AWS Introduction

1. AWS Introduction

Title Slide Information

- Topic: AWS Introduction

Key Point: This introduction sets the stage for learning about AWS, a leading cloud service provider. It highlights that the course or lecture will cover the fundamental concepts, service categories, infrastructure, and shared responsibility model of AWS.

---

2. Amazon Web Services Overview

What is AWS?

- **Leading Cloud Platform**: AWS offers over 200 fully featured services. It has one of the largest market shares in the cloud computing industry.
- **High Availability**: AWS achieves reliability and redundancy by hosting its services in multiple **regions** and **availability zones** around the world. This global footprint ensures minimal downtime and fast response times.
- **Scalability on Demand**: AWS can rapidly scale resources (e.g., compute power, storage, databases) up or down to meet workload demands.
- **Pay-As-You-Go Model**: Instead of paying for and maintaining on-premises hardware, you pay only for what you use on AWS. This cost model often reduces capital expenditure and aligns expenses with usage.

Why is AWS so Popular?

- Rapid deployment of new services
- Broad set of tools for developers and businesses
- Large community and ecosystem support
- Continuous innovation in emerging areas such as AI/ML, IoT, and serverless computing

---

3. History of AWS

- **Initial Launch (2006)**: AWS started with three core services:
  - **Amazon S3** (Simple Storage Service)
  - **Amazon EC2** (Elastic Compute Cloud)
  - **Amazon SQS** (Simple Queue Service)
- **Rapid Growth**: From these initial offerings, AWS quickly expanded to over 200 services covering compute, storage, databases, analytics, machine

learning, and more.

- **Market Share**:
  - AWS: ~33%
  - Microsoft Azure: ~22%
  - Google Cloud Platform (GCP): ~9%
  - Other providers share the remaining portion of the market.
- **Global Reach**: AWS operates large-scale data centers worldwide, generating billions of dollars in revenue and serving a wide variety of customers, from startups to large enterprises.

Key Takeaway: AWS began with a small set of foundational cloud services and evolved into a massive ecosystem that dominates the cloud market.

---

4. AWS Service Categories

AWS groups its many offerings into categories, each addressing a different layer of infrastructure, application development, or data management. Common categories include:

1. **Compute** – Services that provide virtual servers, containers, and serverless functions (e.g., Amazon EC2, AWS Lambda, AWS Fargate).
2. **Storage** – Services to store and manage data (e.g., Amazon S3, Amazon EBS, Amazon EFS).
3. **Databases** – Managed database offerings (e.g., Amazon RDS, Amazon DynamoDB, Amazon Redshift).
4. **Networking & Content Delivery** – Services that help configure virtual networks and deliver content globally (e.g., Amazon VPC, Amazon CloudFront).
5. **Analytics** – Tools for data processing, data warehousing, and real-time analytics (e.g., Amazon EMR, AWS Glue, Amazon Kinesis).
6. **Machine Learning & AI** – Services providing ML platforms and AI-driven APIs (e.g., Amazon SageMaker, Amazon Rekognition).
7. **Security & Identity** – Services to secure AWS environments and manage identities (e.g., AWS IAM, AWS KMS).
8. **Management & Governance** – Tools for monitoring, auditing, and organizing AWS resources (e.g., Amazon CloudWatch, AWS CloudFormation).
9. **Developer Tools** – Services that streamline software development and CI/CD (e.g., AWS CodeCommit, AWS CodeBuild, AWS CodeDeploy).

---

5. Cloud Models

When discussing cloud computing, it is important to distinguish between different service models. Each model shifts certain responsibilities between the cloud provider and the customer:

1. **Infrastructure as a Service (IaaS)**
    - AWS provides fundamental building blocks (compute, storage, networking).
    - Customers manage operating systems, software, and configurations.
    - Example: Amazon EC2 (virtual servers on which you install and manage your own OS and applications).
2. **Platform as a Service (PaaS)**
    - AWS provides an environment for building, testing, and deploying applications.
    - Underlying infrastructure (servers, OS, runtime) is abstracted away.
    - Example: AWS Elastic Beanstalk (you focus on the code; AWS handles deployment and capacity).
3. **Software as a Service (SaaS)**
    - Complete applications provided by AWS or third parties, managed entirely by the service provider.
    - Users simply use the software; there is no need to manage or configure infrastructure or platform.
    - Example: AWS-provided applications or third-party software offered in the AWS Marketplace.

