurs.
- Each data center has redundant power, networking, and connectivity, and is housed in a separate facility.
- AWS data centers are also designed for security.
- A data center typically has 50,000 to 80,000 physical servers.

AWS Compute Services

Instance


Amazon EC2


Amazon EC2 Spot


Amazon EC2 Autoscaling


Amazon Lightsail


AWS Batch

Containers


Amazon ECS


Amazon ECR


Amazon EKS


AWS Fargate

Serverless


AWS Lambda

Edge and hybrid


AWS Outposts


AWS Snow Family


AWS Wavelength


Vmware Cloud on AWS


AWS Local Zones

Cost and capacity management


AWS Savings Plan


AWS Compute Optimizer


AWS Elastic Beanstalk


EC2 Image Builder


Elastic Load Balancing

AWS Storage Services

Storage


Amazon Elastic Block Store (EBS)


Amazon Elastic File System

Amazon FSx

Amazon FSx for Lustre

Amazon FSx for Windows File Server


Amazon Simple Storage Service (S3)


Amazon S3 Glacier


AWS Snowball

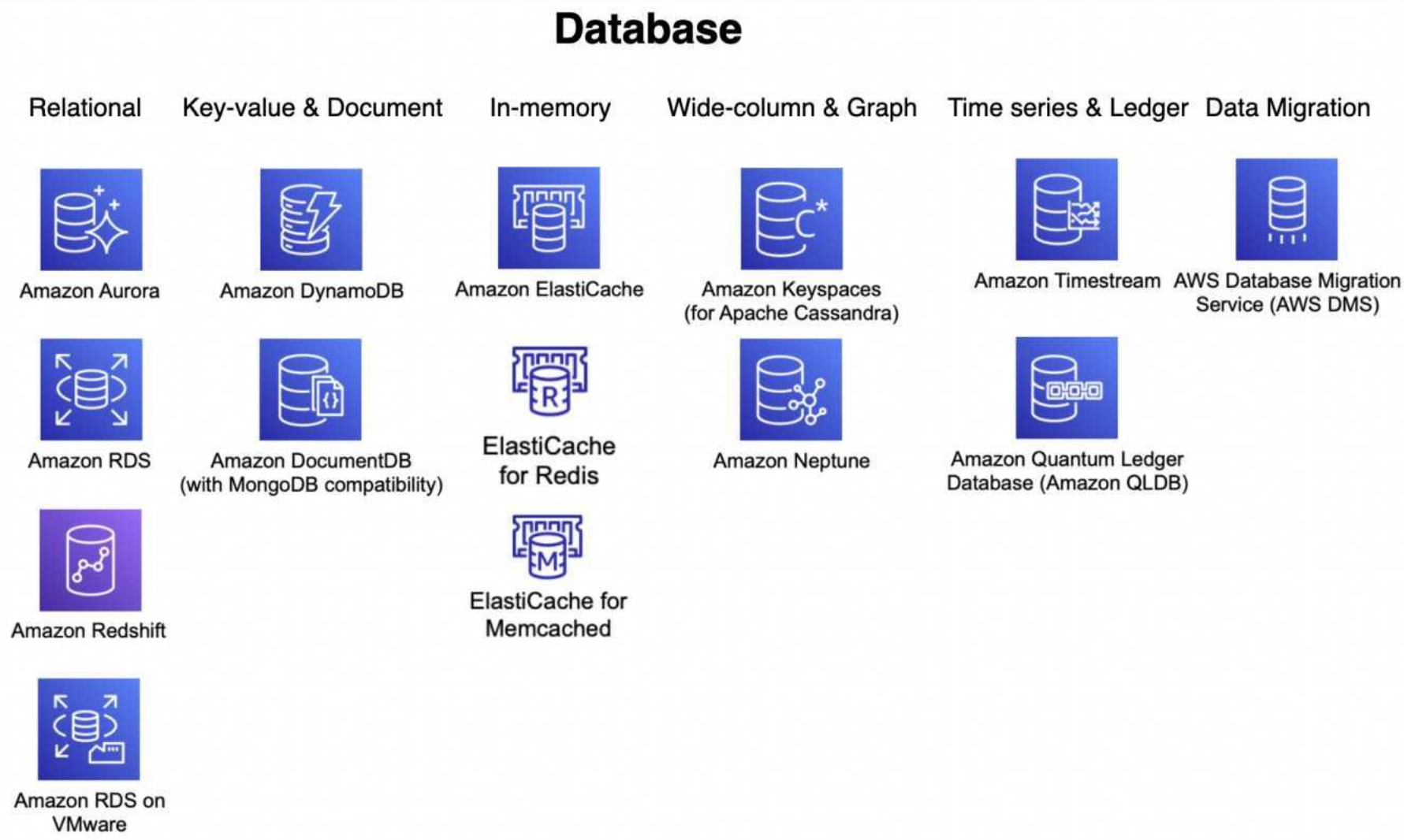

AWS Snowball Edge


AWS Snowmobile


AWS Backup


AWS Storage Gateway

AWS Database Services

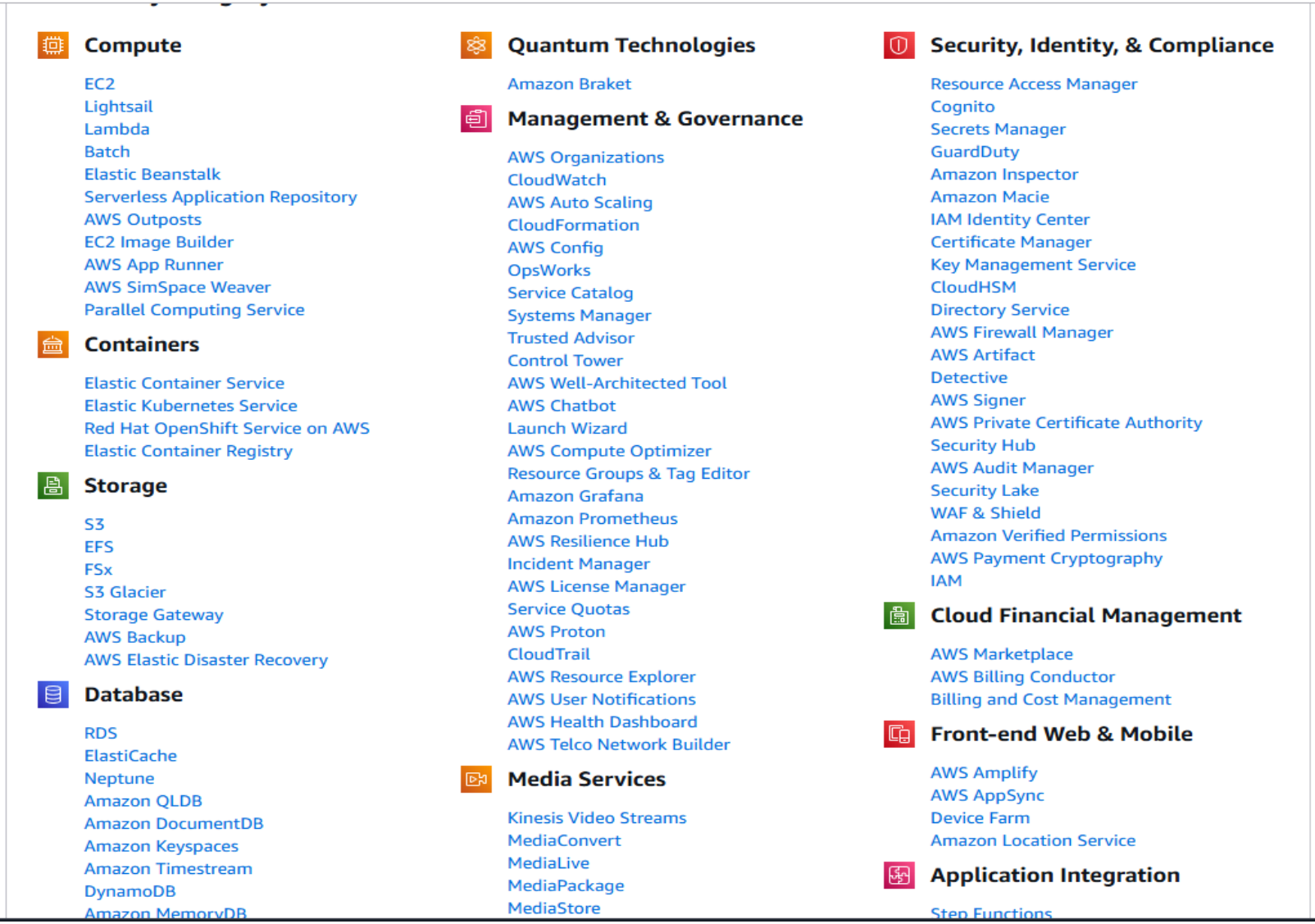


Networking and Content Delivery Service

Security, Identity, and Compliance Service

Photo by Pawel Czerwiński on Unsplash

AWS All Services – By Category



Compute

EC2
Lightsail
Lambda
Batch
Elastic Beanstalk
Serverless Application Repository
AWS Outposts
EC2 Image Builder
AWS App Runner
AWS SimSpace Weaver
Parallel Computing Service

