AWS Flash Notes, Tips and Tricks

A sample document

By: Sujal Thapa

LinkedIn: https://www.linkedin.com/in/sujal-thapa-29386227b/

Instagram: @sujal_thapa__

Cloud Computing

Variations of this Simpsons meme are extremely popular among Cloud engineers in Reddit, Discord and other online spaces.

The development, testing, and deployment of digital products require extensive hardware and software resources. Cloud computing provides a solution by allowing individuals and organizations to lease IT infrastructure, platforms, and software from third-party providers over the internet. This model often utilizes pay-as-you-go pricing, enabling flexible and cost-effective access to computing resources.

Cloud computing services are typically categorized into three main types:

1. **Infrastructure as a Service (IaaS):** Provides virtualized computing resources for the operation of digital applications/products, such as virtual machines, storage, and networks.
2. **Platform as a Service (PaaS):** Offers hardware and software tools, typically used for application development and/or testing.
3. **Software as a Service (SaaS):** Delivers software applications, often on a subscription basis.

AWS

Amazon Web Services is, put it simply, the largest and most widely adopted cloud provider in the world, used by everyone from the smallest of startups and hobbyists to well, the United States government and 90% of Fortune 500 companies.

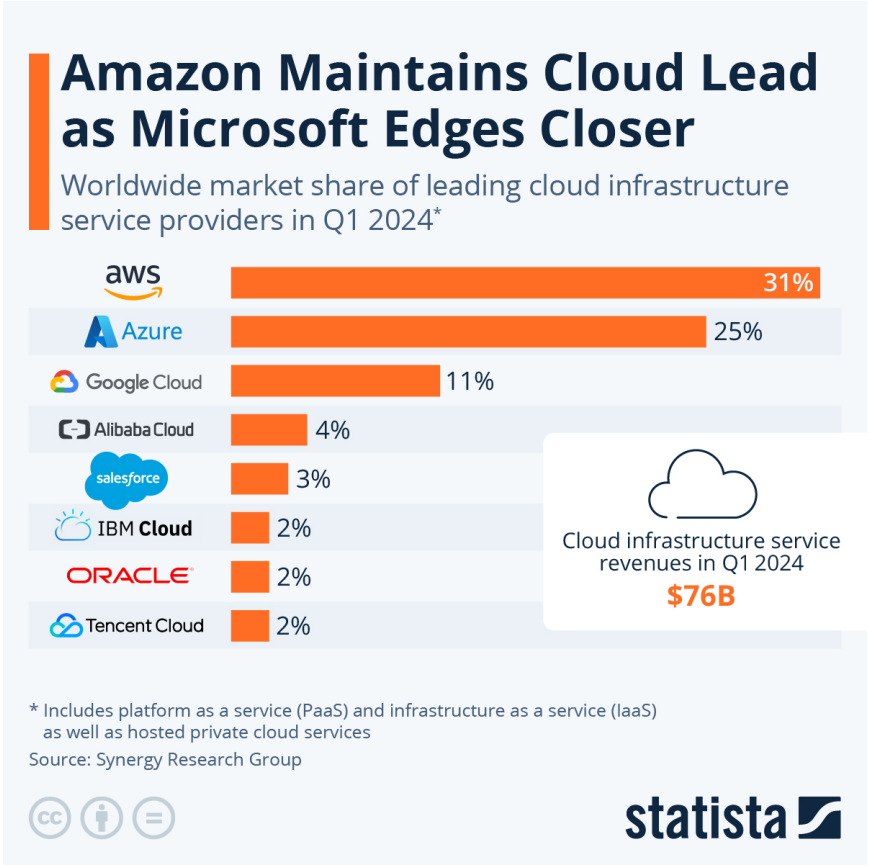
It is a behemoth of a platform with over 200 services (most of which are fully-featured and quite comprehensive in and of themselves) and global geographical presence, with infrastructure in all the seven continents serving 245 countries and territories.

Safe to say, it is a pretty big deal and diving into the AWS ecosystem, especially if you are not from an IT background may be a psychologically daunting task. But do not worry, for I am here to help.

