AWS INTRO

- Leading cloud platform with 200+ services
- Global network of regions and availability zones
- Pay-as-you-use cost model (generally cheaper than traditional infrastructure)
- Started in 2006 with S3 and EC2, expanded rapidly through 2010 and beyond

Cloud Service Models

- **IaaS**: Infrastructure as a Service - basic building blocks, more control
- **PaaS**: Platform as a Service - focus on deploying applications without managing infrastructure
- **SaaS**: Software as a Service - complete applications managed by vendors

Shared Responsibility Model

AWS Handles:

- Physical infrastructure security
- Data center operations (power, HVAC)
- Network equipment and global connectivity
- Virtualization layer and host operating systems
- Managed service maintenance

Customer Handles:

- Data classification, encryption and access policies
- IAM configuration and access management
- Operating systems on self-hosted applications
- Network security within VPCs
- Compliance and governance requirements

AWS Global Infrastructure

- **Regions**: Distinct geographical areas (us-east-1, etc.)
- **Availability Zones**: Multiple isolated data centers per region
- **Edge Locations**: Points of presence for content delivery
- Current scale: 36+ regions, 114+ availability zones, 700+ CloudFront POPs

Key Service Categories

Compute

- EC2 - Virtual machines with various instance types
- Lambda - Serverless compute
- Container services: ECS, EKS, ECR, Fargate

Storage

- S3 - Object storage with multiple storage classes
- EBS - Block storage for EC2 instances
- EFS - Elastic File System for shared storage
- File Cache - High-speed caching
- AWS Backup - Automated data protection

Database

- RDS & Aurora - Relational databases
- DynamoDB - NoSQL key-value database
- MemoryDB & ElastiCache - In-memory databases
- DocumentDB - MongoDB-compatible document database
- Neptune - Graph database

Analytics

- Athena - Query data in S3 using SQL
- EMR - Hadoop/Spark processing
- Glue - ETL and data cataloging
- Redshift - Data warehousing
- Kinesis - Real-time data streaming
- QuickSight - Business intelligence

Machine Learning

- SageMaker - End-to-end ML platform
- Pre-built AI services:
  - Comprehend - Natural language processing
  - Rekognition - Image and video analysis
  - Textract - Document text extraction
  - Translate - Language translation

Implementation Considerations

Instance Selection

- Match instance types to workload requirements
- Consider specialized instances for specific needs
- Leverage auto-scaling for variable workloads

Storage Strategy

- Use appropriate storage class based on access patterns
- Implement lifecycle policies for cost optimization
- Ensure proper backup and redundancy

Database Selection

- Choose database type based on data model
- Consider read/write patterns and scale requirements
- Plan for proper backup and disaster recovery

Cost Management

- Monitor usage with Cost Explorer and Budgets
- Utilize reserved instances for predictable workloads
- Implement tagging for cost allocation
- Consider Savings Plans for flexible commitments

Security Best Practices

- Enable MFA for all accounts
- Implement least privilege access with IAM
- Encrypt data at rest and in transit
- Use security groups and network ACLs effectively
- Monitor with GuardDuty and Security Hub

Resilience Planning

- Design multi-AZ deployments for high availability
- Consider multi-region for critical applications
- Implement proper backup and recovery procedures
- Test disaster recovery regularly

AWS Free Tier

- 12-month access to select services with limitations
- EC2: 750 hours/month of t2.micro instances
- S3: 5GB storage with limited operations
- RDS: 750 hours/month of db.t2.micro instances