Lecture 5: AWS Storage

Maharishi International University

Department of Computer Science

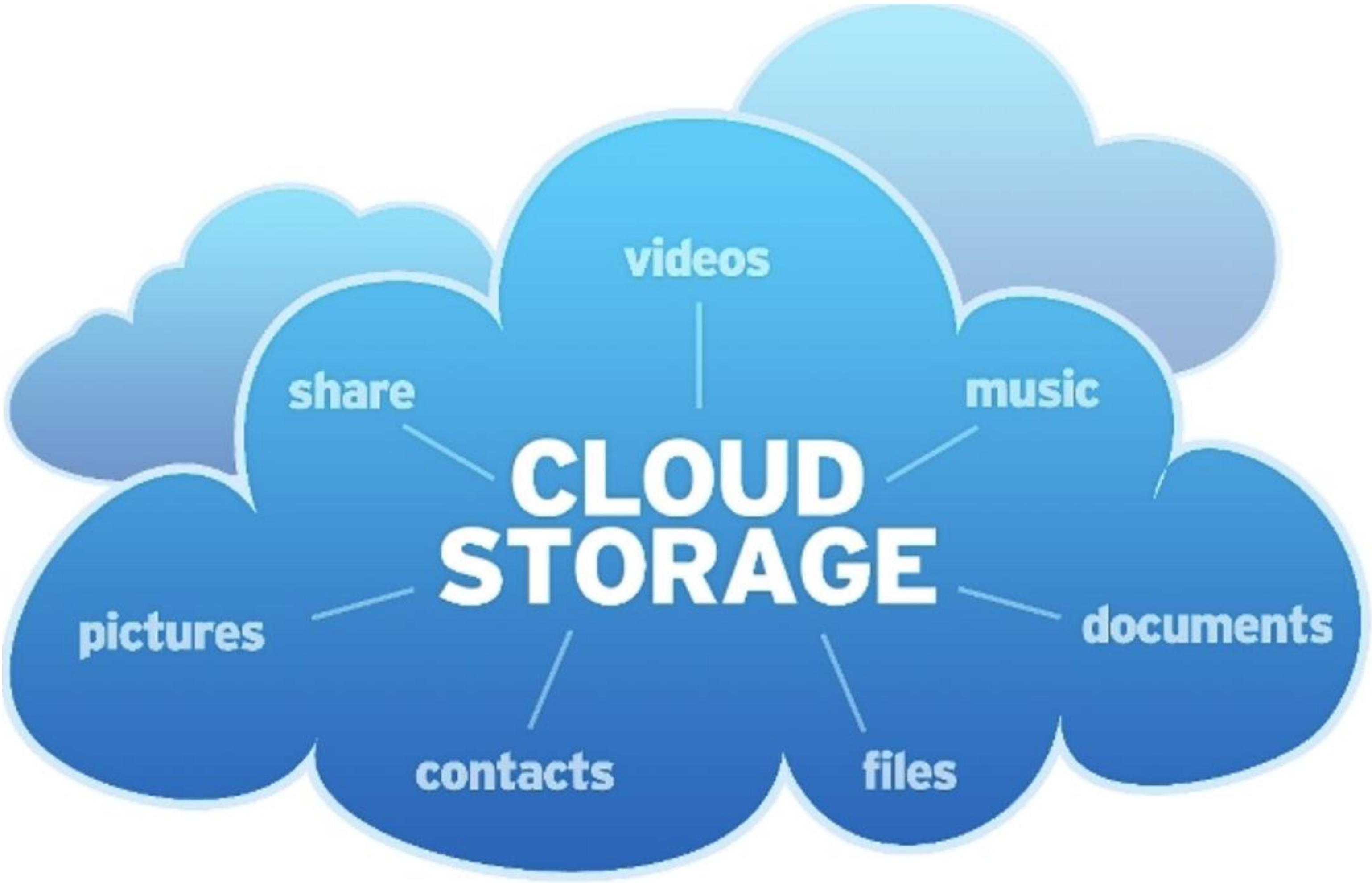
M.S. Thao Huy Vu

Maharishi International University - Fairfield, Iowa

Agenda

- **Introduction to Cloud Storage**
- **Overview of AWS Storage Services**
- **Amazon S3 (Simple Storage Service)**
- **CloudFront**
- **Amazon EBS (Elastic Block Store)**
- **Amazon EFS (Elastic File System)**

Introduction to Cloud Storage



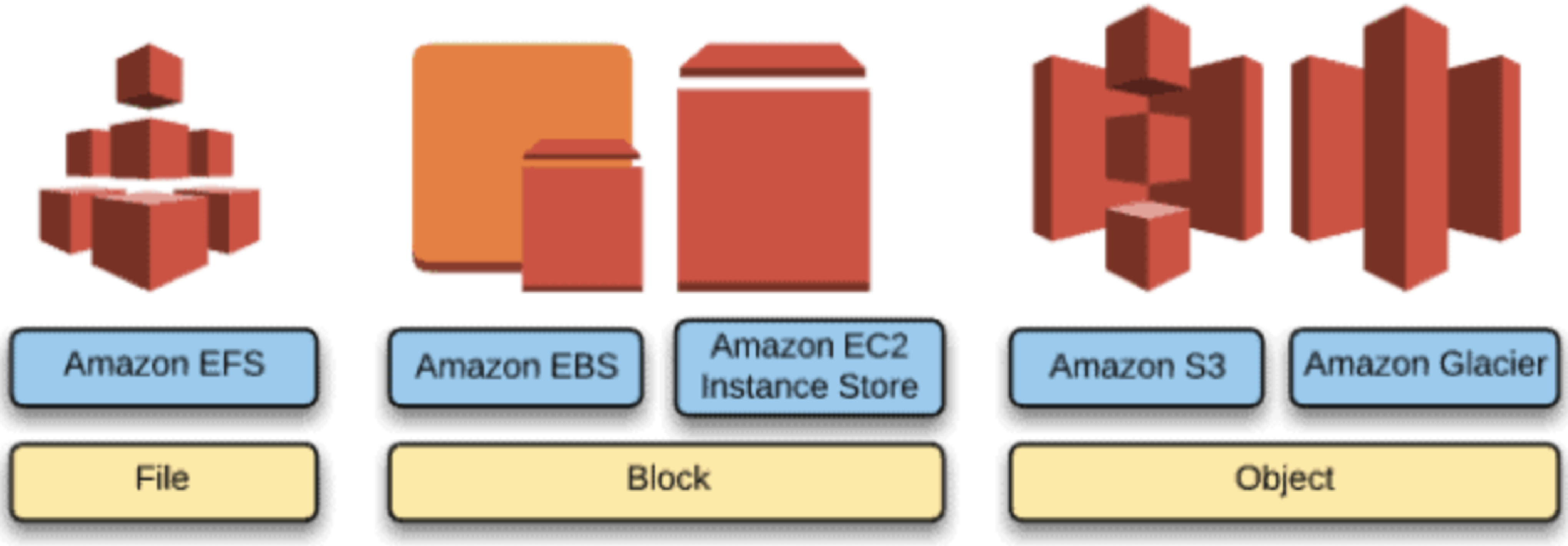
Introduction to Cloud Storage

- Store digital data across multiple virtual servers hosted by third parties
- Scalability
- Cost efficiency
- Accessibility and Availability
- Security
- Disaster recovery

Introduction to Cloud Storage

- Types of Storage
    - Object: Data is managed as objects such as photos, videos.
    - Block: Data is divided into blocks of memory.
    - File: Data is saved as files in a file system.

