Cloud Computing

What Is Cloud Computing?

Cloud Computing is the on-demand delivery of IT resources, particularly compute power, application hosting, database application, networking, and more.

How Does It Work?

To create resources in the cloud, the client (customer) sends a request to the cloud to create/modify/delete resources.

It follows a Client-Server model, with the cloud acting as the "server".

Cloud Computing Models

Cloud	Hybrid	On-Premises
Utilized heavily by startups of all sizes	Some parts of the application are run on the cloud, while other parts are run on-premise	Cloud is not used at all
All aspects of the application are hosted in the cloud	Some existing applications may be migrated to the cloud, while others remain on-premise	All applications are run on their own data centers or on those that are rented out
All existing applications are migrated to the cloud	Most new applications are designed and built for the cloud	Responsible for all security and operation
New applications are built in the cloud only	Fast connection between on-premise and cloud resources	Used by established companies that haven't had a reason to move to the cloud
		Infrastructure already in place to manage the infrastructure themselves
		Companies that need strict control and security over the entire infrastructure

Private Cloud:

- Cloud services used by a single organization, not exposed to the public.
- Complete control
- Security for sensitive applications
- Meet specific business needs

Public Cloud:

- Cloud resources owned and operated by a third-party cloud service provider delivered over the Internet.
- Six Advantages of Cloud Computing

Hybrid Cloud:

- Keep some servers on premises and extend some capabilities to the Cloud
- Control over sensitive assets in your private infrastructure
- Flexibility and cost-effectiveness of the public cloud

What Is AWS?

- Launched in 2006, AWS is the first hyper-scale cloud computing platform.
- Provides access to powerful resources like servers, storage, and databases without owning physical infrastructure.
- Used for hosting websites, storing data, processing large datasets, etc.
- Scalable, flexible, and cost-effective, making it accessible to businesses of all sizes

AWS Core Service Categories

AWS services can be broken down into six major categories.

CATEGORY	DESCRIPTION
Compute	Servers are used to run applications.
Networking and Content Delivery	These are services for managing networking in the cloud.
Storage	These are services used to store data.
Databases	These are services that manage databases.
Security, Identity and Compliance	These handle the security of your AWS infrastructure.
Management and Governance	These ensure that AWS infrastructure is following best practices and meeting regulatory requirements.

CDN stores copies of content on servers worldwide, delivering data from the nearest server to the user.

Benefits of the Cloud

- Trade capital expense (CAPEX) for operational expense (OPEX)
  - Pay On-Demand: don't own hardware
  - Reduced Total Cost of Ownership (TCO) & Operational Expense (OPEX)
- Benefit from massive economies of scale
  - Prices are reduced as AWS is more efficient due to large scale
- Stop guessing capacity
  - Scale based on actual measured usage
- Increase speed and agility
- Stop spending money running and maintaining data centers
- Go global in minutes: leverage the AWS global infrastructure

CAPEX : Gros investissement initial (ex : acheter des serveurs).
OPEX : Coûts flexibles payer seulement ce que vous utilisez sur aws

The Five Characteristics of Cloud Computing

- On-demand self service:
  - Users can provision resources and use them without human interaction from the service provider
- Broad network access:
  - Resources available over the network, and can be accessed by diverse client platforms
- Multi-tenancy and resource pooling:
  - Multiple customers can share the same infrastructure and applications with security and privacy
  - Multiple customers are serviced from the same physical resources
- Rapid elasticity and scalability:
  - Automatically and quickly acquire and dispose resources when needed
  - Quickly and easily scale based on demand
- Measured service:
  - Usage is measured, users pay correctly for what they have used

Elasticité : Ajuster automatiquement les ressources selon la demande.

Scalabilité : Ajouter ou enlever des ressources selon la croissance.

Flexibilité : Adapter le système aux besoin

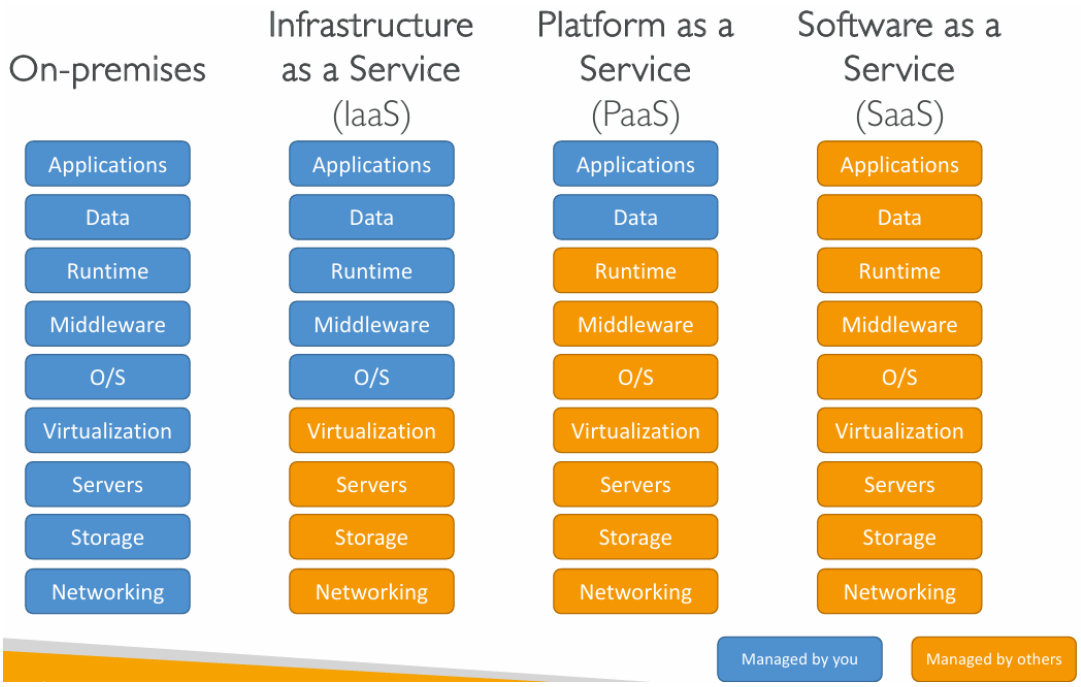
Problems solved by the Cloud

- Flexibility: change resource types when needed
- Cost-Effectiveness: pay as you go, for what you use
- Scalability: accommodate larger loads by making hardware stronger or adding additional nodes
- Elasticity: ability to scale out and scale-in when needed
- High-availability and fault-tolerance: build across data centers
- Agility: rapidly develop, test and launch software applications

Types of Cloud Computing

- Infrastructure as a Service (IaaS)
  - Provide building blocks for cloud IT
  - Provides networking, computers, data storage space
  - Highest level of flexibility
  - Easy parallel with traditional on-premises IT
- Platform as a Service (PaaS)
  - Removes the need for your organization to manage the underlying infrastructure
  - Focus on the deployment and management of your applications
- Software as a Service (SaaS)
  - Completed product that is run and managed by the service provider

Cloud



Economics

AWS has five different pricing models, so you can select the one that saves you on cost for your specific workload.

Free Tier	On-Demand	Reserved	Volume Discounts	Price Drops
- Plus de 100 services disponibles gratuitement - 12 mois de service gratuit	- Payez à l'usage en fonction de la consommation ou de la taille demandée	- Si vous prévoyez d'utiliser un service sur le long terme, vous pouvez le réserver à l'avance (1-3 ans) pour réduire les coûts	- Comme dans la plupart des domaines, plus vous achetez, plus le prix unitaire diminue	- AWS baisse régulièrement ses prix - 129 baisses de prix entre 2006 et début 2023

Pricing of the Cloud

- AWS has 3 pricing fundamentals, following the pay-as-you-go pricing model
- Compute:
  - Pay for compute time
- Storage:
  - Pay for data stored in the Cloud
- Data transfer OUT of the Cloud:
  - Data transfer IN is free
- Solves the expensive issue of traditional IT

AWS Design principles and strategies

PRINCIPLE	DESCRIPTION
Design for Failure	- No single point of failure: no single component or location should cause the entire application to fail.
	- Add redundancy wherever possible.
Decouple Components	- AWS offers the Simple Queue Service (SQS) to move data between different components.
	- When components need to communicate, they send messages through the queue.
	- This allows individual components to go down without losing data.
Implement Elasticity	- Ensure that your application and all its components can scale up and down as load varies.
Think Parallel	- Run multiple instances concurrently to complete tasks as quickly as possible.

Well-Architected Framework

Goals

is a set of best practices and guidelines designed to help cloud architects build secure, high-performing, resilient, and efficient applications on AWS.

