



Solutions Architect Professional

Storage on AWS



Agenda

- Introduction
- Storage Primer
- Block Storage
- Shared File Systems
- Object Store
- On-Premises Storage Integration

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

Block vs File vs Object



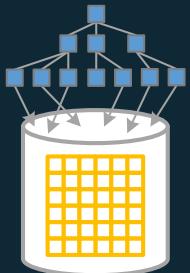
Block Storage

Raw Storage

Data organized as an array of unrelated blocks

Host File System places data on disk

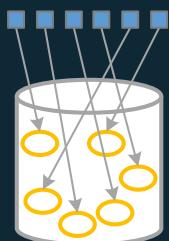
e.g.: Microsoft NTFS, Unix ZFS



File Storage

Unrelated data blocks managed by a file (serving) system

Native file system places data on disk



Object Storage

Stores Virtual containers that encapsulate the data, data attributes, metadata and Object IDs

API Access to data

Metadata Driven, Policy-based, etc

Storage - Characteristics

Some of the ways we look at storage

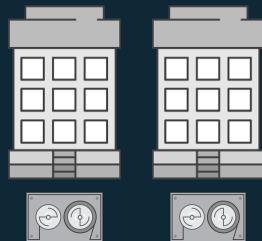
Durability	Availability	Security	Cost	Scalability	Performance	Integration
Measure of expected data loss	Measure of expected downtime	Security measures for at-rest and in-transit data	Amount per storage unit, e.g. \$ / GB	Upward flexibility, storage size, number of users	Performance metrics (bandwidth)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

Understanding Durability



Two copies on one site

designed for
99.99%
durability



Copies on two sites

designed for
99.999%
durability



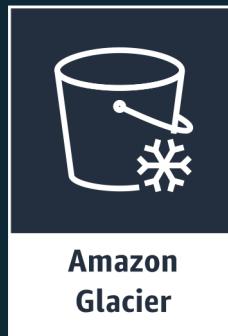
2 copies in each of three AZ

designed for
99.999999999%
durability

AVAILABILITY VS. DURABILITY

%	Availability (Projected Downtime)	Durability (Projected loss)
99.999	5 minutes 15 seconds	1 in 100,000
99.9999	31 seconds	1 in 1,000,000
99.99999	3 seconds	1 in 10,000,000
99.999999999	300 uSeconds	1 in 100,000,000,000

AWS has a variety of storage options



More choice for more applications

Block storage volume types

- General Purpose SSD
- Provisioned IOPS SSD
- Throughput-Optimized HDD
- Cold HDD
- Elastic Volumes



Backup

AWS Backup^{NEW!}



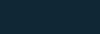
Re-host

Amazon EC2



FSx_a

Amazon FSx
for Lustre



Re-platform



File storage classes

EFS Standard

EFS Infrequent Access^{NEW!}



Object storage classes

S3 Standard

S3 Standard-IA

S3 One Zone-IA

S3 Glacier

S3 Intelligent-Tiering^{NEW!}

S3 Glacier Deep Archive^{NEW!}



Re-architect

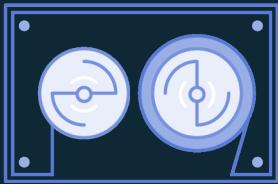


What are customers building?



Backup & Restore

Non-disruptive
Easy place to start
Integrated with all major vendors



Archive & Compliance

Media workflows
Tape replacement
Public Sector, FinServ, Healthcare/Life Sciences



Home Directories

Simple to move
Less latency sensitive
Significant cost savings



Data Lakes

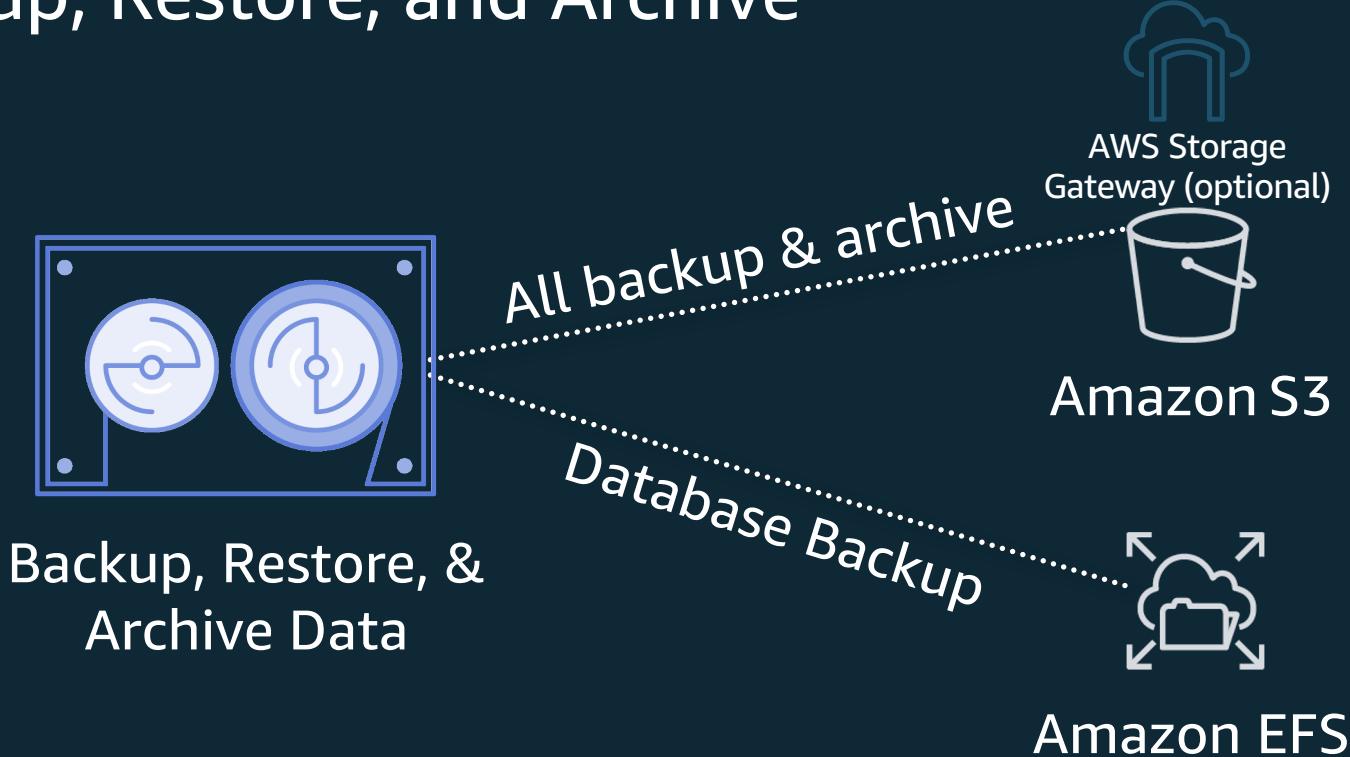
Variety of analytics tools
Foundation for AI/ML
Built for streaming data
Data visualization