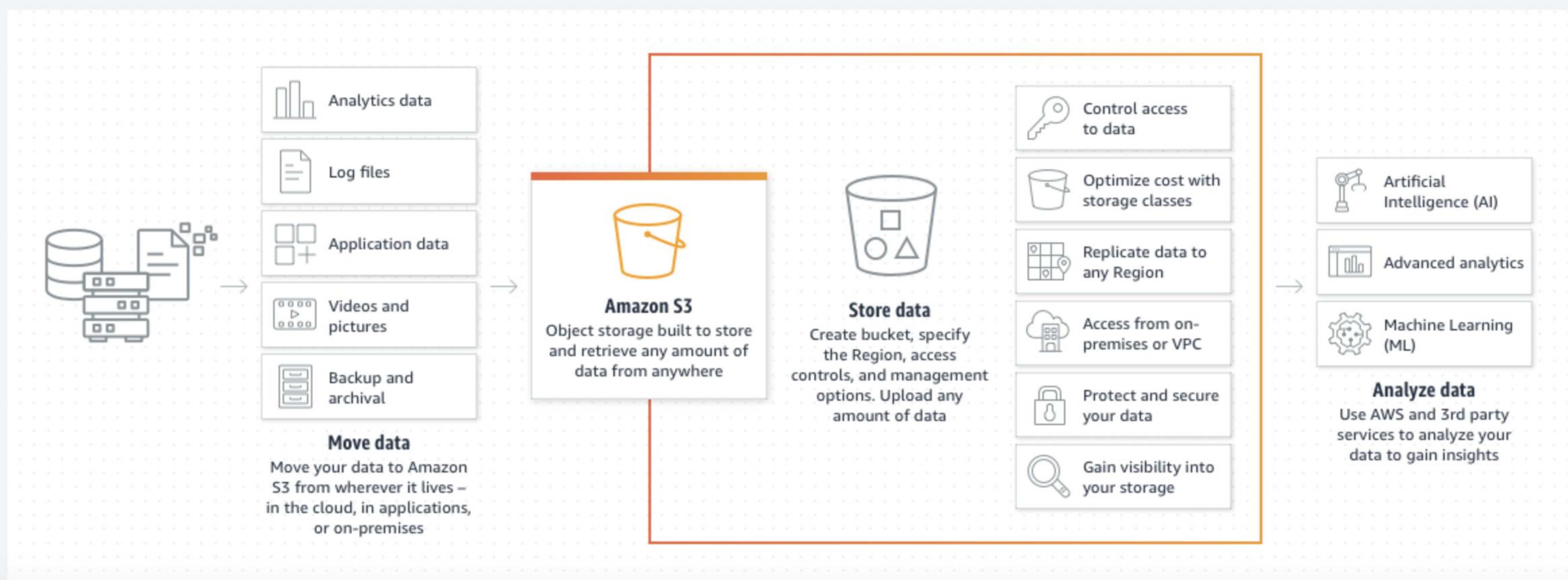
AWS Storage


Amazon EFS
Amazon EBS
Amazon EC2 Instance Store
Amazon S3
Amazon Glacier
File
Block
Object

Amazon S3

- **Definition**: A scalable, durable, and secure object storage service provided by AWS.
- **Storage Type**: Stores data as objects in buckets. Each object consists of data, metadata, and a unique key.
- **Key Features**:
  - **Scalability**: Automatically scales to handle unlimited data.
  - **Durability**: 99.999999999% (11 nines) durability.
  - **Accessibility**: Access via the web, AWS CLI, SDKs, or APIs.
  - **Security**: Supports encryption, IAM policies, bucket policies, and ACLs.
- **Use Cases**:
  - Data backups, archives, and disaster recovery.
  - Hosting static websites.
  - Storing media files, logs, or application data.

Amazon S3



Analytics data

Log files

Application data

Videos and pictures

Backup and archival

Move data

Move your data to Amazon S3 from wherever it lives – in the cloud, in applications, or on-premises

Amazon S3

Object storage built to store and retrieve any amount of data from anywhere

Store data

Create bucket, specify the Region, access controls, and management options. Upload any amount of data

