Introduction to Storage on AWS

Bayard Mason
baymason@amazon.com
November 19, 2021

Agenda

- Introduction
- Storage Primer
- Block Storage
- Shared File Systems
- Object Store
- Backup

Introduction: Why choose AWS for storage

Compelling Economics

Pay as you go

No risky capacity planning

No need to provision for redundancy or overhead

Easy to Use

Self service administration

SDKs for simple integration

No Commitment

Reduce risk

Durable and Secure

Avoid risks of physical media handling

Speed, Agility, Scale

Reduce time to market

Focus on your business, not your infrastructure

Global Scale

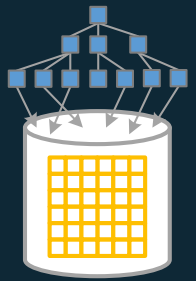
aws

Storage Primer

Block vs File vs Object

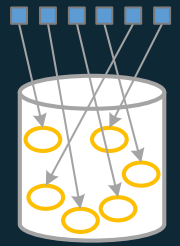
Block Storage

Raw Storage
Data organized as an array of unrelated blocks
Host File System places data on disk
Ex: Hard Disks, Storage Area Network (SAN) Storage Arrays



File Storage

Unrelated data blocks managed by a file (serving) system
Native file system places data on disk
Ex: Network Attached Storage (NAS) Appliances, Windows File Servers



Object Storage

Stores Virtual containers that encapsulate the data, data attributes, metadata and Object IDs
API Access to data
Metadata Driven, Policy-based, etc.
Ex: Ceph, OpenStack Swift

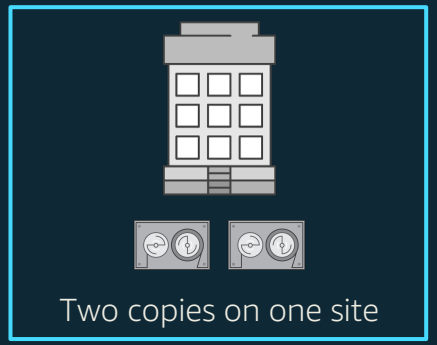
aws

Storage - Characteristics

Some of the ways we look at storage

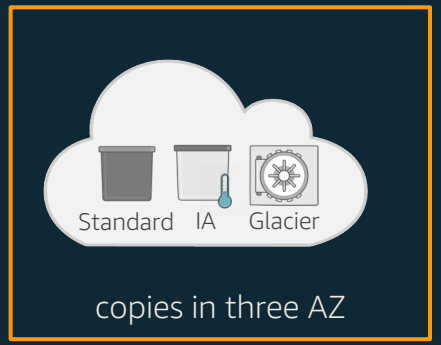
Durability	Availability	Security	Cost	Scalability	Performance	Integration
Measure of expected data loss	Measure of expected downtime	Security measures for at-rest and in-transit data	Amount per storage unit, e.g. $ / GB	Upward flexibility, storage size, number of users	Performance metrics (bandwidth	Ability to interact via API or with other services

Understanding Durability



designed for
99.99%
durability

designed for
99.999%
durability

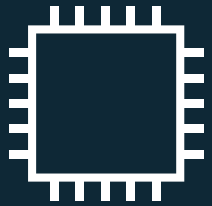


designed for
99.999999999%
durability

aws

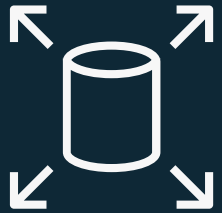
Block Storage

Block storage portfolio



Instance storage

Temporary block-level storage attached to host hardware that is ideal for storage of information that frequently changes or is replicated across multiple instances



Amazon EBS

Easy to use, high performance block storage service designed for use with Amazon Elastic Compute Cloud (EC2) for both throughput and transaction intensive workloads

