Amazon Web Service (AWS)

Tessema Mengistu (Ph. D.)

mengistu@vt.edu

Outline

- Amazon Cloud Infrastructure
- Amazon Web Service

Amazon Cloud Infrastructure

- AWS is the front runner cloud service provider – 32% market share
- The AWS Cloud infrastructure is built around:
  - AWS Regions
    - A geographical area in the world
    - Have 2/3 Availability Zones
  - Availability Zones (AZ)
    - Consist of one or more discrete data centers
      - Each with redundant power, networking, and connectivity
        - Separate facilities
        - Connected by a high-bandwidth, low-latency network, fully redundant fiber based dedicated network
  - 34 Regions, 108 AZs   - as of 09/2024

3

Amazon Cloud Infrastructure

Source: https://aws.amazon.com/about-aws/global-infrastructure/
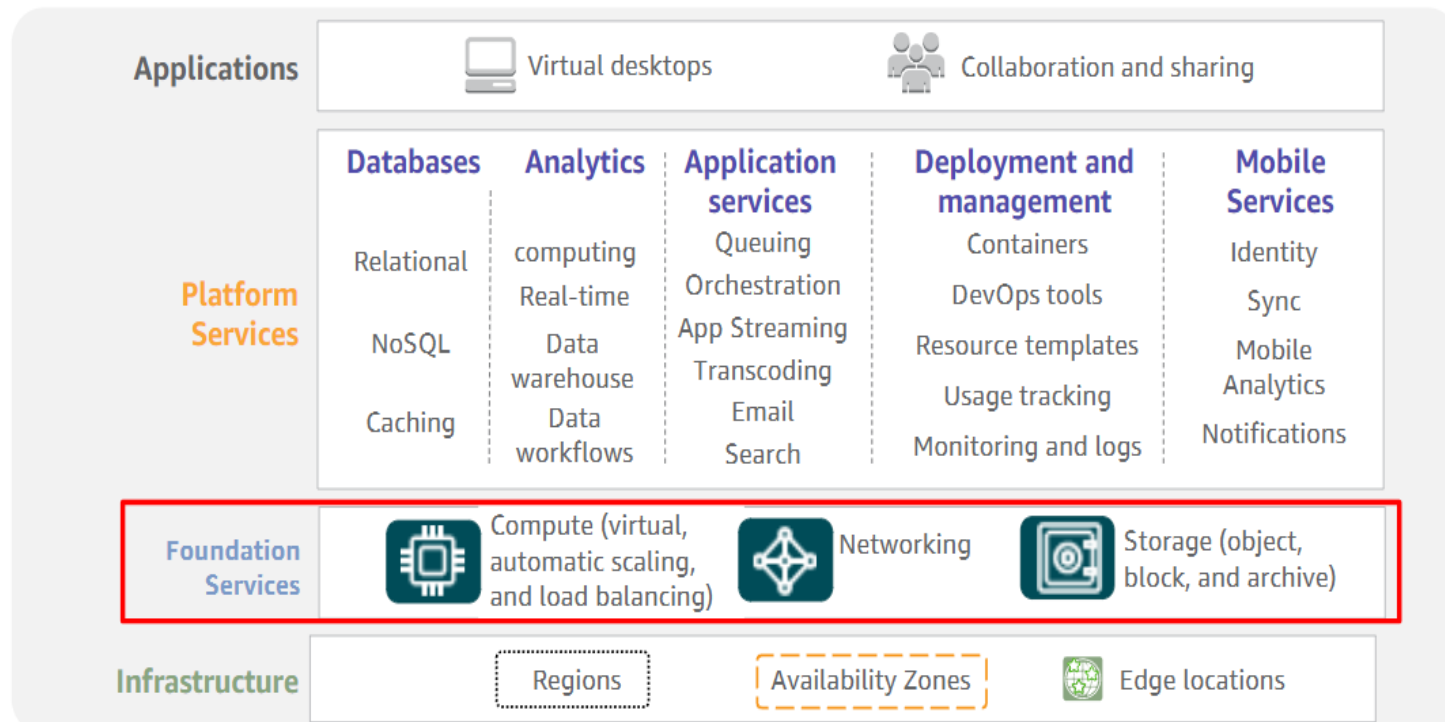
Amazon Cloud Infrastructure