Control access to data

Optimize cost with storage classes

Replicate data to any Region

Access from on-premises or VPC

Protect and secure your data

Gain visibility into your storage

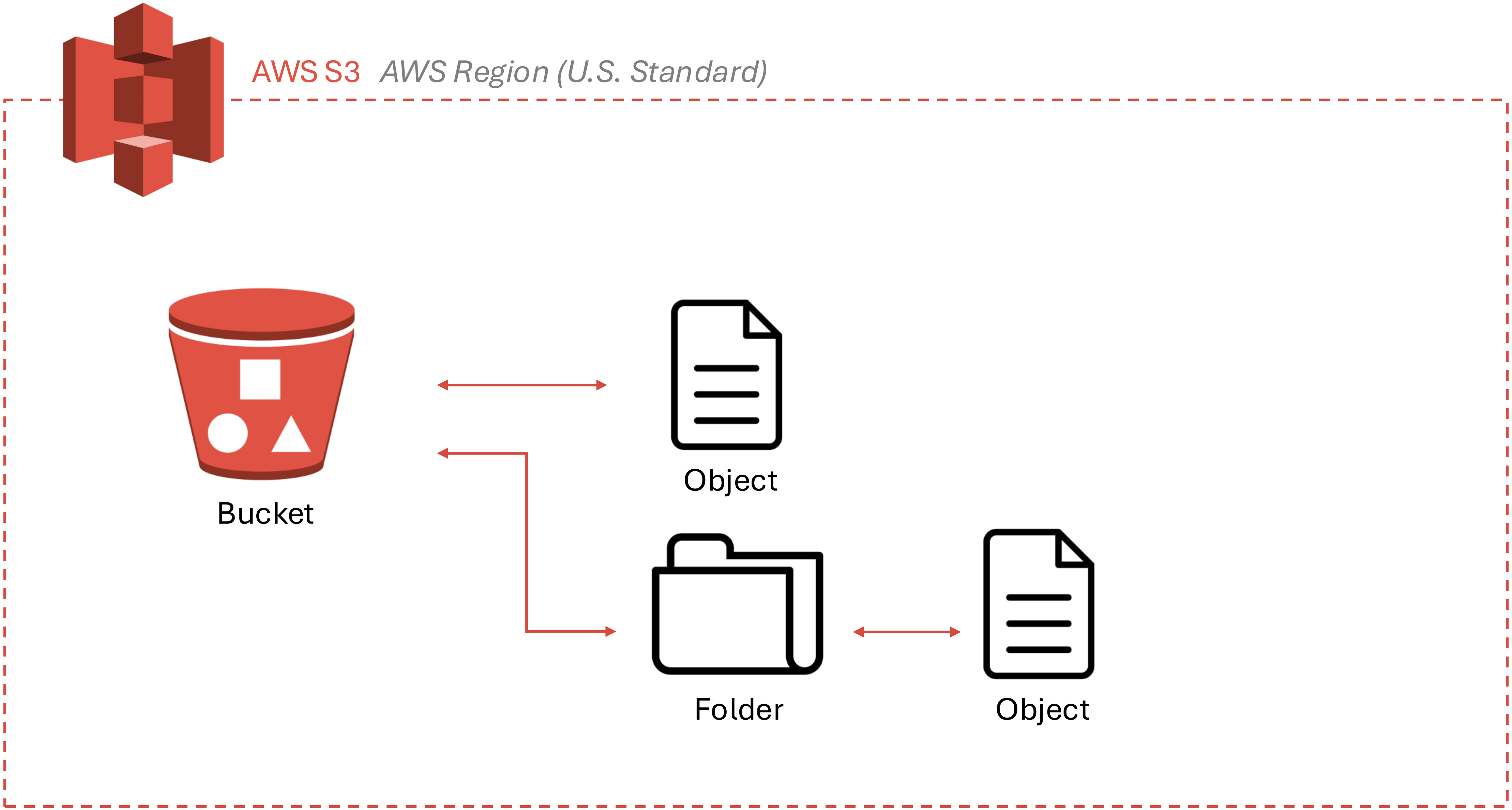
Artificial Intelligence (AI)

Advanced analytics

Machine Learning (ML)

Analyze data

Use AWS and 3rd party services to analyze your data to gain insights



AWS S3 *AWS Region (U.S. Standard)*

Bucket

Object

Folder

Object

AWS S3 Regions

- **Region Selection**: When creating a bucket, you must choose a specific AWS region where your data will be stored in a regional data center.
- **Best Practice**: Select a region closest to your users or applications to minimize latency and improve performance.
- **Global Scalability**: Enable Cross-Region Replication (CRR) to replicate data across multiple regions for high availability and disaster recovery.
- **CloudFront Integration**: Combine S3 with CloudFront to further reduce latency, improve content delivery, and optimize costs.
- **Bucket Naming**: Ensure bucket names are globally unique across all AWS accounts, as common names like "develop" might already be in use.

AWS S3 Objects

- **Definition**: Objects are the fundamental data entities stored in S3 buckets. Each object consists of:
  - **Data**: The actual content (e.g., file, image, video).
  - **Metadata**: Key-value pairs describing the object (e.g., file type, size).
  - **Key**: A unique identifier (name) within the bucket.
  - **Version ID** (optional): Tracks changes if versioning is enabled.
- **Features**:
  - Objects can be up to 5TB in size.
  - Supports tagging for categorization and lifecycle management.
- **Use Cases**: Storing files, backups, media, logs, and more.

AWS S3 Permissions

- **Definition**: Control access to S3 buckets and objects using various mechanisms like policies.
- **Bucket Policies**:
  - JSON-based policies applied to the entire bucket.
  - Useful for granting access to multiple users or accounts.
- **IAM Policies**:
  - Define permissions for IAM users, groups, or roles.
  - Allow or deny actions on specific buckets or objects.
- **Object-Level Permissions**: Permissions can be set for individual objects within a bucket.
- **Public Access Settings**: Block public access to prevent unintentional data exposure.
- **Pre-Signed URLs**: A **temporary, secure link** that allows **someone to access a private object** in your S3 bucket **without making it public**.

AWS S3 Storage Class

- A storage class is assigned to each object in S3 and determines its cost, availability, durability, and access frequency characteristics.
- Storage classes vary in:
  - **Storage Cost**: Different costs per GB based on the class.
  - **Object Availability**: High availability for frequently accessed data (e.g., S3 Standard) versus lower availability for archival storage (e.g., Glacier).
  - **Object Durability**: All storage classes provide 99.999999999% (11 nines) durability.
  - **Frequency of Access**: Classes like Intelligent-Tiering manage access patterns automatically.
- Objects are assigned the **S3 Standard storage class** by default unless specified otherwise.
- Each object must belong to a storage class.

AWS Storage classes

- **S3 Standard**:
  - For frequently accessed data.
  - High performance and durability.
- **S3 Standard - Infrequent Access (IA)**: For data accessed less often but still requires fast retrieval (e.g., backups).
- **S3 One Zone-IA**:
  - For infrequently accessed data stored in a single availability zone.
  - Suitable for re-creatable data.
- **S3 Glacier**:
  - For long-term archival storage.
  - Retrieval times: Minutes to hours.
- **S3 Glacier Deep Archive**:
  - Lowest-cost archival storage.
  - Retrieval times: Hours.
- **S3 Intelligent-Tiering**: Automatically moves data between frequent and infrequent access tiers based on access patterns.

AWS S3 Static Web Hosting

- **Static Content Only**: Supports static files; does not support server-side scripting like PHP or databases.
- **Bucket Configuration**:
  - Must enable the **"Static Website Hosting"** option in the bucket settings.
  - Requires an **index document** (e.g., `index.html`) and optionally a **custom error document** (e.g., `error.html`).
- **Public Access**: Objects must be publicly accessible, and the bucket policy may need to allow public reads.
- **Endpoint**: Website is accessible via a dedicated **S3 website endpoint** (e.g., `http://bucket-name.s3-website-region.amazonaws.com`).

AWS S3 Transfer Acceleration

- **Definition**: A feature that speeds up data uploads and downloads to/from S3 by using AWS's globally distributed **Edge Locations** via **Amazon CloudFront**.
- **How It Works**:
  - Data is routed through the nearest Edge Location.
  - Uses AWS's optimized network to transfer data to the target S3 bucket.
- **Benefits**:
  - Reduces latency for long-distance transfers.
  - Improves upload and download performance, especially for geographically dispersed users.
- **Use Case**: Ideal for applications requiring fast, large-scale file transfers over long distances.
- **Requirement**: Must enable Transfer Acceleration on the S3 bucket.