Snapshots

Incremental, point-in-time copies of your EBS data that can be used to restore new volumes, expand the size of a volume, or move volumes across Availability Zones

aws

EBS is designed for a wide range of workloads on EC2

Enterprise applications

SAP ERP, Oracle ERP, Microsoft SharePoint, Microsoft Exchange

Relational databases

MySQL, PostgreSQL, SQL Server, Oracle DB, SAP HANA

Non-relational/ NoSQL databases

Cassandra, MongoDB, CouchDB

Big data analytics

Kafka, Splunk, Hadoop, Data warehousing

File/media

CIFS/NFS, transcoding, encoding, rendering

LOW LATENCY AND CONSISTENT, HIGH IOPS AND THROUGHPUT

SCALABLE WITHOUT DISRUPTION TO YOUR WORKLOAD

99.999% AVAILABILITY AND AN ANNUAL FAILURE RATE (AFR) OF BETWEEN 0.1%–0.2%

aws

Easy to use, high performance block storage at virtually any scale

Performance for any workload

Up to 256,000 IOPS and single digit millisecond latency

Easy to use

Easily add/remove capacity, or change volume types with Elastic Volumes

High reliability

99.999% availability and an annual failure rate of between 0.1%–0.2%

Virtually unlimited scale

Use a single gigabyte or less, or scale up to petabytes of data

Secure

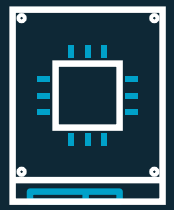
Encrypt all new volumes and data for a region by default with a single setting

Cost-effective

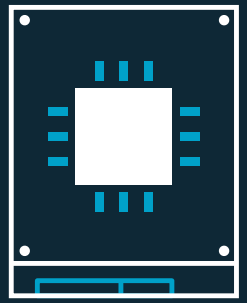
Pay as low as $0.025/GB-month for highly cost-effective dollar per gigabyte block storage

aws

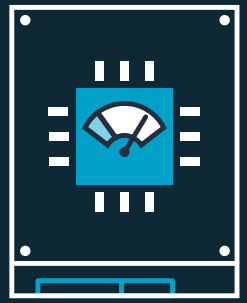
Six different volume types for optimal use