Containers

Elastic Container Service
Elastic Kubernetes Service
Red Hat OpenShift Service on AWS
Elastic Container Registry

Storage

S3
EFS
FSx
S3 Glacier
Storage Gateway
AWS Backup
AWS Elastic Disaster Recovery

Database

RDS
ElastiCache
Neptune
Amazon QLDB
Amazon DocumentDB
Amazon Keyspaces
Amazon Timestream
DynamoDB
Amazon MemoryDB

Quantum Technologies

Amazon Braket

Management & Governance

AWS Organizations
CloudWatch
AWS Auto Scaling
CloudFormation
AWS Config
OpsWorks
Service Catalog
Systems Manager
Trusted Advisor
Control Tower
AWS Well-Architected Tool
AWS Chatbot
Launch Wizard
AWS Compute Optimizer
Resource Groups & Tag Editor
Amazon Grafana
Amazon Prometheus
AWS Resilience Hub
Incident Manager
AWS License Manager
Service Quotas
AWS Proton
CloudTrail
AWS Resource Explorer
AWS User Notifications
AWS Health Dashboard
AWS Telco Network Builder

Media Services

Kinesis Video Streams
MediaConvert
MediaLive
MediaPackage
MediaStore

Security, Identity, & Compliance

Resource Access Manager
Cognito
Secrets Manager
GuardDuty
Amazon Inspector
Amazon Macie
IAM Identity Center
Certificate Manager
Key Management Service
CloudHSM
Directory Service
AWS Firewall Manager
AWS Artifact
Detective
AWS Signer
AWS Private Certificate Authority
Security Hub
AWS Audit Manager
Security Lake
WAF & Shield
Amazon Verified Permissions
AWS Payment Cryptography
IAM