AWS S3 Lifecycle Policy

- **Transition Rules**:
  - **Goal**: Automatically **move** objects to **cheaper storage classes** as they age based on object **creation date** or **last access date**.
  - **Example Workflow:**
    - **Day 0–30:** Store in **S3 Standard** (frequent access)
    - **After 30 days:** Move to **S3 Standard-IA**
    - **After 90 days:** Move to **S3 Glacier**
    - **After 180+ days:** Optionally move to **Glacier Deep Archive**
- **Expiration Rules**:
  - **Goal:** Automatically **delete objects** after a certain period.
  - **Use case:** Clean up outdated logs, old backups, or temp files.
  - **Example:** Delete files **180 days after creation**.

S3 Lifecycle Policy vs. S3 Intelligent-Tiering

Feature	S3 Lifecycle Policy	S3 Intelligent-Tiering
Definition	Automates object transitions or deletions based on predefined time periods or access patterns.	Automatically optimizes storage costs by moving objects between tiers based on actual access patterns.
Cost	Lower cost for storage transitions but requires manual setup and monitoring.	Slightly higher cost due to monitoring and automation fees.
Storage Classes	Moves objects between Standard, Standard-IA, Glacier, and Deep Archive.	Includes frequent and infrequent tiers within Intelligent-Tiering.
Deletion	Can automatically delete objects based on age.	Does not delete objects; focuses only on tier transitions.
Use Case	Suitable for known lifecycle patterns (e.g., backups, logs).	Best for dynamic workloads with unknown or changing access patterns.

AWS S3 Encryption

- **Definition**: Protects data stored in S3 by encrypting it at rest and during transit to ensure security and compliance.
- **SSE-S3**: Encrypts S3 objects using keys handled & managed by AWS. Out-of-box feature. Nothing to do on your side. All objects stored in S3 are encrypted.
- **SSE-KMS**: You can use AWS Key Management Service which provides keys to encrypt your data in S3.
- **SSE-C**: The key is generated by you (the appliance) and use that key to encrypt data.
- **Client-Side Encryption**: Encrypt in your client app before storing in S3.
- Encryption in Transit: Ensures secure data transfer using **HTTPS/TLS**.

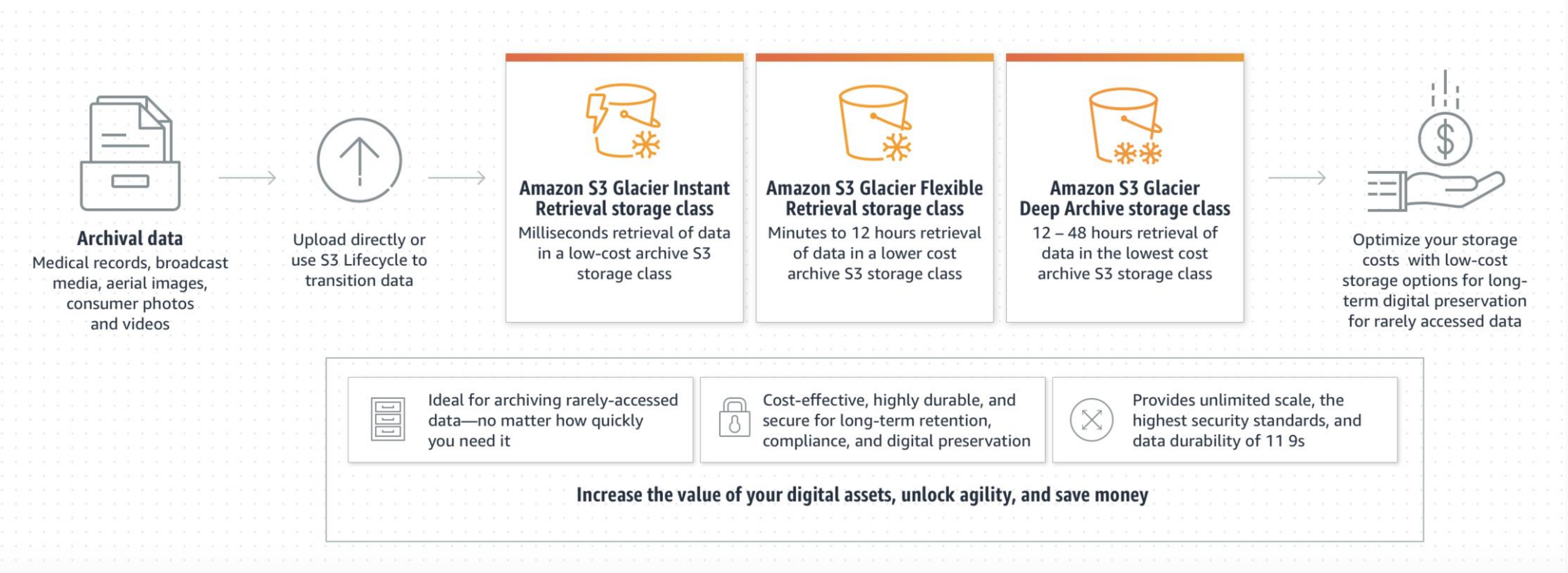
Benefits of S3 Multipart Upload

- **Improved Throughput**: Parts can be uploaded in parallel, speeding up the upload process.
- **Resilience to Network Issues**: Smaller parts minimize the impact of network errors, allowing quick recovery by retrying only failed parts.
- **Pause and Resume**: Uploads can be paused and resumed over time. Multipart uploads have no expiry until explicitly completed or stopped.
- **Flexible Object Creation**: Allows uploading an object before knowing its final size, enabling dynamic creation and upload.

S3 Pre-Signed URL

- **Definition**: A temporary URL generated to provide secure, time-limited access to an object in an S3 bucket.
- Grant temporary access to download or upload files without making the bucket or objects public.
- The URL has an expiration time set during creation (e.g., 15 minutes, 1 hour).
- Only users with the URL and within the specified time window can access the object.
- Created using AWS SDKs, CLI, or APIs by signing the request with valid AWS credentials.
- Use cases:
  - Temporary file sharing (e.g., reports, media).
  - Secure file uploads to an S3 bucket from external applications.
  - Controlled access for external users or systems.

