

Amazon Web Service

Gokul G -
22z221

Hariharan D - **22z222**

Harish K -
22z223

Pranavii K -

INTRODUCTION TO AWS

Hariharan D (22z222)



A meme featuring Woody and Buzz Lightyear from the Toy Story franchise. Woody is on the left, looking concerned. Buzz is on the right, wearing his green and purple space suit and holding a purple lollipop. The background is a simple room with a door and a window.

AWS

AWS EVERYWHERE

Amazon Web Services

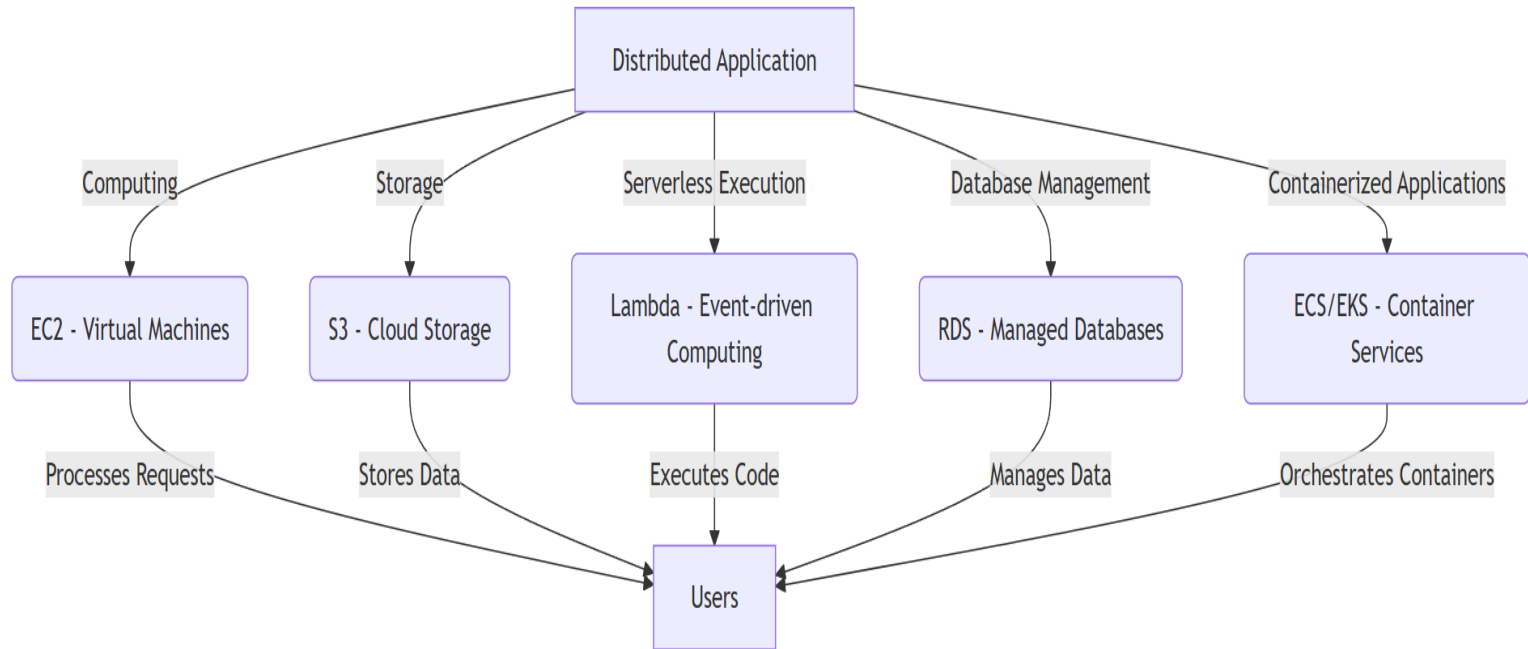
Amazon Web Services (AWS) is a cloud computing platform that provides on-demand computing resources. Instead of buying and maintaining physical servers, businesses can use AWS to store data, run applications, and scale operations efficiently. AWS offers services like computing power, storage, databases, and machine learning tools, enabling



Distributed Computing

Distributed computing is a model where multiple computers work together to process tasks efficiently. Instead of relying on a single machine, workloads are split across multiple systems, improving speed, fault tolerance, and scalability. It is essential for handling large-scale applications like social media platforms, online shopping, and financial systems.

