Agenda

1. Storage services
2. Different storage types
3. Storage services in Aws
4. Intro to S3
5. Storage classes
6. lifecycle policies
7. Bucket types
8. Encryption
9. Event notifications
10. S3 Access
11. Hosting a website using Amplify
12. Cloudfront
13. Gitops for S3

Storage Background

The Beginning

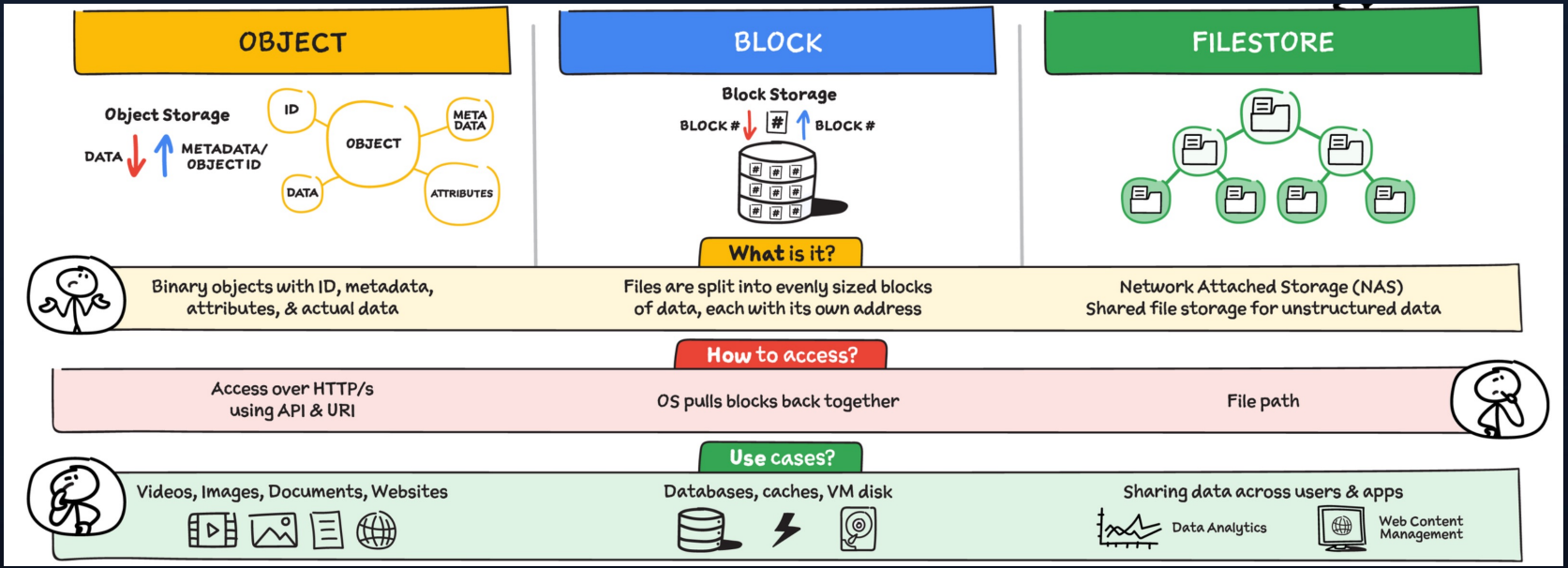
March 14, 2006 - Amazon S3 (Simple Storage Service) became the first AWS service ever launched, revolutionizing how we think about data storage.

What is an Object Storage?

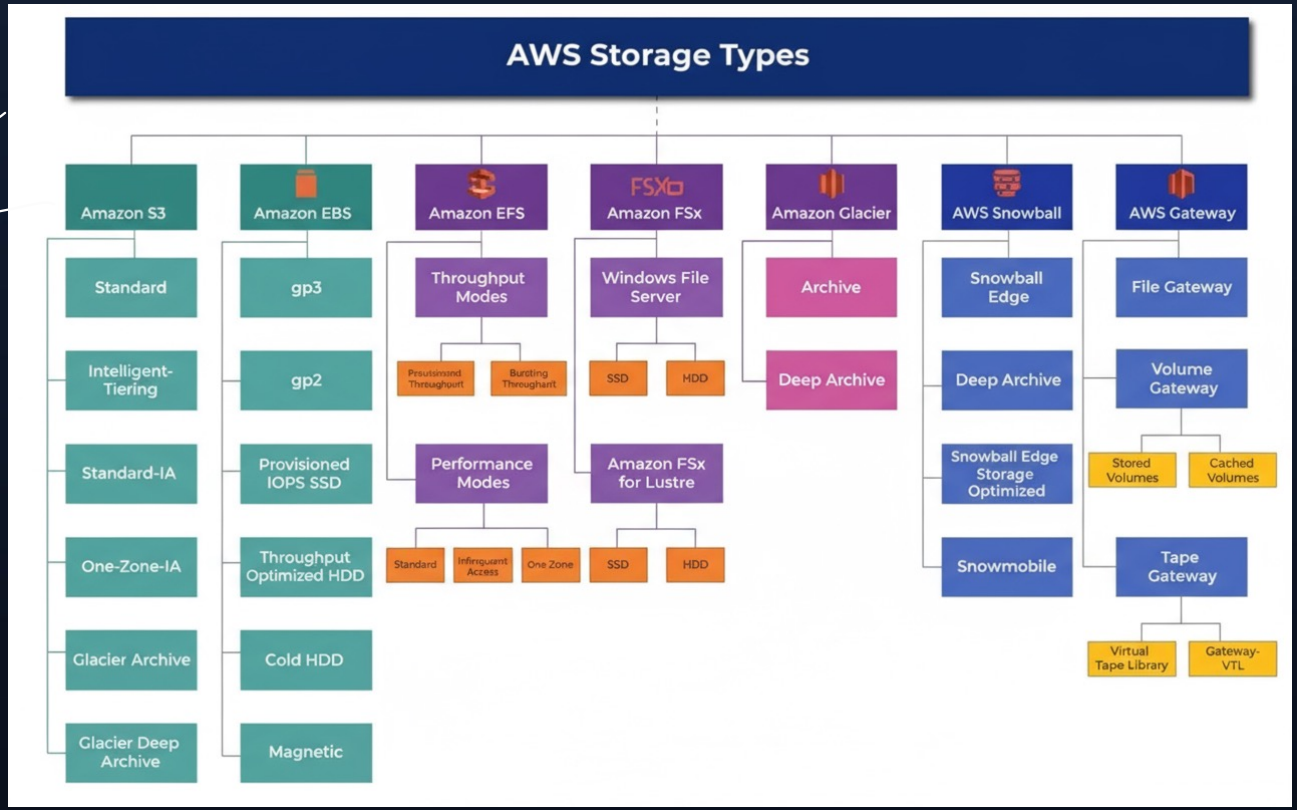
Object storage is a technology that stores and manages data in an unstructured format called objects. Modern organizations create and analyze large volumes of unstructured data such as photos, videos, email, web pages, sensor data, and audio files. Cloud object storage systems distribute this data across multiple physical devices but allow users to access the content efficiently from a single, virtual storage repository. Object storage solutions are ideal for building cloud native applications that require scale and flexibility, and can also be used to import existing data stores for analytics, backup, or archive.