1. **Operational Excellence:**

Run and optimize operations efficiently.

- Perform operations as code (IaC).
- Make frequent, small, reversible changes.
- Automate documentation and operations procedures.
- Anticipate failures and learn from them.

2. **Security:**

Protect data and systems.

- Implement strong identity foundations and follow the principle of least privilege.
- Automate security best practices, such as encryption and data protection.
- Use traceability and preparation for security events (e.g., incident response simulations).

3. **Reliability:** Fiabilité

Ensure workload resilience and recovery.

- Ensure systems can recover from failures and mitigate disruptions.
- Use Auto Scaling and redundancy to handle disruptions.
- Test recovery procedures regularly.
- Use horizontal scaling to avoid single points of failure.

4. **Performance Efficiency:**

Optimize computing resources.

- Utilize advanced technologies that AWS offers, and experiment frequently with different solutions.
- Use serverless architectures to avoid managing servers manually.
- Design for flexibility and scalability as dem[illegible] changes.

5. **Cost Optimization:**

Avoid unnecessary spending.

- Pay only for what you use (adopt a consumption model).
- Use services that scale efficiently (e.g., AWS Lambda, auto-scaling).
- Analyze expenditure and ensure proper attribution using tags.

6. **Sustainability:** Durabilité

Reduce environmental impact.

- Maximize energy efficiency by right-sizing workloads and minimizing idle resources.
- Adopt new, efficient hardware and software offerings over time.
- Use managed services to automate sustainability best practices.

AWS Services for Each Pillar:

1. **Operational Excellence:** AWS CloudFormation, CloudWatch, AWS X-Ray, CodeDeploy.
2. **Security:** IAM, CloudTrail, KMS, Shield, VPC, WAF.
3. **Reliability:** IAM, Auto Scaling, CloudWatch, Route 53, S3 Glacier.
4. **Performance Efficiency:** Lambda, Auto Scaling, RDS, CloudFront, ElastiCache.
5. **Cost Optimization:** AWS Budgets, Cost Explorer, Spot Instances, Reserved Instances.
6. **Sustainability:** EC2 Auto Scaling, Lambda, EC2 Graviton, S3 Glacier, Fargate.

AWS Cloud Adoption Framework

is a structured guide to help organizations adopt cloud computing.

AWS CAF Perspectives & Foundational Capabilities:

- **Business Perspective:** Focus on accelerating digital transformation.
- **People Perspective:** Bridges technology and business, fosters continuous growth.
- **Governance Perspective:** Manages cloud initiatives and minimizes risks.
- **Platform Perspective:** Builds scalable, enterprise-grade hybrid cloud platforms.
- **Security Perspective:** Ensures confidentiality, integrity, and availability of data.
- **Operations Perspective:** Ensures that cloud services meet business needs.

B P G PF S O

AWS CAF Transformation Domains:

- **Technology:** Migrate and modernize legacy infrastructure, applications, and analytics.
- **Process:** Optimize operations with new data and analytics platforms.
- **Organization:** Reorganize around value streams and agile methodologies.
- **Product:** Reimagine business models and create new products and services.

tech proc og pd

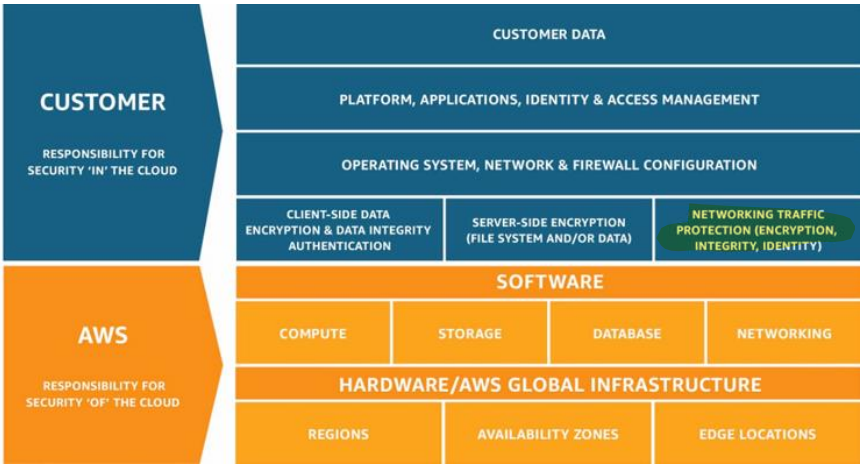
AWS Support

DEVELOPER	• Business hours email access to Cloud Support Associates • General guidance: < 24 business hours • System impaired: < 12 business hours
BUSINESS	• 24x7 phone, email, and chat access to Cloud Support Engineers • Production system impaired: < 4 hours • Production system down: < 1 hour
ENTERPRISE	• Access to a Technical Account Manager (TAM) • Concierge Support Team (for billing and account best practices) • Business-critical system down: < 15 minutes

Security and Compliance

(include monitoring)

AWS Shared Responsibility Model



- AWS responsibility - Security **of** the Cloud
  - Protecting infrastructure (hardware, software, facilities, and networking) that runs all the AWS services
  - Managed services like S3, DynamoDB, RDS, etc.
- Customer responsibility - Security **in** the Cloud
  - For EC2 instance, customer is responsible for management of the guest OS (including security patches and updates), firewall & network configuration, IAM
  - Encrypting application data
- Shared controls:
  - Patch Management, Configuration Management, Awareness & Training

➢ Unmanaged services need to be secured by users.

➢ Managed services offload some of the security responsibilities of a service on AWS.

Shared Responsibility Model for IAM

You

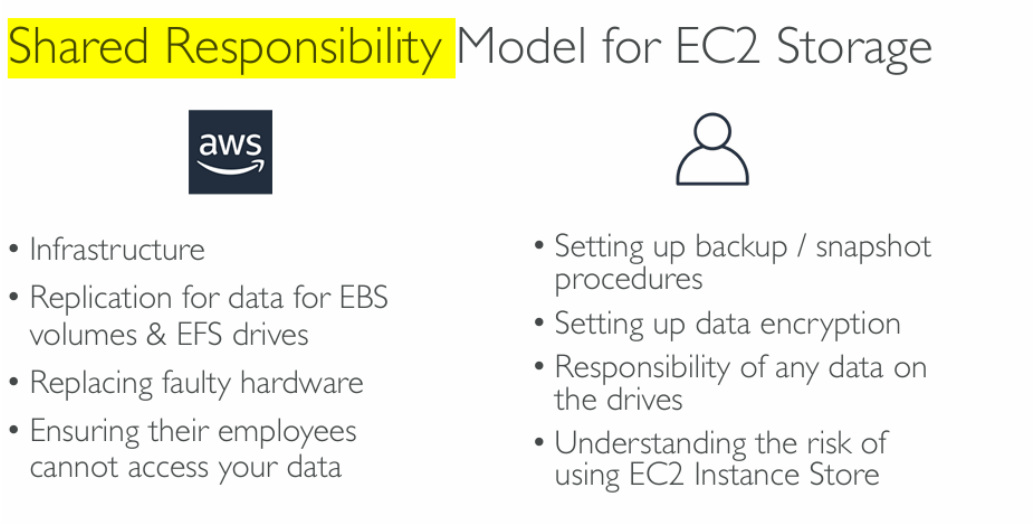
- Infrastructure (global network security)
- Configuration and vulnerability analysis
- Compliance validation

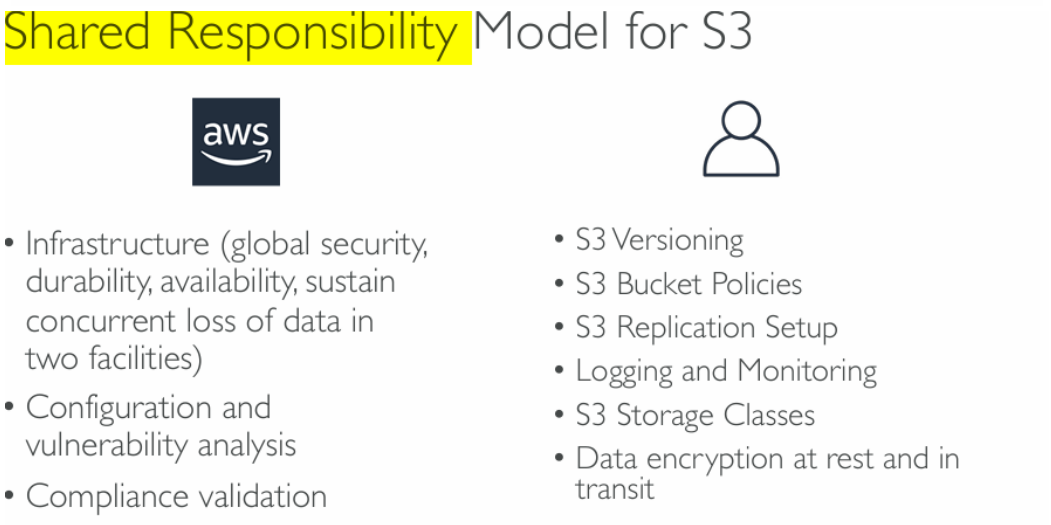
- Users, Groups, Roles, Policies management and monitoring
- Enable MFA on all accounts
- Rotate all your keys often
- Use IAM tools to apply appropriate permissions
- Analyze access patterns & review permissions