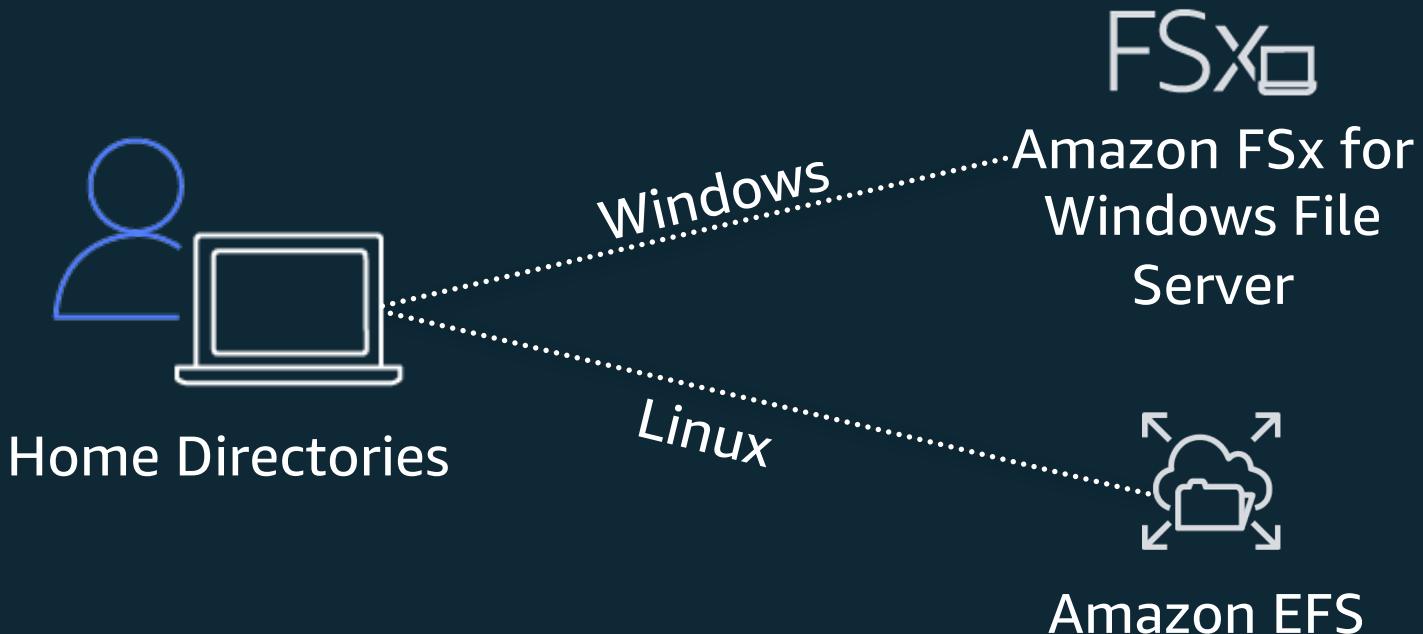
Business-Critical Applications

Integrated with major vendors
Fully managed infrastructure
Lift-and-shift migrations

Backup, Restore, and Archive



Home Directories

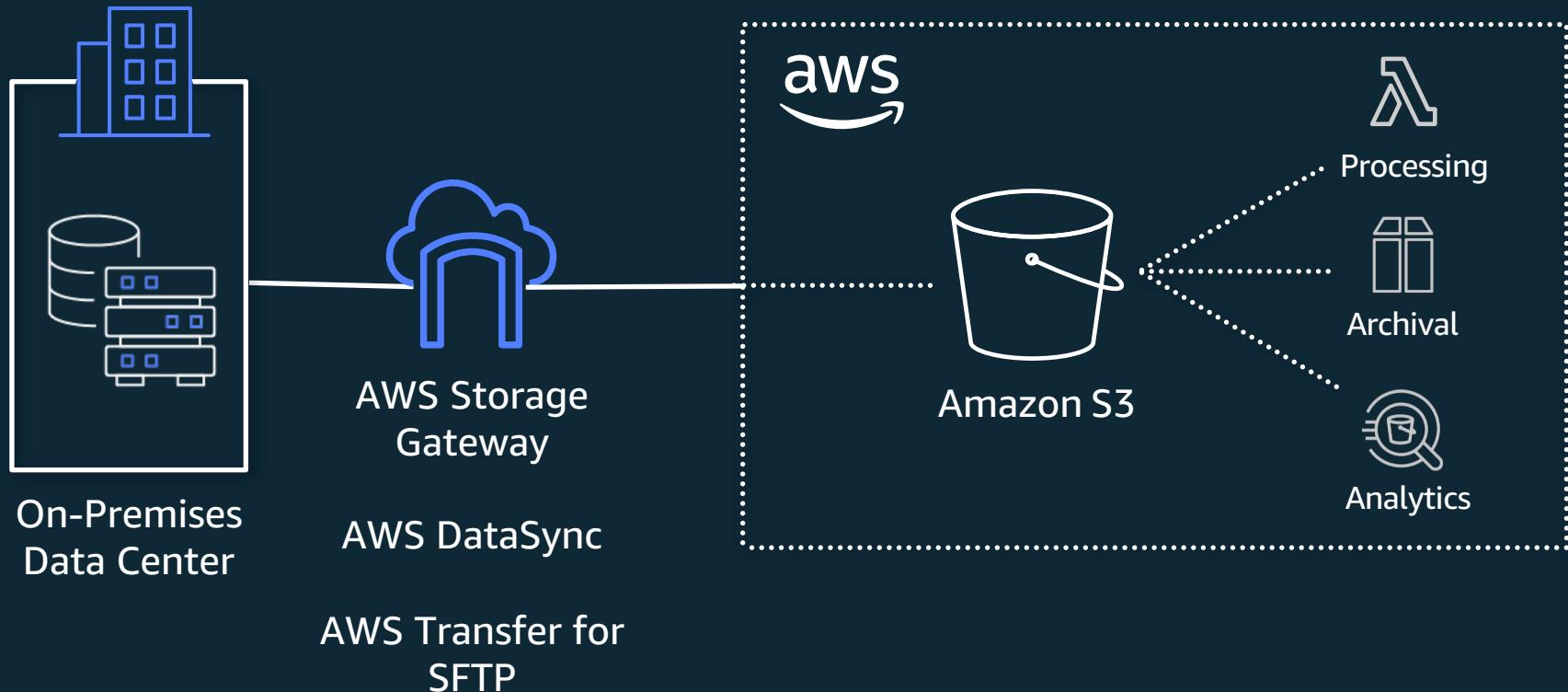


Business-Critical Applications



Block Storage	Linux/NFS File Storage	Windows File Storage
Amazon EBS	Amazon EFS	Amazon FSx for Windows File Server

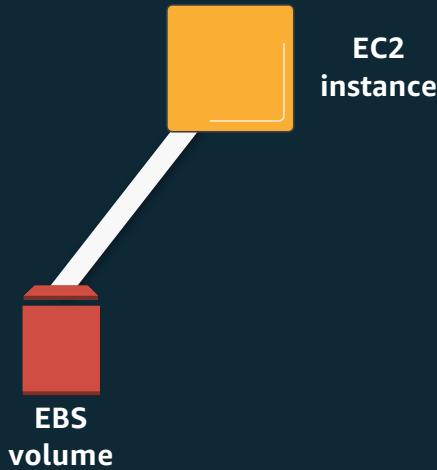
Hybrid Cloud Storage



1

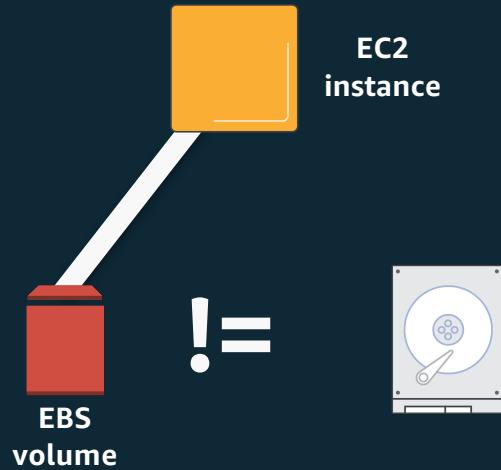
Block Storage

What is Amazon EBS?



- Block storage as a service
- Create, attach volumes through an API
- Service accessed over the network

What is Amazon EBS?