https://aws.amazon.com/s3/

What are different types of Storage



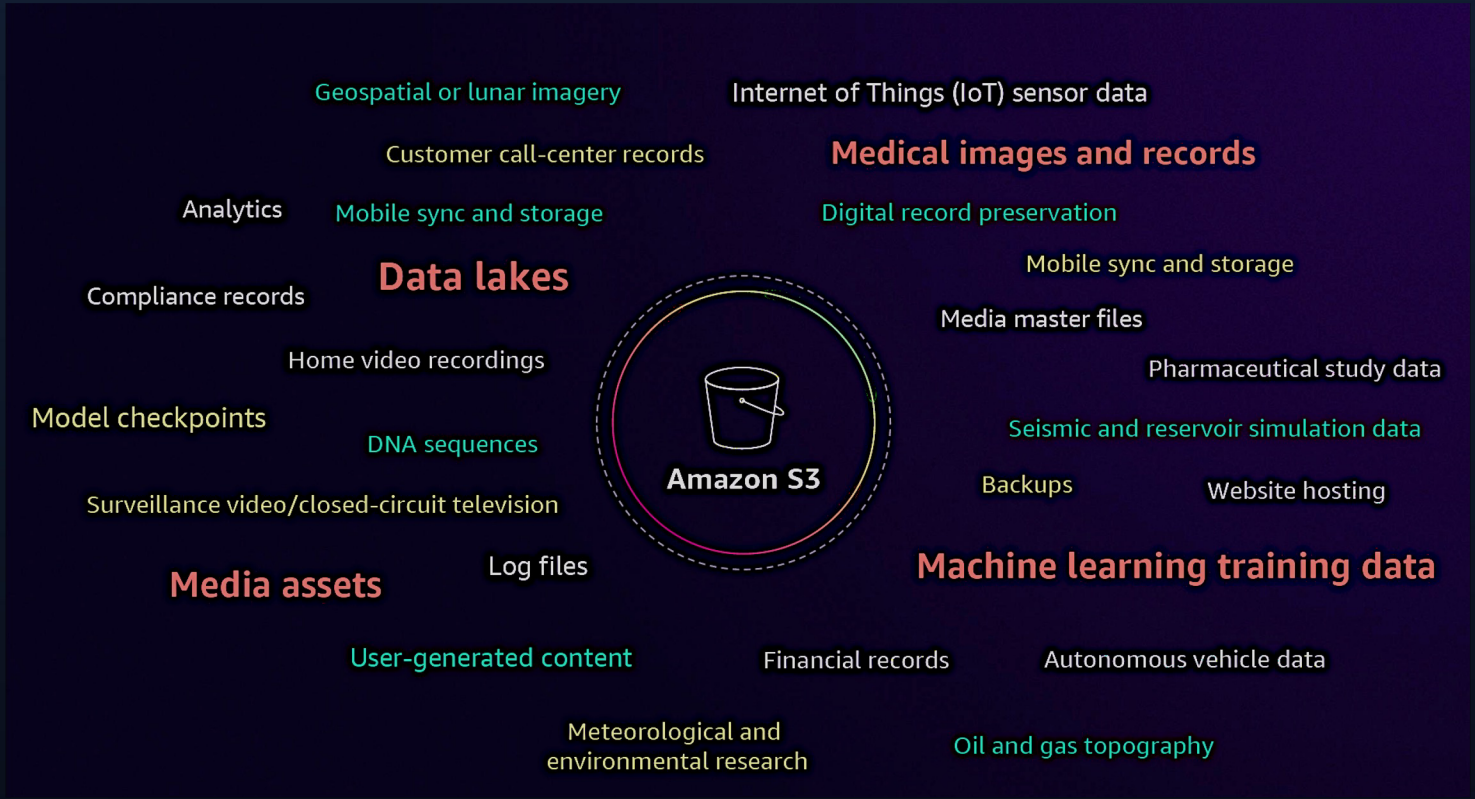
Storage Services in AWS



Introduction to Amazon S3

Amazon Simple Storage Service (Amazon S3) is an object storage service offering industry-leading scalability, data availability, security, and performance. Millions of customers of all sizes and industries store, manage, analyze, and protect any amount of data for virtually any use case, such as data lakes, cloud-native applications, and mobile apps. With cost-effective storage classes and easy-to-use management features, you can optimize costs, organize and analyze data, and configure fine-tuned access controls to meet specific business and compliance requirements.

S3 Use Cases


Geospatial or lunar imagery
Internet of Things (IoT) sensor data
Customer call-center records
Medical images and records
Analytics
Mobile sync and storage
Digital record preservation
Data lakes
Mobile sync and storage
Compliance records
Media master files
Home video recordings
Pharmaceutical study data
Model checkpoints
Seismic and reservoir simulation data
DNA sequences
Amazon S3
Backups
Website hosting
Surveillance video/closed-circuit television
Log files
Media assets
Machine learning training data
User-generated content
Financial records
Autonomous vehicle data
Meteorological and environmental research
Oil and gas topography

Storage Class

S3 Intelligent-Tiering
Automatic cost savings by auto-tiering data with any access pattern

S3 Standard
General purpose storage for active, frequently accessed data

S3 Standard-Infrequent Access (S3 Standard-IA)
Low cost storage for data accessed monthly, and requires milliseconds retrieval

S3 Glacier Instant Retrieval
Low cost storage for long-lived data, with retrieval in milliseconds

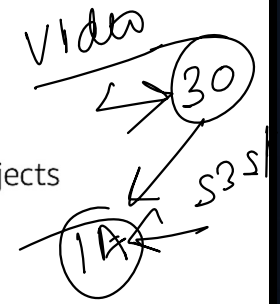
S3 Glacier Flexible Retrieval
Long-term, low-cost storage for backups and archives, with retrieval options from minutes to hours

S3 Glacier Deep Archive
Lowest cost cloud storage for long-term, rarely accessed archive data, with retrieval in hours

S3 One Zone-Infrequent Access (S3 One Zone-IA)
Infrequently accessed data in a single AZ for cost savings

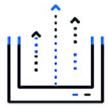
S3 on Outposts
Delivers object storage to on-premises AWS Outposts environments to meet local data processing and data residency needs

Intelligent Tiering



S3 Intelligent-Tiering

Only cloud storage class with automatic cost savings by moving objects between cost-optimized access tiers



Automatically moves objects between the optimal access tiers	Optimizes storage costs based on access patterns	Data lakes and applications with changing or unknown access patterns

S3 Standard

S3 Standard

General purpose storage for active, frequently accessed data with millisecond access

