

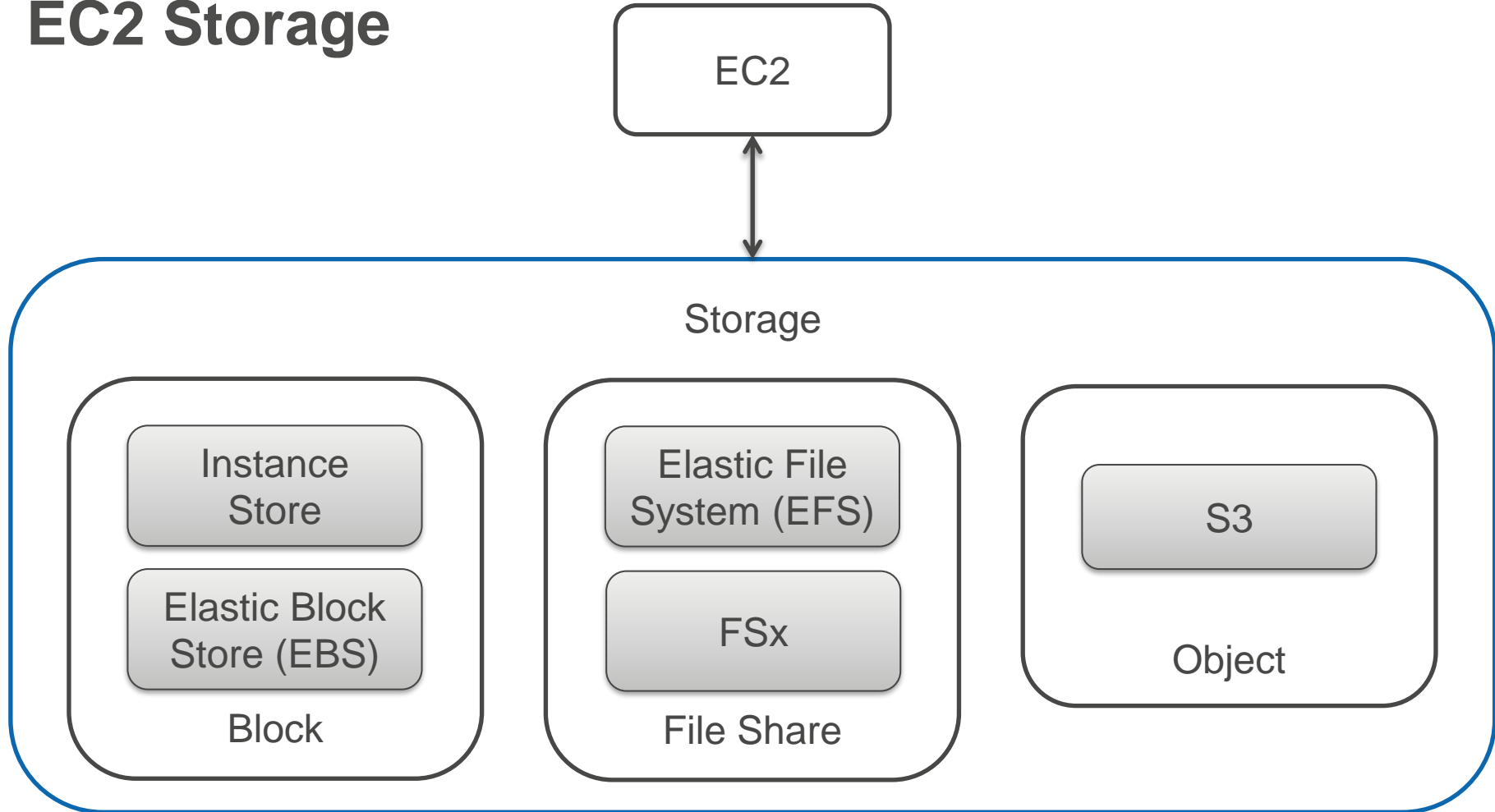
# EC2 Storage

## Options, Usage

Chandra Lingam

Cloud Wave LLC

# EC2 Storage



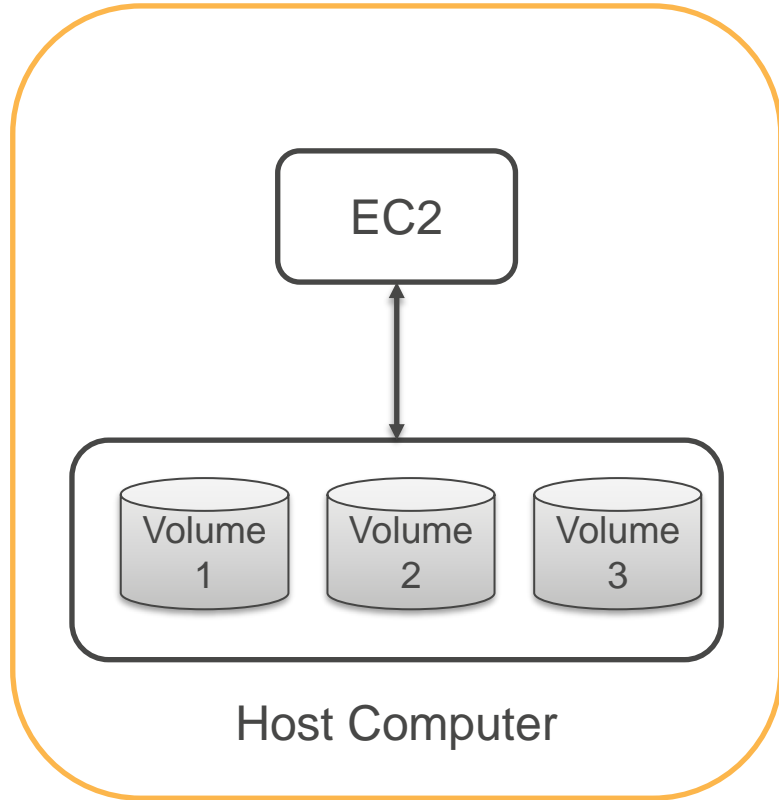
# Storage Units

- In the context of computer memory,
  - 1 KB = 1,024 Bytes ( $2^{10}$ )
  - 1 MB = 1,024 KB
- In the context of SSD/HDD
  - 1 KB = 1,000 Byte ( $10^3$ )
  - 1 MB = 1,000 KB
- Amazon uses KiB (Kibibyte), MiB (Mebibyte), GiB (Gibibyte) Standard units – Matches with memory
  - 1 KiB = 1,024 Bytes
  - 1 MiB = 1,024 KiB

# Block Storage

Instance and Elastic Block Store (EBS)