Cloud Financial Management

AWS Marketplace
AWS Billing Conductor
Billing and Cost Management

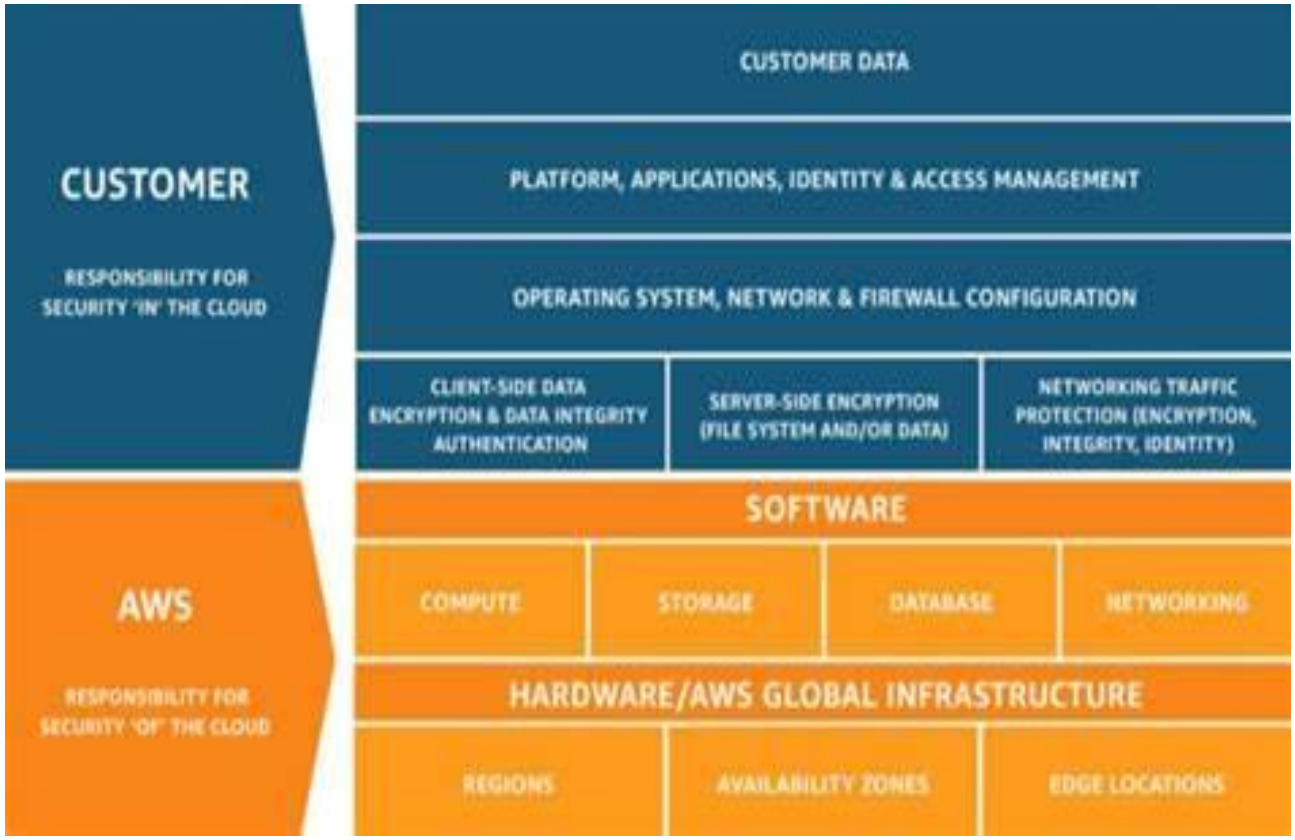
Front-end Web & Mobile

AWS Amplify
AWS AppSync
Device Farm
Amazon Location Service

Application Integration

Step Functions

AWS Shared Responsibility Model



Security of the Cloud

Versus

Security in the cloud

AWS's Responsibility (Security of the Cloud)

- Infrastructure is composed of the hardware, software, networking, and facilities that run the AWS Cloud services.
- Protecting the infrastructure that runs all the services that are offered in the AWS Cloud.
- Operates, manages, and controls the components from the software virtualization layer down to the physical security of the facilities where AWS services operate.

Customer Responsibility (Security in the Cloud)

- Ensure Data confidentiality: data at rest and data in transit.
- Ensure that the network is configured for security and that security credentials and logins are managed safely. Configuration of security groups.
- Ensure safe configuration of the operating system that run on compute instances that they launch (including updates and security patches)

AWS Identity and Access Management (IAM)

- Control **authentication** and **authorization** for your AWS account.
- Use IAM to manage access to AWS resources

– A resource is an entity in an AWS account that you can work with

- Amazon Resource Name (ARN)

– Example resources:

An Amazon EC2 instance,

An Amazon S3 bucket.

- Example–Control who can terminate Amazon EC2 instances

AWS Identity and Access Management (IAM)

Define fine-grained access rights

- Who can access the resource
- Which resources can be accessed and what can the user do to the resource
- How resources can be accessed

IAM is a no-cost AWS account feature

IAM Components

- **IAM User**: A person or application that can authenticate with an AWS account
- **IAM Group**: A collection of IAM users that are granted identical authorization
- **IAM Policy**: The document / set of instructions that defines which resources can be accessed and the level of access to each resource
- **IAM Role**: Useful mechanism to grant a set of permissions for making AWS service requests

