Introduction to Storage on AWS

aws

Agenda

- Introduction

- Storage Primer

- Block Storage

- Shared File Systems

- Object Store

- Data Transfer and Edge Processing

- Backup

aws

O

Storage Primer
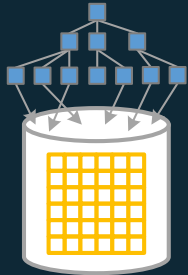
Block vs File vs Object

Block Storage

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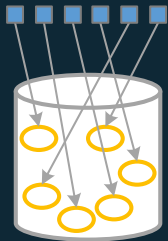Ex: Hard Disks, Storage Area Network (SAN) Storage Arrays

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, iops)	Ability to interact via API or with other services

aws

Understanding Durability
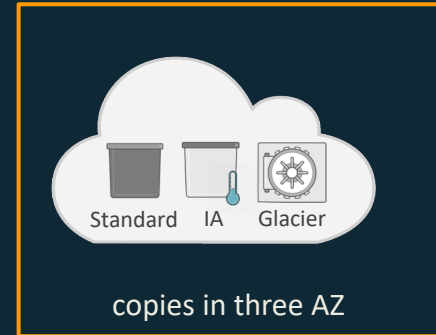

Two copies on one site
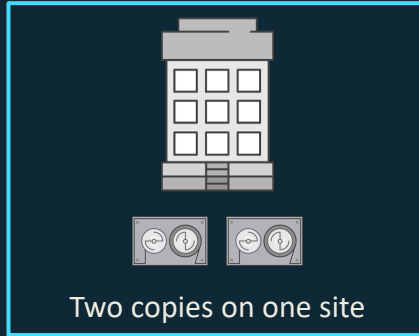
designed for
99.99%
durability

Copies on two sites
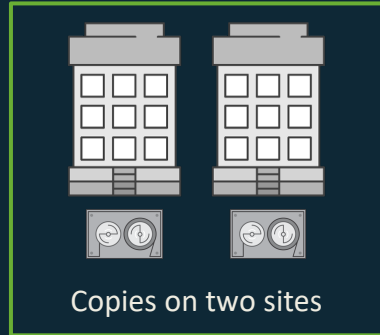
designed for
99.999%
durability

Standard   IA   Glacier

copies in three AZ

designed for
99.999999999%
durability

aws

Availability vs Durability

%	Availability	Durability
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

AWS delivers broadest storage portfolio in industry

OBJECT	BLOCK	FILE			BACKUP

| Amazon S3 | Amazon EBS | Amazon EFS | Amazon FSx for Windows File Server | Amazon FSx for Lustre | AWS Backup |

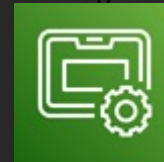## DATA TRANSFER AND EDGE PROCESSING

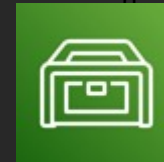| AWS Storage Gateway | AWS DataSync | AWS Transfer Family | AWS Snowmobile | AWS Snowball Edge | AWS Snowcone |

aws

1

Object Stores

Amazon S3

aws

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 (eu-west-1) - $0.023 per GB

aws

Your choice of object storage classes

S3 Standard	**S3 Intelligent-Tiering**	**S3 Standard-IA**	**S3 One Zone-IA**	**S3 Glacier**	**S3 Glacier Deep Archive**

Frequent ←――――――――――― *Access Frequency* ――――――――――→ *Infrequent*

• Active, frequently accessed data	• Data with changing access patterns	• Infrequently accessed data	• Re-creatable, less accessed data	• Archive data	• Archive data
• Milliseconds access	• Milliseconds access	• Milliseconds access	• Milliseconds access	• Select minutes or hours	• Select 12 or 48 hours
• > 3 AZ	• > 3 AZ	• > 3 AZ	• 1 AZ	• > 3 AZ	• > 3 AZ
• $0.0230/GB	• $0.0210 to $0.0125/GB ($0.004 to $0.00099/GB Archive)	• $0.0125/GB	• $0.0100/GB	• $0.0040/GB – ($4.10/TB)	• $0.00099/GB - ($1.01/TB)
	• No retrieval fees	• Retrieval fee per GB	• Retrieval fee per GB	• Retrieval fee per GB	• Retrieval fee per GB
	• Monitoring fee per Obj.	• Min storage duration	• Min storage duration	• Min storage duration	• Min storage duration
	• Min storage duration	• Min object size	• Min object size	• Min object size	• Min object size
	• Min object size				