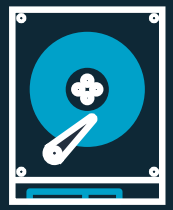
SSD



gp2 – gp3
General Purpose SSD



io1 – io2
Provisioned IOPS SSD



HDD

st1
Throughput Optimized HDD

sc1
Cold HDD

aws

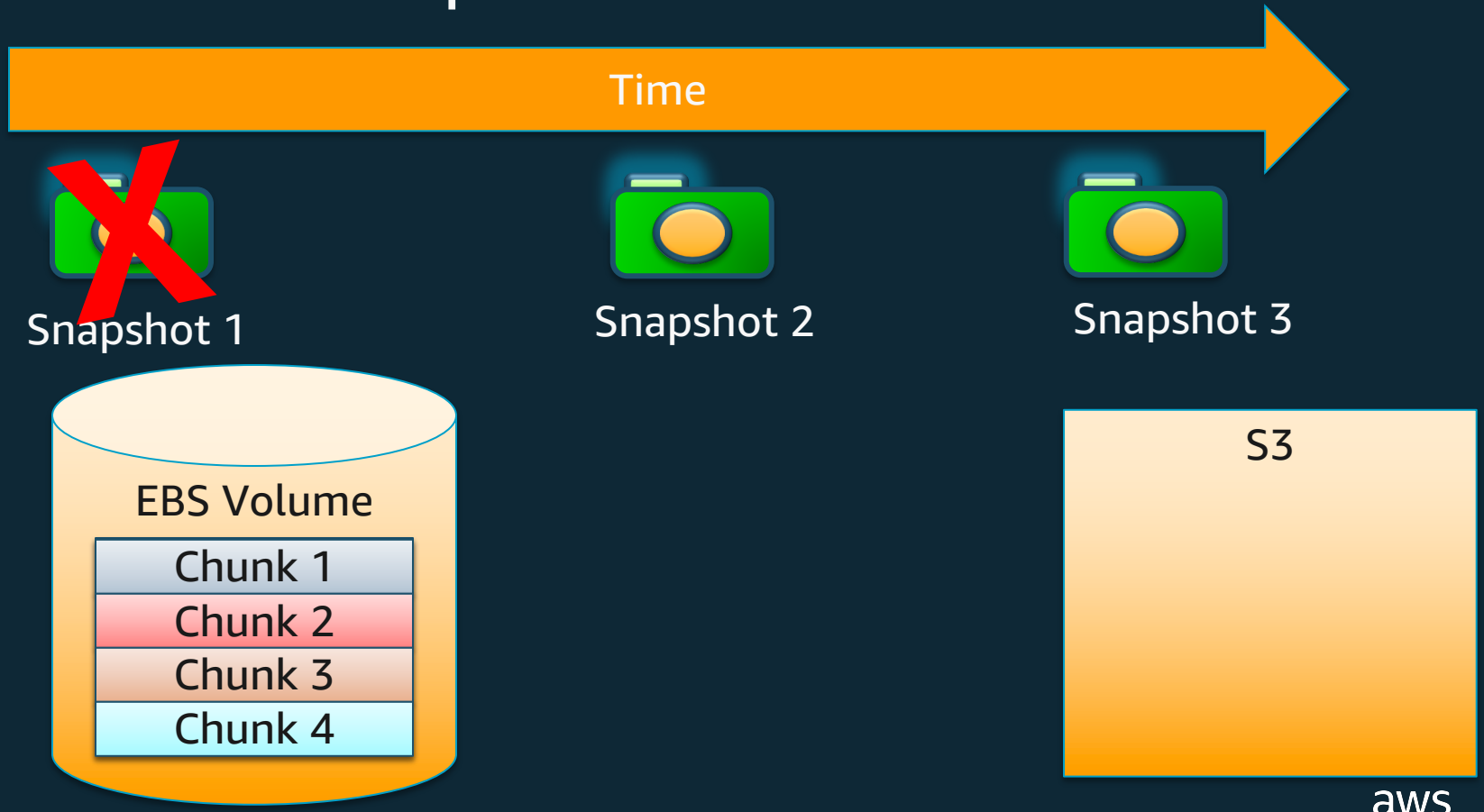
Fully managed backup with EBS Snapshots

Low cost
Incremental backups do not duplicate data and reduce storage costs

Protection
Snapshots are stored in Amazon S3

Agility
Quickly restore volumes across Availability Zones within a region

How Do EBS Snapshots Work?



What is Amazon EC2 instance store?

- Local to instance
- Non-persistent data store
- Available on several EC2 families
- Data is not replicated (by default)
- No snapshot support
- SSD or NVMe

Shared file system

Amazon Elastic File System

Amazon Elastic File System (Amazon EFS)

Simple, serverless, set-and-forget, elastic file system for AWS compute

Serverless shared storage

Serverless and scalable

No provisioning, scale capacity, connections, and IOPS

Full AWS compute integration

EC2 Instances, containers, and serverless

Supports 10,000s of connections

Simple and highly reliable

Elastic

Pay only for capacity used

Performance built-in, scales with capacity

Highly durable and available

Designed for 11 9s of durability

99.99% availability SLA

Performant and cost-optimized

Performant

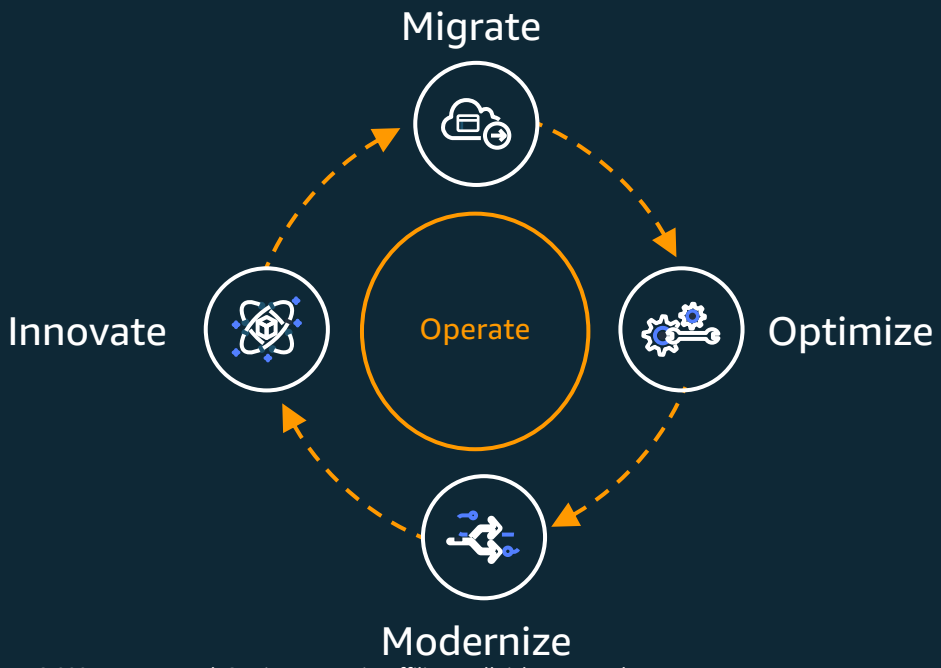
10s of GB/s of throughput and 500,000+ IOPS