Delivers high durability, availability, and performance

Designed for frequent access, low latency, and high throughput

Data lakes, cloud-native applications, websites, and content distribution

S3 Standard IA

S3 Standard-Infrequent Access (S3 Standard-IA)

Lower cost storage for data accessed monthly, with milliseconds retrieval

Infrequently accessed data with rapid retrieval and the durability, availability, and performance of S3 Standard

Low-latency and high throughput of S3 Standard, with a low per GB storage price and per GB retrieval fee

Long-term storage, backups, and disaster recovery

S3 Glacier Instant Retrieval

S3 Glacier Instant Retrieval

Lowest cost storage for long-term archive data, with milliseconds retrieval

Designed for rarely accessed, long-term data that requires immediate retrieval

Save up to 68% on storage costs compared with using the S3 Standard-Infrequent Access storage class

Long-term digital preservation of data that is accessed once per quarter

S3 Glacier Flexible Retrieval

S3 Glacier Flexible Retrieval

Archive or backup data with secure, durable, and low-cost storage

Low-cost, durable archive with low retrieval fees

3 retrieval speeds: expedited (1–5 mins), standard (3–5 hours), and bulk (12 hours)

Data archive with query-in-place capabilities to learn from data-at-rest

S3 Glacier Deep Archive

S3 Glacier Deep Archive

Lowest-cost cloud storage for long-term archives at about $1 per TB/month

Eliminate on-premises tape libraries and the need for hardware refresh cycles

2 retrieval speeds: standard (within 12 hours) and bulk (within 48 hours)

Long-term digital preservation for data that is accessed once per year

S3 One Zone-IA

S3 One Zone-Infrequent Access (S3 One Zone-IA)
Performance of S3 Standard-IA stored in a single AZ at 20% of the cost

Low latency, high throughput storage for infrequently accessed objects that require rapid retrieval

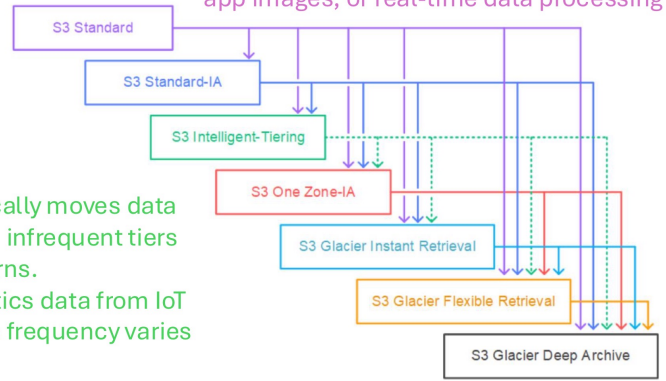
Saves storage costs for data that doesn't require multi-AZ redundancy

Ideal for secondary backups and workloads with easily re-creatable data

Storage Lifecycle Policies Summary

Definition: For data accessed less frequently but requires rapid access when needed.
Prod Example: Monthly reporting data, compliance logs, or backups you retrieve occasionally.

Definition: Default storage class for frequently accessed data with low latency and high throughput.
Prod Example: Hosting dynamic website assets, mobile app images, or real-time data processing inputs.



Definition: Lower-cost infrequent access storage in a single Availability Zone.
Prod Example: Re-creatable data such as image thumbnails or temporary backup copies.

Definition: Automatically moves data between frequent and infrequent tiers based on usage patterns.
Prod Example: Analytics data from IoT devices where access frequency varies unpredictably.

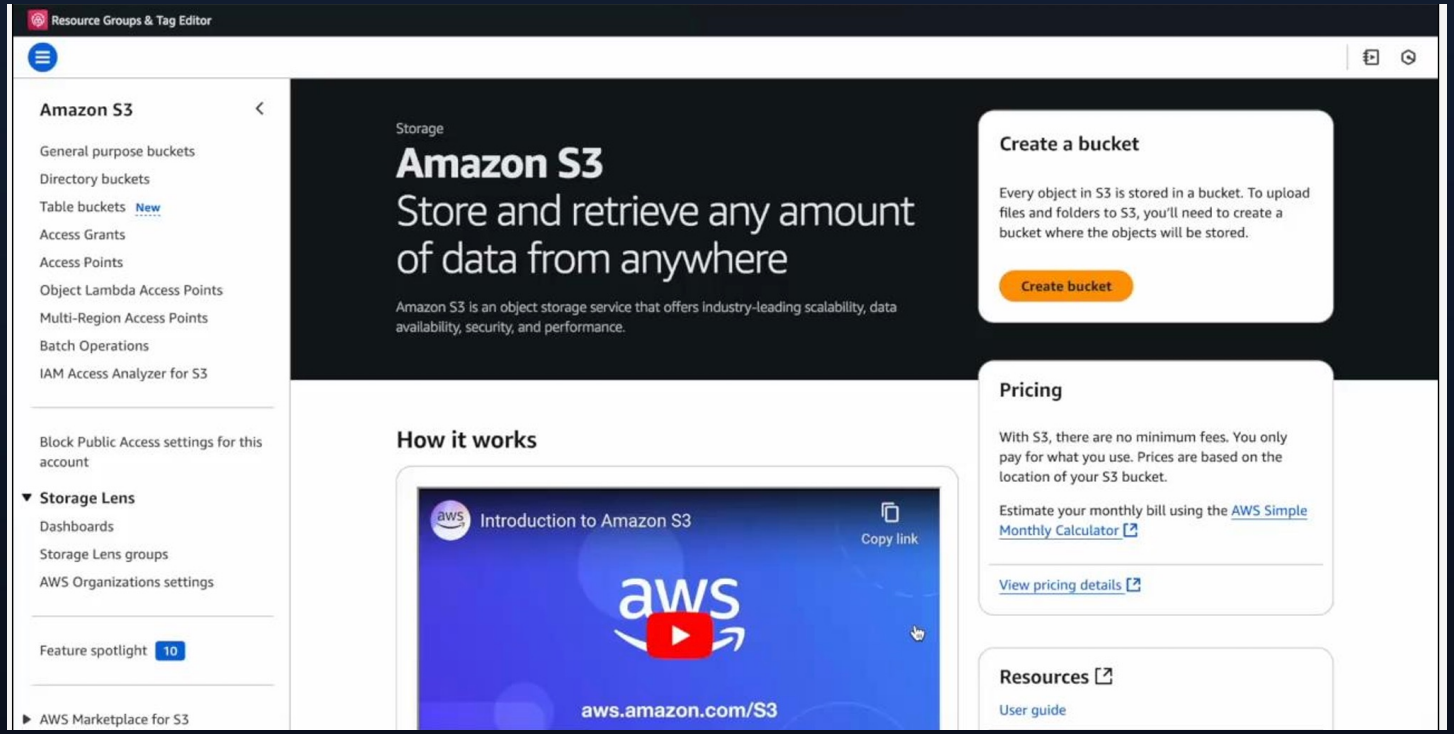
Definition: Cost-effective archive storage with flexible access times (minutes to hours).
Prod Example: Long-term backups of customer data, HR records, or archived video footage

Definition: Low-cost archive class with millisecond access for rarely accessed data.
Prod Example: Medical imaging or financial documents needed instantly for audits.