# Instance Store (Block)



Storage of host computer is assigned to EC2 instance

Temporary Storage

Highest Performance

Storage included as part of instance pricing

# Instance Store Durability

Data persists only for the lifetime of the instance

Reboot – Data Persists

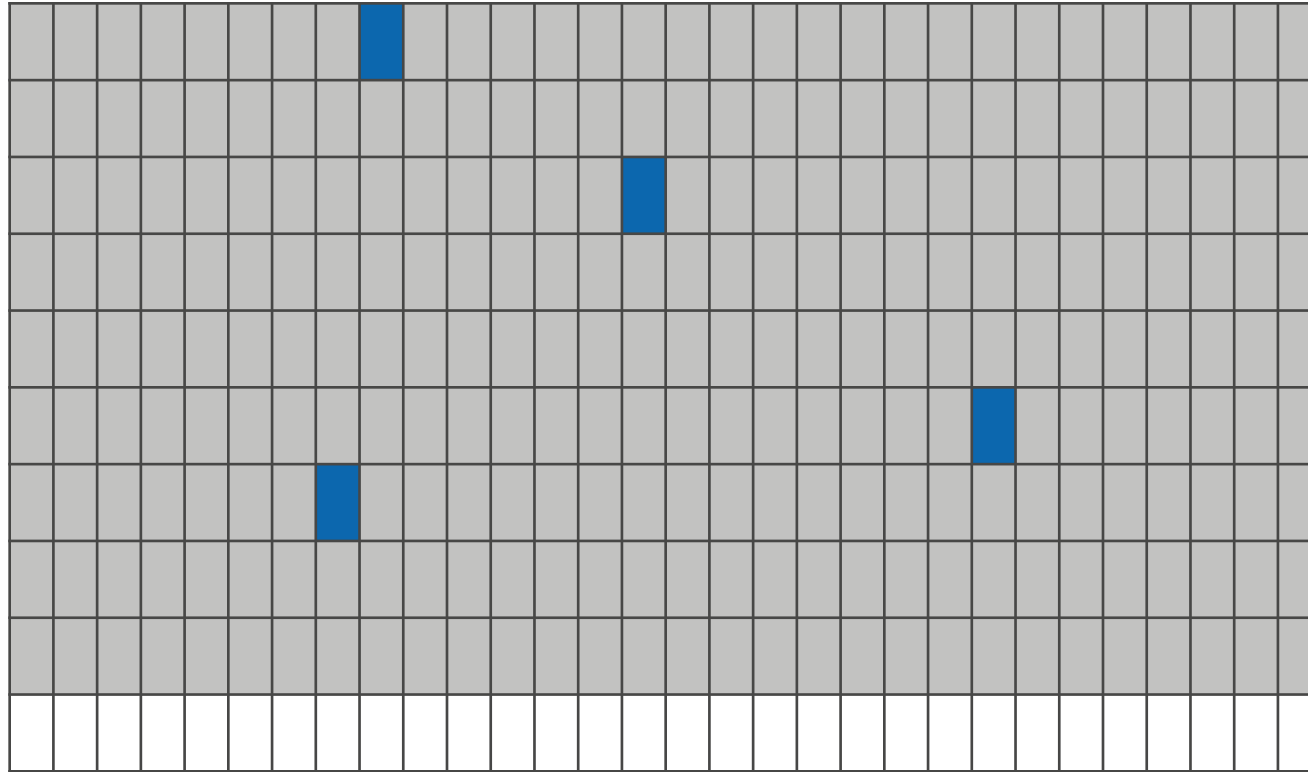
Data is lost – when underlying hardware fails, instance stops, or instance terminates

# Instance Store – Do it yourself High Availability and Durability



1. Replicate Data to multiple instances to prevent data loss.  
Example: Hadoop File System, MongoDB
2. Backup using OS provided software or third-party software

## Random I/O Workload – SSD Preferred



Performance  
Measured in IOPS

SSD based Instance  
Store offers very high  
IOPS

100,000 IOPS to  
Millions of IOPS  
(Based on Instance  
Type)



