AMAZON EC2

ABINAYA SURESH - 22Z203
AKSHARAA P- 2ZZ208
HARSHINI S P - 22Z224
RAMAPRIYA S - 22Z252
RITHANIYAA B - 22Z252
THEETCHANAA RA - 22Z269

INTRODUCTION TO AMAZON EC2

RAMAPRIYA S - 22Z252

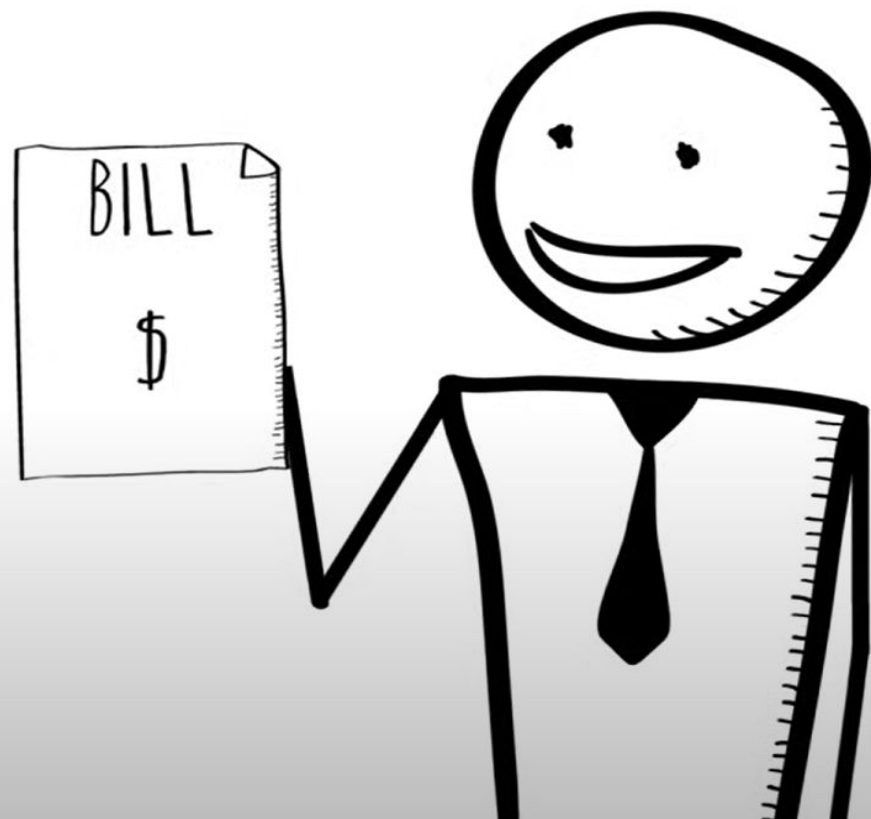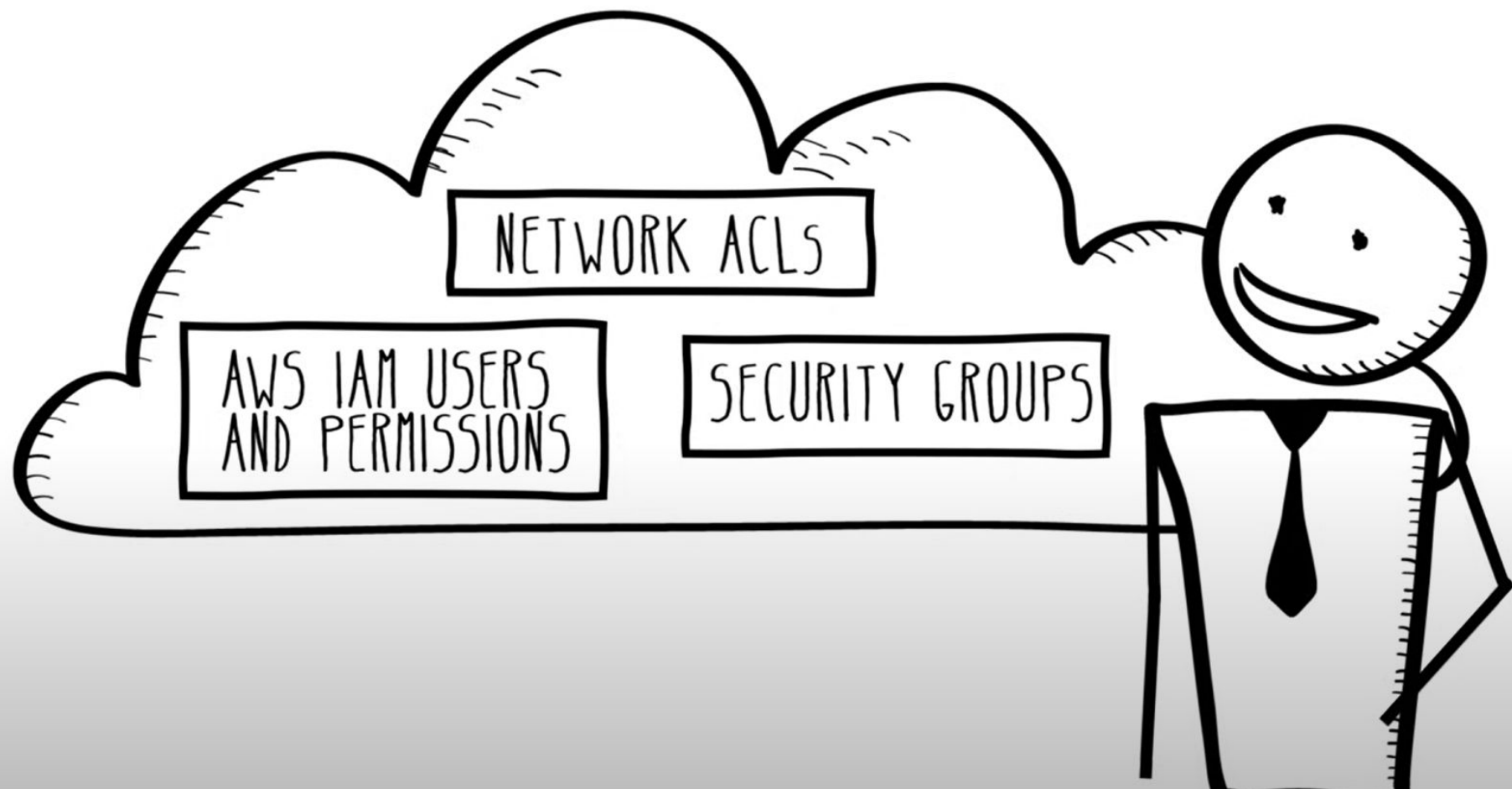# EC2 Architecture & Key Components

