AWS EC2 – Full In-Depth Notes

🔹 What is Amazon EC2?

Amazon Elastic Compute Cloud (EC2) is a **web service** that provides **secure, resizable compute capacity** in the cloud. It allows you to launch virtual servers (instances) on demand.

Use case: Hosting applications, running containers, big data workloads, and more.

🔹 Core EC2 Concepts

Concept	Description
Instance	A virtual server in the cloud
AMI (Amazon Machine Image)	A template for the instance OS and software
Instance Type	Defines compute, memory, storage (e.g., t2.micro, m5.large)
Key Pair	SSH key used for secure access
Security Group	Virtual firewall controlling traffic
EBS Volume	Persistent block storage attached to instances
Elastic IP	Static public IP that you can assign to an instance

---

🔹 EC2 Instance Lifecycle

1. **Pending** → preparing the instance
2. **Running** → instance is active
3. **Stopping** → shutting down gracefully
4. **Stopped** → halted but preserved
5. **Terminated** → deleted and unrecoverable

🔹 EC2 Instance Types (Families)

Family	Use Case
General Purpose (t3, t4g, m5)	Balanced compute, memory, network
Compute Optimized (c5, c6g)	High-performance compute workloads
Memory Optimized (r5, x2idn)	Large in-memory databases, real-time analytics

Family	Use Case
Storage Optimized (i3, d3)	High IOPS workloads (e.g., NoSQL, cache)
Accelerated Computing (p4, inf1)	GPU-intensive apps, ML, HPC

Tip: Use **T3 or T4g** for general usage; **Spot Instances** for batch jobs.

🔹 Amazon Machine Image (AMI)

An **AMI** includes:

- Operating system (e.g., Linux, Windows)
- Application server
- Custom software

Types of AMIs:

- AWS-provided
- Marketplace AMIs
- Custom AMIs (your snapshots)

🔹 EC2 Storage Options

Storage	Description
EBS (Elastic Block Store)	Persistent block storage for EC2 (gp3, io2, etc.)
Instance Store	Ephemeral storage physically attached to host (lost on stop/terminate)
EFS (Elastic File System)	Shared file system for multiple instances
FSx	Windows/Linux native file system support

🔹 EC2 Networking

Component	Description
VPC	Virtual network to launch EC2 in
Subnet	Subdivision of a VPC (public or private)
Security Group	Acts as a stateful firewall
Network ACL	Optional stateless network rules

Component	Description
Elastic IP	Static public IP for instances
ENI (Elastic Network Interface)	Additional NIC you can attach to EC2
Placement Group	Control instance placement (cluster, spread, partition)

🔹 **EC2 Purchasing Options**

Option	Description
On-Demand	Pay per second with no long-term commitment
Reserved Instances	1 or 3-year commitment, up to 75% savings
Spot Instances	Unused capacity at up to 90% discount, may be interrupted
Savings Plans	Flexible pricing model (commit $/hr for compute)
Dedicated Host	Physical server for compliance or licensing needs

---

🔹 **Elastic Load Balancer Integration**

Use **ALB** or **NLB** to distribute traffic across EC2 instances in multiple AZs:

- Improves availability
- Enables health-based routing
- Supports auto scaling integration

🔹 **Auto Scaling Integration**

Auto Scaling allows you to:

- Add/remove instances automatically
- Scale based on CPU, memory, or custom metrics
- Maintain minimum instance count

🔹 **Monitoring & Logging**

Service	Description
CloudWatch	Collect CPU, disk, network metrics; set alarms

Service	Description
CloudTrail	Log API actions related to EC2
EC2 Serial Console	Access for troubleshooting boot/login issues
AWS Systems Manager (SSM)	Manage and patch EC2 at scale (no SSH needed)

🔹 **Security**

Feature	Description
Key Pair (SSH)	For Linux instance login
Password (RDP)	For Windows instances
Security Group	Whitelists IPs and ports
IAM Roles	Assign permissions to EC2 to access other AWS services
SSM Agent	For secure access without SSH (via Systems Manager)

🔹 **EC2 Pricing Considerations**

You pay for:

- Compute time (per second)
- EBS volumes
- Elastic IP (if unused)
- Data transfer out (to internet or other regions)
- Additional services (SSM, CloudWatch metrics, etc.)

💡 Always stop or terminate unused instances!

🔹 **EC2 Backup & Recovery**

- **EBS Snapshots** – Backup volumes to S3
- **AMI Creation** – Save configured instance images
- **Recovery** – Use snapshots/AMIs to restore or launch new instances

---

🔹 **EC2 Hands-On Features**

Feature	Description
User Data	Run scripts at instance launch (e.g., install software)
Cloud Init	Automate setup during boot (on Linux)
Elastic IP	Maintain static IP even if instance changes
Instance Metadata	Access instance-specific info via HTTP (169.254.169.254)
Nitro System	New virtualization stack for better performance and security

---

🔹 **EC2 Best Practices**

✅ Use **IAM roles** instead of access keys on instances
✅ Enable **detailed monitoring** for production workloads
✅ Use **placement groups** for low-latency or fault tolerance
✅ Turn off **unused Elastic IPs** to avoid charges
✅ Apply **auto scaling** for availability and cost savings
✅ Encrypt **EBS volumes** at rest
✅ Use **SSM Agent** to avoid SSH in production environments
✅ Tag resources for cost tracking and automation

🔹 **Common EC2 Use Cases**

- Web app hosting (Apache, NGINX)
- Backend services (APIs, microservices)
- Batch processing
- Machine learning training
- Game servers
- CI/CD runners (e.g., Jenkins agents)