aws

S3 Management Features
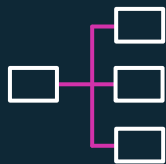
Organize

S3 Tagging

S3 Prefixes

S3 Versioning

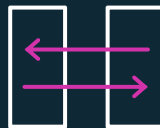## Monitor

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

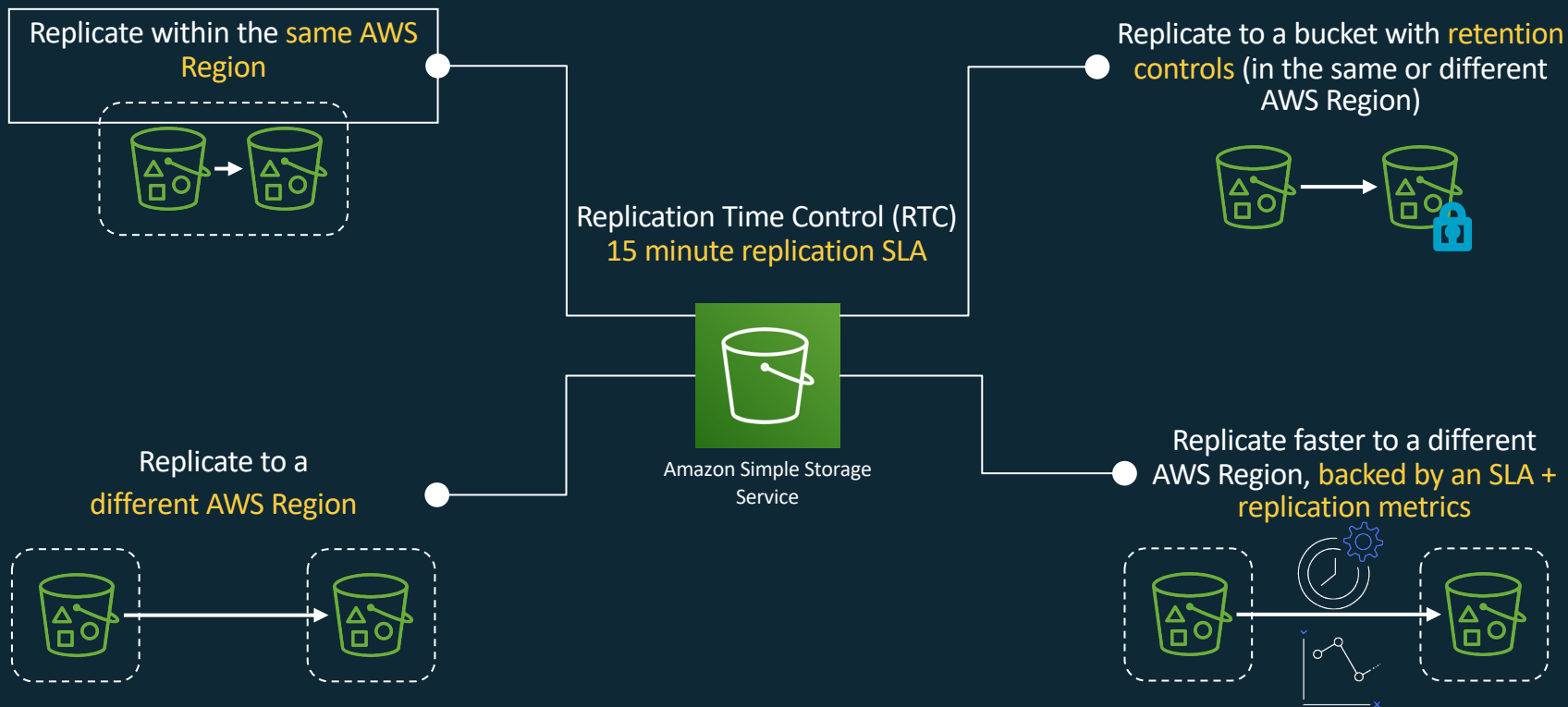Cross-Region Replication

Replication Time Control (RTC)