Shared Responsibility Model for EC2

- Infrastructure (global network security)
- Isolation on physical hosts
- Replacing faulty hardware
- Compliance validation

- Security Groups rules
- Operating-system patches and updates
- Software and utilities installed on the EC2 instance
- IAM Roles assigned to EC2 & IAM user access management
- Data security on your instance





AWS Compliance and governance

- **Aws Compliance** The set of policies, programs, and controls AWS implements to meet industry standards, legal requirements, and regulatory frameworks. It includes certifications, audits, and security measures to help customers maintain compliance in the cloud.
- **Governance** provides tools to manage security, compliance, and costs, using services like AWS Organizations and AWS IAM.

AWS Compliance Resources

AWS Compliance Center

A hub for understanding cloud compliance, offering access to standards, laws, industry solutions, and audit tools.

AWS Audit Manager

is a service that helps organizations **automate** the collection and organization of compliance data. It generates **audit-ready reports**, tracks **configuration changes over time**, and simplifies regulatory compliance efforts.

AWS Artifact:

is a portal for accessing AWS compliance documentation and agreements. Customers can use AWS Artifact to review, download, and accept compliance documents to meet regulatory requirements

Aws config

Service helps audit and record compliance of AWS resources, tracking configuration changes over time.

Monitoring services

Aid in compliance and security.

Amazon CloudWatch

monitors AWS resources in real time, tracking **metrics, logs, alarms, and events** to enhance performance and security.

Amazon CloudWatch Alarms surveille les métriques AWS et déclenche des actions (notifications, arrêt/redémarrage d'instances, Auto Scaling) selon des seuils définis.

- **Metrics:** Monitor AWS service performance and billing.
- **Alarms:** Automate actions based on metrics.
- **Logs:** Collect logs from EC2, servers, and Lambda. Enables real-time monitoring of logs
- **Events (or EventBridge):** Schedule tasks and trigger actions based on event patterns. Integrates with AWS services like EC2, S3, and Lambda.

AWS CloudTrail:

provides **governance, compliance, and auditing** by tracking **API calls and events** across AWS accounts.

AWS X-Ray:

Analyses

offers deep insights into **application performance** and **service dependencies**, helping developers identify bottlenecks and optimize application behavior.

Amazon CodeGuru:

- **CodeGuru Reviewer:** Automates **code reviews** to improve security and best practices.
- **CodeGuru Profiler:** Analyzes **application performance** to optimize costs and resource usage.

AWS Health Dashboard:

- **Service Health:** Displays the global status of AWS services.
- **Account Health:** Provides **personalized notifications** about the impact of service events on your AWS resources.

Security Resources

DDoS Protection

A. Web Application Firewall (WAF)

Web Application Firewall is a service that protects web applications from Layer 7 attacks like SQL injection and XSS.

- Web ACL: Set rules based on IP, headers, body, or URI.
- Includes rate-based rules for DDoS protection and geo-match for blocking countries

B.AWS Shield

is a service that protects against **DDoS** attacks by detecting and mitigating malicious traffic to ensure the availability of AWS resources.

Limitation du trafic malveillant

- **Shield Standard**: Free, basic protection against DDoS attacks.
- **Shield Advanced**: Premium protection with 24/7 support.

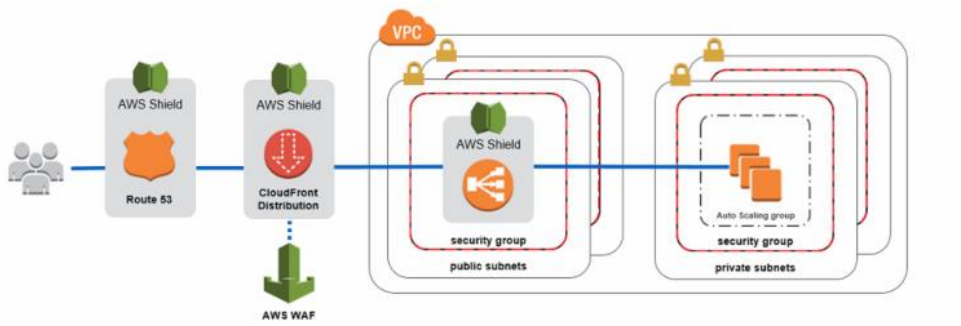
C. CloudFront & Route 53

Leveraging AWS's global network, CloudFront and Route 53 provide advanced protection against network-based attacks through AWS infrastructure.

D.AWS Auto Scaling: Helps mitigate traffic spikes caused by DDoS attacks.

Surcharges

Sample Reference Architecture for DDoS Protection



AWS Network Firewall

It is a stateful managed firewall and intrusion detection and prevention service for VPCs. It monitors traffic going into and out of a VPC.

AWS Penetration Testing

AWS a allows testing on 8 services, such as EC2, RDS, CloudFront, API Gateway, Lambda, and Elastic Beanstalk, without requiring prior approval. However, certain activities are prohibited, including **DNS zone walking** on Route 53, **DoS/DDoS attacks**, and **port, protocol, and request flooding**.

Encryption options

- **Data at Rest**: Stored data (e.g., on EFS, S3, RDS, Glacier).
- **Data in Transit**: Data moving over a network (e.g., EC2 to DynamoDB, on-premises to AWS).
- **Encryption**: Protects data in both states using encryption keys.

AWS Inspector

Automated Security Assessments for: find software vulnerabilities in EC2, ECR Images, and Lambda function

- **EC2 Instances**: Checks network accessibility and OS vulnerabilities.
- **Container Images (ECR)**: Scans images upon push.
- **Lambda Functions**: Identifies vulnerabilities in code and dependencies.

integrates with **AWS Security Hub** & **Amazon EventBridge** for reporting and Send findings to Amazon Event Bridge

Amazon GuardDuty

uses machine learning and anomaly detection to monitor **CloudTrail logs**, **VPC flow logs**, and other AWS data, protecting against threats, including cryptocurrency attacks.

AWS Macie

is a managed service that uses machine learning to discover and protect sensitive data in AWS, (ex **personally identifiable information (PII))**

Amazon Detective

Amazon Detective analyzes, investigates, and quickly identifies root cause of security issues or suspicious activities (using ML and graphs)
Automatically collects and processes events from VPC Flow Logs, CloudTrail, GuardDuty and create a unified vie

AWS Security Hub

- **Centralized security tool** to manage security across multiple AWS accounts.
- **Aggregates alerts** from AWS services like **GuardDuty**, **Macie**, and **Inspector**.
- Must enable **AWS Config** first to integrate with **Security Hub**.

Security Lake

It collects security logs and events from multiple sources including on-premises, AWS services, and third-party services. It transforms the data into storage and query-efficient Parquet format.

Firewall Manager

It helps manage security services (WAF, security groups, network firewall) across multiple AWS accounts. It configures rules just once and has them available across all accounts.

Resource Access Manager

It helps you securely share resources across accounts, organizations, and OUs. So, you can create a resource once and have the Resource Access Manager make that resource usable by other accounts.

AWS Trusted Advisor

is a management tool that provides real-time guidance to help you provision your AWS resources following best practices. It offers insights across various areas, including security.

AWS Certificate Manager (ACM)

simplifies creating, storing, and renewing public and private SSL/TLS certificates for AWS websites and applications. It eliminates the need for third-party certificates by generating free certificates for use with services like ELB and API Gateway.

AWS Private Certificate Authority

offers a private CA for issuing certificates to authenticate internal users, computers, and applications, trusted only within your organization, without the need to manage your own internal CA.

Key Management Service (KMS)

manages cryptographic keys for encrypting/decrypting data, with granular access control and key rotation capabilities.

CloudHSM

provides dedicated, tamper-resistant hardware security modules for encryption, with user-controlled keys stored securely on the HSM, offering full key management control unlike KMS.

AWS Abuse

Report abusive activities such as spam, DoS attacks, and intrusion attempts.