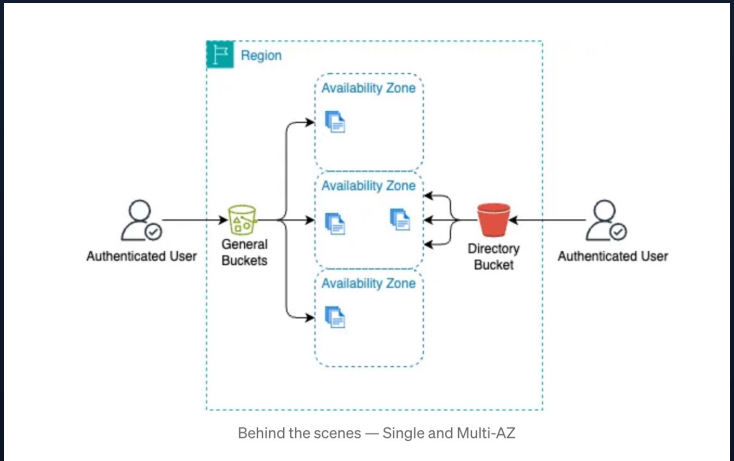
Definition: Lowest-cost storage for archival with retrieval time from 12 hours or more.
Prod Example: Regulatory compliance data retention for 7–10 years (e.g., tax records, legal archives).

Storage Class	Use Case	Access Latency	Notes	Pricing (as of October 2024)
S3 Standard	Frequently accessed data	Millisecond retrieval	High durability, multi-AZ storage	$0.023/GB
S3 Intelligent-Tiering	Dynamic access patterns	Millisecond retrieval	Automatically optimizes cost	$0.023 - $0.00099 per GB/month
S3 Standard-IA (Infrequent Access)	Less frequently accessed data	Millisecond retrieval	Lower cost, with retrieval fees	$0.006 per GB/month
S3 Glacier Instant Retrieval	Archive data with occasional access	Millisecond retrieval	Best for quick archival retrievals	$0.01 per GB/month
S3 Glacier Flexible Retrieval	Long-term archival storage	Minutes to hours	Lower retrieval cost than Instant	$0.004 per GB/month + retrieval fee
S3 Glacier Deep Archive	Deep cold storage with rare access	Up to 12 hours	Cheapest storage class for archiving	$0.00099 per GB/month + retrieval fee

S3 Bucket via AWS Console



Bucket Types (Generic)

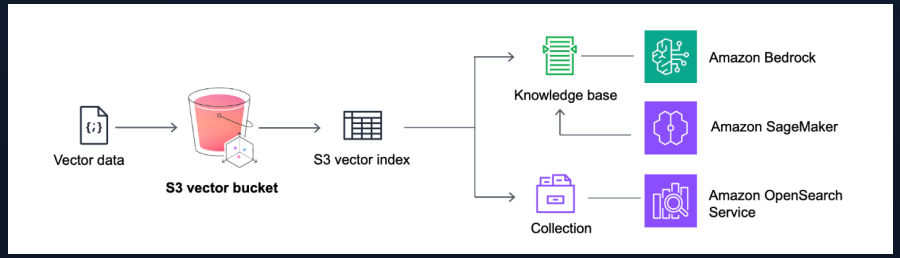


Behind the scenes — Single and Multi-AZ

Feature	AWS S3 Directory Buckets	Regular S3 Buckets
Storage Class	S3 Express One Zone	Standard, Intelligent-Tiering, Glacier, etc.
Data Storage	Single Availability Zone	Multiple Availability Zones
Latency	Low	Variable
Durability	Lower (Single AZ)	Higher (Multiple AZs)
Cost	Lower	Variable, generally higher
Use Cases	Low-latency applications, temporary data storage	General-purpose storage, high durability needs

Bucket Types (Advanced)