Four storage classes

Automatic lifecycle-based cost optimization

aws

Amazon EFS meets you where you are today and tomorrow



Migrate: lift and shift to AWS cloud without refactoring application

Optimize: enable cost efficiency

Modernize: build micro-services into application with common data platform

Innovate: improve development efficiency, build new features, enter new markets

aws

Use cases for Amazon EFS

Home directories

DevOps

Application dev. & test

Enterprise apps

Database backups

Web serving & content mgmt.

Analytics

Machine learning

Media workflows

Business Criticality

Metadata-intensive jobs

Scale-out jobs

Low latency and serial I/O

High throughput and parallel I/O

aws

Automatic cost optimization

Using EFS storage classes and lifecycle management

$0.043/GB-Month*
Effective storage cost

$0.08/GB-Month*
Effective storage cost

EFS One Zone

$0.16/GB-month*

EFS Standard

$0.30/GB-month*

EFS One Zone-IA

Cost-optimized for less accessed files
$0.01333/GB-month* for storage
$0.01/GB* for access

EFS Standard-IA

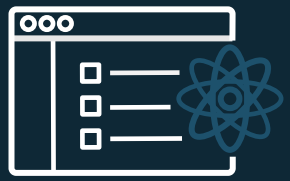
Cost-optimized for less accessed files
$0.025/GB-month* for storage
$0.01/GB* for access

*Pricing in the US East (N. Virginia) Region. Assumes 80% of the files are infrequently accessed