Key AWS Services That Support Distributed Computing



AWS Storage and Database Solutions

**Harish K -
22z223**



Amazon **RDS**

Introduction to AWS Storage and Databases

Why storage and databases matter in distributed computing?

- Scalability, availability, and performance
- Ensuring data persistence and redundancy

AWS Storage Solutions Categories:

1. **Object Storage** – Amazon S3
2. **Block Storage** – Amazon EBS
3. **Relational & NoSQL Databases** – AWS RDS, DynamoDB
4. **Data Warehousing** – Amazon Redshift

Amazon S3 – Distributed Object Storage

Object storage is a data storage architecture where files (data) are stored as objects instead of traditional file system hierarchy (folders).

Overview:

- Highly scalable, durable, and secure object storage
- Ideal for backup, data lakes, and big data analytics

Key Features:

- **Unlimited storage:** Objects up to 5TB in size
- **Storage Classes:** Standard, Intelligent-Tiering, Glacier (for archival)
- **Data replication** across multiple AWS regions for durability (99.999999999% durability)

Use Cases:

- Amazon S3 allows **hosting static websites** without needing a web server. It supports **HTML, CSS, JavaScript, images, and videos**, making it ideal for lightweight, low-cost web hosting.
- Amazon S3 provides **secure and durable storage** for backup and disaster recovery, ensuring that data is **protected against loss or failure**.
- Amazon S3 is widely used for **big data processing, data lakes, and machine learning** because of its ability to store **large-scale structured and unstructured data**.

Amazon EBS (Elastic Block Store) – Persistent Storage

Block storage is a type of data storage where data is divided into fixed-sized **blocks** and stored separately with unique addresses.

Overview:

- Block storage for EC2 instances
- Provides low-latency, high-performance storage

Key Features:

- **Persistent storage:** Data remains even if EC2 instances are stopped
- **Scalability:** Resize storage without downtime
- **Snapshot backups** for disaster recovery
- **Encryption** for secure data storage



Use Cases:

- **MySQL, PostgreSQL, Oracle, and Microsoft SQL Server** require low-latency, high-throughput storage. Amazon EBS provides persistent storage that maintains database integrity.
- Amazon EBS **delivers low-latency, high-throughput storage** for high-performance computing applications.
- Amazon EBS provides a **scalable, high-performance solution** for log storage and real-time data processing

AWS RDS (Relational Database Service) & DynamoDB

AWS RDS (Relational Database Service)

- Fully managed relational database
- Supports **MySQL, PostgreSQL, MariaDB, SQL Server, and Amazon Aurora**
- Automated backups, failover, and patching
- Read replicas for load balancing

Amazon DynamoDB – NoSQL Database

- Fully managed, key-value, and document-based NoSQL database
- **Single-digit millisecond latency**
- **Auto-scaling** for large workloads
- **Serverless** – no need to provision infrastructure



Use Cases:

- Most dynamic web applications need a **reliable relational database** to store structured data **Amazon RDS supports MySQL, PostgreSQL, MariaDB, SQL Server, and Oracle**, making it ideal for web applications.
- DynamoDB is designed for **real-time, high-speed applications** that require **single-digit millisecond latency**

Amazon Redshift – Data Warehousing

Overview:

- Cloud-based, fully managed **data warehouse**
- Used for **big data analytics** and business intelligence

Key Features:

- **Columnar storage** for fast query performance
- **Massively parallel processing (MPP)** for handling large datasets
- **Integration with BI tools** (Tableau, Power BI)
- **Data lake integration** with Amazon S3

Use Cases:

- Amazon Redshift enables organizations to **store, analyze, and visualize** this data for strategic decision-making in business intelligence and reporting
- Redshift can **ingest and process streaming data** to generate real-time business intelligence.
- Amazon Redshift is widely used for **fraud detection and customer segmentation** by analyzing large volumes of transactional and behavioral data

| Scenario | Recommended AWS Database | Reason |
|--|-------------------------------------|---|
| E-commerce website managing orders and inventory | Amazon RDS (MySQL/PostgreSQL) | Ensures structured, transactional data with ACID compliance for order management and stock updates . |
| Social media platform storing likes, comments, and user interactions | Amazon DynamoDB | Handles high-speed, low-latency NoSQL data for real-time interactions and massive scalability. |
| Business analyzing customer behavior for marketing campaigns | Amazon Redshift | Designed for large-scale data analytics and business intelligence (BI) . Processes petabyte-scale data for insights. |
| Media streaming service storing videos, images, and documents | Amazon S3 | Provides highly durable, scalable, and cost-effective object storage for media content. |
| IoT-based smart home system storing sensor data | Amazon DynamoDB + Amazon Timestream | DynamoDB for real-time NoSQL data storage, Timestream for time-series analytics of IoT data. |

AWS Security, Monitoring, and Performance Optimization

**Pranavji K -
22z245**



AWS Identity and Access Management (IAM)

- AWS IAM is a service that securely manages access to AWS resources.
- Enables authentication and authorization for users, groups, and roles.