Allocated



Used



Read/Write





Search in this guide

English

Sign In to the Console

AWS > Documentation > Amazon EC2 > User Guide for Linux Instances

Feedback Preferences

## Amazon Elastic Compute Cloud

User Guide for Linux Instances

### What Is Amazon EC2?

Setting Up

Getting Started

Best Practices

### Tutorials

### Amazon Machine Images

### Instances

#### Instance Types

#### General Purpose Instances

Compute Optimized Instances

Memory Optimized Instances

**Storage Optimized Instances**

#### Accelerated Computing Instances

Finding an Instance Type

Changing the Instance Type

Getting Recommendations

#### Instance Purchasing Options

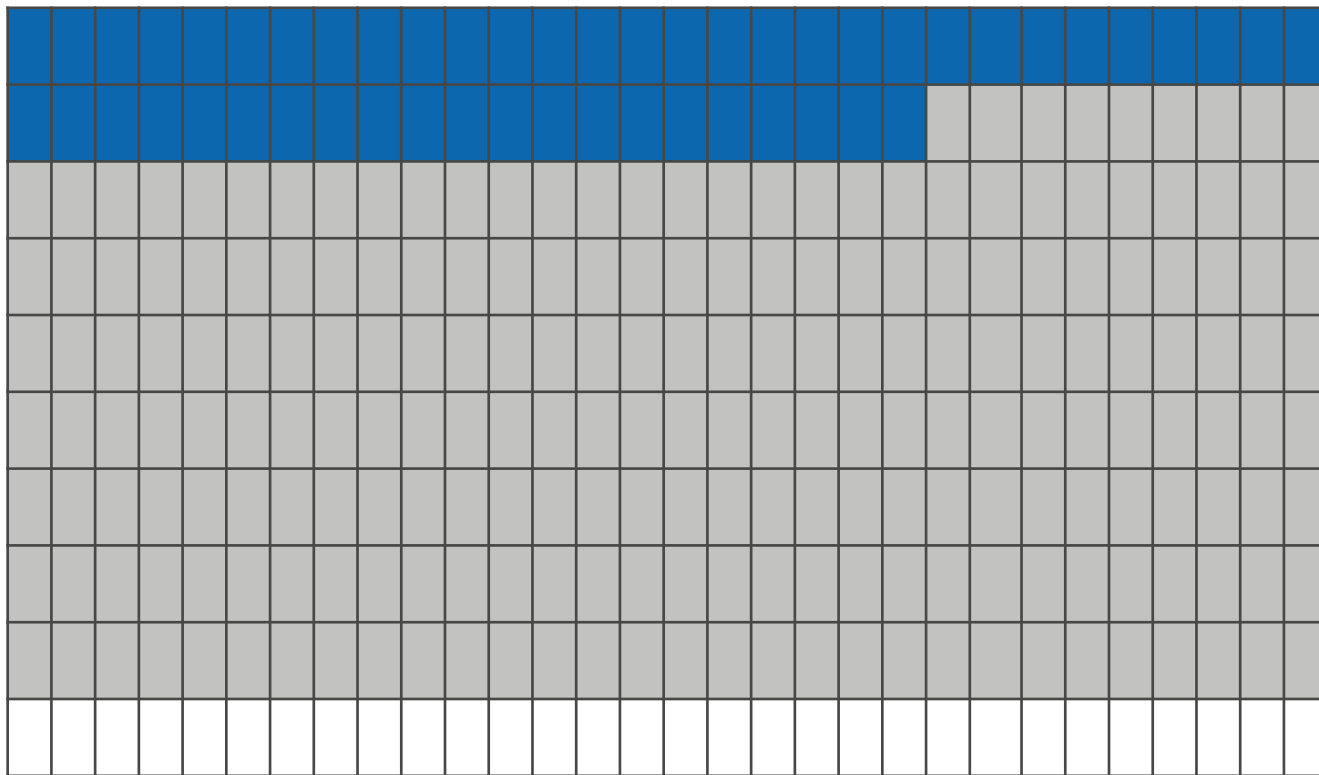
#### Instance Lifecycle

## SSD I/O Performance

If you use a Linux AMI with kernel version 4.4 or later and use all the SSD-based instance store volumes available to your instance, you get the IOPS (4,096 byte block size) performance listed in the following table (at queue depth saturation). Otherwise, you get lower IOPS performance.

Instance Size	100% Random Read IOPS	Write IOPS
i3.large *	100,125	35,000
i3.xlarge *	206,250	70,000
i3.2xlarge	412,500	180,000
i3.4xlarge	825,000	360,000
i3.8xlarge	1.65 million	720,000
i3.16xlarge	3.3 million	1.4 million
i3.metal	3.3 million	1.4 million
i3en.large *	42,500	32,500
i3en.xlarge *	85,000	65,000
i3en.2xlarge *	170,000	130,000
i3en.3xlarge	250,000	200,000
i3en.6xlarge	500,000	400,000
i3en.12xlarge	1 million	800,000
i3en.24xlarge	2 million	1.6 million

## Sequential I/O Workload – Magnetic/HDD Preferred



Performance  
Measured in  
Throughput (MiB/s)

Magnetic based  
Instance Store offers  
very high throughput

Lower Cost

SSD can be used – at  
a higher cost

 Allocated  Used  Read/Write

- 1. Choose AMI
- 2. Choose Instance Type
- 3. Configure Instance
- 4. Add Storage
- 5. Add Tags
- 6. Configure Security Group
- 7. Review

## Step 2: Choose an Instance Type

Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instances are virtual servers that can run applications. They have varying combinations of CPU, memory, storage, and networking capacity, and give you the flexibility to choose the appropriate mix of resources for your applications. [Learn more](#) about instance types and how they can meet your computing needs.