aws

Amazon FSx

Amazon FSx for Windows File Server



Fully managed file storage built on Windows Server

Easy migration to AWS

Fully managed Windows file storage means you no longer have to …

Managed hardware

Plan capacity

Procure and purchase hardware

Set up storage servers and volumes

Detect and address hardware failures

Incur high upfront costs

Managed software

Install and configure server software

Set up and configure file systems

Apply Windows updates

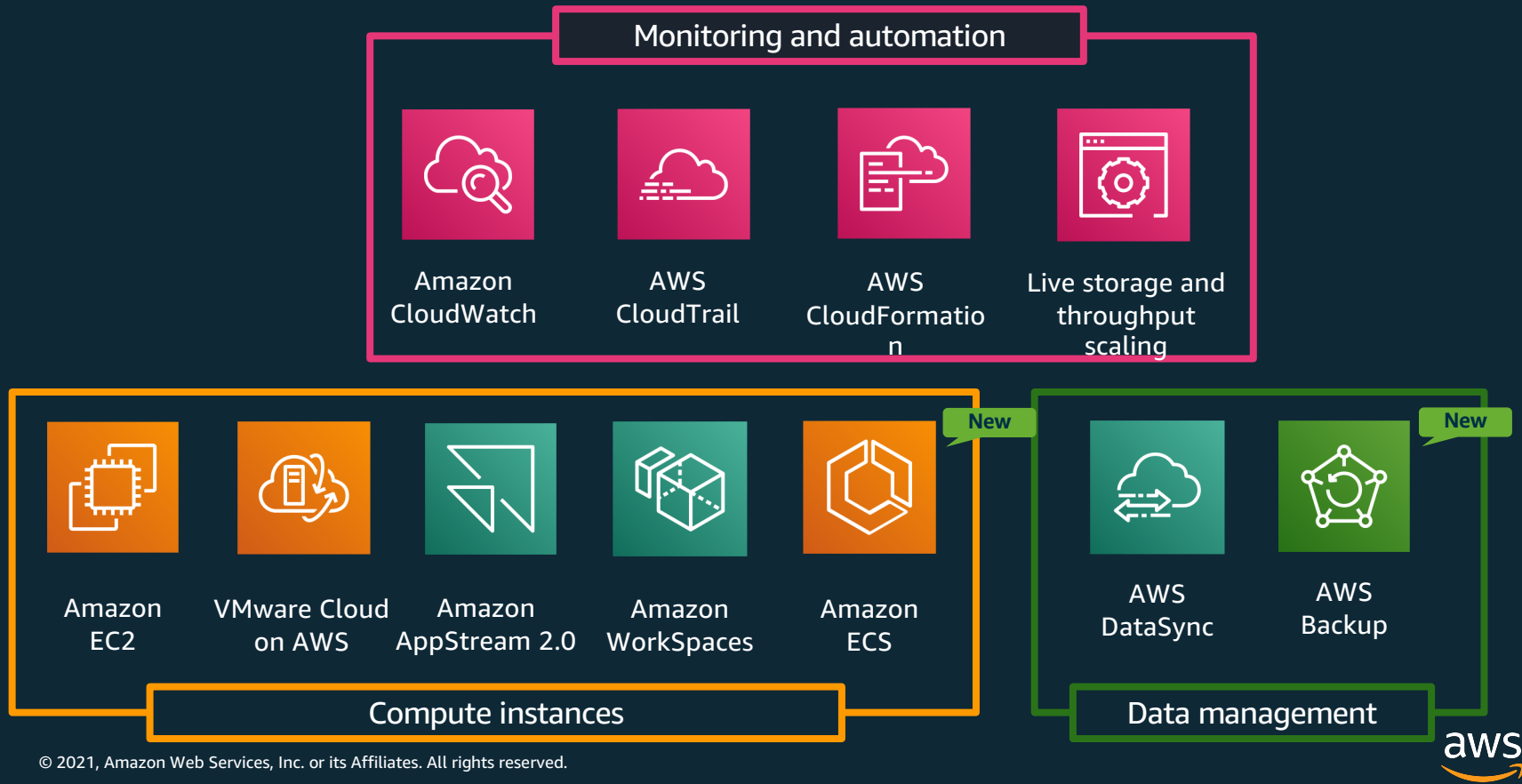
Manage software licenses

Manage backups

Monitor security

aws

Agile, scalable, and automated



Flexible price and performance options

Storage type flexibility (SSD / HDD)

Deployment type flexibility (Single-AZ / Multi-AZ)

Select throughput and storage independently

Choice of in-line snapshots and backups stored in S3

Data deduplication and compression

aws

Fully featured, secure, reliable, and scalable

Accessibility

- ✓ Full SMB protocol support
- ✓ Windows Server 2008+, Windows 7+, Linux, and MacOS
- ✓ EC2, WorkSpaces and AppStream 2.0
- ✓ VMware Cloud on AWS
- ✓ Amazon ECS and Amazon EKS containers
- ✓ Cross-VPC / Account / Region access
- ✓ On-premises access (DirectConnect / VPN)

Administration

- ✓ Active directory integration
- ✓ Managing file shares
- ✓ Monitoring user sessions and open files
- ✓ Restoring locked files
- ✓ User storage quotas
- ✓ Monitoring actions via AWS CloudTrail

Availability and durability

- ✓ High availability – automatic recovery
- ✓ High durability – automatic replication
- ✓ Multi-AZ deployment option
- ✓ SMB continuous availability (CA)

Performance and scale

- ✓ Consistent, sub-millisecond latencies
- ✓ PB-scale storage scalability
- ✓ Tens of GB/s throughput scalability
- ✓ Millions of IOPS scalability
- ✓ Select throughput and storage independently
- ✓ Server-side and client-side caching
- ✓ SMB Multichannel
- ✓ Performance monitoring via CloudWatch
- ✓ Live scaling of throughput capacity

Cost optimization

- ✓ Storage type flexibility (SSD / HDD)
- ✓ Deployment type flexibility (Single-AZ / Multi-AZ)
- ✓ Live scaling of storage capacity
- ✓ Data deduplication and compression

Data protection

- ✓ Snapshots (with end-user file restore)
- ✓ Backups

Security and compliance

- ✓ Encryption at rest and in transit
- ✓ Kerberos authentication
- ✓ Access controls via NTFS ACLs, share ACLs, VPC, and IAM
- ✓ PCI DSS, ISO, SOC, GDPR, IRAP, and HIPAA compliances

aws

Object Stores

Amazon S3

Amazon S3 (Simple Storage Service)