- Factors to select a region
  - Data governance and compliance
  - Energy cost

Amazon Web Service

- Amazon Web Service (AWS)
  - A secure cloud platform that provides a broad range of global cloud-based products

                                    - AWS Academy

  - Categorized into service types

Amazon Web Service

AWS foundational services

Applications	Virtual desktops		Collaboration and sharing	

Platform Services	**Databases**	**Analytics**	**Application services**	**Deployment and management**	**Mobile Services**
	Relational	computing	Queuing	Containers	Identity
		Real-time	Orchestration	DevOps tools	Sync
	NoSQL	Data warehouse	App Streaming	Resource templates	Mobile Analytics
			Transcoding	Usage tracking	
	Caching	Data workflows	Email	Monitoring and logs	Notifications
			Search		

Foundation Services	Compute (virtual, automatic scaling, and load balancing)	Networking	Storage (object, block, and archive)

Infrastructure	Regions	Availability Zones	Edge locations

Amazon Web Service

Compute services –
- Amazon EC2
- AWS Lambda
- AWS Elastic Beanstalk
- Amazon EC2 Auto Scaling
- Amazon ECS
- Amazon EKS
- Amazon ECR
- AWS Fargate

Storage services –
- Amazon S3
- Amazon S3 Glacier
- Amazon EFS
- Amazon EBS

Management and Governance services –
- AWS Trusted Advisor
- AWS CloudWatch
- AWS CloudTrail
- AWS Well-Architected Tool
- AWS Auto Scaling
- AWS Command Line Interface
- AWS Config
- AWS Management Console
- AWS Organizations

Database services –
- Amazon RDS
- Amazon DynamoDB
- Amazon Redshift
- Amazon Aurora

Security, Identity, and Compliance services –
- AWS IAM
- Amazon Cognito
- AWS Shield
- AWS Artifact
- AWS KMS

Networking and Content Delivery services –
- Amazon VPC
- Amazon Route 53
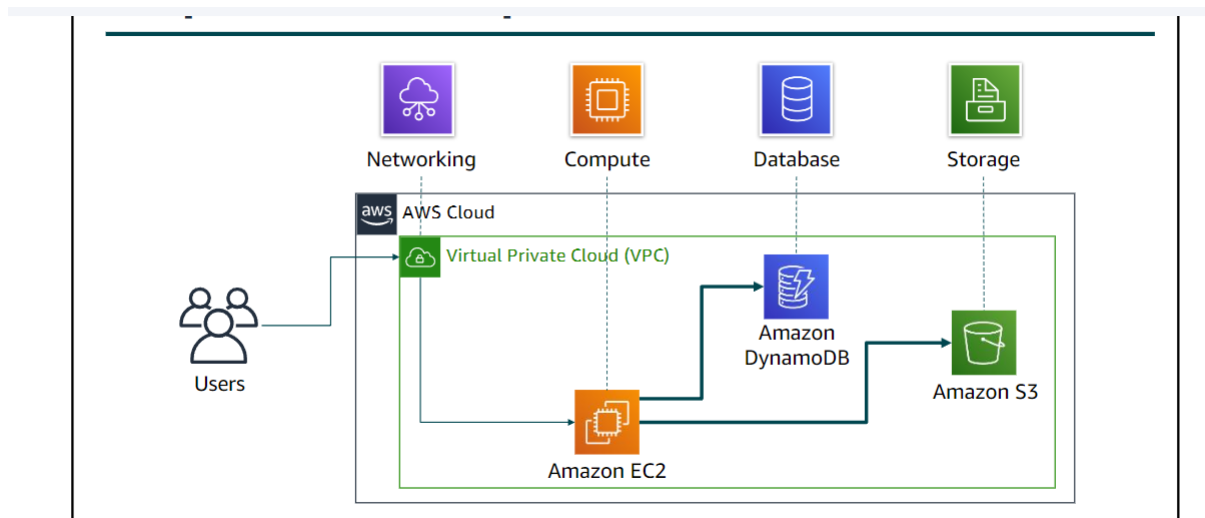- Amazon CloudFront
- Elastic Load Balancing