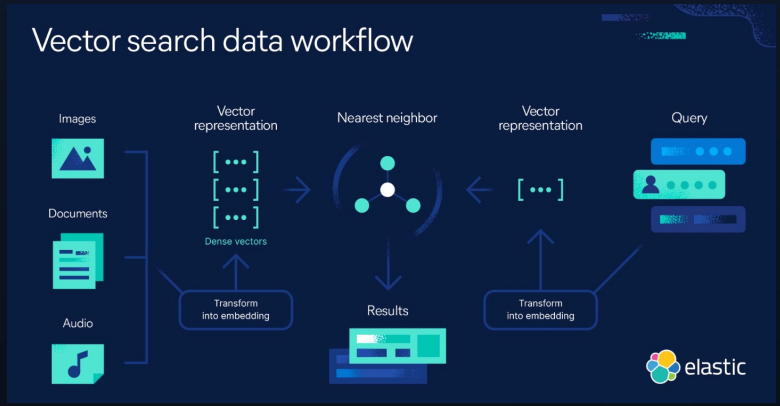
Vector Buckets



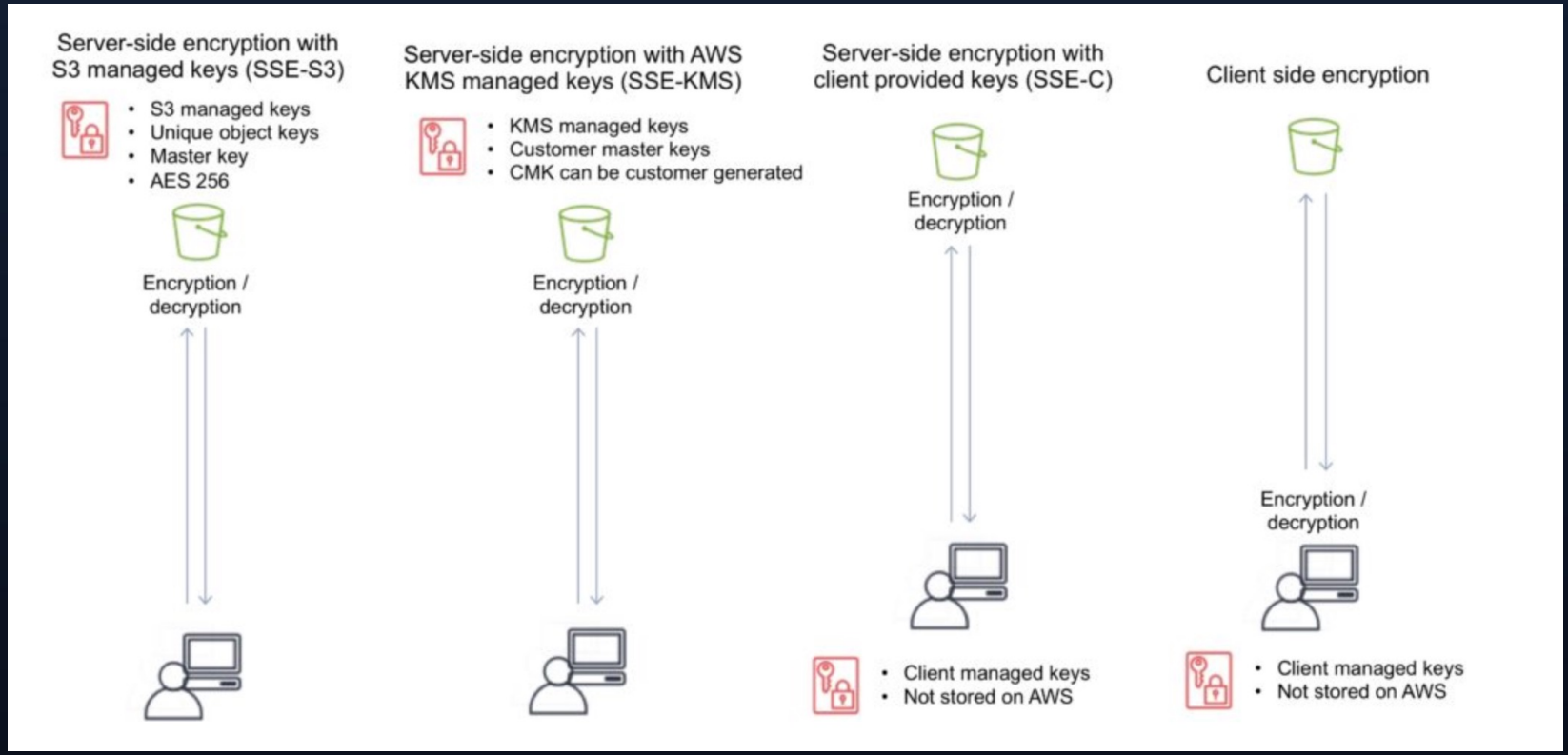
Semantic search is a search engine technology that interprets the meaning of words and phrases.

What is a Vector DB?

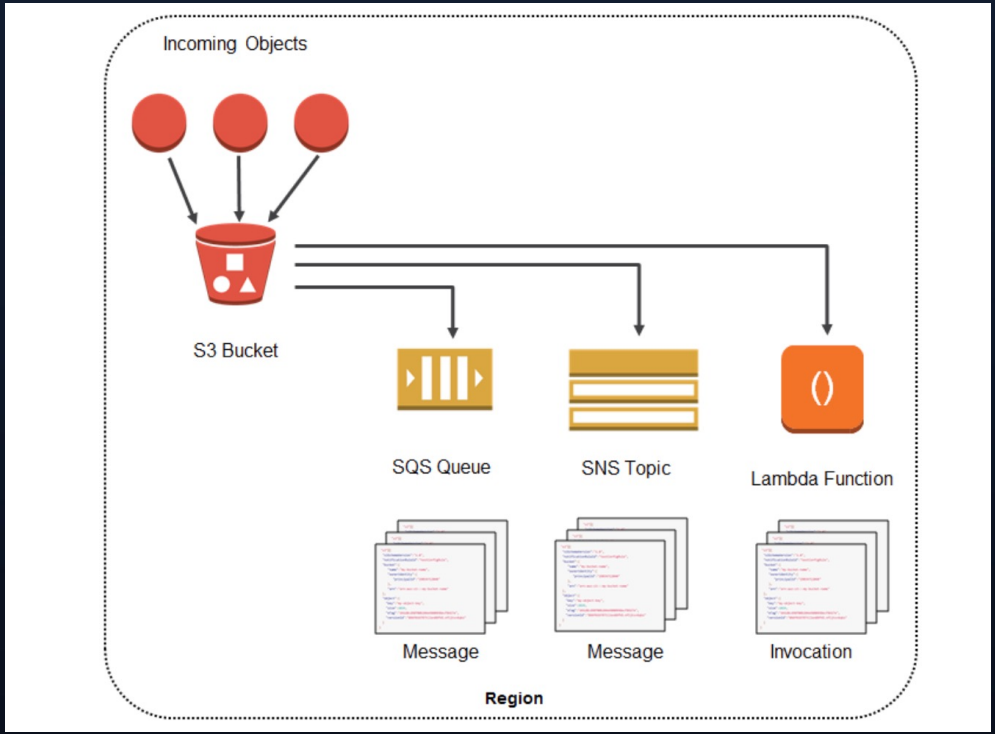
A vector space is a set of vectors that can be played with each other. You can add or multiply them, but need to follow some certain rules.



Encryption in S3



Event Notifications



Recap

- Durability: Amazon S3 offers 99.999999999% (11 nines) durability, ensuring extremely reliable data storage with minimal risk of data loss.
- Availability: S3 provides 99.99% availability for standard access, ensuring data is accessible when needed.
- Storage Classes: Multiple storage classes are available, such as S3 Standard, S3 Intelligent-Tiering, S3 Glacier variants, and the high-performance Express One Zone for different access patterns and cost optimizations.
- Strong Consistency: S3 guarantees strong read-after-write consistency, so any changes to objects are immediately visible for reads across the system.
- Security and Access Control: Supports fine-grained access policies, encryption options (SSE-S3, SSE-KMS, SSE-C), and features like Multi-Factor Authentication (MFA) Delete to prevent accidental or malicious deletions.
- Replication: Cross-Region Replication (CRR) and Same-Region Replication (SRR) enable automatic copying of objects for disaster recovery, compliance, and latency reduction.
- Scalability: Automatically scales to handle any amount of data and any number of requests without performance degradation.
- Performance: Supports high request rates (thousands of requests per second per prefix), parallel access without needing random object key names for performance optimization.
- Data Lifecycle Management: Allows automated transition of data through various storage classes based on rules to optimize cost and data retention.
- Integration and Analytics: Integrates with AWS analytics, big data tools (Athena, Redshift), and provides storage analytics and logging features for usage insights and auditing