Authenticate as an IAM user to gain access

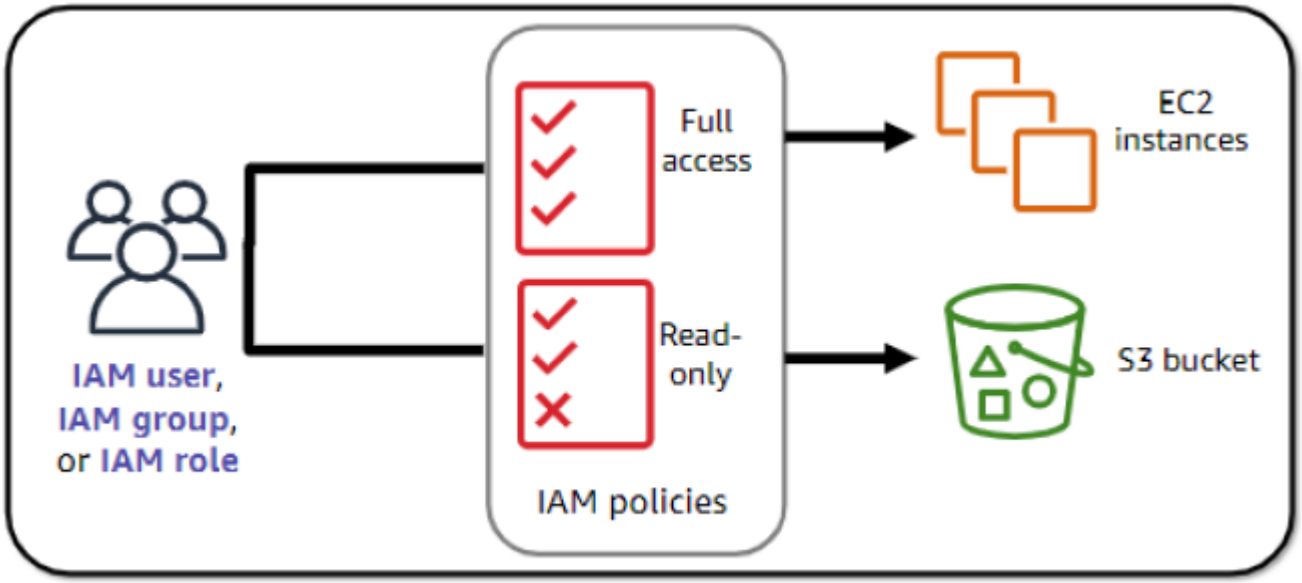
- Programmatic access
- – Authenticate using: Access key ID, Secret access key
- – Provides AWS CLI and AWS SDK access
- AWS Management Console access
- – Authenticate using:

  12-digit Account ID or alias

  IAM user name

  IAM password
- – If enabled, Multi-Factor Authentication (MFA) prompts for an authentication code

Authorization: What actions are Permitted



IAM: Authorization

Assign permissions by creating an IAM policy.

– Permissions determine which resources and operations are allowed:
– All permissions are implicitly denied by default.
– If something is explicitly denied, it is never allowed.

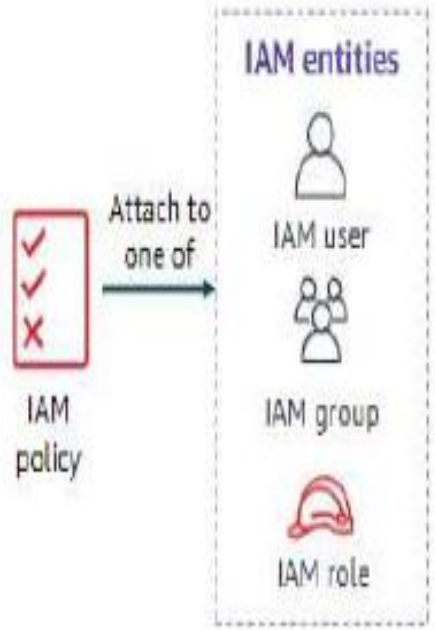
Best practice: Follow the principle of least privilege.

Note: The scope of IAM service configurations is global.
Settings apply across all AWS Region

IAM: Policies

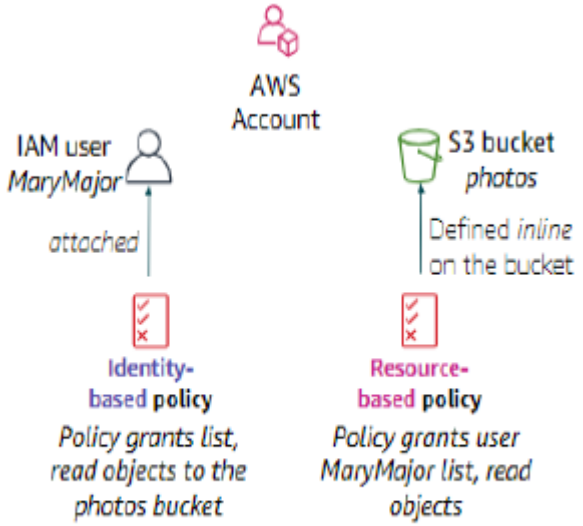
- An IAM policy is a document that defines permissions
  – Enables fine-grained access control

- Policies specify:
  - Actions that may be performed by the entity
  - Actions that may not be performed by the entity
  - A single policy can be attached to multiple entities.
  - A single entity can have multiple policies attached to it