- Block storage as a service
- Create, attach volumes through an API
- Service accessed over the network (SAN like)

AWS EBS Features

Durable

Designed for 99.99 reliability

Redundant storage across multiple devices within an AZ

Secure

Identity and Access Policies

Encryption

Scalable

Capacity when you need it

Easily scale up and down

Performance

Low-latency SSD

Consistent I/O Performance

Stripe multiple volumes for higher I/O performance

Backup

Point-in-time Snapshots

Copy snapshots across AZ and Regions

Amazon EBS

Network attached block device

- Independent data lifecycle
- Multiple volumes per EC2 instance
- **Only one EC2 instance at a time per volume**
- Can be detached from an instance and attached to a different one

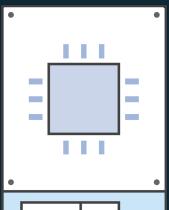
Raw block devices

- Unformatted block devices
- Ideal for databases, filesystems

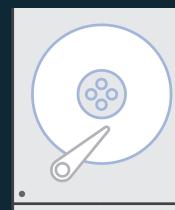
Multiple Drive Types

- SSD (iops) and Magnetic (throughput)

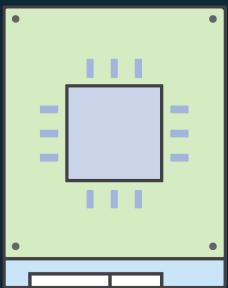
Amazon EBS volume types



SSD

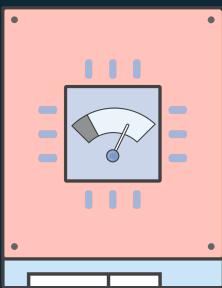


HDD



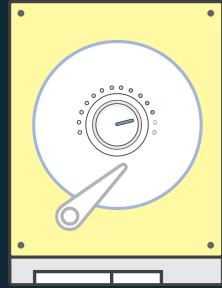
gp2

General Purpose
SSD



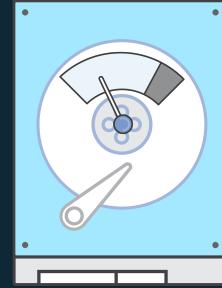
io1

Provisioned IOPS
SSD



st1

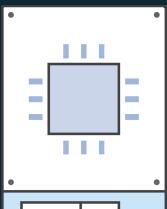
Throughput Optimized HDD



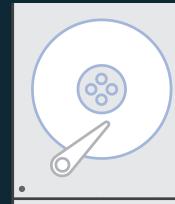
scl

Cold HDD

Amazon EBS use cases



SSD



HDD



Relational Databases

MySQL, SQL Server,
PostgreSQL, SAP,
Oracle



NoSQL Databases

Cassandra, MongoDB,
CouchDB



Big Data , Analytics

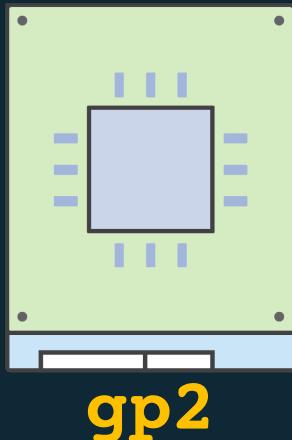
Kafka, Splunk,
Hadoop, Data
Warehousing



File / Media

CIFS/NFS,
Transcoding,
Encoding, Rendering

Amazon EBS volume types: General Purpose SSD



General Purpose SSD

Baseline: 100 to 16,000 IOPS; 3 IOPS per GiB

Burst: 3,000 IOPS (for volumes up to 1,000 GiB)