Filter by: Storage optimized Current generation [Show/Hide Columns](#)

Currently selected: t2.micro (Variable ECUs, 1 vCPUs, 2.5 GHz, Intel Xeon Family, 1 GiB memory, EBS only)

	Family	Type	vCPUs	Memory (GiB)	Instance Storage (GB)	EBS-Optimized Available	Network Performance
<input type="checkbox"/>	Storage optimized	i3en.24xlarge	96	768	8 x 7500 (SSD)	Yes	100 Gigabit
<input type="checkbox"/>	Storage optimized	i3en.metal	96	768	8 x 7500 (SSD)	Yes	100 Gigabit
<input type="checkbox"/>	Storage optimized	d2.8xlarge	36	244	24 x 2048	Yes	10 Gigabit
<input type="checkbox"/>	Storage optimized	i3en.12xlarge	48	384	4 x 7500 (SSD)	Yes	50 Gigabit
<input type="checkbox"/>	Storage optimized	d2.4xlarge	16	122	12 x 2048	Yes	High
<input type="checkbox"/>	Storage optimized	h1.16xlarge	64	256	8 x 2000	Yes	25 Gigabit
<input type="checkbox"/>	Storage optimized	i3.16xlarge	64	400	8 x 1000 (SSD)	Yes	25 Gigabit

# Uses - Storage Optimized Instances

D2, H1 – Magnetic

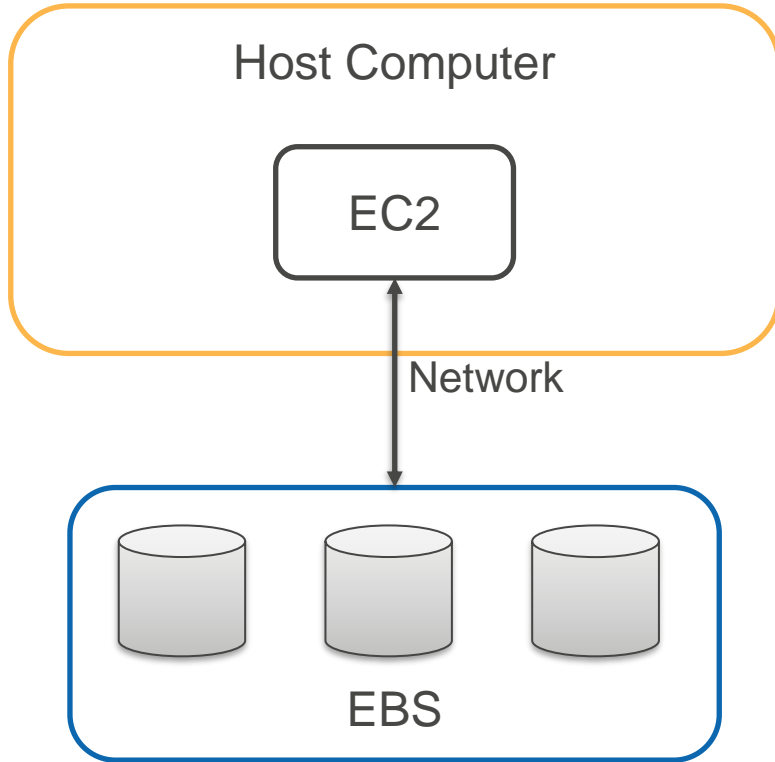
- High Throughput Access to Large Datasets
- Massively Parallel Processing Data Warehouse (MPP)  
For Example, a Redshift like solution
- MapReduce, Hadoop Distributed Compute
- Log or data processing

# Uses - Storage Optimized Instances

## I3 – SSD

- High frequency Online Transaction Processing Systems (OLTP)
- NoSQL Databases
- Relational Databases
- Data warehousing
- Cache for in-memory databases like Redis
- Distributed File Systems