Modify

S3 Event Notifications + Lambda

S3 Batch Operations

S3 Object Lock

S3 Object Lambda

aws

S3 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

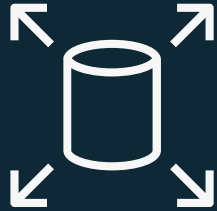Replicate faster to a different AWS Region, backed by an SLA + replication metrics
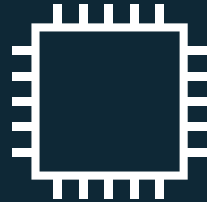
aws

2

Block Storage

Block storage portfolio

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

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

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

Relational databases

Non-relational/ NoSQL databases

Big data analytics

File/media

SAP ERP, Oracle ERP, Microsoft SharePoint, Microsoft Exchange

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

Cassandra, MongoDB, CouchDB
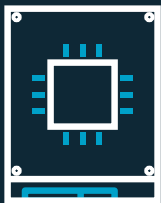
Kafka, Splunk, Hadoop, Data warehousing

CIFS/NFS, transcoding, encoding, rendering

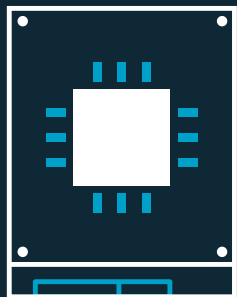LOW LATENCY AND CONSISTENT, HIGH IOPS AND THROUGHPUT

SCALABLE WITHOUT DISRUPTION TO YOUR WORKLOAD

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

aws

Six different volume types for optimal use

SSD

HDD

gp2 – gp3
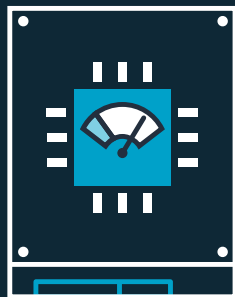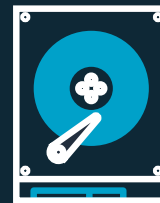General Purpose
SSD

io1 – io2
Provisioned IOPS
SSD

st1
Throughput Optimized
HDD

sc1
Cold
HDD

aws

Fully managed backup with EBS Snapshots

Backup

Restore

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

aws

What is Amazon EC2 instance store?

EC2 instances

Instance Store

or

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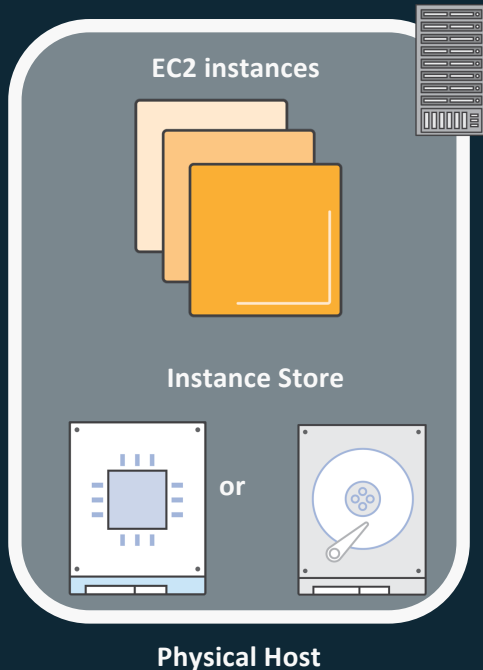
- Local to instance
- Non-persistent data store
- Available on several EC2 families
- Data is not replicated (by default)
- No snapshot support
- SSD or NVMe

aws

③ Shared file system

Amazon Elastic File System

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

$0.36/GB-Month*
Standard Storage

$0.0266/GB-Month*
Standard IA Access

EFS One Zone
$0.192/GB-month*

EFS Standard
$0.36/GB-month*

EFS One Zone-IA
Cost-optimized for less accessed files
$0.0142/GB-month* for storage
$0.01/GB* for access

EFS Standard-IA
Cost-optimized for less accessed files
$0.0266/GB-month* for storage
$0.01/GB* for access