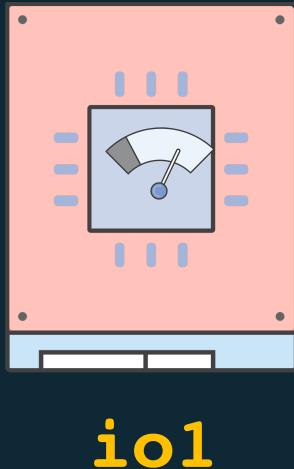
Throughput: Up to 250 MiB/s

Latency: Single-digit ms

Capacity: 1 GiB to 16 TiB

Great for boot volumes, low-latency applications, and bursty databases

Amazon EBS volume types: Provisioned IOPS



Provisioned IOPS

Baseline: 100–64,000 IOPS

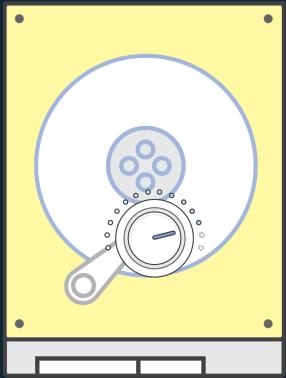
Throughput: Up to 1,000 MiB/s

Latency: Single-digit ms

Capacity: 4 GiB to 16 TiB

Ideal for critical applications and databases with sustained IOPS

Amazon EBS volume types: Throughput Provisioned



Baseline: 40 MiB/s per TiB up to 500 MiB/s

Burst: 250 MiB/s per TiB up to 500 MiB/s

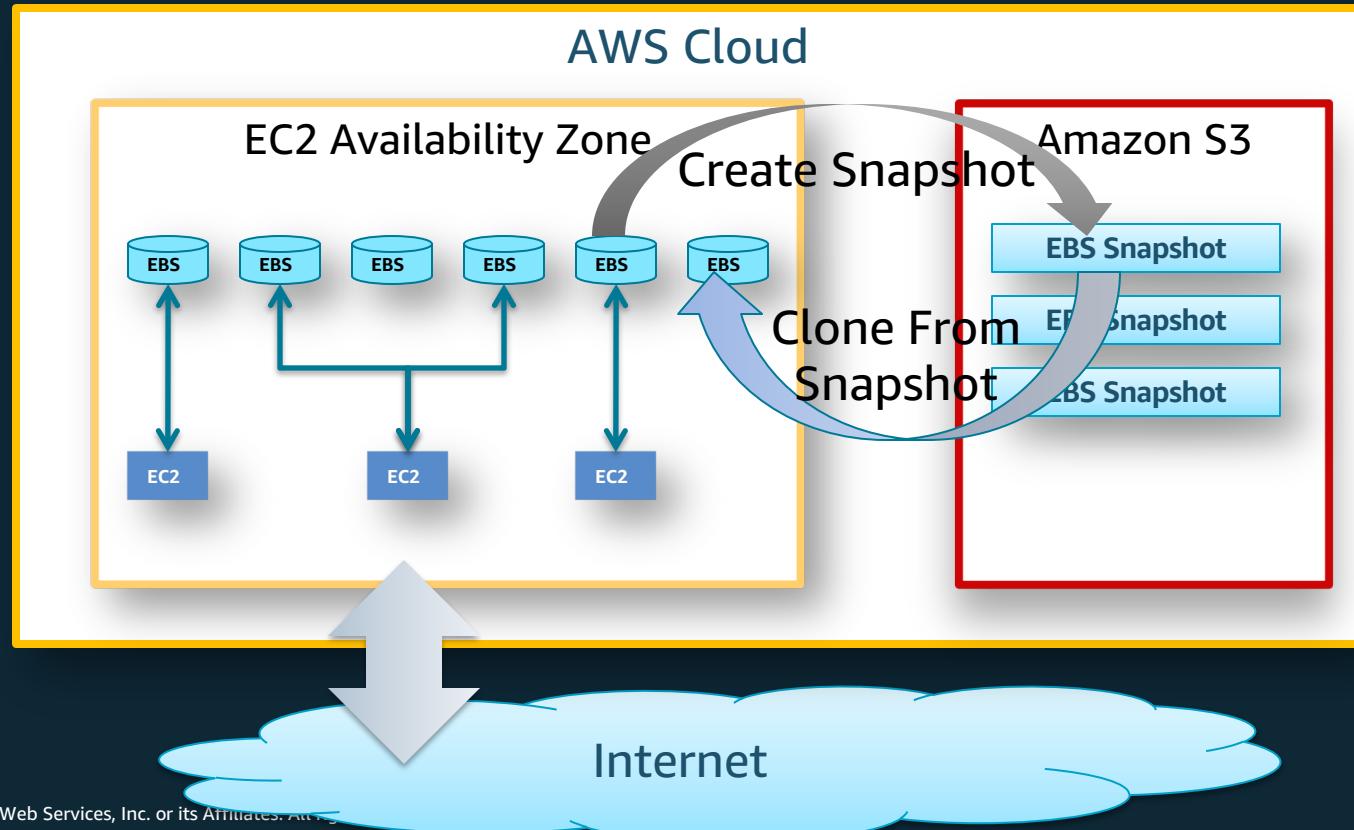
Capacity: 500 GiB to 16 TiB

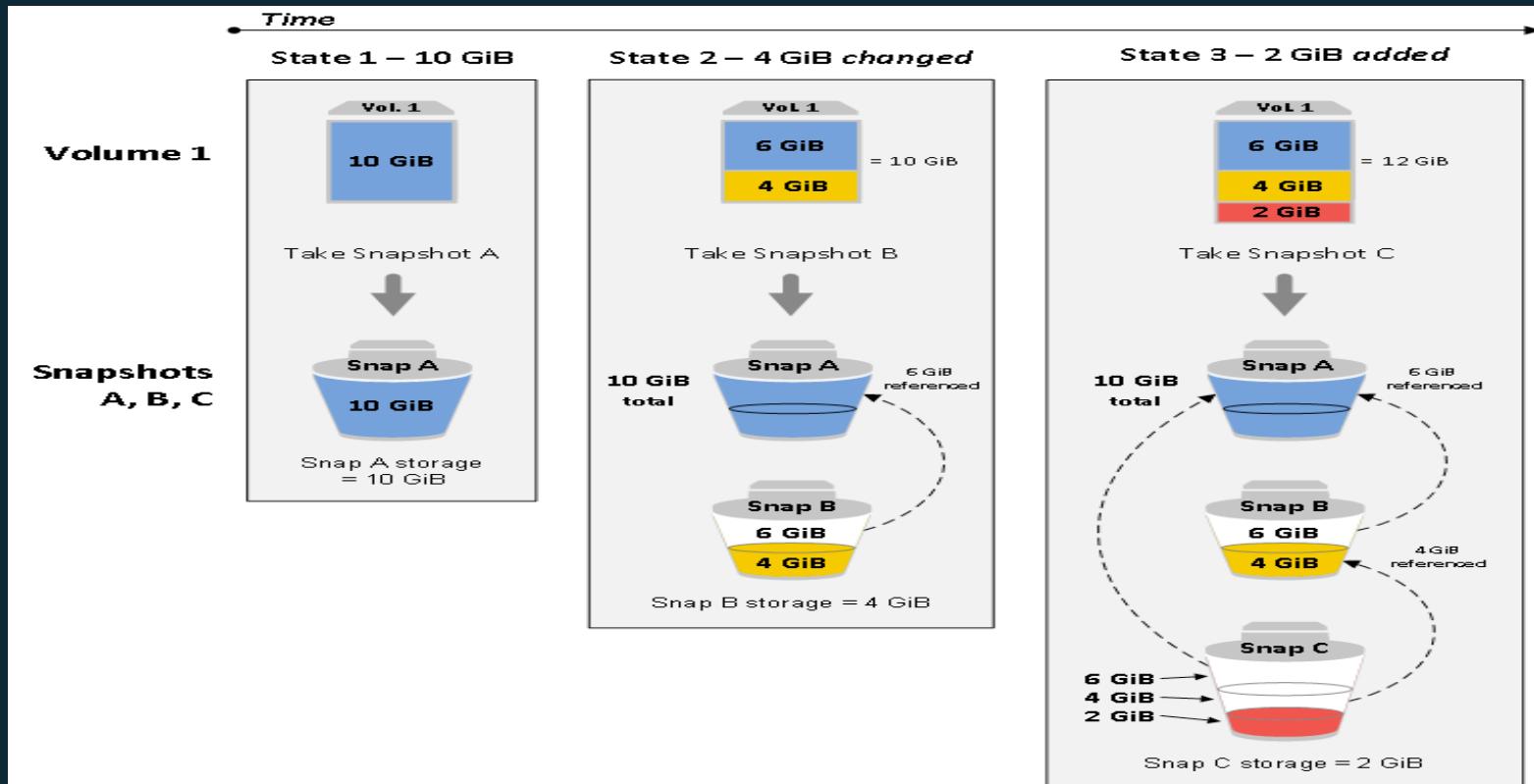
Ideal for large-block, high-throughput sequential workloads

st1

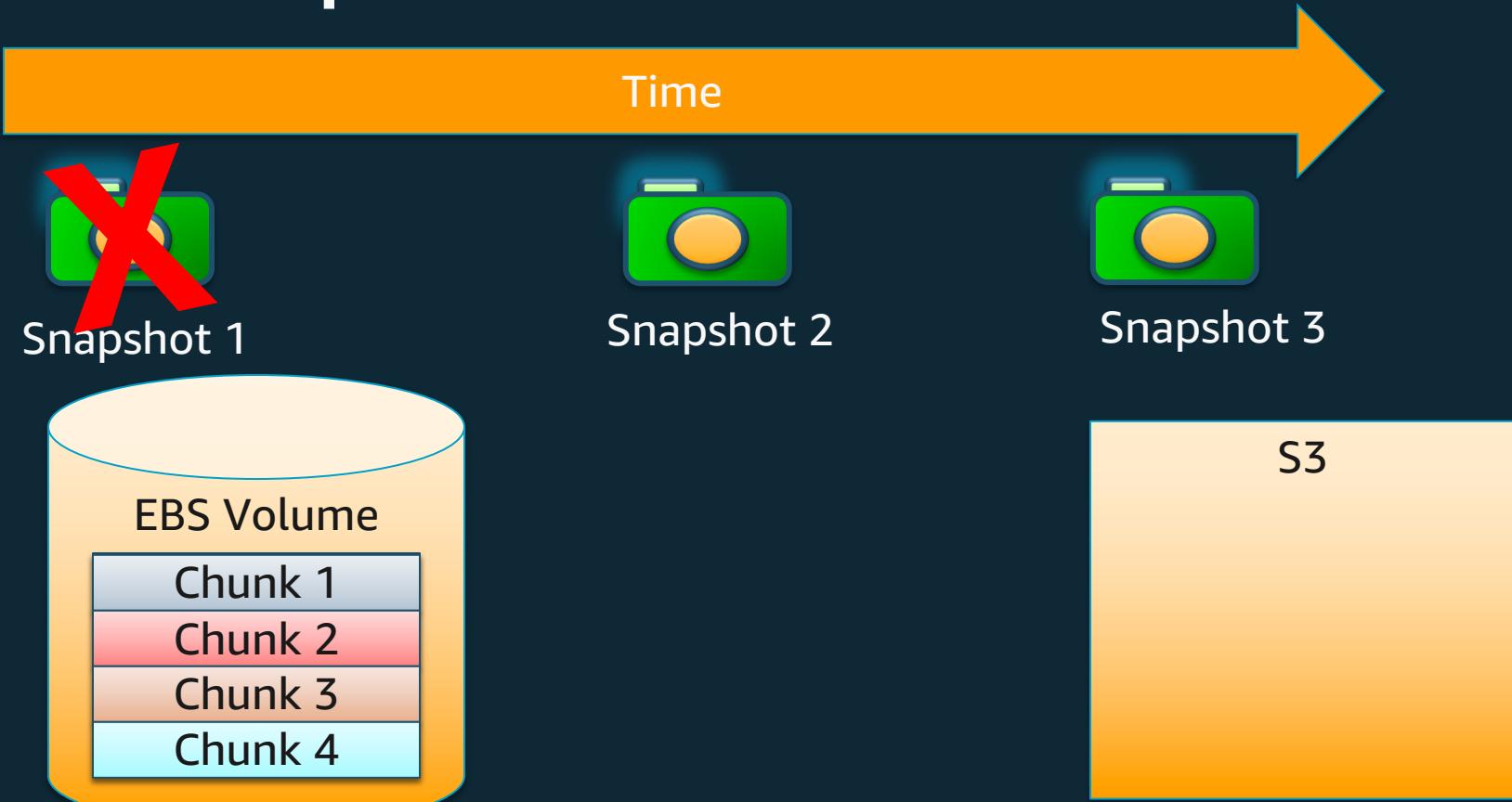
Throughput Optimized HDD

EBS Snapshots





How Do Snapshots Work?



What is Amazon EC2 instance store?