AWS Cost Management services –
- AWS Cost & Usage Report
- AWS Budgets
- AWS Cost Explorer

Amazon Web Service

Amazon Web Service

- Three ways of communication with AWS
  - AWS Management Console
    - Rich graphical interface
  - AWS CLI
    - Suite of utilities launched from scripts
  - SDK
    - Accessing AWS from different programming languages

AWS Compute Services

- Can be classified as IaaS, Serverless, container based, and PaaS
- Different services
  - Amazon Elastic Compute Cloud (EC2)
  - Amazon EC2 Auto Scaling
  - Amazon Elastic Container Service (ECS)
  - AWS Elastic Beanstalk
  - AWS Lambda
  - . . .

AWS Compute Services

- Amazon Elastic Compute Cloud (EC2)
  - An IaaS offered by AWS powered by a huge infrastructure
  - Computing capacity is provided in the form of virtual machines or server instances by booting Amazon Machine Images **(AMI)**
  - The infrastructure is virtualized by a custom-developed, minimized hypervisor based on KVM - 2017
    - Nitro System
  - Supports different workloads: web, game, database, application server, . . .

AWS Compute Services

- Amazon Machine Images (AMI)
  - Information required to launch an EC2 instance
    - Operating system and everything that is installed in the OS
    - Launch permission
  - Can be user defined or AWS provided

AWS Compute Services

EC2 instance type naming and sizes

Instance type naming

- Example: **t3.large**
  - T is the family name
  - 3 is the generation number
  - Large is the size

Example instance sizes

Instance Name	vCPU	Memory (GB)	Storage
t3.nano	2	0.5	EBS-Only
t3.micro	2	1	EBS-Only
t3.small	2	2	EBS-Only
t3.medium	2	4	EBS-Only
t3.large	2	8	EBS-Only
t3.xlarge	4	16	EBS-Only
t3.2xlarge	8	32	EBS-Only

AWS Compute Services

Select instance type: Based on use case

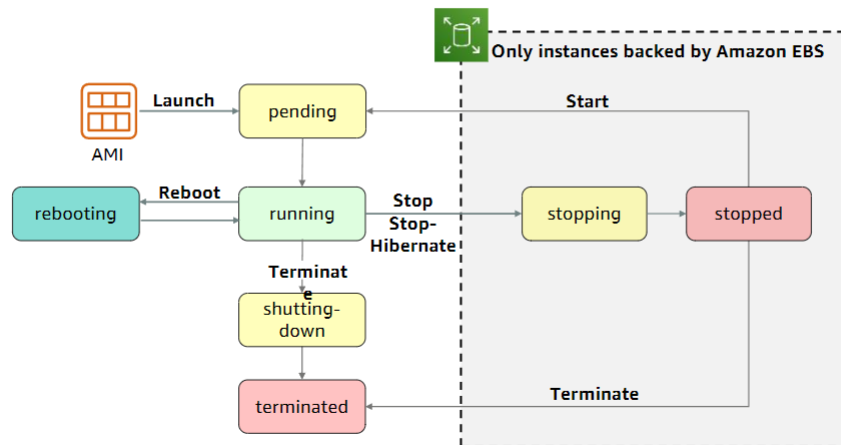|	General Purpose	Compute Optimized	Memory Optimized	Accelerated Computing	Storage Optimized
Instance Types	a1, m4, m5, t2, t3	c4, c5	r4, r5, x1, z1	f1, g3, g4, p2, p3	d2, h1, i3
Use Case	Broad	High performance	In-memory databases	Machine learning	Distributed file systems

AWS Compute Services

Amazon EC2 instance lifecycle

AWS Compute Services

- Pricing models
    - On-Demand Instances
    - Spot Instances
    - Reserved Instances
    - Dedicated instances
    - Dedicated Hosts

AWS Storage Services

- AWS provides multiple storage services
  - Protocol
    - Block
    - File-based
    - Object
  - Client type
    - Windows
    - Linux/Unix
  - Migration
    - Online
   - Offline
  - . . .