Amazon Glacier



Archival data

Medical records, broadcast media, aerial images, consumer photos and videos

Upload directly or use S3 Lifecycle to transition data

Amazon S3 Glacier Instant Retrieval storage class

Milliseconds retrieval of data in a low-cost archive S3 storage class

Amazon S3 Glacier Flexible Retrieval storage class

Minutes to 12 hours retrieval of data in a lower cost archive S3 storage class

Amazon S3 Glacier Deep Archive storage class

12 – 48 hours retrieval of data in the lowest cost archive S3 storage class

Optimize your storage costs with low-cost storage options for long-term digital preservation for rarely accessed data

Ideal for archiving rarely-accessed data—no matter how quickly you need it

Cost-effective, highly durable, and secure for long-term retention, compliance, and digital preservation

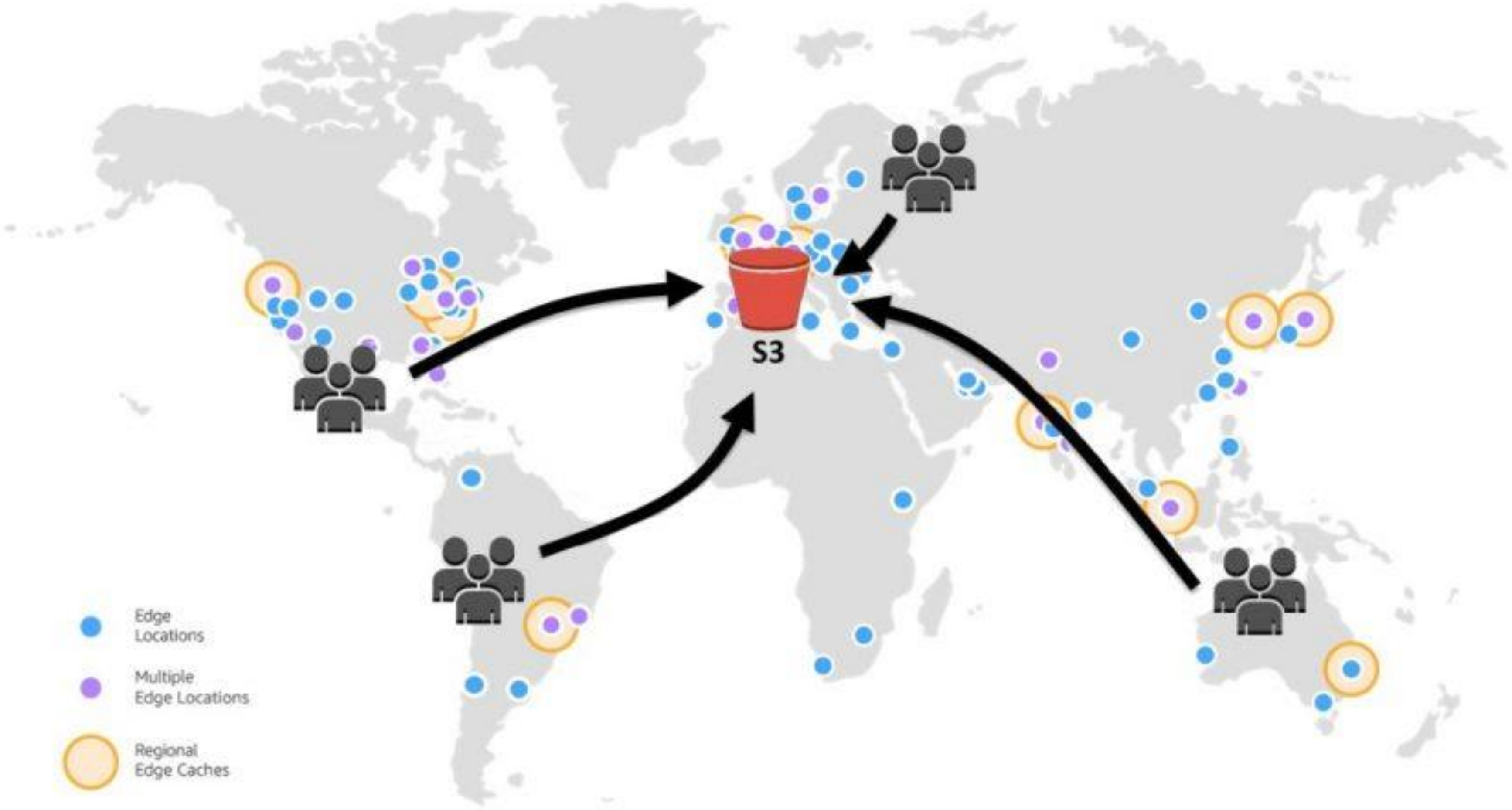
Provides unlimited scale, the highest security standards, and data durability of 11 9s