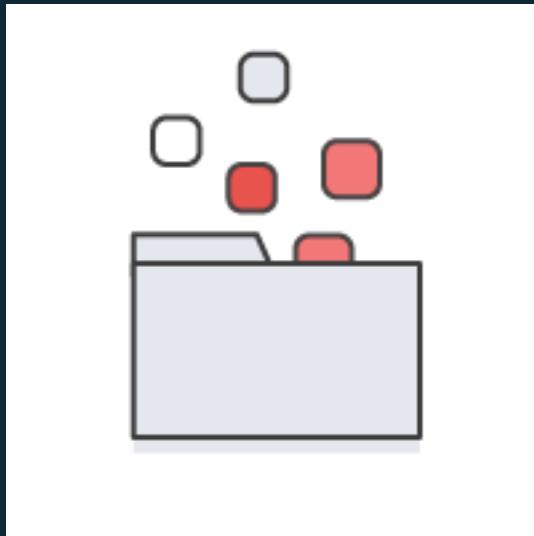


- Local to instance
- Physically attached disk
- Ideal for temporary storage
- Non-persistent data store
(Data is lost: instance stops or terminates; or when disk fails)
- Data not replicated (by default)
- No snapshot support
- SSD or HDD
- >80,000 iops
- >1,750 MB/S



Shared file system

Amazon EFS is Simple



Fully managed

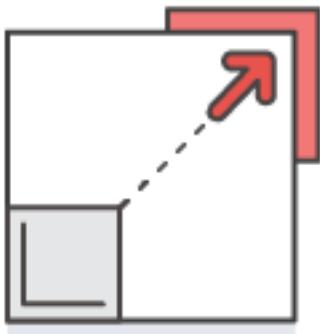
- No hardware, network, file layer
- Create a scalable file system in seconds!

Seamless integration with existing tools and apps

- NFS v4.1—widespread, open
- Standard file system access semantics
- Works with standard OS file system APIs

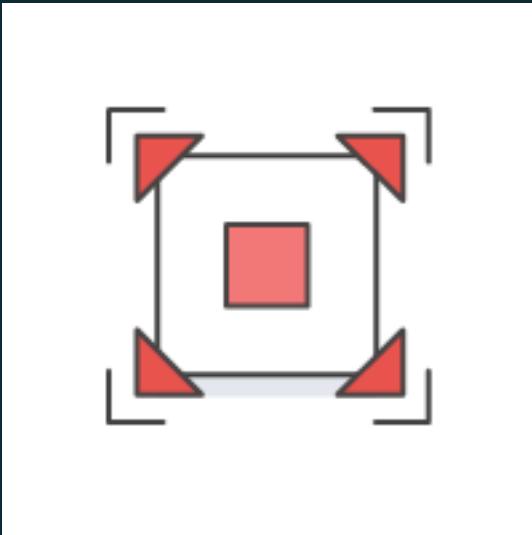
Simple pricing = simple forecasting

Amazon EFS is Elastic



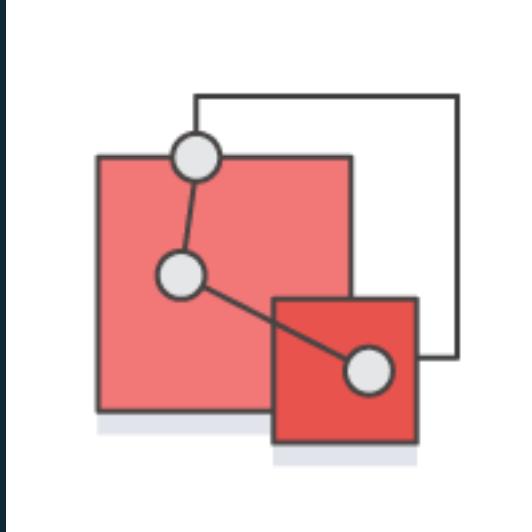
- File systems grow and shrink automatically as you add and remove files
- No need to provision storage capacity or performance
- You pay only for the storage space you use, with no minimum fee

Amazon EFS is Scalable



- File systems can grow to petabyte scale
- Throughput and IOPS scale automatically as file systems grow
- Consistent low latencies regardless of file system size
- Support for thousands of concurrent NFS connections

Highly Durable and Highly Available



- Designed to sustain AZ offline conditions
- Resources aggregated across multiple AZ's
- Superior to traditional NAS availability models
- Appropriate for Production / Tier 0 applications

Example use cases

Big Data Analytics

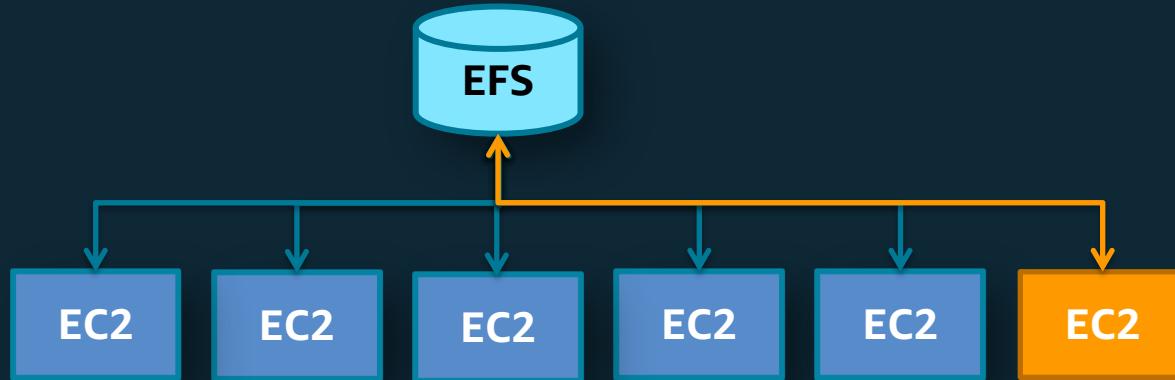
Media Workflow Processing

Web Serving

Content Management

Home Directories

EFS – Mounting



EFS DNS Name

availability-zone.file-system-id.efs.aws-region.amazonaws.com

Mount on machine

```
sudo mount -t nfs4 mount-target-DNS:/ ~efs-mount-point
```

FSx for Windows



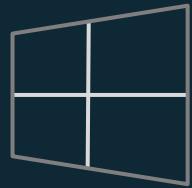
Fully managed Windows
file systems ...

... built on Windows
Server



Integrated with
AWS

Native Windows compatibility and features



Native Windows compatibility



NTFS



Native SMB*
2.0 to 3.1.1



Integrates with
Microsoft AD
and supports
Windows ACLs



Distributed File
System (DFS)
Namespaces
and
DFS Replication
(Multi-AZ)



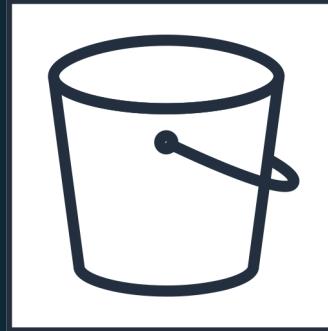
Windows Server



Object Stores

Amazon S3 (Simple Storage Service)

- Web accessible object store (through API or HTTPS)
- Highly durable (99.99999999% design)
- Limitlessly scalable
- Multiple Tiers to match your workload
- Data Lifecycle Rules
- Static Website Hosting



Amazon Simple Storage Service (S3)