Key Idea: Each model offers different levels of control and responsibility for the user. As you move from IaaS to PaaS to SaaS, you manage fewer details but also have less flexibility in customizing the environment.

———

6. The Shared Responsibility Model (AWS Responsibilities)

AWS uses a **shared responsibility model** for security and compliance:

- **AWS Responsibilities (Security of the cloud)**
    - Protecting the infrastructure that runs AWS services (hardware, software, networking, facilities).
    - Maintaining physical security of data centers.
    - Managing the host operating system and virtualization layer for AWS compute services.
    - Patching and fixing flaws within the AWS global infrastructure.
    - Ensuring redundancy and availability of core systems.

In other words, AWS takes care of the underlying cloud environment so that customers can focus on their applications and data.

---

7. The Shared Responsibility Model (Client Responsibilities)

- **Client Responsibilities (Security in the cloud)**
  - **Configuration of Services**: Correctly configuring security groups, access policies, encryption settings, and network parameters.
  - **Managing Data**: Ensuring that data is classified properly, encrypted if needed, and backed up according to the organization's requirements.
  - **Operating System and Application Maintenance**: For services like EC2, the customer must patch and maintain the operating system and installed software.
  - **Identity and Access Management**: Defining IAM policies and best practices, such as using least privilege, rotating keys, and enabling multi-factor authentication (MFA).
  - **Compliance**: Ensuring that internal policies, industry regulations, and legal requirements are met at the application and data level.

Key Takeaway: While AWS secures the underlying infrastructure, the customer must secure and manage everything they build on top of it.

---

8. The AWS Global Infrastructure

- **Regions**: Distinct geographical areas around the world (e.g., US East (N. Virginia), EU (Ireland), Asia Pacific (Sydney)). Each region contains multiple Availability Zones.
- **Availability Zones (AZs)**: Data centers that are physically isolated from each other within a region. AZs provide redundancy and fault tolerance; if one AZ fails, others remain available.
- **Edge Locations**: Content delivery network (CDN) endpoints for services like Amazon CloudFront. By caching content closer to end users, AWS reduces latency and speeds up data delivery.

Importance of Choosing the Right Region:

- **Latency**: Deploy services in regions closer to your users to reduce response times.
- **Compliance**: Some regions have specific data sovereignty or regulatory requirements.

- **Cost**: Pricing can differ slightly by region, so choosing the most cost-effective region may reduce expenses.

---

9. Global Infrastructure Map

A typical AWS global infrastructure map shows:

- Number of launched regions worldwide
- Number of Availability Zones across those regions
- Edge locations for services like CloudFront

Key Point: AWS is continually expanding into new regions and adding new availability zones to increase resilience and global reach.

---

10. Compute Services

AWS offers several ways to run workloads, each suited to different needs:

1. **VM-based**:
   - **Amazon EC2 (Elastic Compute Cloud)**: Provides resizable virtual machines.
   - **Amazon EC2 Spot**: Use spare compute capacity at discounted rates, though instances can be interrupted if capacity is needed elsewhere.
2. **Container-based**:
   - **Amazon ECS (Elastic Container Service)**: Orchestrates Docker containers across a cluster of EC2 instances.
   - **Amazon EKS (Elastic Kubernetes Service)**: Managed Kubernetes for container orchestration.
3. **Serverless**:
   - **AWS Lambda**: Runs code without provisioning or managing servers. You pay only for the compute time you consume.
   - Ideal for event-driven applications (e.g., responding to S3 uploads, API Gateway requests).

---

11. Storage Services

1. **Amazon S3 (Simple Storage Service)**
   - Object storage for the internet.
   - Highly scalable, durable (designed for 99.999999999% durability), and secure.

- Ideal for static assets, backups, and big data.

2. **Amazon EBS (Elastic Block Store)**
   - Block-level storage volumes for use with Amazon EC2 instances.
   - Persistent storage that remains even if the EC2 instance is stopped or terminated.
3. **Amazon EFS (Elastic File System)**
   - Fully managed file storage for use with AWS compute services.
   - Scales automatically, provides a file system interface.
4. **Amazon S3 Glacier**
   - Low-cost storage for data archiving and long-term backup.
   - Retrieval times vary (minutes to hours), optimized for infrequent access.
5. **AWS Storage Gateway**
   - Hybrid storage service enabling on-premises applications to use AWS cloud storage seamlessly.