- Web accessible object store (through API or HTTPS)
- Highly durable (99.999999999% design)
- Limitlessly scalable
- Multiple Tiers to match your workload
- Data Lifecycle Rules
- Static Website Hosting
- Security, Compliance, and Audit capabilities
- Standard Storage Pricing (us-east-1) - $0.023 per GB

aws

Your choice of object storage classes

S3 Standard

- Active, frequently accessed data
- Milliseconds access
- ≥ 3 AZ
- $0.0210/GB

S3 Intelligent-Tiering

- Data with changing access patterns
- Milliseconds access
- ≥ 3 AZ
- $0.0210 to $0.0125/GB ($0.004 to $0.00099/GB Archive)
- No retrieval fees
- Monitoring fee per Obj.
- Min storage duration
- Min object size

S3 Standard-IA

- Infrequently accessed data
- Milliseconds access
- ≥ 3 AZ
- $0.0125/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

S3 One Zone-IA

- Re-creatable, less accessed data
- Milliseconds access
- 1 AZ
- $0.0100/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

S3 Glacier

- Archive data
- Select minutes or hours
- ≥ 3 AZ
- $0.0040/GB – ($4.10/TB)
- Retrieval fee per GB
- Min storage duration
- Min object size

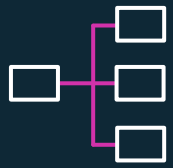
S3 Glacier Deep Archive

- Archive data
- Select 12 or 48 hours
- ≥ 3 AZ
- $0.00099/GB - ($1.01/TB)
- Retrieval fee per GB
- Min storage duration
- Min object size

Frequent ← *Access Frequency* → *Infrequent*

aws

S3 Management Features



Organize

S3 Tagging

S3 Prefixes

S3 Versioning

Monitor

CloudWatch

CloudTrail

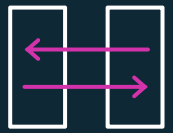
S3 Event Notifications

S3 Inventory

S3 Glacier Restore Notifications

S3 Storage Lens

AWS Config



Replicate & Tier

S3 Lifecycle

S3 Storage Class Analysis

S3 Intelligent-Tiering

Cross-Region Replication

Replication Time Control (RTC)

Modify

S3 Event Notifications + Lambda

S3 Batch Operations

S3 Object Lock

S3 Object Lambda

aws

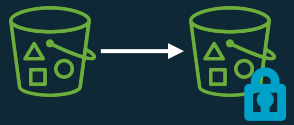
S3 Access Management & Security

- Deep integration with AWS Identity and Access Management (IAM)
- Access Control Lists (ACLs), S3 bucket policies, and S3 Access Points
- Query String Authentication
- Audit Logs
- S3 supports both server-side & client-side encryption
- S3 Block Public Access to ensure S3 buckets and objects do not public access
- Amazon Macie to discover, classify, and protect sensitive data stored in Amazon S3
- Access Analyzer for S3
- Amazon S3 Object Lock
- AWS PrivateLink for S3
- Amazon GuardDuty for S3

S3 Replication

Replicate within the same AWS Region

Replicate to a bucket with retention controls (in the same or different AWS Region)



Replication Time Control (RTC) 15 minute replication SLA

Amazon Simple Storage Service

Replicate to a different AWS Region

Replicate faster to a different AWS Region, backed by an SLA + replication metrics