IAM: Policies

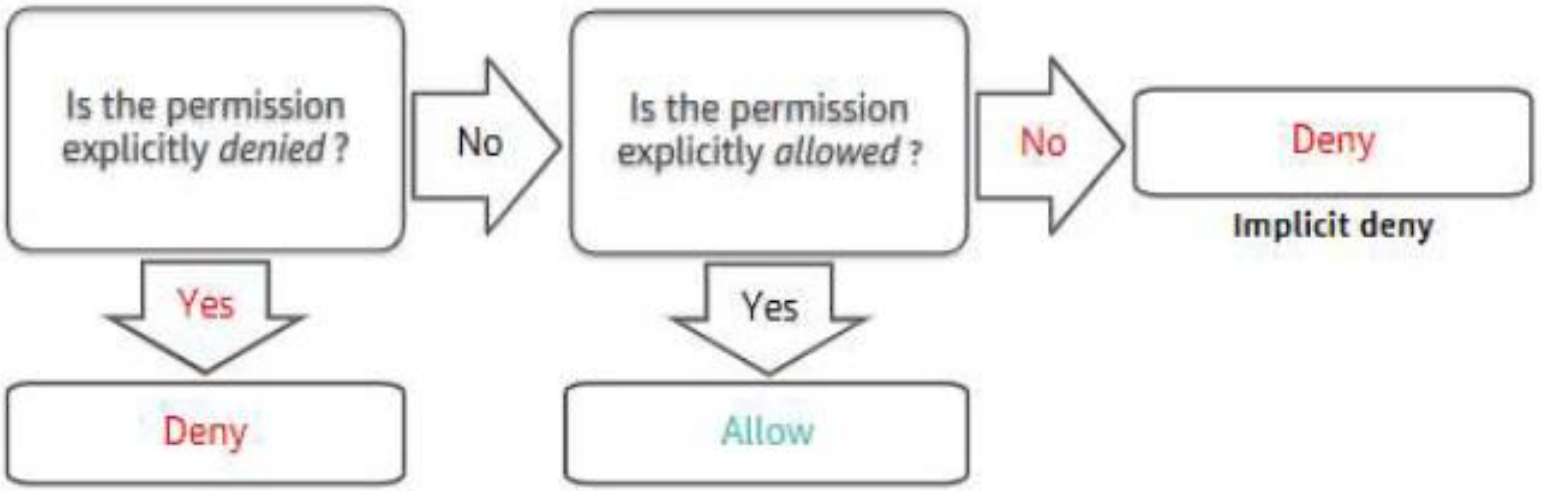
- Two types of policies:
  1. **Identity-based**
  2. **Resource based**



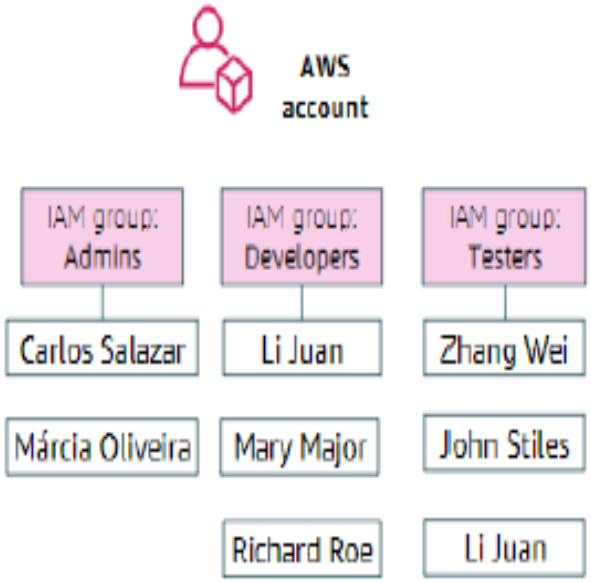
- **Identity-based policies**
- Attach a policy to any IAM entity
- An IAM user, an IAM group, or an IAM role

- **Resource-based policies**
- Attached to a resource (such as an S3 bucket)
- Specifies who has access to the resource and what actions they can perform on it

IAM Permissions



IAM Permissions



- An IAM group is a collection of IAM users
- A group is used to grant the same permissions to multiple users
- Permissions granted by attaching IAM policy or policies to the group
- A user can belong to multiple groups
- There is no default group
- Groups cannot be nested

Summary

- Key Functions of IaaS
  - Cloud Bursting, Multi Tenancy, Resource Pooling, Hypervisor
- Amazon Web Services
  - Overview – Infrastructure, Regions, Availability Zones, Data Centers
  - Variety of Services – Compute, Storage, Database, Networking
  - Shared Responsibility Model
  - IAM Service: User, Groups, Role, Policy

IaaS for you

*Thanks, I feel so "**Clouded**" now*

References

- docs.aws.amazon.com
- T1 – Chapter 3