After all, it is precisely to make that daunting task easy that I prepared these resources.

The Big Three

Though there are actually many cloud providers, the cloud computing market is mostly dominated by three large players, namely: Amazon Web Services (AWS), Microsoft Azure and Google Cloud Platform (GCP) with AWS having a little more than 30% of the market all to itself at the time of writing.



Autoscaling Group and Spanning

Auto Scaling Groups (ASGs) in AWS provide dynamic scaling and management for EC2 instances, ensuring applications maintain performance and availability. ASGs automatically adjust the number of instances based on demand, scaling out to handle increased load and scaling in during low demand, optimizing cost and efficiency.

Key features of ASGs include health checks to automatically replace unhealthy instances, integration with Elastic Load Balancing (ELB) to distribute traffic evenly, and scheduled scaling to anticipate load changes. ASGs also support scaling policies driven by CloudWatch alarms, allowing fine-tuned responsiveness to real-time metrics. This automation ensures applications remain resilient, scalable, and cost-effective.

Important note for the exam:

ASG can span multiple Availability Zones.

Autoscaling tip

When a scenario requires say, “2 as the minimum capacity”, it is always best to construct an architecture that gives more than 2 instances (i.e. at least 3 instances) at all times because if an AZ outage happened, ASG will launch a new instance on the unaffected AZ.

This provisioning does not happen instantly, which means that for a certain period of time, there will only be 1 running instance left.

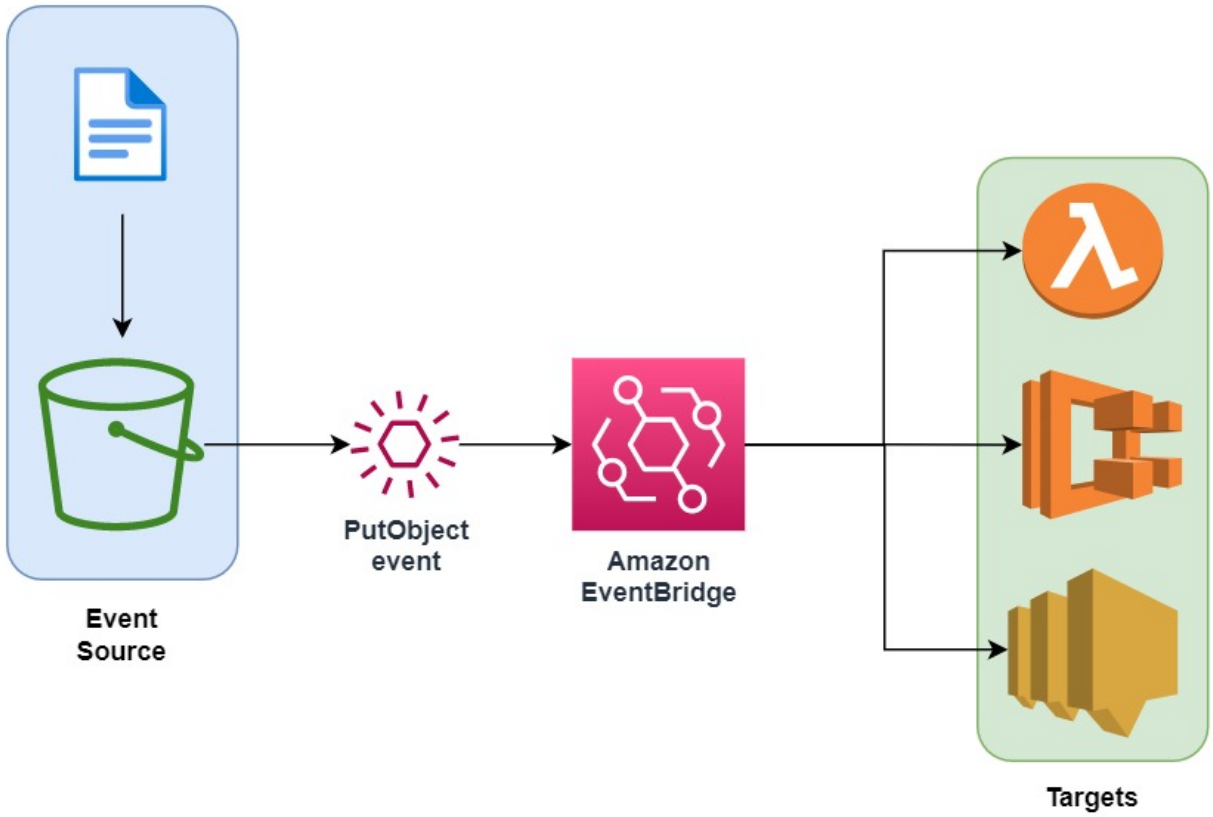
For example:

If a company needs to deploy at least 2 EC2 instances to support the normal workloads of its application and automatically scale up to 6 EC2 instances to handle the peak load and the architecture is processing mission-critical workloads.

Then, as the Cloud Architect of the company, create an architecture that has at least 3 instances up and running at all times.

Amazon EventBridge

Amazon EventBridge (Amazon CloudWatch Events) is a serverless event bus that makes it easy to connect applications together. It uses data from your own applications, integrated software as a service (SaaS) applications, and AWS services. This simplifies the process of building event-driven architectures by decoupling event producers from event consumers. This allows producers and consumers to be scaled, updated, and deployed independently. Loose coupling improves developer agility in addition to application resiliency.



You can use Amazon EventBridge (Amazon CloudWatch Events) to run Amazon ECS tasks when certain AWS events occur.

You can set up an EventBridge rule that runs an Amazon ECS task whenever a file is uploaded to a certain Amazon S3 bucket using the Amazon S3 PUT operation for example.

Compliance vs Governance Object Lock

Object Lock is a feature of Amazon S3 that helps enforce compliance by preventing deletion or modification of objects for a specified retention period. It ensures that data remains immutable and protected against accidental or malicious deletion, supporting regulatory requirements and data governance policies.

The two types of Object Lock in Amazon S3 are:

1. **Retention Periods**: This mode sets a fixed retention period during which objects cannot be deleted or altered. It ensures data immutability for compliance with regulations like SEC Rule 17a-4(f) and the General Data Protection Regulation (GDPR).
2. **Legal Hold**: This mode allows you to place legal holds on objects, which prevents them from being deleted by any user until the hold is removed. It is typically used for preserving data relevant to legal or investigative proceedings.

Also, object lock has two modes namely:

Compliance Object Lock ⇒ Stuff cannot be changed, even by governing authorities.

Governance Object Lock ⇒ Stuff can be changed by governing authorities.