AWS Storage Services

- Amazon Elastic Block Store (Amazon EBS)
  - Persistent block storage volumes for use with Amazon EC2 instances in the AWS Cloud
  - Throughput- and transaction- intensive workloads
  - Mountable storage and only on EC2 instance can mount at a time
  - Automatically replicated within the AZ
  - Example use cases:
    - HD for EC2

AWS Storage Services

- Amazon Simple Storage Service (S3)
  - An object storage service
  - It supports a minimal set of functions
    - write, read, and delete
  - Objects store in **buckets**
    - Logical containers with universally unique names
  - Common use cases:
    - Static website hosting
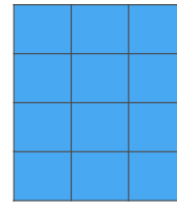   - backup/disaster recovery

AWS Storage Services

- Block vs. Object
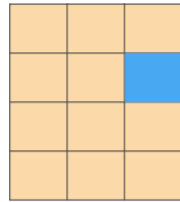
Block storage

Change one block (piece of the file)
that contains the character

Object storage

Entire file must be updated

AWS Storage Services

- ## Amazone EC2 Instance Store
  - Temporary block level storage

- ## Amazon Elastic File System(EFS)
  - Network File System (NFS) for instances and applications
  - Shared file system

- ## Amazon S3 Glacier
  - A secure, durable, and extremely low-cost storage service for data archiving and long-term backup

Amazon Database Services

- Amazon Relational Database Service (Amazon RDS)
  - Managed database service
    - Set up, operate, and scale a relational database in the cloud
  - Database instance
  - The basic building block
  - An isolated database environment
  - Supports: MySQL, Oracle, Microsoft SQL Server, PostgreSQL, . . .

- Amazon Aurora
  - A MySQL and PostgreSQL compatible relational database engine

Amazon Database Services

- Amazon DynamoDB
  - A a key-value and document database

- Amazon Neptune
  - fully-managed graph database service

- Amazon Timestream
  - fully managed time series database service for IoT and operational applications

- Amazon DocumentDB (with MongoDB compatibility)
  - fully managed document database service that supports MongoDB workloads

Networking and Security Services

- Private Virtual Cloud (VPC)
  - Provide a logical isolated section of the AWS cloud where AWS resources can be launched in a private network
  - Free service

- Identity and Access Management (IAM)
  - Controls user's access to AWS resources and services
  - Free service

Amazon VPC

- Private Virtual Cloud (VPC)
    - Provide a logical isolated section of the AWS cloud where AWS resources can be launched in a private network
    - Free service
    - Provides
        - IP address selection, creation of subnets, routing tables and network gateways configuration, ...
    - Belongs to a single region and spans multiple availability zones
    - Contains one or more subnets