Collect

Move Data via API, HTTPS,
SDK

Multiple Encryption Options

Automated cost reduction
tools



Store

Designed for
99.99999999% durability

Parallel I/O for Max Speed

Replication options across
regions



Analyze

On-demand analytics

Built-in support for SQL
expressions with S3 Select

Detailed data on usage
patterns and access

Object storage classes



Standard



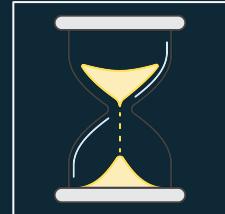
Infrequent Access



1 Zone - IA



Intelligent



Glacier

Active data
Millisecond access
Min 3 AZs
\$0.023

30 day min duration
Millisecond access
Min 3 AZs
\$0.0125

30 day min duration
Millisecond access
Min 1 AZ
\$0.01

ML to optimize
Storage costs
Min 3 AZs
*scan cost

Archive data
Minutes to Hours
\$0.001 - \$0.004
Min 3 AZs

Pricing is per GB per month in the US East (N. Virginia) region

Amazon Glacier



Secure

Regulatory compliance
certifications

Vault Lock

Locking, encryption,
audit and alerting tools



Archive

Designed for
99.999999999%
durability

Replication options
across regions



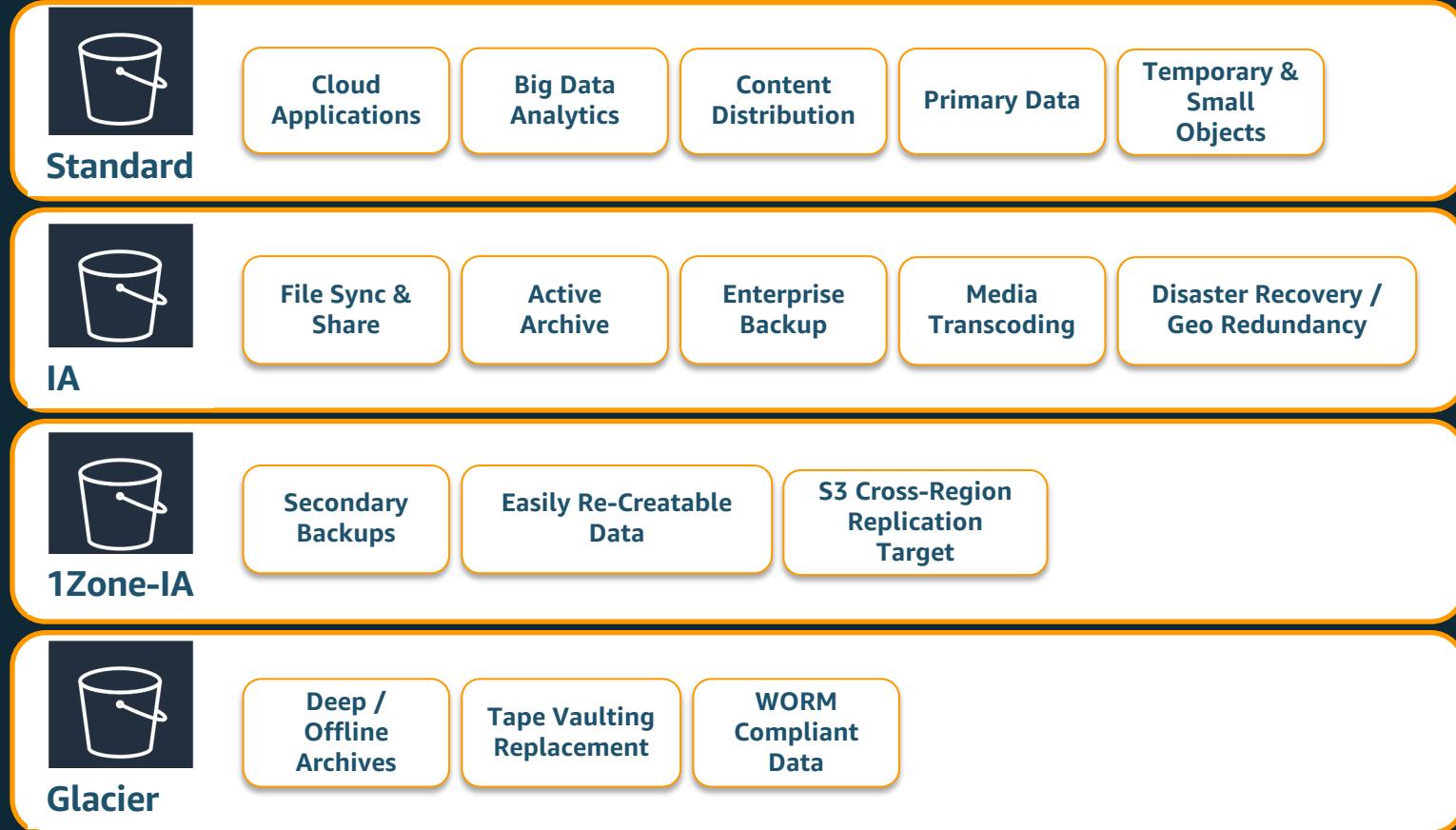
Cost-effective

Query-in-place
analytics

Expedited and bulk
retrievals

Deep Archive

Object Storage Use Cases



Storage Tiered To Your Requirements



S3

"Hot" Data
Active and/or
Temporary Data

Starts at \$0.023 / GB per month

> 0K

≥ 0 Days



S3-IA

"Warm" Data
Infrequently
Accessed Data

\$0.0125 / GB per month

≥ 128K

≥ 30 Days

\$0.01/GB retrieval



Glacier

"Cold" Data
Archive and
Compliance Data

\$0.001 / GB per month
\$0.004 / GB per month

> 0K

≥ 90 Days

1-5 mins

3-5 hrs

5-12 hrs

Expedited
\$0.03 / GB

Standard
\$0.01 / GB

Bulk
\$0.0025 / GB

Lifecycle

Durable
99.99999999%

Available
S3: 99.99%
S3-IA: 99.9%

Performant
Low Latency
High Throughput

Scalable
Elastic capacity
No preset limits



S3 Storage Management Features

S3 Object Tagging

manage and control access for Amazon S3 objects.

S3 Analytics, Storage Class Analysis

Analyze storage access patterns and transition the right data to the right storage class.

S3 Inventory

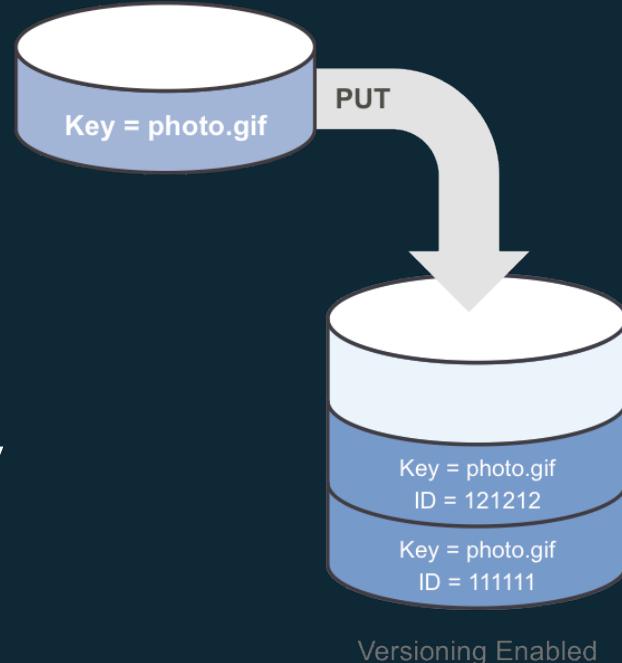
Simplify and speed up business workflows and big data jobs

S3 CloudWatch Metrics

Understand and improve the performance of your applications that use S3

S3 versioning

- Preserve, retrieve, and restore every version of every object stored in your bucket
- S3 automatically adds new versions and preserves deleted objects with delete markers
- Easily control the number of versions kept by using lifecycle expiration policies
- Easy to turn on in the AWS Management Console

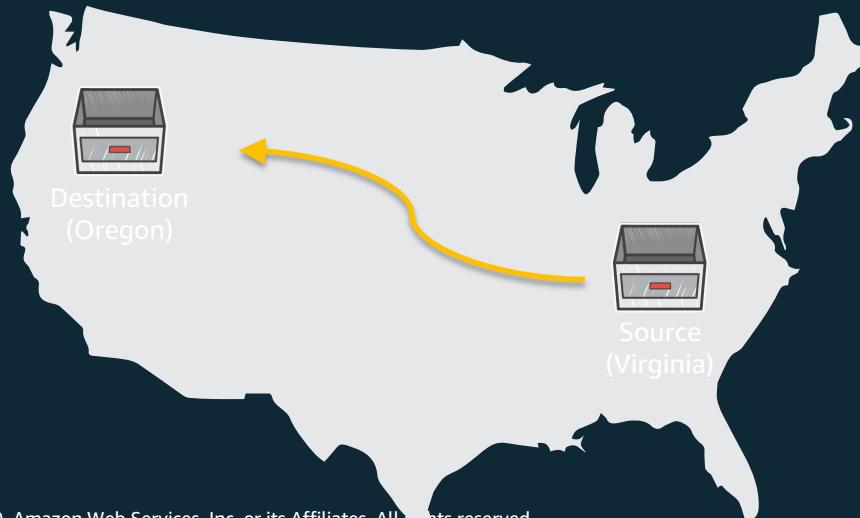


S3 cross-region replication

Automated, fast, and reliable asynchronous replication of data across AWS regions