ABINAYA SURESH - 22Z203

EC2 Architecture & Key Components

Amazon EC2 (Elastic Compute Cloud) provides scalable computing power in the AWS cloud. It is a key component of cloud-based applications, supporting high availability, security, and performance.

- Amazon Cloud-front
- Elastic Load Balancer
- Security Management
- Elastic Caches
- Elastic Block Storage
- Storage & Backups
- EC2 Instances
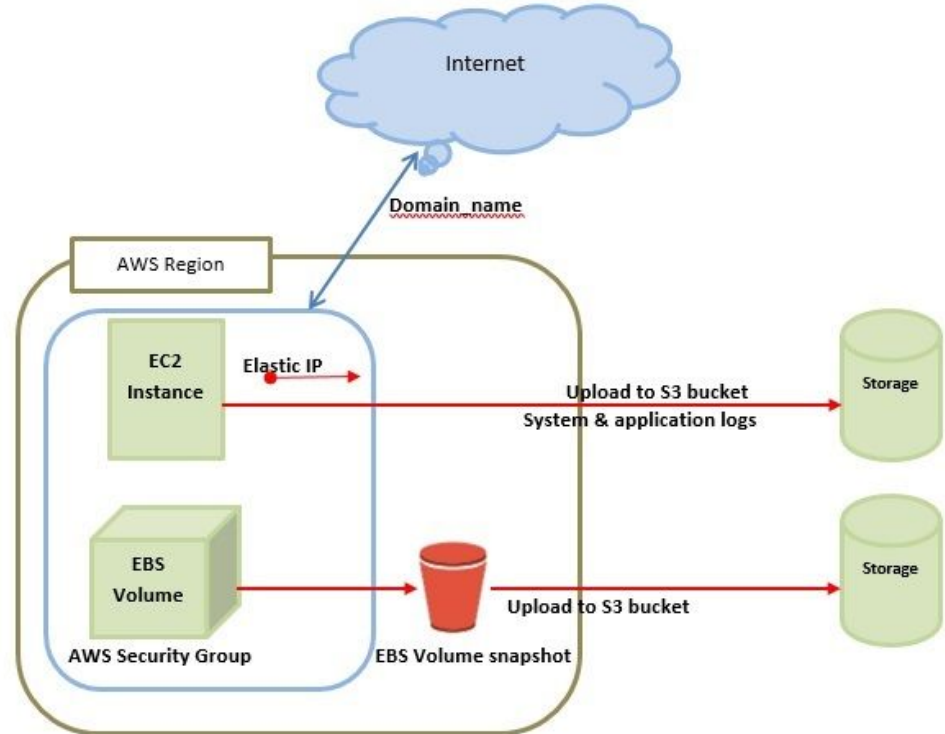
Amazon CloudFront