Key Components

- **Users** – Individual AWS identities with specific credentials.
- **Groups** – Collection of users with shared permissions.
- **Roles** – Temporary permissions for users, services, or applications.
- **Policies** – JSON-based documents defining permissions.

Security Features

- **Least Privilege Principle** – Grant only necessary permissions.
- **Multi-Factor Authentication (MFA)** – Adds extra security layer.
- **Access Control Policies** – Define what users and roles can do.



AWS Identity and Access Management (IAM)

Use Cases

- Secure access to AWS resources.
- Assign roles to applications without sharing credentials.
- Helps control permissions efficiently while ensuring security.



AWS CloudWatch & AWS CloudTrail for Monitoring

AWS provides monitoring and logging services for tracking system performance and security.

- **Amazon CloudWatch** – Monitors AWS resources and applications.
- **AWS CloudTrail** – Tracks API activity and user actions.



Amazon CloudWatch



AWS CloudTrail

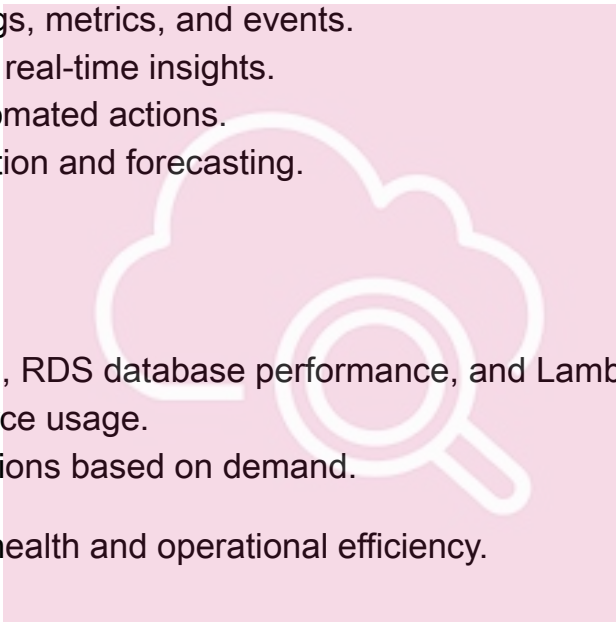
AWS CloudWatch (Performance Monitoring)

- Collects and monitors logs, metrics, and events.
- Provides dashboards for real-time insights.
- Enables alarms and automated actions.
- Supports anomaly detection and forecasting.

Use Cases:

- ✓ Monitor EC2 CPU utilization, RDS database performance, and Lambda execution time.
- ✓ Set up alerts for high resource usage.
- ✓ Automatically scale applications based on demand.

CloudWatch ensures system health and operational efficiency.



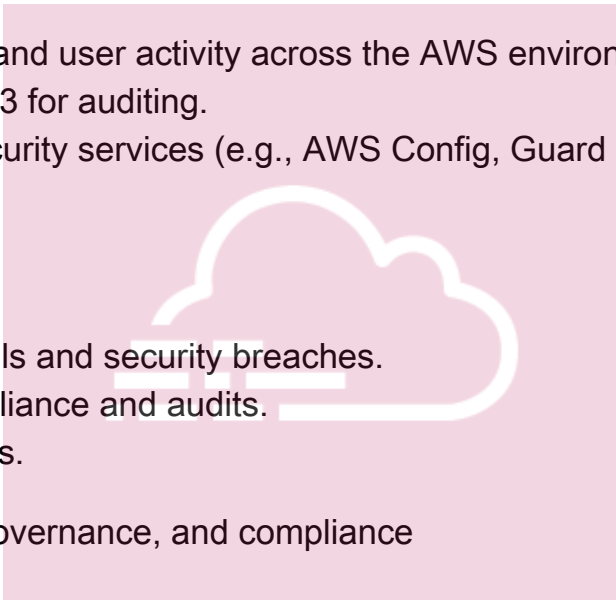
AWS CloudTrail (Security & Compliance Monitoring)

- Records AWS API calls and user activity across the AWS environment.
- Stores logs in Amazon S3 for auditing.
- Integrates with AWS Security services (e.g., AWS Config, Guard Duty).