aws

Backup

AWS Backup

AWS Backup – meeting the challenges

Backup operations unified across AWS services

Complexity

Simple & Performant

Policy- and tag-based backup solution

Automated backup scheduling

Compliance

Reliable & Secure

Centralized backup activity monitoring and logs

Backup encryption

Backup access policies

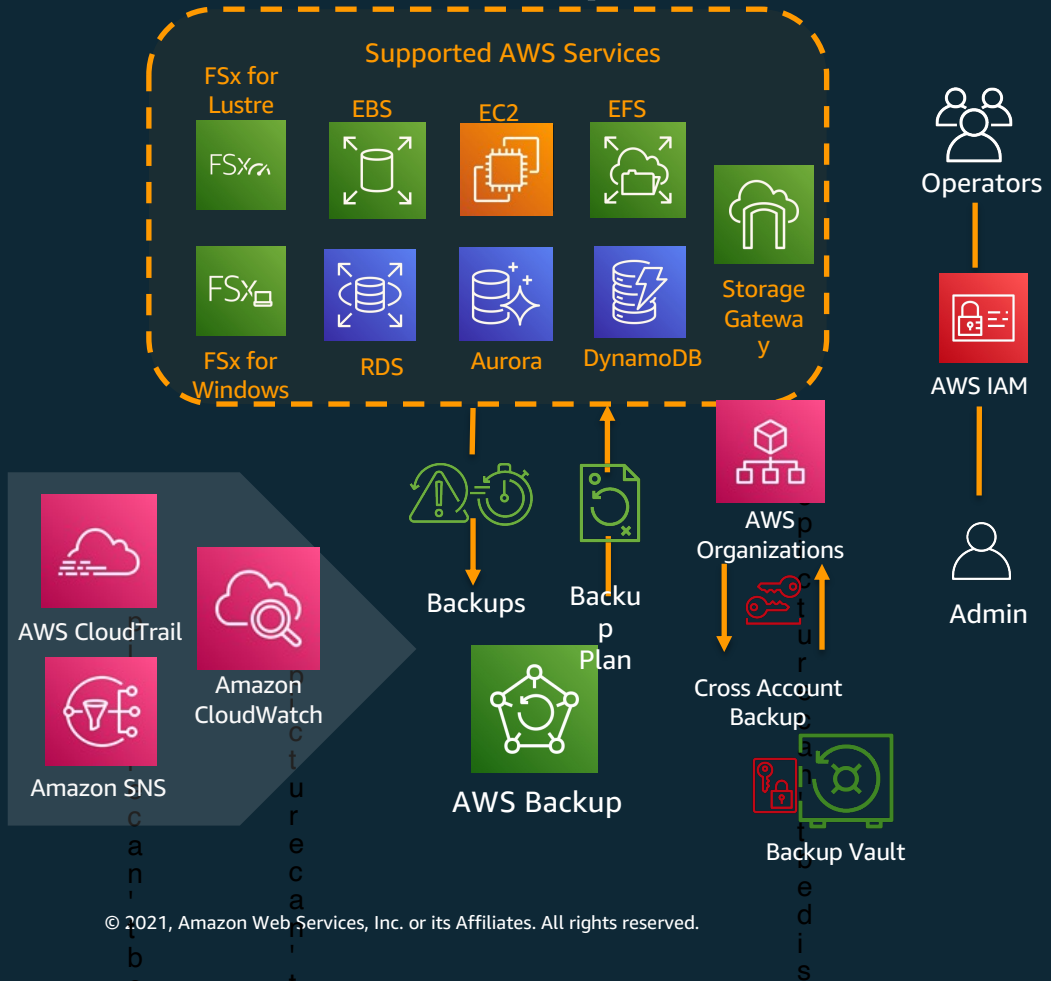
Cost

Cost Effective

Automated backup retention management

No added cost for orchestration

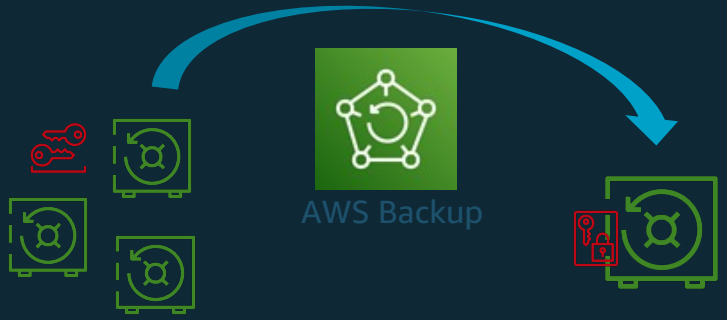
How AWS Backup works



With AWS Backup, you can:

- Create backup plans that orchestrate backup operations per policy
- Secure backups and manage restore processes
- Control access for managed resources via IAM
- Setup notifications/logging via Amazon SNS, AWS CloudTrail & AWS CloudWatch
- Scale through AWS Organizations
- Protect EC2 Windows/SQL with app consistency
- Backup to other accounts to protect against multiple threats

DR & Ransomware Recovery with AWS Backup



Vault characteristics:

- Backups are highly efficient incremental forever
- Backup copies cannot be changed or encrypted
- Manage with vault specific CMK/KMS best practices
- Air-gapped backups using vault access policies
- Prescriptive guidance for vault account access provided

Recovery options:

- Supports 1-to-many, many-to-many, many-to-1, etc.
- Recover from same account locally or from across region
- Recover from cross-account locally or across region
- Recover from RPOs that are hours, days, weeks or months old
- Simple workflow to apply any forensic analysis

Any Questions?