AWS access management and identity

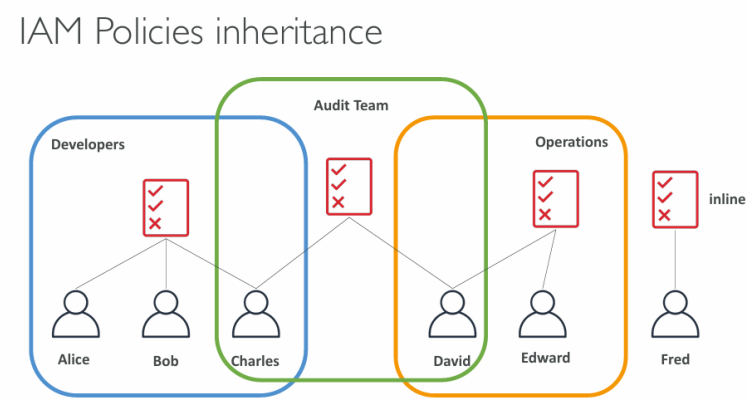
Identity Access Management (IAM)

IAM stands for **Identity and Access Management**. It is a service used to control access to AWS resources securely.

- **Root Account:** The root account is created by default and should not be used regularly. It has full permissions and unrestricted access to all AWS resources.
- **Users:** Individual entities that need access to AWS resources. They can be grouped for easier management.
- **Groups:** Collections of users. Policies are applied to groups, not directly to users.
- **Roles:** IAM roles allow AWS services (like EC2 or Lambda) to perform actions on your behalf.
- **Policies:** Policies are JSON documents that specify what actions are allowed or denied for users, groups, or roles.

IAM Users & Groups:

- Users can belong to multiple groups, but groups can only contain users (not other groups).
- Users can be assigned specific permissions through IAM policies.



IAM Policies:

- Policies define permissions in JSON format. The **least privilege** principle should be applied, meaning only the minimum necessary permissions should be granted.
- A policy includes:
    - **Version:** The version of the policy language.
    - **Statement:** One or more permissions defining:
        - **Effect:** (Allow/Deny)
        - **Action:** (What services the policy applies to)
        - **Resource:** (Which AWS resources the actions apply to)

Password Policies:

- Password policies enforce secure password practices for IAM users.
- You can configure minimum password length, require specific characters, set expiration, and prevent password reuse.

Root User Privileges:

- The root user has full access to all AWS resources.
- **Best Practice:** Lock away the root user's keys and avoid using the root account for daily tasks.

Multi-Factor Authentication (MFA):

- MFA provides an additional layer of security by requiring both a password and a second form of authentication (e.g., a security device).
- There are **virtual MFA devices** (e.g., Google Authenticator, Authy) and **hardware MFA devices** (e.g., YubiKey).
- It is recommended to use MFA for both the root account and IAM users to enhance security.

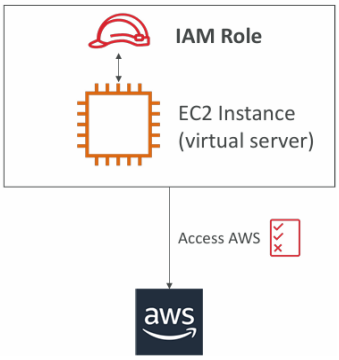
Access Keys:

Rotate access keys regularly to minimize security risks.

- To access AWS, you have three options:
  - AWS Management Console (protected by password + MFA)
  - AWS Command Line Interface (CLI): protected by access keys
  - AWS Software Developer Kit (SDK) - for code: protected by access keys
- Access Keys are generated through the AWS Console
- Users manage their own access keys

IAM Roles for Services:

- IAM roles grant AWS services permissions to perform actions on your behalf.
- For example, an **EC2 Instance Role** allows an EC2 instance to access other AWS resources (like S3) without requiring embedded access keys.



IAM Security Tools:

- **IAM Credentials Report:** Provides a report of all IAM users and the status of their credentials (e.g., password age, MFA status).
- **IAM Access Advisor:** Helps understand the permissions granted to a user and whether they are actively used.

IAM Best Practices:

- Do not use the root account for everyday tasks.
- Use groups to manage user permissions.
- Apply the principle of **least privilege**—only grant the permissions necessary for users to perform their job.
- Use **MFA** on all IAM accounts to improve security.
- Regularly review and rotate access keys.
- Enable and enforce password policies and MFA for better security.

IAM Identity Center

It simplifies managing users across multiple AWS accounts. It also allows you to manage sign-in security for your users centrally and grant them access across all AWS accounts and resources
one login for multiple AWS accounts & applications

one login for multiple AWS accounts & applications
centralizes user authentication and authorization for AWS resources.

Secrets Manager

The Secrets Manager helps retrieve and rotate secrets and other sensitive data like credentials and Auth tokens.

AWS Cognito

AWS Cognito helps implement customer identity and access management for mobile and web applications.

Security Token Service (STS):

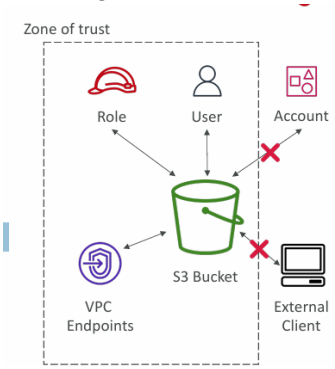
temporary, limited-privileges credentials to access AWS resources

Directory Services: integrate Microsoft Active Directory in AW

IAM Access Analyzer

Find out which resources are shared externally

- Define Zone of Trust = AWS Account or AWS Organization
- Access outside zone of trusts => findings



Organizations

- **Manages multiple AWS accounts** under a **master account**.
- **Benefits**: consolidated billing, cost savings, shared reserved instances.
- **SCPs** to restrict account privileges.

Multi-Account Strategies

- Separate accounts by **department, environment, compliance**.
- **Resource isolation (VPC)** and **independent service limits**.
- **Centralized logging** with CloudTrail and CloudWatch.

Organizational Units (OU)

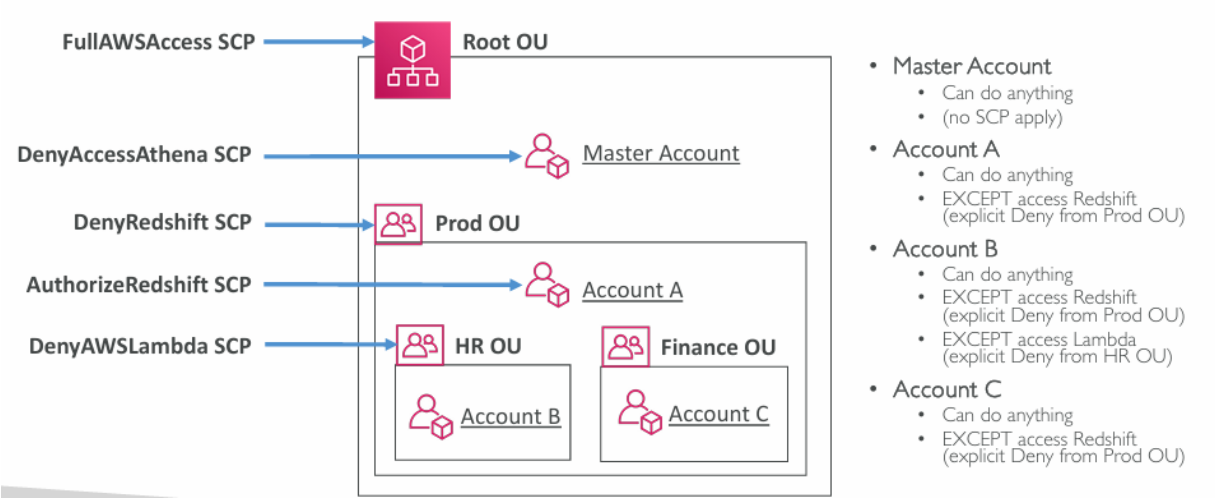
- Organize by **Business Unit, environment, or project**.

Service Control Policies (SCPs)

- **Whitelist or blacklist IAM actions** at the **OU or account level**.
- **Does not apply to the master account**.
- **Default deny unless explicitly allowed**.
- Examples: block **EMR**, enforce **PCI compliance**.

SCP Hierarchy

- **Master account: full access**.
- **Other accounts: restrictions based on applied SCPs** (e.g., deny Redshift or Lambda).