Amazon VPC
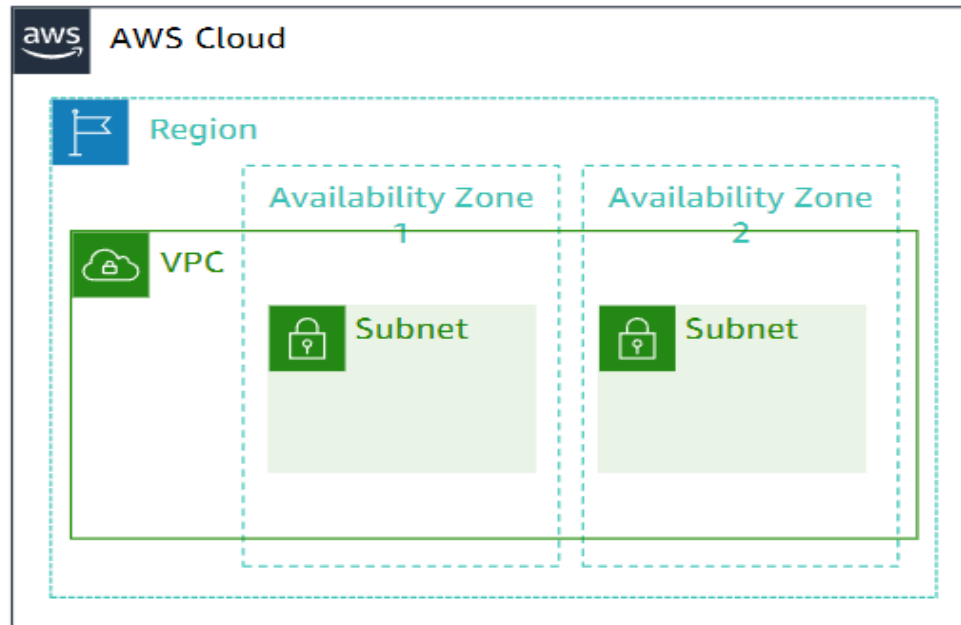
Amazon VPC

- Subnet
  - Has Classless Inter-domain routing (CIDR) block
  - Belongs to one AZ
  - Can be private or public

Amazon VPC

- Elastic network interface
  - A virtual NIC that can be attached to an instance in VPC
  - Has private IP address
- Internet gateway
  - Scalable highly available
  - Allows instances to communicate with each other and the Internet
- NAT gateway
  - Allows private subnets to communicate with the Internet
  - Resides in a public subnet

Amazon VPC

Public IP address types

Public IPv4 address

- Manually assigned through an Elastic IP address
- Automatically assigned through the auto-assign public IP address settings at the subnet level

Elastic IP address

- Associated with an AWS account
- Can be allocated and remapped anytime
- Additional costs might apply

Amazon VPC

- Amazon Route 53
  - Highly available and scalable DNS service
  - Supports different routing algorithms

Identity and Access Management (IAM)

- Controls access to AWS services and resources
- Authentication and authorization
- Components:
  - User
    - Person or application defined in AWS account
    - Unique
  - Group
    - Collection of users
  - Policy
    - A document that defines permissions that grant/deny access to AWS resource to users
  - Role
    - Grant temporary access to specific AWS resource

Identity and Access Management (IAM)

- Authentication
  - Proof of Identity
  - Two access types
    - Management console access
      - Account Id
      - Username/password
    - Programmatic access
      - Access key/secret key

Identity and Access Management (IAM)

- Policy
  - Written in JSON format
  - Principle of least privilege
  - Identity based vs Resource based
  - Managed vs inline

Identity and Access Management (IAM)

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "FullAccess",
            "Effect": "Allow",
            "Action": ["s3:*"],
            "Resource": ["*"]
        },
        {

            "Sid": "DenyCustomerBucket",
            "Action": ["s3:*"],
            "Effect": "Deny",
            "Resource": ["arn:aws:s3:::customer", "arn:aws:s3:::customer/*" ]
        }
    ]
}
```

https://docs.aws.amazon.com/IAM/latest/UserGuide/access_policies_policy-summary-examples.html

Security

- The AWS Cloud enables a **shared responsibility** model
  - AWS manages security **of** the cloud
    - Virtualization layer , infrastructure layer, . . .
  - Customer is responsible for security **in** the cloud
    - Data security, network configuration , application software, . . .
- Security services:
  - AWS Shield
    - DDoS protection
  - ASW Trusted adviser
    - Analyze and give guidance
  - AWS Key Management Service (KMS)

References

- https://aws.amazon.com/whitepapers/
  - https://d1.awsstatic.com/whitepapers/aws-overview.pdf