Use Cases:

- ✓ Detect unauthorized API calls and security breaches.
- ✓ Track user actions for compliance and audits.
- ✓ Investigate security incidents.

CloudTrail ensures security, governance, and compliance



CloudWatch vs CloudTrail

| Feature | CloudWatch | CloudTrail |
|------------|---------------------------------------|-------------------------------------|
| Purpose | Performance & resource monitoring | Security & compliance monitoring |
| Tracks | Metrics, logs, and events | API calls and user activities |
| Real-time? | Yes | No (logs recorded after actions) |
| Use Case | System health, alarms, and automation | Audit logs, security investigations |

AWS Auto Scaling & Cost Optimization

AWS Auto Scaling (Elasticity & Performance)

- **Dynamically adjusts** compute capacity to maintain performance.
- Works with EC2, ECS, DynamoDB, Aurora, etc.
- Uses policies based on metrics (CPU usage, requests per second).
- Ensures **high availability** and **efficient resource utilization**.

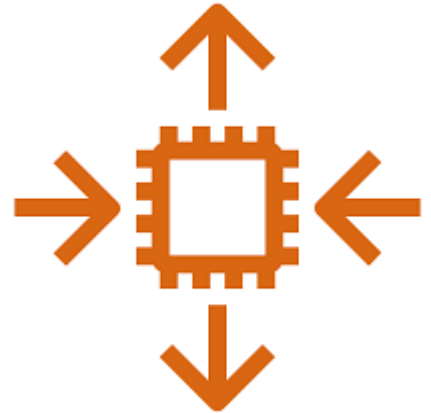
Benefits:

- ✓ Reduces downtime and improves fault tolerance.
- ✓ Enhances application scalability based on real-time demand.

Example: EC2 Auto Scaling

Cost Optimization Framework:

Identify **unused resources** → Select **right pricing models** → Use **Auto Scaling** → Monitor & Optimize.



AWS Use Cases and Real-World Applications in Distributed Computing

Goku G - 22z221



AWS Use Cases and Real-World Applications in DC

Big Data Processing with AWS Services (EMR, Glue)

- **Amazon EMR** : Enables processing of vast datasets using frameworks like Apache Spark and Hadoop.
- **AWS Glue**: A fully managed ETL (Extract, Transform, Load) service for preparing and integrating data.
- **Amazon Redshift**: A petabyte-scale data warehouse for fast analytics.
- **AWS Lake Formation**: Facilitates secure data lake creation and management.

Real-World Applications

- **Netflix**: Uses Amazon EMR and AWS Glue to process billions of events daily for content recommendation.
- **Airbnb**: Leverages AWS Glue and Redshift for scalable data warehousing and analytics.
- **Uber**: Processes trip data using Spark on EMR for route optimization and demand forecasting.

AWS Use Cases and Real-World Applications in DC

Microservices and Containerization with AWS ECS and EKS:

- **Amazon ECS** : A fully managed container orchestration service for Docker applications.
- **Amazon EKS** : A managed Kubernetes service for deploying and scaling containerized applications.
- **AWS Fargate**: A serverless compute engine for running containers without managing infrastructure.
- **AWS App Mesh**: Facilitates communication between microservices.

Real-World Applications:

- **Spotify**: Uses AWS ECS to manage its backend microservices architecture.
- **Expedia**: Leverages EKS to deploy scalable and resilient travel services.
- **Capital One**: Migrated to microservices using AWS to enhance application agility and security.



AWS Use Cases and Real-World Applications in DC

Case Studies of Companies Leveraging AWS for DC:

1. **NASA:** Uses AWS to process satellite images, reducing analysis time from weeks to hours.
2. **Lyft:** Processes real-time ride requests and traffic data using AWS Lambda and Kinesis.
3. **Coca-Cola:** Migrated to AWS to optimize supply chain operations using AI-driven analytics.
4. **Slack:** Uses AWS ECS and EKS to manage its collaboration platform at scale.
5. **General Electric (GE):** Analyzes industrial sensor data using AWS IoT and analytics services.