Quick Recap of S3

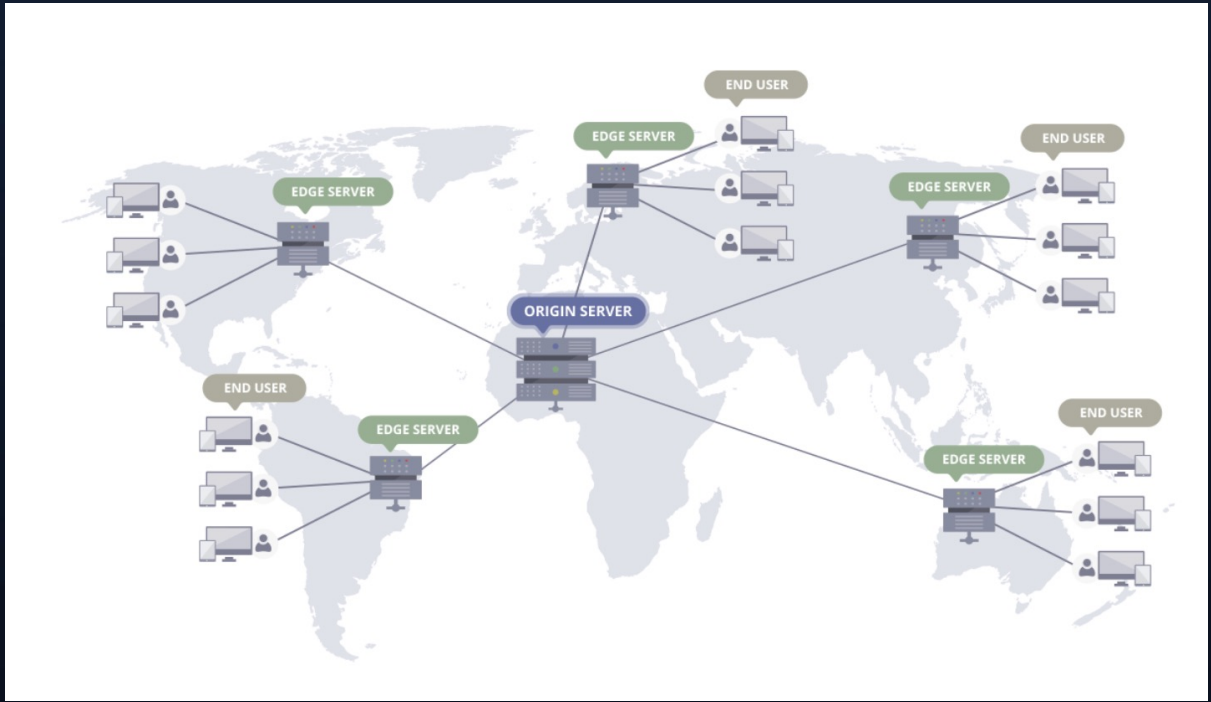
Hosting a website using Amazon S3

Method 0 – Directly exposing the website

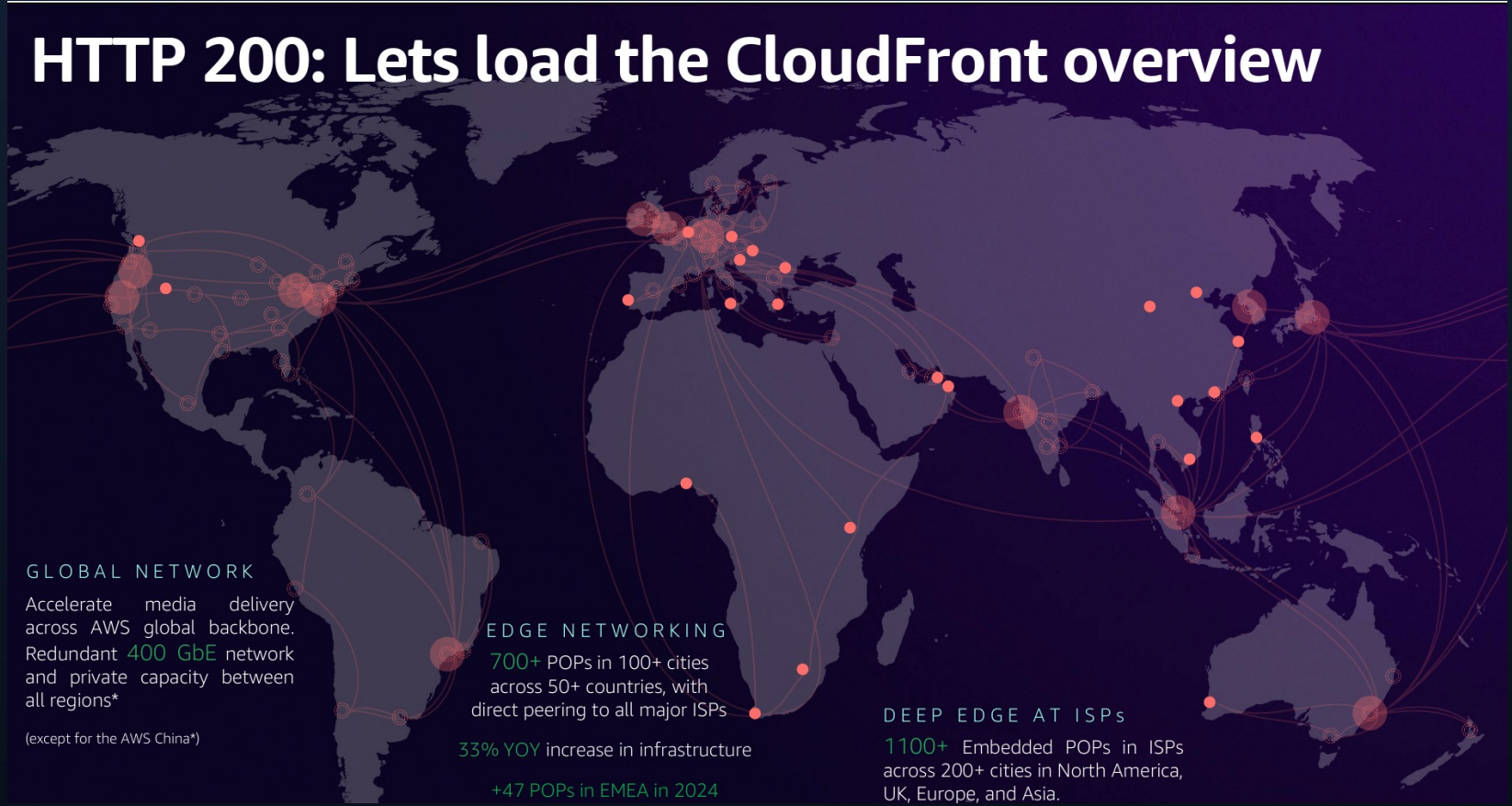
Method 1 – Maintaining the website using AWS Amplify