Increase the value of your digital assets, unlock agility, and save money

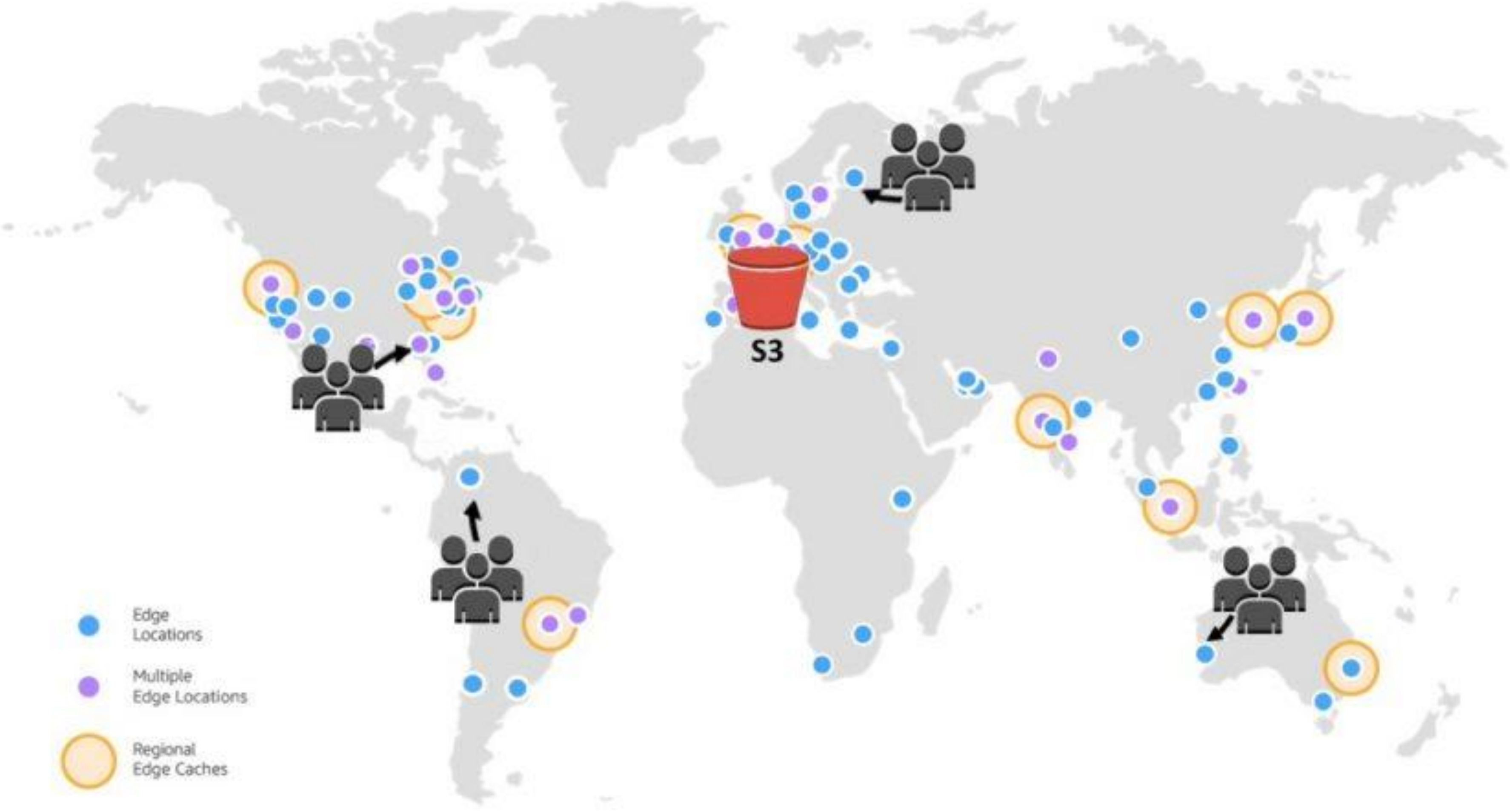
S3 Pricing Factors

- **Storage Class**: Costs vary by the storage class (e.g., S3 Standard, Standard-IA, Glacier).
- **Storage Amount**: Charged per GB stored per month.
- **Requests and Data Retrieval**:
  - Costs for PUT, GET, LIST, DELETE, and other requests.
  - Retrieval costs for storage classes like Glacier and Glacier Deep Archive.
- **Data Transfer**:
  - **Inbound Transfers**: Free for uploads into S3.
  - **Outbound Transfers**: Charged for data leaving S3 to the internet or other AWS regions.
- **Lifecycle Transitions**: Costs for moving objects between storage classes using lifecycle policies.
- **Replication**: Additional charges for replicating objects across regions.
- **Other Features**: Costs for S3 Transfer Acceleration, Object Lock, and analytics.

Hosting static contents on S3



Hosting static contents on S3 with CloudFront



Amazon CloudFront

- CloudFront is a CDN (Content Delivery Networks) that delivers content or API securely on AWS.
  - Frontend: Speeds up distribution of static and dynamic web content, such as .html, .css, .js, and image files.
  - Backend: Accelerate and secure the backend API.
- Edge Location: The data centers where content is cached and served.
- Features:
  - Speeds up delivery of static and dynamic web content (e.g., .html, .css, .js, images).
  - Routes user requests to the nearest **edge location** for low latency and high performance.
  - Delivers content via a global network of data centers, improving user experience.

Amazon CloudFront

- Content Distribution Process: The CloudFront delivers content by routing requests to the nearest edge location, which improves speed and reduces latency.
- Caching Mechanism: Data is cached across multiple global edge locations and the options available for invalidating or updating cached content.
- Security:
    - AWS Shield for DDoS
    - IAM for access control
    - SSL/TLS to encrypt transit data

CloudFront Invalidation

- **Definition**: CloudFront invalidation is the process of removing cached content from edge locations in the Content Delivery Network (CDN), ensuring that updated content is served to users.
- **Scope**: Invalidate specific files (e.g., `/index.html`) or all files using a wildcard (`/*`).
- **Impact**: Ensures that users access the latest content but may temporarily increase latency as new files are retrieved from the origin.
- **Costs**: A limited number of invalidation requests are free each month (1,000 paths). Additional requests incur charges.
- **Alternatives**: Use versioned file names (e.g., `style-v2.css`) to avoid invalidations by making updates automatically bypass caches.

Amazon CloudFront with Route 53

- **Enhanced DNS Routing**: Route 53 directs users to the nearest CloudFront distribution, improving content delivery speed and reducing latency.
- **Alias Records**:
  - Use Route 53 alias records for routing traffic to CloudFront distributions.
  - Benefits:
    - Avoids DNS query charges.
    - Faster DNS resolution compared to CNAME records.
- Health Checks and Failover:
  - Route 53 can monitor the health of CloudFront distributions.
  - Implements failover to redirect users to alternative distributions if the primary one is unavailable.

AWS Global Accelerator

- **Definition**: AWS Global Accelerator is a networking service that improves the availability and performance of applications globally by routing user traffic through AWS's global network infrastructure.
- **Performance**:
  - Routes traffic through AWS’s optimized global network, reducing latency and jitter.
  - Directs users to the nearest application endpoint (e.g., EC2, ALB) for faster responses.
- **High Availability**: Automatically redirects traffic to healthy endpoints in case of endpoint or regional failures.
- **Static IPs**: Provides static IP addresses that act as a fixed entry point for your application, simplifying DNS management.
- **Traffic Distribution**: Supports weighted traffic distribution across multiple endpoints.
- **Health Checks**: Monitors endpoint health to ensure requests are routed only to healthy instances.