Billing

General Billing

There are three main drivers of billing:

- **Compute** that was used/requested
  - ➔ I want a server with 4 CPUs and 32 GB of RAM
  - ➔ That server was run for 8 hours
- **Storage** that was used/requested
  - ➔ I want 50 GB of fast disk storage
- **Network** that was used/requested (only in the outbound direction)
  - ➔ I transferred 20 GB of data out to my office
- Ensure to understand all the aspects of pricing for a service
- Choose the correct sizing of service to optimize costs
- Make use of AWS' Optimize and Save tools when possible
- Scale up only when needed and make sure to scale back down
  - ➔ Utilize auto-scaling when possible

AWS Pricing Models

Free Tier	On-Demand	Reserved	Savings plan	Spot

- Over 100 services available for free - 12 months of free service - Some services are always free	- Pay for what you use or the size you request	- If you know you will be using a service for a long time, you can reserve it ahead of time (for 1-3 years) to save on cost	- Offers discounted prices on services in exchange for a commitment to spend a certain period of time	- If Amazon has spare capacity, they'll offer it at a discounted rate Capacité AWS inutilisée à prix réduit (jusqu'à 90%), mais récupérable à tout moment.

C'est un engagement de dépenses sur 1 ou 3 ans, offrant des réductions flexibles. Contrairement aux instances réservées, il s'adapte à différents types d'instances et régions.

EC2 Billing

factors

- **Virtual Machine Size**: Number of vCPUs and memory.
- **Billing Unit**: Charges per second or per hour.
- **Licensing Type**: Instance licensing model.
- **Features**: Additional services or features enabled.
- **Instance Status**: Running or stopped (stopped instances may still incur small costs).

EC2 Pricing Options:

- On-Demand Instances – short workload, predictable pricing, pay by second
- Reserved (1 & 3 years)
  - Reserved Instances – long workloads
  - Convertible Reserved Instances – long workloads with flexible instances
- Savings Plans (1 & 3 years) –commitment to an amount of usage, long workload
- Spot Instances – short workloads, cheap, can lose instances (less reliable)
- Dedicated Hosts – book an entire physical server, control instance placement
- Dedicated Instances – no other customers will share your hardware
- Capacity Reservations – reserve capacity in a specific AZ for any duration

RDS Billing

- **RDS Flavor**: Aurora or Aurora Serverless.
- **SQL Engine**: Oracle, MSSQL, MariaDB, MySQL, or PostgreSQL.
- **Memory Size**: Size of the database.
- **Storage Disk Type**: General purpose or provisioned IOPS.
- **Additional Features**: For example, Multi-AZ or backup retention.
- **Reserved Instances**: RDS offers reserved instances like EC2, but lacks Spot, Dedicated, or Savings plans

VPC Billing

- **Charge per VPC**: Includes base components.
- **Data Transfer**: Charges apply for outbound data transfers, including cross-region traffic. No charges for inbound transfers or EC2 to S3 transfers within the same region.
- **Add-ons**: Additional services for VPCs.
- **NAT Gateways**: Charges apply for the existence of the gateway.
- **Free Components**: No charges for subnets, Network ACLs (NACLs), Security Groups, or IP ranges.

Lambda Billing

Lambda pricing is based on

Size Duration Frequency

- The more often your functions run, the more you pay
- The longer it runs, the more you pay
- The more memory it uses, the more you pay
- Any additional features like hot provisioning will cost extra

Other Services

EBC pricing factors:

- Volume – size and type over time
- Snapshots – size over time
- EBS Fast Snapshot Restores
- EBS Direct APIs for Snapshots

S3 pricing factors:

- Type of storage class
- Number and size of objects stored
- Type of requests made to S3
- Charged for outbound data
- Other backup and management features

DynamoDB cost factors:

- Reading, writing, or storing data
- Optional features
- Charges based on read request units and write request units

CloudFront cost factors:

- Charges based on amount of data taken from CloudFront
- HTTP/HTTPS requests
- Invalidation requests

Macie cost factors:

- Amount of data that needs to be scanned
- S3 charges like reading objects and listing buckets as Macie scans

Kinesis cost factors:

- For how long is data stored?
- How much data is stored?

Billing Account Structure

Single Account	Multiple Accounts
Receives a **single bill** for all resources in this account If you have a **second account**, the billing for that account is separate	One account acts as the **"payer"** account ➔ Receive a **single bill** from AWS for all accounts ➔ View detailed billing from each account ➔ **Apply** Reservation and Savings Plans across **all accounts**

Multiple Accounts Within an AWS Organization	Multiple Accounts in Control Tower
Works identical to consolidated billing ➔ One account acts as the "payer" account	Works identical to consolidated billing ➔ One account acts as the "payer" account

resources for billing, budget, and cost management.

Estimating Costs in the Cloud:

- **Pricing Calculator**
  this tool helps estimate the cost for your solution architecture

Tracking Costs in the Cloud:

- **Billing Dashboard**
  Provides a comprehensive view of your AWS costs and usage.
- **Cost Allocation Tags**
  Track your AWS costs at a detailed level using:
  Tags help organize resources such as EC2 instances, RDS, and VPC resources.
- **Cost and Usage Reports**
  Offers in-depth cost and usage data, including metadata on AWS services, pricing, and reservations. This report can be integrated with Athena, Redshift, or QuickSight.
- **Cost Explorer**
  Visualize, understand, and manage AWS costs and usage over time. You can create custom reports at a high level or by month, hour, or resource. It also allows you to forecast usage up to 12 months and optimize your Savings Plan.

Monitoring Against Cost Plans:

- **Billing Alarms**
  Stored in CloudWatch, these alarms track actual AWS costs, not projections. They provide notifications for billing thresholds.
- **AWS Budgets**
  Create budgets and set alarms when costs exceed set limits. It supports different types of budgets (Usage, Cost, Reservation, Savings Plans) and allows filtering by service, region, instance type, etc. AWS Budgets offers up to 5 SNS notifications per budget.
- **AWS Cost Anomaly Detection**
  Uses machine learning to monitor your cost and usage, detecting unusual spending patterns. It sends alerts and provides root-cause analysis.
- **AWS Service Quotas**
  Alerts you when service quotas are nearing their limit. You can set up CloudWatch alarms and request quota increases as needed.
- **Trusted Advisor**
  Provides a high-level account assessment and recommendations for cost optimization, security, performance, fault tolerance, and service limits.
- **AWS Compute Optimizer**
  Recommends optimal AWS resources for your workloads, helping lower costs and improve performance by right-sizing resources using machine learning.

Savings Plans:

- A cost-saving option based on long-term AWS usage. It offers a simple way to save money by committing to certain resource usage over a period of time.

Technology

Deployment Methods

AWS Console	Web GUI used to manage AWS resources ❖ Great for people who want to visually see their infrastructure ❖ Ideal for monitoring logs, alerts, and metrics in nicely presented graphs ❖ Lots of menus, so provisioning resources will involve a lot of clicking and navigating
AWS CLI	Command line utility (CLI) for managing AWS resources ❖ Engineers naturally prefer working on command line ❖ Very easy to manage resources and commands can be copied and pasted

	❖ Some settings/knobs on resources can only be toggled through the CLI
AWS SDK	Provides APIs in most programming languages to manage and interact with AWS ❖ Allows applications to create resources in AWS

Global Infrastructure

Regions are locations across the globe to which services can be deployed.

- ❖ All services are not available in all regions
- ❖ Pricing can be different between regions

Availability Zones (AZ) are isolated and independent data centers inside regions.
Edge Locations are smaller points of presence across the globe.

- ❖ Allows you to get services closer to end-users to minimize latency
- ❖ Limited services available; mainly used for CDN
- ❖ Services available – CloudFront, Route 53, AWS WAF

Local Zones are extensions of AWS regions located near users in select metropolitan areas.

- ❖ Have isolated infrastructure but are connected to parent AWS regions through highbandwidth network links
- ❖ Provide a subset of services like EC2 and EBS

- **Global Services:** Operate globally, accessible everywhere, like **Route 53**, **IAM**, and **CloudFront**.
- **Scoped Services:** Limited to a single region, like **EC2**, **S3**, and **RDS**. These services are "limited" to one region, meaning to use them across multiple regions, you need to configure and manage separate resources in each region.

Global Applications

Routing & Acceleration

- **Route 53** → A **global DNS service** that routes users to the nearest endpoint.
- **AWS Global Accelerator** → Optimizes **connectivity** and **reduces latency** using AWS's global network.

Content Distribution & Fast Transfers

- **CloudFront** → A **CDN** that caches content at **AWS Edge Locations** for faster delivery.
- **S3 Transfer Acceleration** → Speeds up **file transfers** to Amazon S3 via an optimized network.