Method 2 – Hosting the website using Amazon Cloudfront

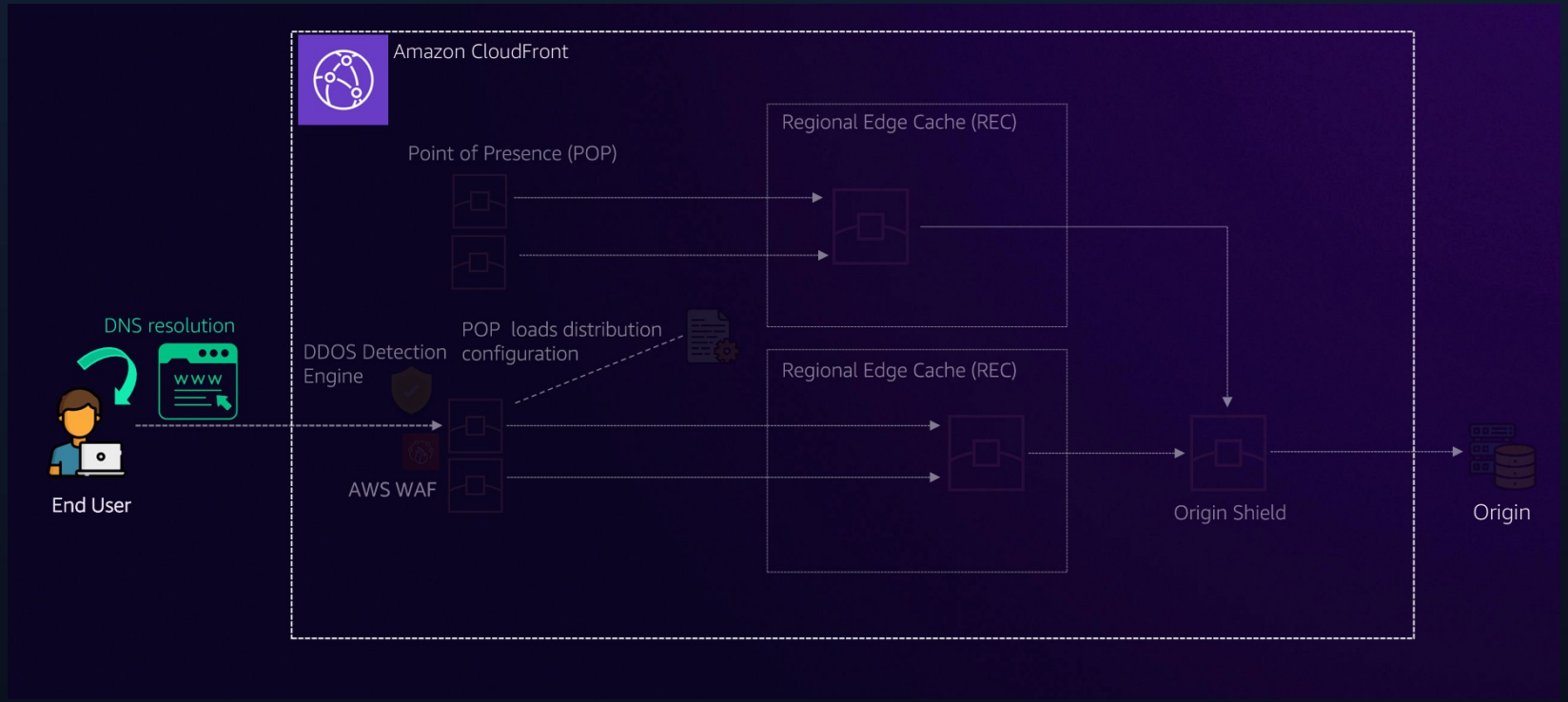
Introduction to Amazon CloudFront



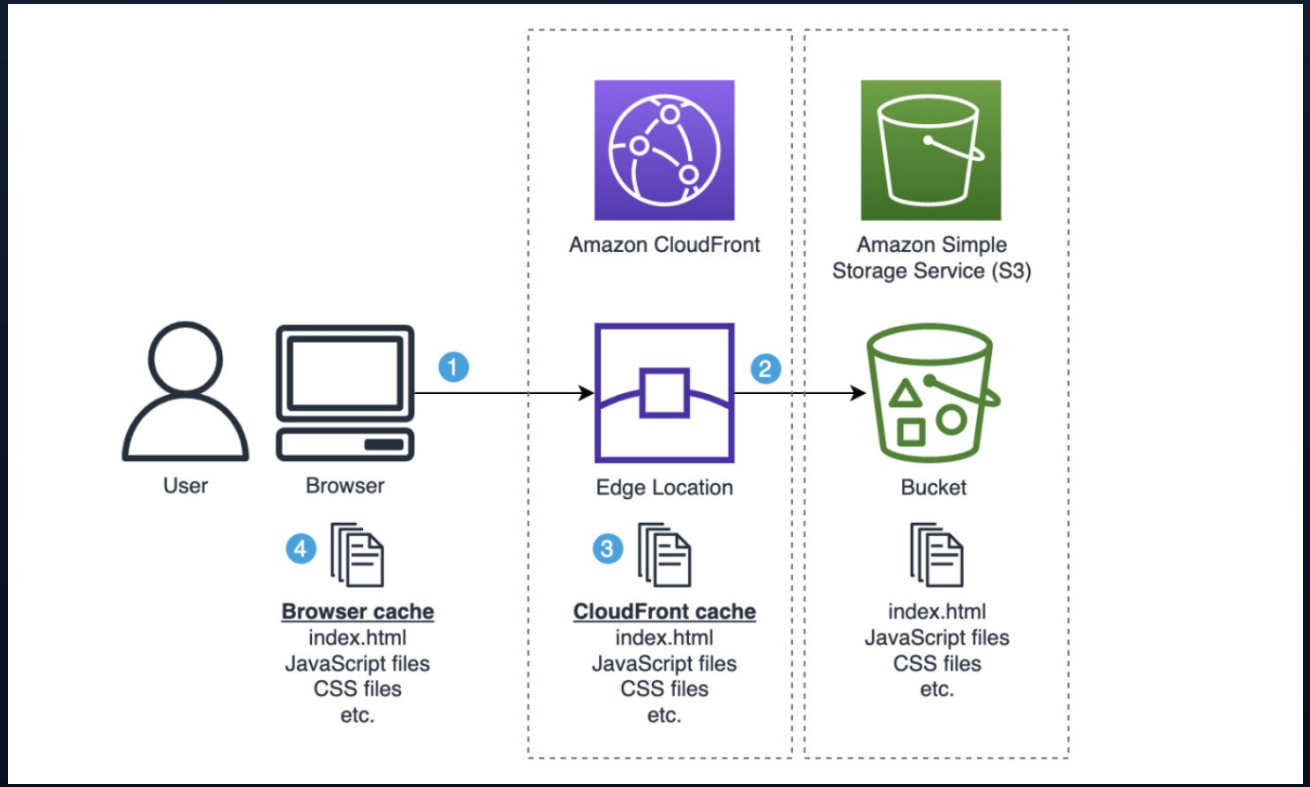
Overview


HTTP 200: Lets load the CloudFront overview
GLOBAL NETWORK
Accelerate media delivery across AWS global backbone. Redundant 400 GbE network and private capacity between all regions*
(except for the AWS China*)
EDGE NETWORKING
700+ POPs in 100+ cities across 50+ countries, with direct peering to all major ISPs
33% YOY increase in infrastructure
+47 POPs in EMEA in 2024
DEEP EDGE AT ISPs
1100+ Embedded POPs in ISPs across 200+ cities in North America, UK, Europe, and Asia.