AWS Global Accelerator

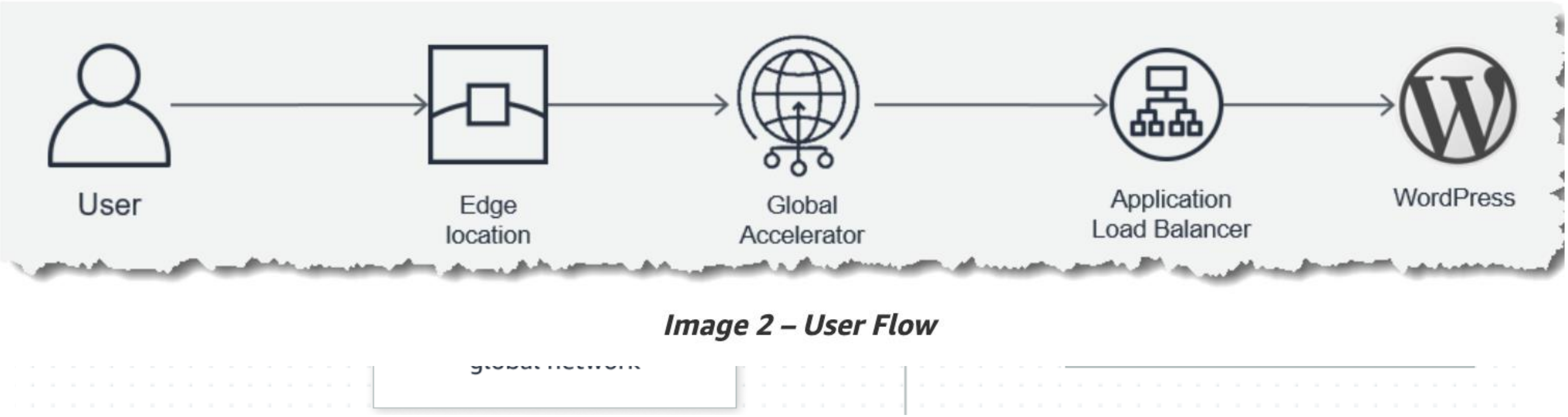


Image 2 – User Flow

AWS Global Accelerator vs. CloudFront

Feature	AWS Global Accelerator	Amazon CloudFront
Purpose	Optimizes global performance for applications (e.g., APIs, gaming).	Accelerates delivery of static and dynamic web content.
Static IPs	Provides static IPs for consistent entry points.	No static IPs; uses domain-based URLs.
Caching	No caching; directly routes traffic to application endpoints.	Caches static and dynamic content at edge locations.
Use Case	Optimizing global traffic for APIs, gaming, or applications requiring low latency and high availability.	Distributing static and dynamic content globally, such as websites and streaming media.
Cost	Charged per hour and data transfer; ideal for high-performance, low-latency applications.	Charged based on data transfer and requests; ideal for content-heavy applications.

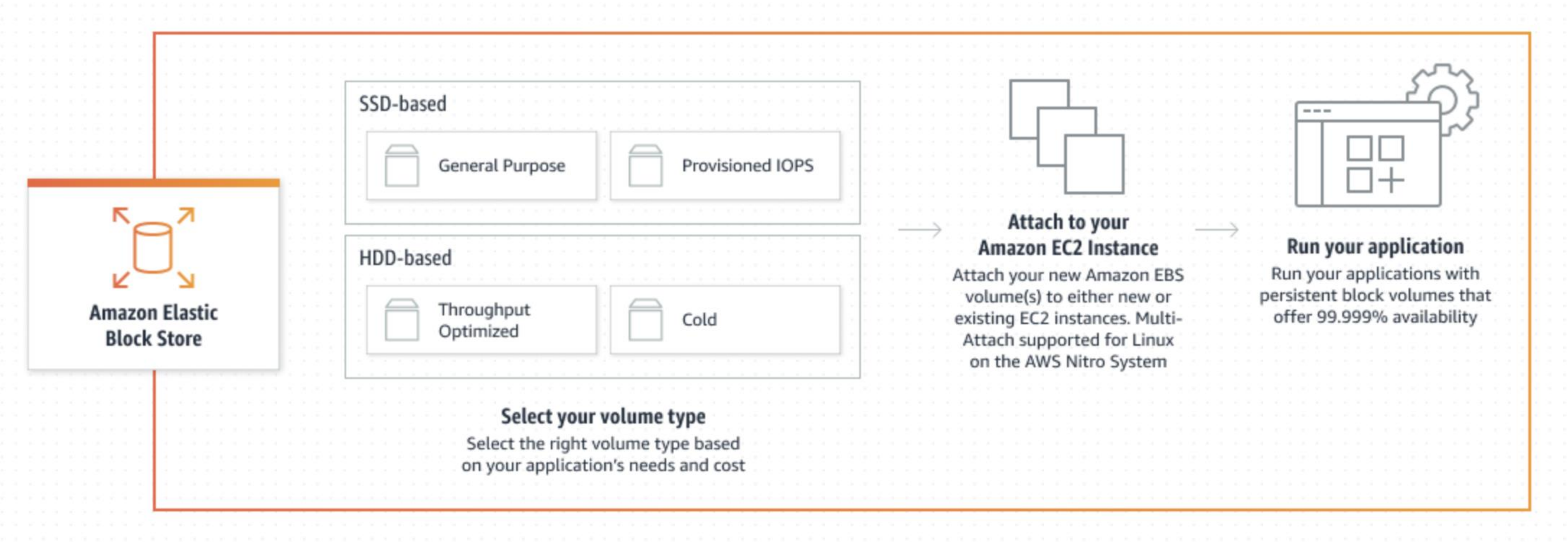
Amazon Elastic Block Storage (EBS)

- **Definition**: Amazon EBS is a high-performance, block storage service for Amazon EC2 instances, providing persistent and reliable storage.
- **Durability**:
  - Data is automatically replicated within the same Availability Zone (AZ).
  - Ensures data persistence even if the associated EC2 instance is stopped or terminated.
- **Performance**:
  - Provides low-latency storage for mission-critical applications.
  - Supports various volume types for different workloads (e.g., general-purpose SSD, provisioned IOPS SSD, HDD).
- **Scalability**: Volumes can be resized dynamically to accommodate growing storage needs.
- **Backup and Recovery**:
  - Supports snapshots to create backups of volumes.
  - Snapshots are stored in Amazon S3 and can be used to create new volumes.
- **Encryption**: Supports encryption at rest and in transit for secure data storage.

Amazon EBS

- Zone-wide: Attach to an EC2 instance in the same AZ.
- Up to 16 TB per volume.
- Use cases:
  - Build mission-critical, I/O intensive applications in the cloud
  - Run Relational or NoSQL databases
  - Right-size big data analytics engines

Amazon EBS



Select your volume type

Select the right volume type based on your application's needs and cost

Attach to your Amazon EC2 Instance

Attach your new Amazon EBS volume(s) to either new or existing EC2 instances. Multi-Attach supported for Linux on the AWS Nitro System

Run your application

Run your applications with persistent block volumes that offer 99.999% availability

Amazon EBS

- **SSD-Based Volumes**:
  - **General Purpose SSD (gp3, gp2)**:
    - Balanced performance for a wide range of workloads (e.g., boot volumes, small databases).
    - Cost-effective with baseline performance and burst capability.
  - **Provisioned IOPS SSD (io2, io1)**:
    - High-performance volumes designed for latency-sensitive, transactional workloads (e.g., databases).
    - Provides predictable, consistent IOPS with high durability.
- **HDD-based Volumes:**
  - **Throughput Optimized HDD (st1)**:
    - Designed for large, sequential workloads requiring high throughput (e.g., big data, log processing).
    - Lower cost per GB compared to SSD.
  - **Cold HDD (sc1)**:
    - Lowest-cost option for infrequently accessed data (e.g., archival storage).
    - Suitable for workloads where performance is not critical.