*Pricing in the EU Central (Frankfurt) Region. Assumes 80% of the files are infrequently accessed

aws

Accelerate modernization and innovation

Highly integrated, serverless shared access

Amazon FSx

aws

Amazon FSx for Windows File Server
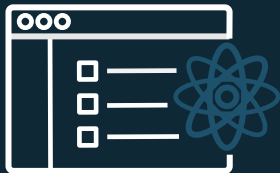
Fully managed file storage built on Windows Server

Easy migration to AWS

aws

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

Manage hardware

Plan capacity

Procure and purchase hardware

Set up storage servers and volumes

Detect and address hardware failures

Incur high upfront costs

Manage software

Install and configure server software

Set up and configure file systems

Apply Windows updates

Manage software licenses

Manage backups

Monitor security

aws

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

Customers continue to increase the size of their workloads on AWS across industry verticals and application areas

Industries and example use cases

Financial services:
Modeling and analytics

Life Sciences:
Genome analysis

Media and Entertainment:
Rendering and transcoding

Automotive:
ECU simulations and
object detection
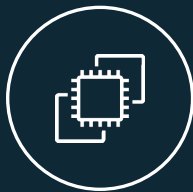
Semiconductor:
Electronic design
automation

Oil and gas:
Seismic data processing

Application areas

Big data
analytics

Machine
learning

High-performance
computing

For every $1 spent on high performance computing, businesses see $463 in incremental revenues and $44 in incremental profit[1]

aws

[1] High Performance Computing on AWS Redefines What is Possible

4

Data Transfer and Edge Processing

Many Options for Data Transfer

AWS
Direct Connect

Amazon
Kinesis
Firehose

Amazon Kinesis
Data Streams

Amazon Kinesis
Video Streams

Amazon S3
Transfer
Acceleration

AWS
Storage
Gateway

AWS
Database
Migration
Service

AWS
Snowcone

AWS
Snowball Edge

AWS
Snowmobile

AWS
DataSync

AWS Transfer
Family

aws

Amazon Snow Family

aws

Amazon Snowball Edge and Snowcone

- Terabyte scale data transport

- Uses secure appliances

- Faster than Internet for significant data sets

- Import into S3

- HIPAA Compliant

aws

How fast is Snowball?

- Less than 1 day to transfer 250TB via 4x10G connections with 4 Snowballs, less than 1 week including shipping

- Number of days to transfer 250TB via the Internet at typical utilizations

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

aws

What is Snowball?
Terabyte scale data transport

Ruggedized case "8.5G Impact"

Rain & dust resistant

E-ink shipping label

Tamper-resistant case & electronics

All data encrypted end-to-end

80 TB 10G network

aws

Introducing AWS Snowcone

Small, portable, rugged, and secure edge computing and data transfer device

- Military-grade security
- 4.5 pounds (2.1 kg)
- Portable computing, anywhere
- Withstands harsh environments
- Offline & online data transfer
- 8 TB of storage
- 2 CPU, 4 GB compute

Use cases

Industrial IoT, healthcare IoT, content distribution, content aggregation, data migration, logistics, autonomous vehicles, and transportation

AWS Snow Family for data collection & data movement

	Snowcone	Snowball Edge Storage Optimized	Snowmobile
Migration size	Up to 24 TB, online and offline	Up to petabytes, offline	Up to exabytes offline
Form factor	Rugged 8.5 G impact cases that are rain and dust resistant, E Ink label for shipping automation		45-foot container, scheduled delivery
Security	256-bit encryption, tamper detection		Encryption, security staff, GPS tracking, video surveillance, alarms
Storage capacity	8 TB usable	80 TB usable	<100 PB
DataSync agent	Pre-installed	-	-
Compute	2 vCPU, 4 GB RAM usable	40 vCPU, 80 GB RAM, 1 TB SSD usable	-
Onboard computing options	AWS IoT Greengrass functions Amazon EC2 AMIs		
Wireless	Wi-Fi	-	-
Portable or Mobile use	Battery based operation	-	-
Clustering	-	Up to 15 nodes	-

Visit the AWS Snow Family page for a more detailed comparison »

Any Questions?

aws

Lab 2: Amazon S3 Lab

https://vlad.cloud/mirri.html

aws