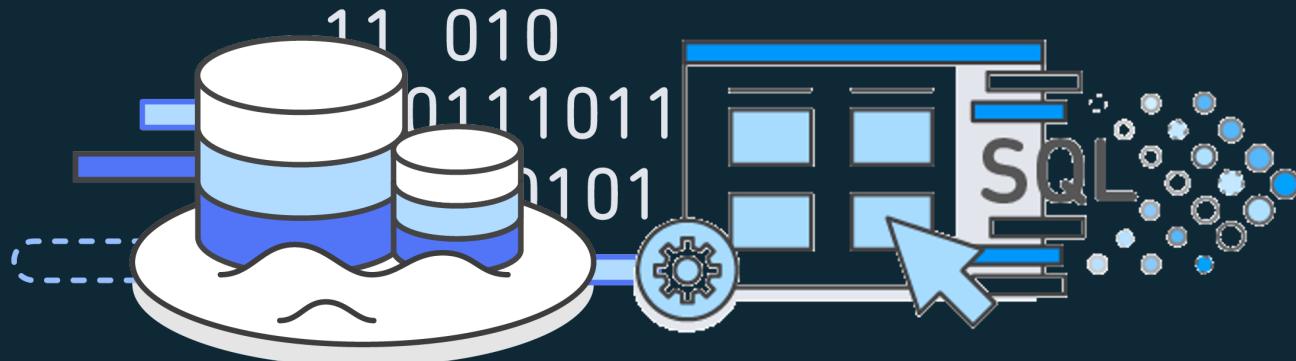
Use cases:

- Compliance—store data hundreds of miles apart
- Lower latency—distribute data to regional customers)
- Security—create remote replicas managed by separate AWS accounts



- Only replicates new PUTs. Once S3 is configured, all new uploads into a source bucket will be replicated
- Entire bucket or prefix based
- 1:1 replication between any 2 regions
- Versioning required

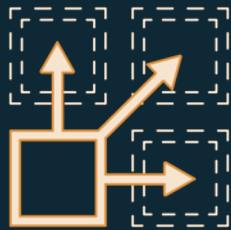
Amazon S3 Select and Amazon Glacier Select



Select subset of data from an object based on a SQL expression

Amazon S3 Select

Select contents from object instead of retrieving the object



Familiar
Work and scales like GET requests

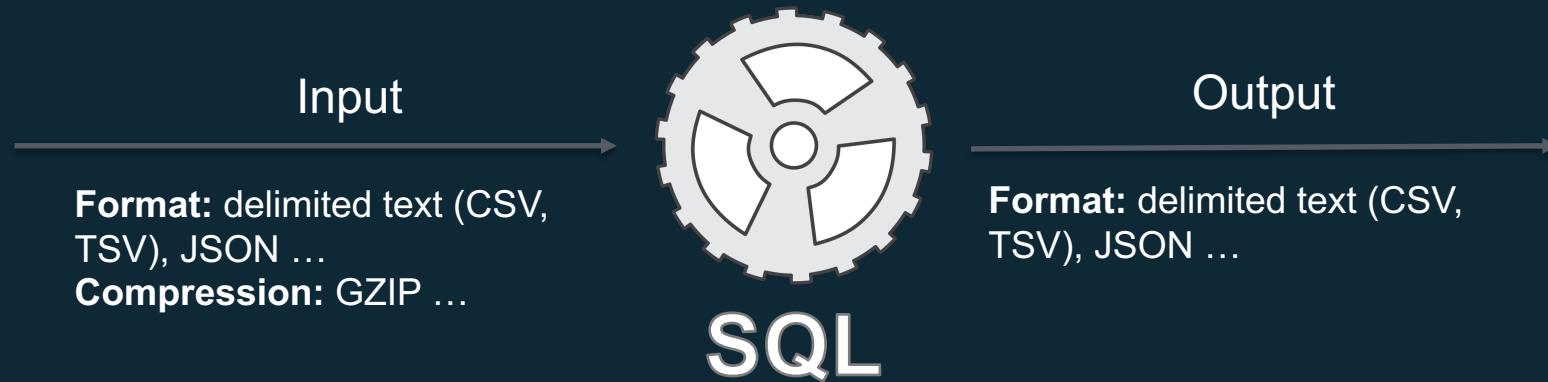


Integrated
AWS SDK and Presto
(others coming soon)



Simple to use
Standard SQL expression

Amazon S3 Select



Clauses	Data types	Operators	Functions
Select	String	Conditional	String
From	Integer, Float, Decimal	Math	Cast
Where	Timestamp	Logical	Math
	Boolean	String (Like,)	Aggregate

Sample S3 Select Query

Dataset:

```
{ "Rules": [ {"id": "1"},  
            {"expr": "y > x"},  
            {"id": "2", "expr": "z = DEBUG"}  
        ]  
    }  
    { "created": "June 27", "modified": "July 6" }
```

S3 Select query:

```
SELECT id FROM S3Object[*].Rules[*].id
```

Resulting dataset:

```
{"id":"1"}  
{}  
{"id":"2"}  
{}
```

S3 use cases

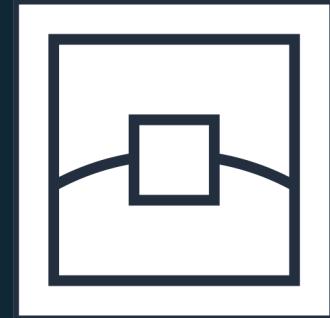
- Web-scale storage capacity and performance for web applications
- Single-origin store with delivery through Amazon CloudFront
- Staging and persistent store for Big Data applications
- Storage target for backup and active archive

Amazon CloudFront

Easy-to-use Content Delivery Network (CDN)

Pay-as-you-go pricing

Multiple origins: S3, EC2, on-premise



Worldwide network of 216 + edge locations and growing

- Video streaming
- Geo Restriction
- Custom SSL Certificates
- Dynamic Content
- Supports POST/PUT