Extending AWS to Local Infrastructure & 5G Networks

- **AWS Outposts** → Allows running **AWS services in on-premises data centers**.
- **AWS Wavelength** → Integrates AWS with **5G networks** for **ultra-low latency** applications.
- **AWS Local Zones** → Brings **AWS resources (compute, storage, databases, etc.)** closer to end-users.

Networking

VPC (Virtual Private Cloud)

A VPC is a private, isolated network in AWS used to deploy resources.

Subnets

- Public Subnet: Connected to the Internet via an Internet Gateway (IGW).
- Private Subnet: No direct Internet access, often used for databases.

Internet Gateway (IGW)

Allows instances in a public subnet to access the Internet.

NAT Gateway

Allows instances in a private subnet to connect to the Internet without being directly exposed.

Security Groups: Firewall attached to the EC2 instance (only allows defined traffic). all inbound traffic is blocked by default and all outbound traffic is authorised by defaul
Network ACLs (NACLs): A firewall at the subnet level (can allow or deny traffic).
Elastic IP :A static public IP that can be assigned to an EC2 instance.
VPC Flow Logs: Captures metadata about network traffic for monitoring and troubleshooting.
VPC Peering:A private connection between two VPCs (no transit between multiple VPCs).
Transit Gateway: Enables communication between multiple VPCs and on-premises networks, unlike VPC Peering.

VPC Endpoints

- **Interface Endpoint (AWS PrivateLink)**: Provides private access to AWS services without using the Internet.
- **Gateway Endpoint**: Used for access to S3 and DynamoDB without requiring an Internet connection.

VPN

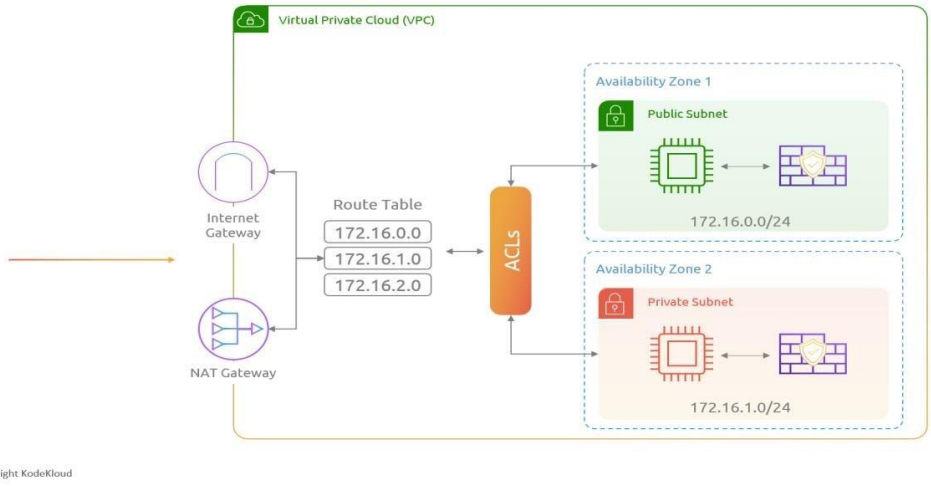
- **Client VPN**: Allows remote users to securely access a VPC via a VPN connection.
- **Site-to-Site VPN**: An encrypted connection between an on-premises network and a VPC over the Internet.

- **Direct Connect**: A dedicated physical connection between an on-premises network and AWS, offering lower latency and better stability than a VPN.

Amazon Route 53
AWS Global Accelerator
Additional Details

- **VPC**: Isolates computing resources in the cloud and gives full control over networking.
- **Subnetting**: Determines IP addressing, routing, and firewalls.
- **Gateways**: Provide access to the Internet or other private networks.

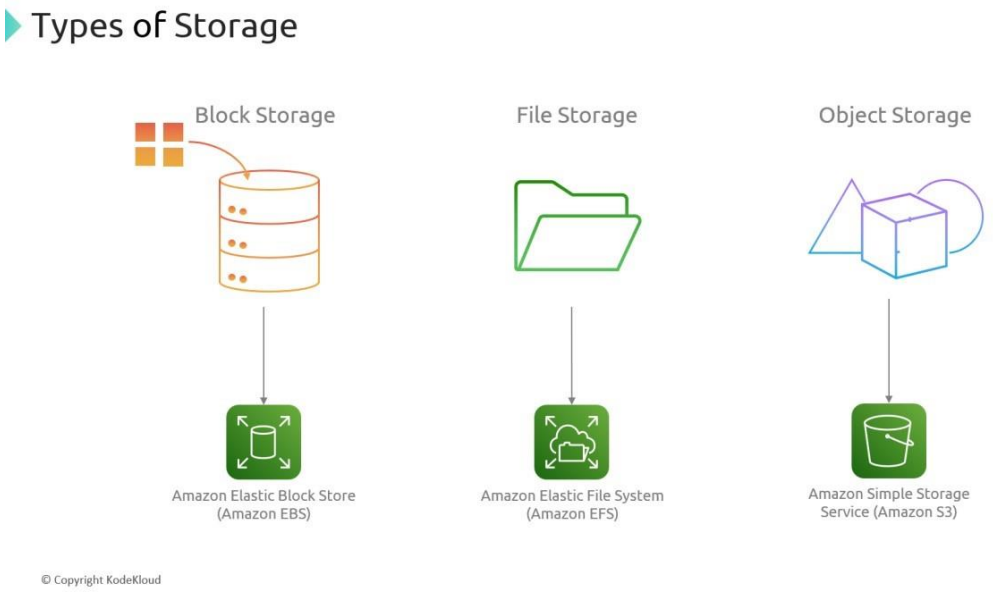
By default, each region has a default VPC with default subnets, security groups, and NACLs. These subnets have outbound access to the Internet, and security groups allow this traffic.



Security & Identity

See security section

Storage



Instance stores

are temporary block-level storage, for instance.

- Data is lost if the instance is stopped and started

Block Storage (EBS) network drives

breaks up data into blocks and presents a collection of blocks as a volume or hard drive to the operating system.

- Can be mounted and booted (OS can be installed on it)
- EBS is Availability Zone specific
- Servers in one AZ cannot attach an EBS volume in a different AZ
- Can use EBS Snapshots for backups / transferring EBS volumes across AZ

File storage (EFS) stores data in a hierarchical structure of files and folders.

- Filesystem that is accessible remotely
- Multiple machines can connect to EFS volume at once
- Cannot be used as a boot volume (can't install OS)

Object storage (s3)

- Object storage (S3) stores objects in a in "buckets" (directories)
  - ➔ Cannot boot or mount object storage
- Great for storing static websites, media files, logs, traces, and audit reports
- Storage classes impact accessibility, resiliency, and cost
  - ➔ **S3 Standard** – Default storage class and most expensive
  - ➔ **Standard-IA** – Same reliability as S3 standard but has a retrieval fee
  - ➔ **One Zone-IA** – Same as Standard-IA but only hosted in one AZ
  - ➔ **Glacier instant** – Millisecond retrieval, ideal for data accessed quarterly. Minimum storage: 90 days.
  - ➔ **Glacier flexible** – Cheaper than Glacier Instant, but data retrieval is time consuming

➔ **Glacier Deep Archive** – Cheapest storage option; has the longest wait time for retrieval

- **Buckets vs Objects:** global unique name, tied to a region
- **S3 security:** IAM policy, S3 Bucket Policy (public access), S3 Encryption
- **S3 Websites:** host a static website on Amazon S3
- **S3 Versioning:** multiple versions for files, prevent accidental deletes
- **S3 Replication:** same-region or cross-region, must enable versioning

Storage Gateway: hybrid solution to extend on-premises storage to S3

AWS Backup:

- A fully managed backup service for AWS services such as EC2, RDS, EFS, DynamoDB, and more.
- Allows centralized backup management, automation, and compliance enforcement.
- Ensures backups are scheduled, retained, and securely stored for disaster recovery.

Amazon FSx:

- A fully managed Windows File System and Lustre File System.
- Provides scalable and high-performance file storage.
    - **FSx for Windows File Server:** Optimized for Windows-based applications.
    - **FSx for Lustre:** Optimized for workloads that require high throughput, like high-performance computing (HPC) or machine learning.

Compute

EC2

- EC2 allows you to provision a server in AWS within minutes
- AMIs are templates for deploying EC2 instances
- AWS has a variety of different instance types

Instance Type	Description	Use Case
General Purpose	Balanced mix of compute, memory, and networking resources.	Versatile, suitable for a wide range of workloads.
Compute Optimized	High-performance CPUs.	Ideal for compute-heavy workloads like batch processing and machine learning.
Memory Optimized	Designed for memory-intensive workloads.	Best for databases and other memory-intensive applications.
Storage Optimized	Optimized for workloads requiring high IOPS (Input/Output Operations Per Second).	Suitable for applications with high data throughput requirements.