- A Content Delivery Network (CDN) for fast delivery of dynamic, static, and streaming content.
- Uses a global network of edge locations for lower latency and improved performance.
- Works seamlessly with AWS services like S3 & EC2 and supports non-AWS origin servers.
- **Pay-as-you-go** pricing with no upfront costs.

Elastic Load Balancer (ELB)

- Distributes incoming traffic across multiple EC2 instances for **high availability & fault tolerance**.
- Supports **auto-scaling** based on traffic fluctuations.
- Types of Load Balancers:
    - **Application Load Balancer (ALB)** – HTTP/HTTPS traffic
    - **Network Load Balancer (NLB)** – High-performance TCP/UDP traffic
    - **Classic Load Balancer (CLB)** – Basic load balancing

Security Management in EC2

- **Security Groups:** Acts as a virtual firewall controlling inbound & outbound traffic.
- **Network ACLs:** Provides an extra layer of security at the subnet level.
- **IAM (Identity & Access Management):** Controls permissions for users and services.
- **AWS Shield:** DDoS protection for AWS applications.

Elastic Caches

- **Amazon ElastiCache** manages in-memory caching for **fast data retrieval**.
- Reduces database load and improves **application performance**.
- Supports **Memcached** and **Redis** for caching frequently accessed data.

Storage & Backups

- **Amazon S3:** Scalable object storage for backup & archival.
- **Amazon EBS (Elastic Block Storage):** Persistent storage for EC2, supporting snapshots & redundancy.
- **AWS Backup:** Centralized service for automated backups & compliance.

EC2 Instances

- Virtual servers providing **scalable compute power** in AWS cloud.

EC2 INSTANCE TYPES & USE CASES

THEETCHANAA RA - 22Z269

AWS EC2 offers different instance types optimized for various workloads.

- General Purpose
- Compute Optimized
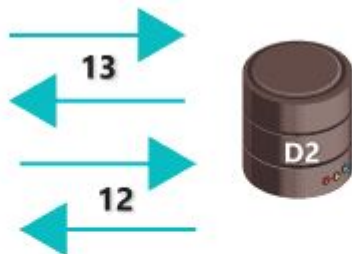- Memory Optimized
- Storage Optimized
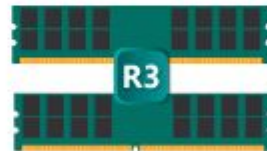- Accelerated Computing

General purpose — T2, M4

Compute optimized — C5, C4, C3

Storage and I/O optimized — D2, 13, 12

Memory optimized — X1, R4, R3

GPU or FPGA enabled — F1, P2, G2

General-Purpose Instances

Balanced in Compute, Memory & Networking
Best for: Gaming servers, small databases, web apps, personal projects

Examples:

- **T2.micro**: 1 CPU, 1 GB RAM (Free tier)
- **M6a.large**: 2 CPUs, 8GiB RAM
- **M5.large**: Intel Xeon-powered

Features:

- AWS Graviton3 Processors - Offers better performance at a lower cost.
- EBS Optimized by default - Faster disk performance.
- Lightweight Hypervisor & Dedicated Hardware - Improves security and performance.
- Higher Bandwidth than other types

Applications: Web hosting, Dev/Test Environments, Content Delivery

Compute-Optimized Instances

Optimized for High CPU Performance
Best for: Machine Learning, Gaming, Web servers, Batch Processing

Examples:

- **C5d.24 xlarge**: 96 CPUs, 192 GiB RAM, 12 Gbps Network

Features:

- AWS Graviton3 Processors
- DDR5 Memory (50% more bandwidth than DDR4)
- Enhanced Networking
- EBS Optimization enabled - for fast storage performance

Applications: Real-time multiplayer gaming, Machine Learning training

Memory-Optimized Instances

Optimized for Large Memory Workloads
Best for: High-performance databases, Real-time processing

Examples:

- **R7g.medium**: 1 CPU, 8 GiB RAM
- **X1**: 64 vCPU, 976 GiB RAM

Features:

- Elastic Fabric Adapter (EFA) Support - for low-latency networking
- DDR5 Memory (50% more bandwidth than DDR4)
- 20% improved networking over R6g instances

Applications: In-memory databases, Big Data Processing

Storage-Optimized Instances

Designed for Fast Data Access
Best for: Data Warehousing, High-frequency OLTP, Distributed File Systems

Examples:

- **Im4gn.large**: 2 CPUs, 8 GiB RAM, EBS Storage

Features:

- AWS Graviton2 Processors for best price-performance
- Optimized for workloads needing 4GB RAM per vCPU
- 100 Gbps Enhanced Networking (ENA) for high-throughput data processing
- Supports Local NVMe Storage (I3, D3 instances) for ultra-fast access

Applications: Amazon EC2 C5d/I3 instances, High-storage applications

Accelerated Computing Instances

Uses GPUs & Hardware Accelerators for Faster Processing
Best for: Graphics rendering, Machine Learning, Simulations

Examples:

- **P4**: 8 GPUs, 96 CPUs, 1152 GiB RAM

Features:

- Intel Xeon Scalable Processors (3.0 GHz)
- NVIDIA A100 Tensor Core GPUs
- 400 Gbps RDMA & Elastic Fabric Adapter (EFA)

Applications: EC2 P3 for Deep Learning, EC2 G4 for Gaming & Video Processing

EC2 PRICING MODELS AND COST MANAGEMENT

HARSHINI S P - 22Z224

INTRODUCTION

➢ Amazon EC2 provides scalable virtual computing resources in the cloud.
➢ AWS offers multiple pricing models to accommodate different usage patterns and budgets.
➢ Users pay only for the compute resources they consume.
➢ **Importance of Cost Management**
   ○ Helps prevent unnecessary expenses by optimizing resource allocation.
   ○ Ensures efficient usage of cloud resources to maximize cost savings.
   ○ Enables scalability while maintaining control over cloud spending.

EC2 Pricing Models

1. On-Demand Instances

- Pay for compute capacity per second or hour with no long-term commitment.
- Suitable for unpredictable workloads or applications with short-term needs.
- Higher cost compared to other pricing models but offers flexibility.

2. Reserved Instances (RI)

- Provides significant cost savings for workloads with steady-state usage.
- Requires a commitment for one or three years.
- Offers three payment options: All upfront, Partial upfront, No upfront.

3. Spot Instances

- Allows using unused AWS capacity at a discounted rate.
- Best for fault-tolerant workloads like batch processing and big data analysis.
- Instances can be terminated if AWS needs the capacity back.

4. Savings Plans

- Offers flexible discounts based on a commitment to consistent usage.
- Compute Savings Plans apply to any EC2 instance type.
- EC2 Instance Savings Plans provide better discounts but are specific to instance families.

5. Dedicated Hosts and Dedicated Instances

- Dedicated Hosts allow running workloads on physical servers reserved for a single customer.
- Dedicated Instances run on hardware that is not shared with other customers.
- Useful for compliance and licensing requirements.

6. Capacity Reservations

- Guarantees EC2 capacity in a specific Availability Zone.
- Ideal for critical applications requiring high availability.
- Charges apply even when instances are not running

Factors affecting EC2 Costs

➢ **Instance Type and Size** – Costs vary based on vCPUs, memory, and configuration. Choosing the right instance optimizes expenses.

➢ **Region and Availability Zone** – Prices differ across AWS regions due to demand and infrastructure costs. Selecting the right region helps reduce spending.

➢ **Storage Costs** – EBS volumes, snapshots, and backups add to costs. Choosing the right storage type minimizes expenses.

➢ **Data Transfer Costs** – Inbound data is free, but outbound transfers, especially between regions or the internet, incur charges.

➢ **Load Balancing and Networking** – ELB and NAT Gateways increase costs. Optimizing network architecture helps control spending.