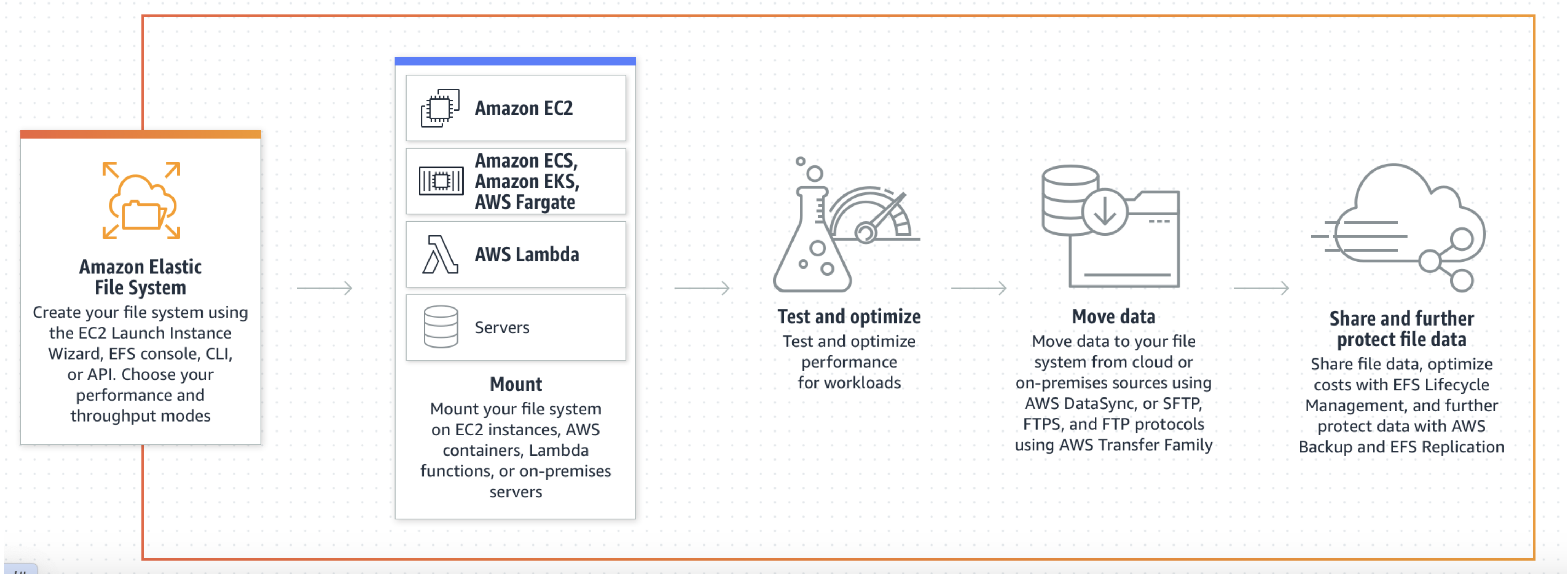
AWS EBS Snapshot

- **Incremental Backup**: Only changes made since the last snapshot are saved, reducing storage costs and time.
- **Durability**: Snapshots are stored in Amazon S3, ensuring high durability.
- **Cross-Region and Cross-AZ**: Snapshots can be copied to different AWS regions or restored to volumes in other Availability Zones.
- **Restore Volumes**: Snapshots can be used to create new EBS volumes or recover data from a previous point in time.
- **Automation**: Supports automated snapshots using AWS Backup or scheduled scripts.
- **Encryption**: Snapshots inherit the encryption status of the source EBS volume.

Amazon Elastic File System

- **Definition**: A fully managed, scalable file storage service designed for use with AWS compute resources like EC2, Lambda, and containers.
- **Scalability**: Automatically scales storage capacity up or down as files are added or removed.
- **Shared Access**: Provides concurrent access for multiple instances and applications across Availability Zones.
- **Performance Modes**:
  - **General Purpose**: Low latency for latency-sensitive applications.
  - **Max I/O**: High throughput for large-scale, data-heavy workloads.
- **Storage Classes**:
  - **Standard**: For frequently accessed data.
  - **Infrequent Access (IA)**: Lower-cost option for infrequently accessed files.
- **High Durability and Availability**: Stores data across multiple Availability Zones for fault tolerance.
- **POSIX (**Portable Operating System Interface) **Compliance**: Supports standard file system semantics (e.g., file locking, directories) across different systems.

Amazon EFS



Amazon Elastic File System

Create your file system using the EC2 Launch Instance Wizard, EFS console, CLI, or API. Choose your performance and throughput modes

Mount

Mount your file system on EC2 instances, AWS containers, Lambda functions, or on-premises servers

Test and optimize

Test and optimize performance for workloads

Move data

Move data to your file system from cloud or on-premises sources using AWS DataSync, or SFTP, FTPS, and FTP protocols using AWS Transfer Family

Share and further protect file data

Share file data, optimize costs with EFS Lifecycle Management, and further protect data with AWS Backup and EFS Replication

Amazon EFS

- Region-wide: Able to attach to multiple EC2 in the same region
- Virtually unlimited storage capacity
- **Use Cases**:
  - Web serving and content management.
  - Big data analytics and machine learning.
  - Shared storage for containers or serverless applications.

EBS vs. EFS

Feature	Amazon EBS	Amazon EFS
Storage Type	Block storage for a single EC2 instance.	Shared file storage accessible by multiple instances.
Maximum Storage	16 TiB per volume (multiple volumes can be attached).	Virtually unlimited, scales automatically.
Performance	Up to 64,000 IOPS and 1,000 MiB/s throughput per volume.	Up to 10 GiB/s throughput; scales with workload.
Use Case	High-performance, low-latency storage for single-instance applications (e.g., databases).	Shared storage for multiple instances or applications requiring POSIX compliance.
POSIX Compliance	Depends on the file system used (e.g., ext4, NTFS).	Fully POSIX-compliant out of the box.
Backup	Supports snapshots for backup and recovery.	Automatic replication across multiple AZs for durability.
Scalability	Manual scaling by adding or resizing volumes.	Automatic scaling based on usage.

Conclusion

- **Amazon S3 (Simple Storage Service):**
  - Scalable, secure object storage for any type of data.
  - Ideal for backup, archiving, and serving static content.
- **Amazon CloudFront:**
  - A fast, secure, global Content Delivery Network (CDN).
  - Caches and delivers content to users with low latency.
- **Amazon EBS (Elastic Block Store):**
  - High-performance block storage for EC2 instances.
  - Best suited for databases and applications requiring persistent storage.
- **Amazon EFS (Elastic File System):**
  - Scalable, shared file storage for multiple EC2 instances.
  - Ideal for modern applications and containerized workloads.

References

- https://docs.aws.amazon.com/
- ChatGPT: https://chatgpt.com/
- Google AI: https://gemini.google.com/app