Lambda

Overview	AWS Lambda is a compute service that lets you run code without provisioning or managing servers. It is AWS' serverless offering, with servers managed entirely by AWS.
Server Management	Servers are required to run the code but are managed completely by AWS (maintenance, scaling, provisioning, and logging). You only upload the code, AWS handles the rest.
Use Cases	- File processing - Stream processing - Web applications - Mobile/Web backend
Components	1. **Function:** A traditional function in any programming language. 2. **Trigger:** An event that triggers the function (e.g., file uploaded to S3, HTTP request, DynamoDB update, etc.). 3. **Event Info:** Information about the event that triggered the function, passed to the function.
Benefits	- No servers to manage - Auto scaling to handle traffic - Pay only for usage
Downsides	- No local state - Maximum execution time of 15 minutes (not suitable for long-running tasks) - Impacted by cold starts
Pricing Factors	- Number of function invocations - Duration of function execution - Memory/CPU usage

Containers

- Containers are a tool that allows you to package an application and all of the necessary files, libraries, and dependencies the application needs to run.
- Container orchestrators are the brains of a containerized environment.
    - ➢ Deploying containers across all available servers
    - ➢ Sending load-balancing requests to containers
    - ➢ Providing container-to-container connectivity
    - ➢ Restarting failed containers
    - ➢ Moving containers when hosts fail

ECS

is a fully managed container orchestration service that helps manage and scale containerized applications.

- ➢ AWS manages ECS, which handles all the orchestration
- ➢ Containers run on EC2 instances or on Fargate

ECR: Private Docker Images Repository

Kubernetes

- is an open-source container orchestrator.
- Kubernetes cluster has two types of nodes :master /worker

EKS

Elastic Kubernetes Service is a fully managed service that makes it easy to run Kubernetes on AWS without needing to install or operate your own Kubernetes control plane.

AWS Batch
is a fully managed service for running batch computing workloads on AWS.
Ideal for processing large datasets or performing complex computations like rendering and simulations.

Lightsail
Amazon Lightsail is a simple, predictable & low pricing cloud platform for launching and managing VPS.

Fargate:
Serverless offering , run Docker containers without provisioning the infrastructure

Database

Feature	SQL Databases	NoSQL Databases
Data Structure	Structured data with tables and columns	Flexible data (key-value, documents, etc.)
Schema	Fixed schema, predefined structure	Flexible schema, no predefined structure
Scaling	Vertical scaling (larger servers)	Horizontal scaling (more servers/nodes)
Query Language	SQL (Structured Query Language)	Varies (e.g., DynamoDB, MongoDB queries)
Use Cases	Complex queries, structured data (e.g., finance)	Fast, large-scale, unstructured data (e.g., social media, IoT)

Managed vs. Self-Managed Databases:

- **Self-managed databases**: Provide full control but require managing backups, availability, and security.
- **Managed databases** (e.g., AWS RDS, Aurora, DynamoDB, etc.): AWS manages infrastructure, backups, and scaling, reducing operational overhead.

AWS Managed Database Services:

RDS (Relational Database Service):
- Managed SQL databases (MySQL, MariaDB, PostgreSQL, Oracle, Microsoft SQL Server).
- Automated provisioning, backups, scaling, and patching.

Aurora:
- AWS's high-performance database (MySQL and PostgreSQL-compatible).
- Faster than traditional MySQL and PostgreSQL on RDS.
- Serverless option for scaling based on demand.

RedShift:
- Managed data warehousing for online analytics processing (OLAP).
- Designed for large-scale data analysis (petabytes of data).
- Serverless option for data warehousing.

DynamoDB:
- AWS’s flagship NoSQL key-value database.
- Supports high write and read performance for unstructured data.

- Designed for scalability and low-latency access.

DocumentDB:

- MongoDB-compatible NoSQL service for document storage.
- Designed for high availability and handling millions of requests per second.

Neptune:

- Graph database for connected data (social networks, recommendation engines).
- Optimized for fast, complex queries with low-latency.

ElastiCache:

- In-memory caching service for Redis and Memcached.
- Improves application performance by reducing database load.

OpenSearch:

- AWS's version of Elasticsearch for search and analytics.

QLDB (Quantum Ledger Database):

- Managed ledger database with cryptographically verifiable transaction logs.

Timestream:

- Serverless time-series database designed for handling time-based data.

Amazon Managed Blockchain:

- Managed blockchain with Hyperledger Fabric and Ethereum.

Category	Database Type	Service Name
SQL Databases	Relational Database	RDS (MySQL, PostgreSQL, Oracle, etc.)
		Aurora
NoSQL Databases	Key-Value Store	DynamoDB
	Document Store	DocumentDB (MongoDB-compatible)
	Graph Database	Neptune
In-Memory Caching	In-memory Cache	ElastiCache (Redis, Memcached)
Data Warehousing	OLAP (Online Analytical Processing)	RedShift
Time-Series Databases	Time-Series Data	Timestream
Ledger Databases	Transactional Ledger	QLDB
Blockchain	Blockchain	Amazon Managed Blockchain

AWS Analytics Services

dashboards on your data

- **Amazon QuickSight**
  Service de business intelligence sans serveur, propulsé par l'apprentissage automatique, permettant de créer des tableaux de bord interactifs.
  Cas d'utilisation : Analyse d'affaires, création de visualisations, analyse ad-hoc, et extraction d'informations commerciales.
  i**ntégrations** : RDS, Aurora, Athena, Redshift, S3.

query data on Amazon S3 (serverless & SQL)

- **Amazon Athena**
  Service de requête sans serveur pour analyser les données stockées dans Amazon S3 en utilisant SQL.
  Formats supportés : CSV, JSON, ORC, Avro, Parquet.
  Cas d'utilisation : Business intelligence, rapports, analyse des logs.

- **Amazon Redshift**
  Data warehouse OLAP (Online Analytical Processing) basé sur PostgreSQL.
  Optimisé pour l'analyse et l'entreposage de données, avec prise en charge de l'évolutivité massive des données.
  Utilise un stockage en colonnes et l'exécution de requêtes en parallèle massif (MPP).
- **Amazon EMR (Elastic MapReduce)**
  Clusters Hadoop managés pour traiter de vastes volumes de données.
  Prend en charge Apache Spark, HBase, Presto, Flink.
  Cas d'utilisation : Traitement de big data, apprentissage automatique, indexation web.

- **Kinesis:** real-time data streaming, persistence and analysis
- **AWS Glue**
  Service ETL (Extract, Transform, Load) entièrement géré, facilitant la préparation des données pour l'analyse.
- **Data Catalog** : Répertoire central pour les métadonnées.
- **ETL Jobs** : Automatisation de l'extraction, transformation et chargement des données.

Serverless : Évolutif automatiquement sans gestion de serveur.

Application Integration

AWS services designed to help connect and replicate messages, events, or traffic between different application components, often acting as a buffer or queue.

Simple Notification Service (SNS)

A publish/subscribe service where one component publishes a message to a topic, and any component subscribed to that topic will receive the message.

- **Use Case**: Distribute messages to multiple consumers.
- **Behavior**: SNS does not persist messages; if you miss them, they're gone.
- **Analogy**: Think of a topic as a news channel where people tune in to get updates on events.

Simple Queue Service (SQS)

A messaging queue used for sending, storing, and receiving messages between components.

- **Persistence**: SQS can store messages in the queue for days.
- **Queue Types**: FIFO (First In, First Out) and standard queues.

- **AppFlow** Used to integrate data without code
- **EventBridge** – Used to build event-driven applications
- **MQ** – Message broker service for ActiveMQ and RabbitMQ users

Cloud Monitoring

Sec section

Management Services

Management services are services that help manage, provision, or optimize other services.

- **AWS Organizations** – Used to manage multiple AWS accounts

- **Service Catalog** – Allows you to provide CloudFormation and Terraform templates to customers who can use them to deploy resources they need quick self-service portal to launch a set of authorized products pre-defined by admins
- **AWS Control Tower** – Helps you set up AWS Organizations in a secure best practice way, with auditing, logging, and compliance rules in place , Easily setup multiple accounts with best-practices with
- **CloudTrail** – A service that tracks and records all user and API activity in your AWS account
- **Trusted Advisor** to get insights, Support Plan adapted to your needs

Deploying and Managing Infrastructure

Deplyment services

- **CloudFormation** – Infrastructure as code tool that allows you to create templates to provision services
- **OpsWorks** – Allows you to automate server configuration by providing a managed instance of Chef and Puppet
    - ❖ Chef/Puppet are automation platforms that allow you to generate server configs through code
- **Beanstalk:** (AWS only) • Platform as a Service (PaaS), limited to certain programming languages or Docker • Deploy code consistently with a known architecture: ex, ALB + EC2 + RDS
- **Systems Manager** – A secure end-to-end management solution for resources (services) on AWS and on-premise environments
- **CodeDeploy**
  Automates the deployment of applications to various environments such as EC2, Lambda, or on-premises servers.