EC2 Cost Optimization Strategies

➢ **Right-Sizing Instances** – Select optimal instance types based on workload needs using AWS Compute Optimizer.

➢ **Auto-Scaling** – Adjusts instances automatically based on demand, reducing idle costs.

➢ **Spot Instances** – Use for batch processing, testing, and analytics, combined with On-Demand or Reserved Instances.

➢ **Reserved Instances & Savings Plans** – Commit to long-term usage for cost savings on predictable workloads.

➢ **Monitor & Analyze Costs** – Use AWS Cost Explorer and Budgets to track spending and set alerts.

➢ **AWS Compute Optimizer** – Recommends instance types to avoid underutilization or over-provisioning.

AWS Cost Management Tools

➢ **Cost Explorer** – Analyzes spending trends with filtering options.
➢ **AWS Budgets** – Sets spending limits and sends alerts when exceeded.
➢ **Cost & Usage Report** – Provides detailed billing insights for cost optimization.
➢ **Trusted Advisor** – Suggests cost-saving improvements by identifying unused resources.
➢ **Billing Dashboard** – Displays a breakdown of service costs and usage.

Real-World Use Cases

➢ **Web Hosting** – Use On-Demand for deployment, Reserved for stable loads, and Auto-Scaling for peak traffic.
➢ **Big Data Processing** – Use Spot Instances, store data in S3, and process with EC2 and AWS EMR.
➢ **Development & Testing** – Use Spot Instances, monitor budgets, and schedule shutdowns to save costs.
➢ **High Availability & Disaster Recovery** – Deploy across multiple AZs, use Reserved Instances, and back up with AWS S3.

Security, Monitoring, and Backup in EC2

RITHANIYAA B - 22Z252

Security in Amazon EC2

Key Security Features:

IAM (Identity & Access Management) – Controls who can access EC2 and what actions they can perform by defining user roles and permissions.

Security Groups & Network ACLs – Acts as a firewall, filtering incoming and outgoing traffic at both the instance and subnet levels to prevent unauthorized access.

Encryption – Protects stored data (EBS, S3, RDS) and data in transit using strong encryption protocols, ensuring information remains secure.

AWS Shield & GuardDuty – Shields EC2 instances from DDoS attacks and security threats by continuously monitoring and automatically mitigating risks.

Monitoring in Amazon EC2

Amazon CloudWatch – Tracks **EC2 performance metrics** (CPU, memory, network activity) in real-time and sends alerts for unusual behavior.

AWS CloudTrail – Logs **all user activities and API actions** taken on EC2, helping with security audits and troubleshooting unauthorized access.

AWS Trusted Advisor – Provides **recommendations for security, cost optimization, and performance** to improve EC2 efficiency and reliability.

Backup & Disaster Recovery in EC2

EBS Snapshots – Creates point-in-time backups of EC2 storage, allowing quick data restoration in case of failure or accidental deletion.

Amazon Machine Images (AMI) – Saves a full copy of an EC2 instance, enabling quick recovery or scaling of pre-configured setups.

AWS Backup Service – Automates scheduled backups of EC2 and other AWS services, ensuring data is consistently protected.

Disaster Recovery Strategies – Uses Multi-AZ deployment, Cross-Region Replication, and Load Balancing to maintain availability and prevent downtime.

EC2 BENEFITS & CHALLENGES

AKSHARAA P- 2ZZ208

BENEFITS

Scalability:

- Instantly scale up during traffic spikes or business growth
- Scale down during low-demand periods to reduce costs
- Auto Scaling groups automate the process based on predefined metrics

Cost-effectiveness:

- No upfront hardware investments required
- Various pricing models to optimize for your specific needs
- Detailed cost analysis and budgeting tools

Flexibility:

- Over 400 instance types optimized for different workloads
- Support for Windows, Linux, and macOS
- Bare metal instances available for specialized workloads

Security:

- Default encryption for data at rest and in transit
- Security groups act as virtual firewalls
- Regular security patches and updates

CHALLENGES

Cost Management:

- Unexpected costs from forgotten or idle resources
- Complex pricing models can be difficult to predict
- Data transfer costs can accumulate quickly
- Need for dedicated cost monitoring and optimization

Dependency on Internet Connectivity:

- Reliance on stable internet connections
- Potential single point of failure
- Bandwidth limitations can affect performance
- Latency issues for certain applications
- Need for redundant connectivity solutions