Legal Hold vs Retention Periods

Legal Hold vs. Retention Period

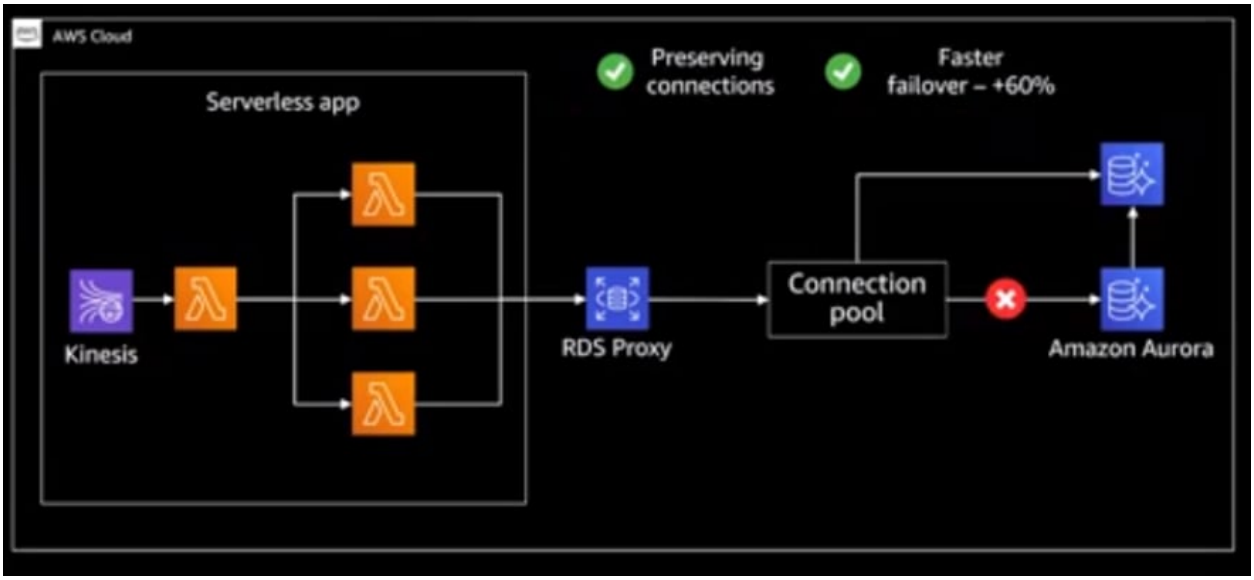
With Object Lock, you can also place a legal hold on an object version. Like a retention period, a legal hold prevents an object version from being overwritten or deleted. However, a legal hold doesn't have an associated retention period and remains in effect until removed. Legal holds can be freely placed and removed by any user who has the **s3:PutObjectLegalHold** permission.

Legal holds are independent from retention periods.

Object lock enforces WORM.

RDS Proxy

AWS RDS Proxy is a fully managed, highly available database proxy service offered by Amazon Web Services. It acts as an intermediary between your application and your Amazon RDS database instances, enhancing performance, scalability, and security.



It also helps when timeout errors because too many customers.

Used to pool database connections together. Helps when too many connections error.

RDS Proxy vs Read Replica actually tbh.

AWS CloudHSM

AWS CloudHSM

- **Purpose**: AWS CloudHSM provides hardware security modules in the cloud, allowing you to generate and use your own encryption keys within a highly secure, dedicated hardware device. It gives you complete control over your key management, including key creation, management, and use for cryptographic operations.
- **Independence from CloudTrail**: While AWS CloudHSM can be integrated with other AWS services, its primary function is to secure cryptographic keys and perform cryptographic operations. Auditing in the context of CloudHSM typically refers to its ability to log and report on cryptographic operations that occur within the HSM itself, which is separate from the kind of API call logging provided by AWS CloudTrail.

IAM Roles

AWS Identity and Access Management (IAM) is a key service that enhances security and access control within the Amazon Web Services ecosystem. It allows administrators to manage users and groups, and set fine-grained permissions to securely control access to AWS resources. IAM ensures that only authorized users can perform specific actions, maintaining robust security and compliance.

Key features of IAM include multi-factor authentication (MFA), integration with identity providers for federated access, and the use of IAM roles for secure cross-account and service-to-service communication. IAM policies enable precise access control, supporting the principle of least privilege to minimize unauthorized access risks and strengthen overall security.

IAM Roles allow you to give permissions to Amazon Services to access other services/perform operations on your behalf.

Ex. Give EC2 instances an IAM Role that grants permission to access S3.

KMS and multi-region keys

AWS Key Management Service (KMS) is a vital service for managing cryptographic keys and controlling their use across a wide range of AWS services and applications. It provides centralized management of encryption keys, enhancing data security both at rest and in transit. By integrating with various AWS services, KMS ensures seamless encryption and decryption processes, facilitating strong security measures without adding operational complexity.

KMS supports several types of keys, including:

- **Customer Managed Keys (CMKs)**: Created and managed by the user, offering full control over permissions and lifecycle.
- **AWS Managed Keys**: Automatically created and managed by AWS services such as S3, RDS, and EBS, providing ease of use without manual key management.
- **AWS Owned Keys**: Managed entirely by AWS and used across multiple AWS accounts, offering a simplified, cost-effective solution for less sensitive data.

Important tip for the exams:

When dealing with info stored in multiple regions, use the multi-region KMS key.