---

12. Database Services

1. **Relational Databases**:
   - **Amazon RDS (Relational Database Service)** supports MySQL, PostgreSQL, MariaDB, Oracle, Microsoft SQL Server, and Amazon Aurora.
   - Automated backups, patching, and scaling.
2. **NoSQL Databases**:
   - **Amazon DynamoDB**: Key-value and document database, fully managed, highly scalable, low latency.
3. **Document Databases**:
   - **Amazon DocumentDB**: Compatible with MongoDB, for JSON-based workloads.
4. **Graph Databases**:
   - **Amazon Neptune**: Fully managed graph database service for highly connected data (e.g., social networks, knowledge graphs).

---

13. Analytics Services

1. **Amazon Athena**
   - Interactive query service that analyzes data in Amazon S3 using standard SQL.
   - Serverless; you pay only for the queries you run.
2. **AWS Glue**
   - ETL (Extract, Transform, Load) service that prepares data for

analytics.
    - Automatically discovers schema and creates metadata in a data catalog.
3. **Amazon EMR (Elastic MapReduce)**
    - Managed Hadoop framework for processing large amounts of data at scale.
    - Integrates with big data frameworks like Spark and Hive.
4. **Amazon Kinesis**
    - Real-time data ingestion and streaming service.
    - Collect, process, and analyze real-time streaming data (e.g., logs, IoT data).
5. **Amazon QuickSight**
    - Scalable business intelligence (BI) service.
    - Allows creation of interactive dashboards and visualizations.

---

14. Machine Learning (ML) and AI Services

1. **Amazon SageMaker**
    - Fully managed ML platform for building, training, and deploying machine learning models.
    - Provides hosted Jupyter notebooks, built-in algorithms, and automatic model tuning.
2. **AWS AI Services (Pre-trained Models)**
    - **Amazon Rekognition**: Image and video analysis (object detection, facial recognition, etc.).
    - **Amazon Transcribe**: Automated speech-to-text.
    - **Amazon Translate**: Neural machine translation for many languages.
    - **Amazon Comprehend**: Natural language processing for text analysis (sentiment, key phrases).

    These AI services enable developers to add advanced capabilities to applications without deep ML expertise.

---

15. Important Services for Data Analytics/Engineering

For anyone focusing on data analytics or data engineering, the following services are especially relevant:

- **Amazon EC2** and **AWS Lambda** (compute for data processing)
- **Amazon S3** (data lake, object storage)
- **Amazon RDS** and **Amazon DynamoDB** (structured and NoSQL

databases)

- **AWS Glue** (ETL and data catalog)
- **Amazon Athena** (serverless SQL queries on S3)
- **Amazon EMR** (big data processing with Hadoop/Spark)
- **Amazon Redshift** (data warehousing and analytics)

---

16. AWS Free Tier

AWS offers a **Free Tier** that provides limited access to certain services for 12 months. This allows users to experiment and learn without incurring large costs:

- **Amazon EC2**: 750 hours/month of t2.micro or t3.micro instances (Linux/Windows)
- **Amazon S3**: 5 GB of standard storage (20k GET requests, 2k PUT requests)
- **Amazon RDS**: 750 hours/month of certain database engines (with size limits)
- **Others**: Many services have free or trial components under certain usage thresholds

Key Takeaway: The Free Tier is a great way to gain hands-on experience with AWS, learn about its services, and prototype applications without worrying about high bills.

Conclusion

Amazon Web Services is a comprehensive cloud platform that covers a vast range of services from compute and storage to analytics, machine learning, and more. Understanding the shared responsibility model, the global infrastructure, and the different service models (IaaS, PaaS, SaaS) is crucial to effectively leveraging AWS in a secure and cost-efficient manner.

Whether you are building simple websites, data analytics pipelines, or enterprise-scale applications, AWS offers tools and services that can help streamline development, reduce overhead, and enable innovative solutions. By taking advantage of AWS's Free Tier and carefully designing architectures to use the right services, businesses and developers can harness the power of the cloud while optimizing both performance and cost.