Developer services

- **CodeDeploy**
  Automates the deployment of applications to various environments such as EC2, Lambda, or on-premises servers.
- **CodeCommit**
  A fully managed source control service to store code in private Git repositories (version-controlled).
- **CodeBuild**
  A fully managed build service to compile, test, and package code in AWS.
- **CodePipeline**
  A fully managed continuous integration and delivery (CI/CD) service for automating the build, test, and deploy phases of applications.
- **CodeArtifact**
  A fully managed artifact repository for storing software packages and dependencies.
- **CodeStar**
  Provides a unified interface for managing software development projects, supporting CI/CD workflows and collaboration.
- **Cloud9**

A cloud-based Integrated Development Environment (IDE) that provides a collaborative environment for writing, running, and debugging code.

- **AWS CDK (Cloud Development Kit)**
  A framework for defining cloud infrastructure using familiar programming languages (e.g., TypeScript, Python, Java)
- **AWS X-Ray** – Analyze and debug distributed applications.
- **AWS AppConfig**: A service for managing and deploying application configurations in real time without restarting the app, allowing separation of logic and configuration.
- **AWS CloudShell**: An integrated cloud-based terminal for running commands and scripts directly in AWS, with a pre-configured development environment

End-user computing services

- **Amazon AppStream 2.0** : Service de diffusion d'applications depuis le cloud vers un navigateur, permettant d'accéder à des applications sans les installer localement.
- **Amazon WorkSpaces** : Bureau virtuel dans le cloud accessible depuis n'importe quel appareil, pour remplacer les postes de travail physiques.
- **Amazon WorkSpaces Web** : Accès sécurisé aux applications web internes via un navigateur, protégeant les données d'entreprise.

Frontend web and mobile services

- **AWS Amplify** : Plateforme pour développer, déployer et héberger des applications web et mobiles, avec des services intégrés comme l'authentification et le stockage.
- **AWS AppSync** : Service pour créer des API GraphQL sécurisées et gérer les données en temps réel entre les applications web et mobiles
- **AWS Device Farm** A service for testing web and mobile applications on real physical devices, providing a wide range of tests to ensure application quality.

Migration Services

Management services are services that help with migrating services from on-prem to AWS

- **Migration Hub** – Allows you to centralize and see all migrations you have in place via AWS services
- **Snow Family** – Used to transfer data into AWS
  - **Snowcone** – most compact an portable device. Comes in SSD & HDD options
  - **Snowball** – Medium sized data transfer (80TB) comes in compute and storage optimized devices
  - **Snowmobile** – Exabyte-scale data migration device used to move large amounts of data to AWS
- **OpsHub**: desktop application to manage Snow Family devices

■ Migrate up to 100PB in a 45 foot long ruggedized shipping container

- **Online Data transfer**
  - FTP, SFTP, FTPS,
  - AS2 – send/receive messages into S3 backend

- **DataSync** – secure, online service that automates and acclerates moving data between on premises and AWS storage services
- **Application Discovery Service** - helps you plan cloud migration projects by gathering information about your on-premises data centers
- **Application Migation Service (MGN)**– Does the actually moving of applications from onpremise into AWS
- **Database Migration Service** - is a managed migration and replication service that helps move your database and analytics workloads to AWS quickly, securely, and with minimal downtime and zero data loss
- **Elastic Disaster Recovery** - minimizes downtime and data loss with fast, reliable recovery of on-premises and cloud-based applications using affordable storage, minimal compute, and point-in-time recovery
- **Mainframe Modernization** – assists in migrating mainframe applications to the cloud
- **Migration Evaluator**Helps build a business case for migrating to AWS by providing a baseline of current infrastructure and using tools to analyze the current state and plan the migration.

AWS has six methods/patterns for migration

- **Rehosting** – also known as “lift-and-shift" involves movign applications without changes
  - Allows companies to carryout migrations and scale quickly as possible

Optimisation légère sans modifier l’architecture.

- **Replatforming** - make a few cloud (or other) optimizations in order to achieve some tangible benefit, but you aren’t otherwise changing the core architecture of the application
- **Refactoring** - involves reimagining how an application is architected and developed by using cloud-native features. Refactoring is driven by a strong business need to add features, scale, or performance that would otherwise be difficult to achieve in the application’s existing environment.
- **Repurchasing** - involves moving from a traditional license to a software-as-a-service model. For example, a business might choose to implement the repurchasing strategy by migrating from a customer relationship management (CRM) system to Salesforce.com
- **Retaining** - consists of keeping applications that are critical for the business in the source environment. This might include applications that require major refactoring before they can be migrated, or, work that can be postponed until a later time

Supprimer les composants inutiles lors de la migration.

- **Retiring** – When migrating to the cloud, companies may find certain components of their infrastructure are no longer needed, in which case they are retired and not moved to the cloud

Artificial Intelligence / Machine Learning (AI/ML)

- **Rekognition**: Face detection, object labeling, celebrity recognition.
- **Transcribe**: Converts audio to text (e.g., subtitles).
- **Polly**: Converts text to speech/audio.
- **Translate**: Translates text between languages.
- **Lex**: Builds conversational bots (chatbots).
- **Connect**: Cloud contact center solution.
- **Comprehend**: Natural language processing (NLP) for understanding text.

- **SageMaker**: Machine learning platform for developers and data scientists.
- **Forecast**: Builds highly accurate forecasts using machine learning.
- **Kendra**: Machine learning-powered search engine for improved search results.
- **Personalize**: Real-time personalized recommendations for users.
- **Textract**: Detects and extracts text and data from documents.

IoT (Internet of Things)

AWS IoT Core

- A platform that enables secure connection of Internet of Things (IoT) devices to AWS.
- Allows devices to communicate with cloud applications and other devices securely and at scale.

AWS IoT Greengrass

- Provides local compute, messaging, and data management for connected devices.
- Enables devices to act locally on the data they generate, while still using the cloud for management, analytics, and storage.

AWS IoT Analytics

- A fully managed service to analyze IoT data.
- Provides insights into device performance and behavior, helping to optimize IoT systems.

Customer Engagement And Business Applications

- **Amazon Connect**: A cloud-based contact center service that enables businesses to provide customer support and engagement through voice, chat, and other channels.
- **Amazon SES (Simple Email Service)**
  - Email sending service for marketing, transactional, and notification emails.
  - Allows businesses to send and receive emails securely and cost-effectively
- **AWS IQ**: A service connecting businesses with AWS-certified experts for on-demand project work and professional services.
- **AWS Managed Services (AMS)**: A fully managed service that helps organizations operate their AWS infrastructure with 24/7 support, ensuring performance, security, and compliance.
- **Amazon Pinpoint**
  Scalable marketing communications service for email, SMS, push, voice, and in-app messaging.
- **AWS Support**: A range of support plans offering access to AWS experts, best practices, and resources to help businesses optimize and troubleshoot their AWS environments.

Others

Amazon Elastic Transcoder

A service that converts media files stored in Amazon S3 into formats optimized for client devices. Useful for videos, audios, or other media files in various formats.

AWS Fault Injection Simulator (FIS)

A service for simulating disruptions in AWS workloads to test system resilience. Based on Chaos Engineering principles, it helps identify bugs and bottlenecks.

AWS Step Functions

A service for creating visual workflows to orchestrate Lambda functions. It handles sequences, parallel actions, conditions, and errors, and integrates with other AWS services like EC2, ECS, and API Gateway.

AWS Ground Station

A fully managed service for controlling satellite communications and processing data directly in AWS. Use cases include weather forecasting, communications, and live video broadcasts.

Scalability:

- **Vertical**: Increase instance size (e.g., t2.micro to t2.large).
- **Horizontal**: Add more instances to handle load.

High Availability:

- Deploy across **multiple Availability Zones** for fault tolerance.

Elasticity:

- Automatically scale resources based on demand (pay-per-use).

Elastic Load Balancer (ELB):

- Distribute traffic across instances to prevent overload.
- Types: **ALB** (HTTP/HTTPS), **NLB** (TCP), **GWLB** (Firewall).

Auto Scaling Groups (ASG):

- Auto-adjust EC2 instances based on load.
- **Scaling**: Dynamic (based on demand), Predictive (forecasted load).