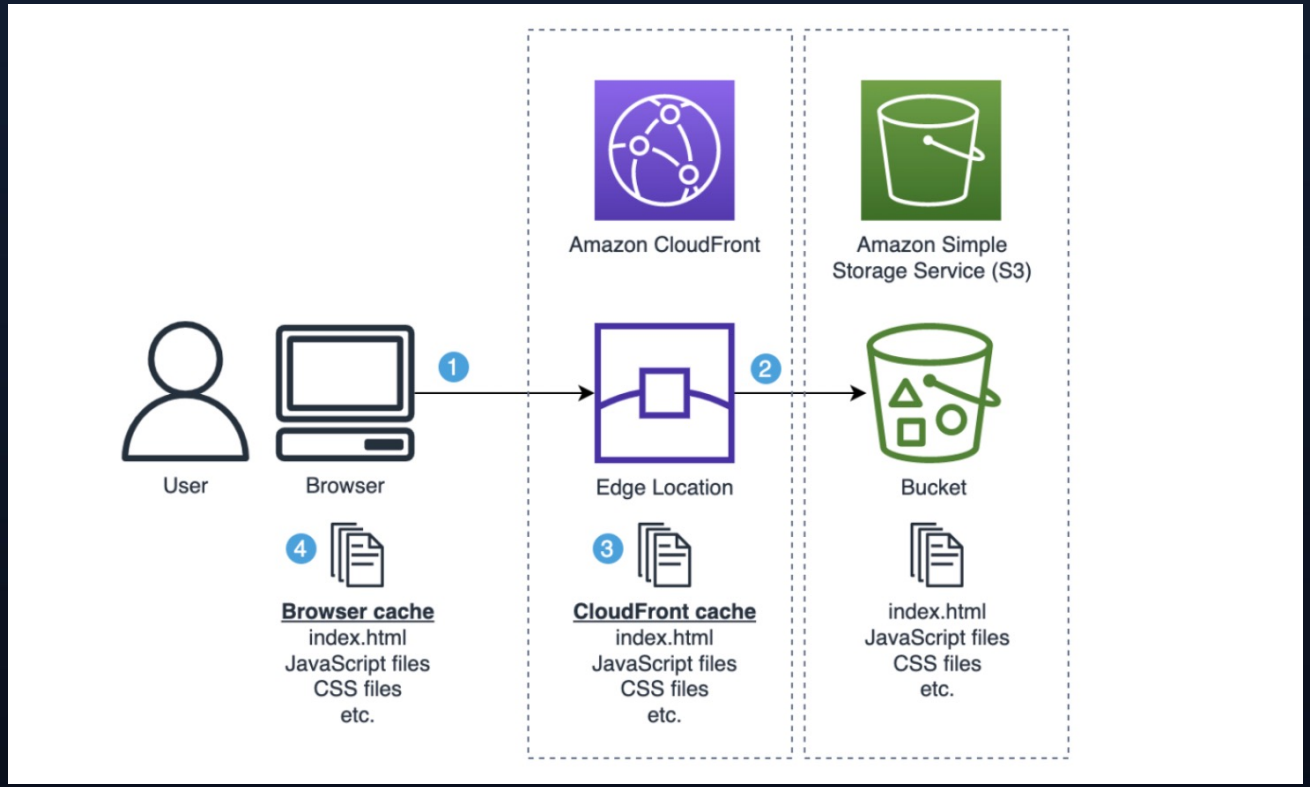
Working of CloudFront



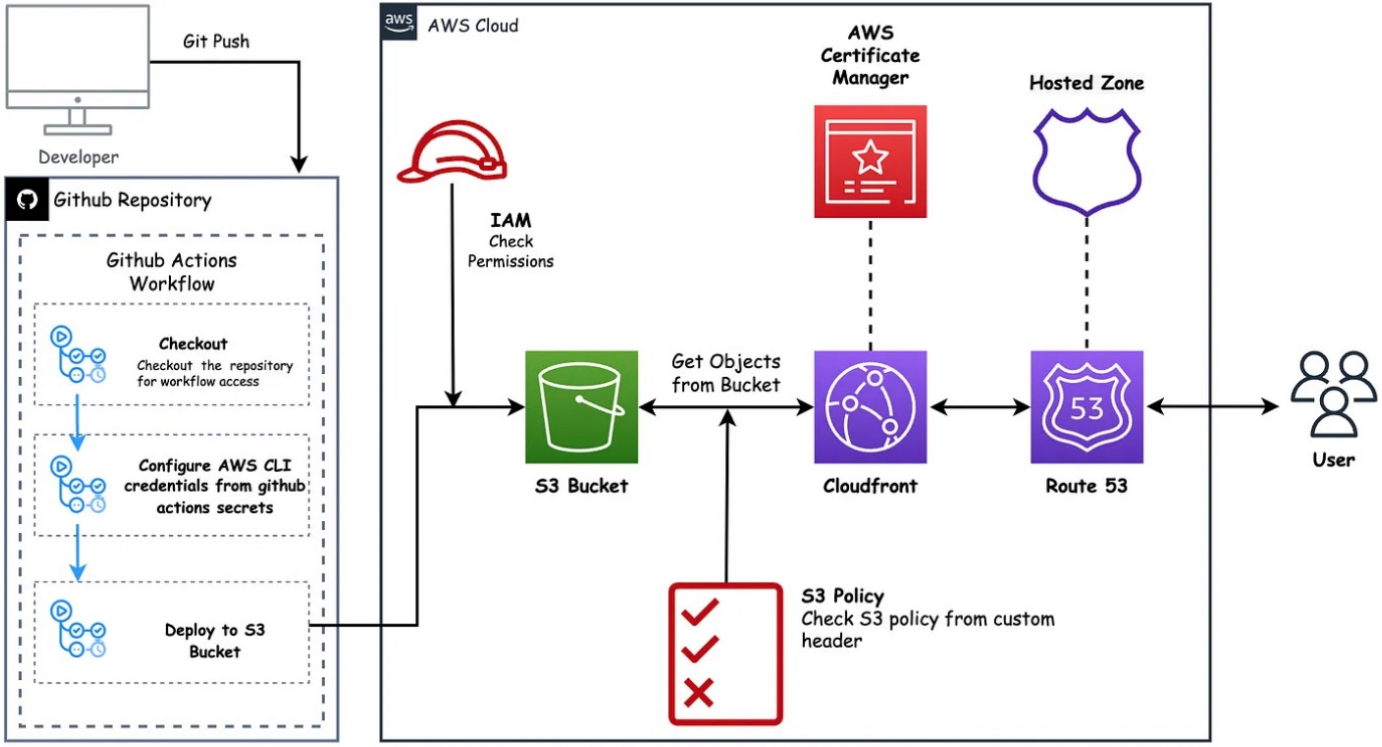
CloudFront S3 Architecture



CloudFront S3 Architecture



GitOps and S3 Hosting



GitOps and S3 Hosting