4

On-Premises Storage Integration

Hybrid cloud storage use cases

Provide on-premises access to virtually unlimited cloud storage



Move backups to
the cloud



Shift on-premises storage
to cloud-backed file shares

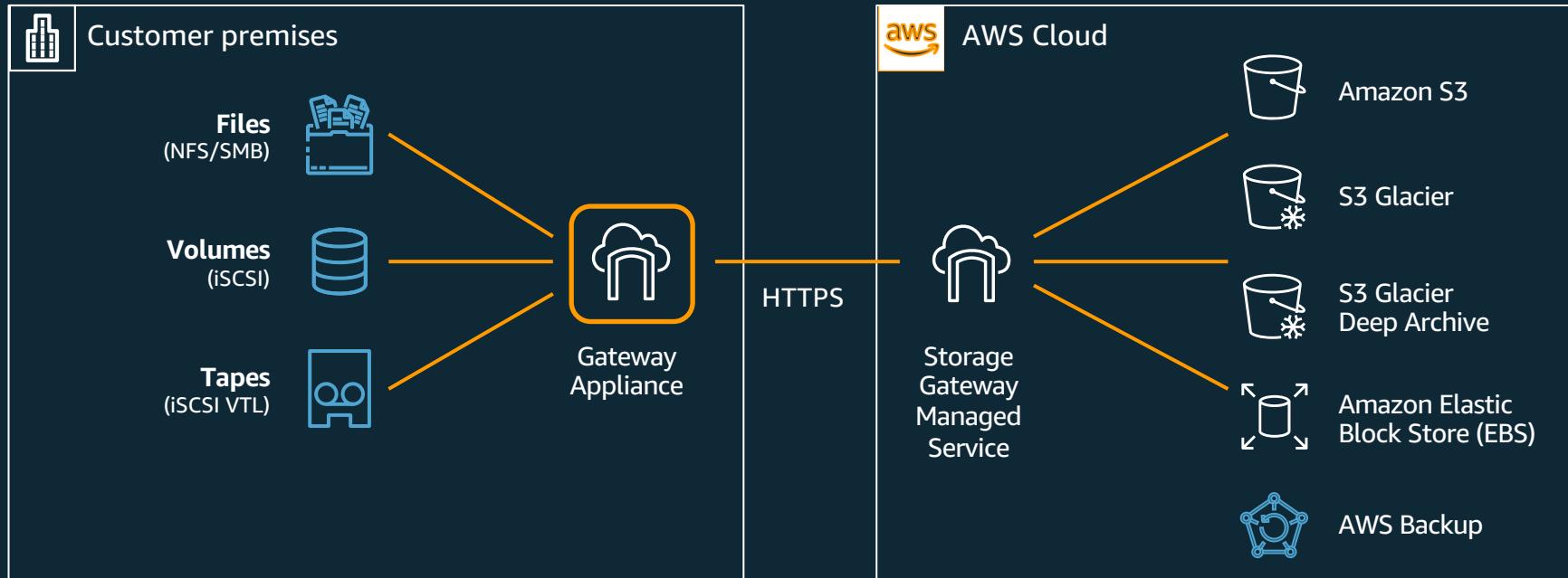


Low latency access for
on-premises applications to
cloud data

Regardless of your cloud adoption stage

AWS Storage Gateway

On-premises access to virtually unlimited cloud storage



Appliance Configuration: VMware, Hyper-V, EC2, Hardware

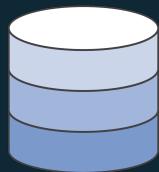
Integrated with IAM, KMS, CloudTrail, CloudWatch services

Storage Gateway – Files, volumes, and tapes



File gateway NFS (v3 and v4.1) interface

On-premises file storage backed by Amazon S3 objects



Volume gateway iSCSI block interface

On-premises block storage backed by S3 with EBS snapshots



Tape gateway iSCSI virtual tape library interface

Virtual tape storage in Amazon S3 and Glacier with VTL management

Storage Gateway – Common capabilities

-  **Standard storage protocols** integrate with on-premises applications
-  **Local caching** for low-latency access to frequently used data
-  **Efficient data transfer** with buffering and bandwidth management
-  **Native data storage** in AWS
-  **Stateless virtual appliance** for resiliency
-  **Integrated with AWS** management and security

Tape gateway

Virtual tape storage in Amazon S3 and Glacier with VTL management



Virtual tape storage in S3 and Glacier accessed via tape gateway

Data compressed in-transit and at-rest

Unlimited virtual tape storage, with up to 1PB of tapes active in library

Supports leading backup applications:

VERITAS™

DELL EMC

VEEAM

Hewlett Packard
Enterprise

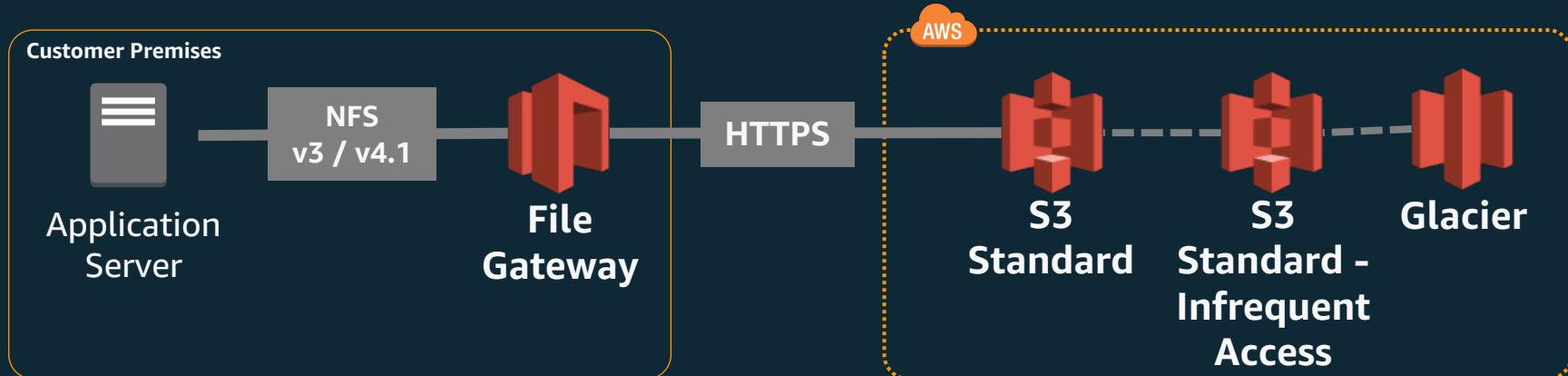
Microsoft®
System Center
Data Protection Manager 2012

arcserve



File gateway

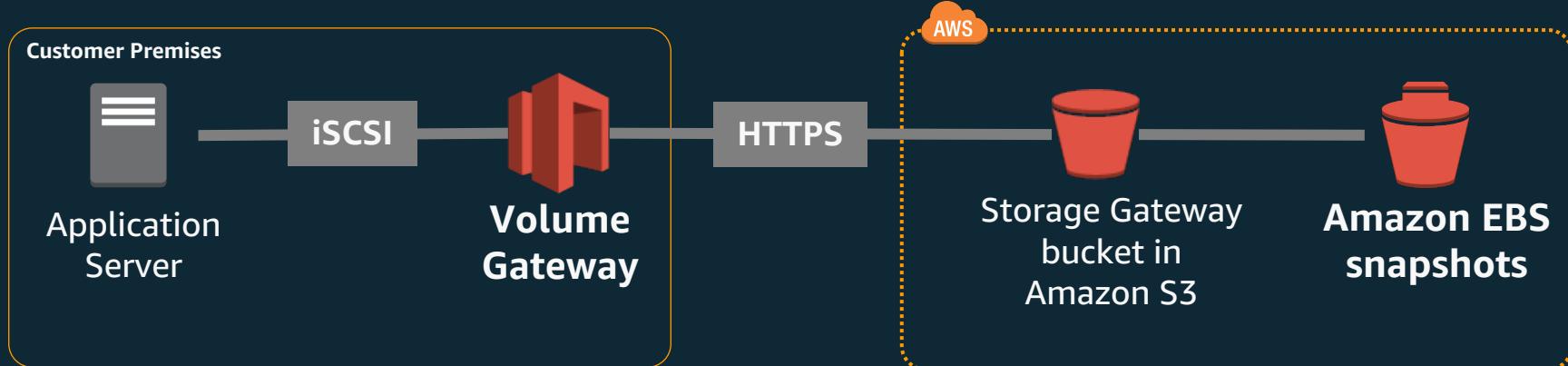
On-premises file storage maintained as objects in Amazon S3



- Data stored and retrieved from your S3 buckets
- One-to-one mapping from files-to-objects
- File metadata stored in object metadata
- Bucket access managed by IAM role you own and manage
- Use S3 Lifecycle Policies, versioning, or CRR to manage data

Volume gateway

On-premises volume storage backed by Amazon S3 with EBS snapshots



Block storage in S3 accessed via the volume gateway

Data compressed in-transit and at-rest

Backup on-premises volumes to EBS snapshots

Create on-premises volumes from EBS snapshots

Up to 1PB of total volume storage per gateway

Hybrid storage use cases with Storage Gateway



Enabling cloud workloads

Move data to AWS storage for Big Data, cloud bursting, or migration



Backup, archive, and disaster recovery

Cost effective storage in AWS with local or cloud restore



Tiered cloud storage

Easily add AWS storage to your on-premises environment

Amazon Snowball & Snowball Edge

- Terabyte scale data transport
- Uses secure appliances
- Faster than Internet for significant data sets
- Import into S3
- HIPAA Compliant



What is Snowball?

Terabyte scale data transport



Ruggedized
case
“8.5G Impact”

E-ink shipping
label



80 TB
10G network



Rain & dust
resistant

Tamper-resistant
case & electronics

All data encrypted
end-to-end



How fast is Snowball?

- Less than 1 day to transfer 250TB via 5x10G connections with 5 Snowballs, less than 1 week including shipping
- Number of days to transfer 250TB via the Internet at typical utilizations

	Internet Connection Speed			
Utilization	1Gbps	500Mbps	300Mbps	150Mbps
25%	95	190	316	632
50%	47	95	158	316
75%	32	63	105	211

Amazon Snowmobile

<https://www.youtube.com/watch?v=8vQmTZTq7nw>



Any Questions?