# Elastic Block Store (EBS)



EBS is a managed block storage service

Storage volume is outside of host computer – Long term persistence

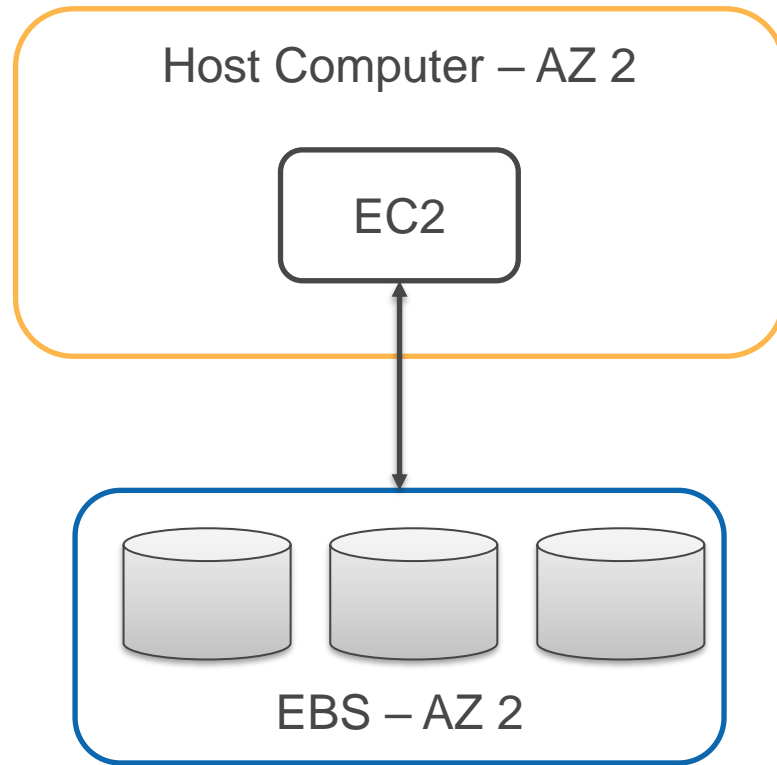
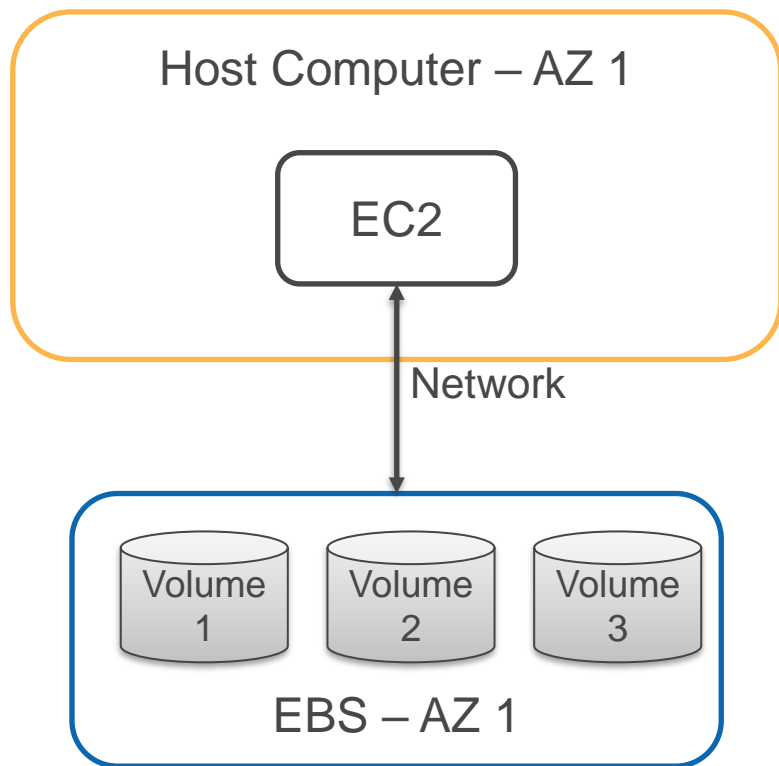
EC2 instance use EBS storage volume as a block device

You need to pay for allocated EBS storage

# EC2 Benefits with EBS

- Stop-Start EC2 Instance
- Persist EBS volumes for terminated instances
- Detach and attach volume to a different instance in the same Availability zone
- Built-in Snapshot Capability for incremental backup to S3
- Create Amazon Machine Image (AMI) from Snapshots to launch new EC2 instances

# Elastic Block Store (EBS) – Availability Zone Level





# EBS Features

EBS volumes are created at Availability Zone level

Highly available and durable - Volume is replicated within the Availability Zone – to protect against hardware failure

Built-in Snapshot Capability for incremental backup to S3

Create volumes from Snapshot (any Availability Zone in the Region)

Copy Snapshots to another Region (disaster recovery, expansion)

# EBS - Uses

Enterprise Applications

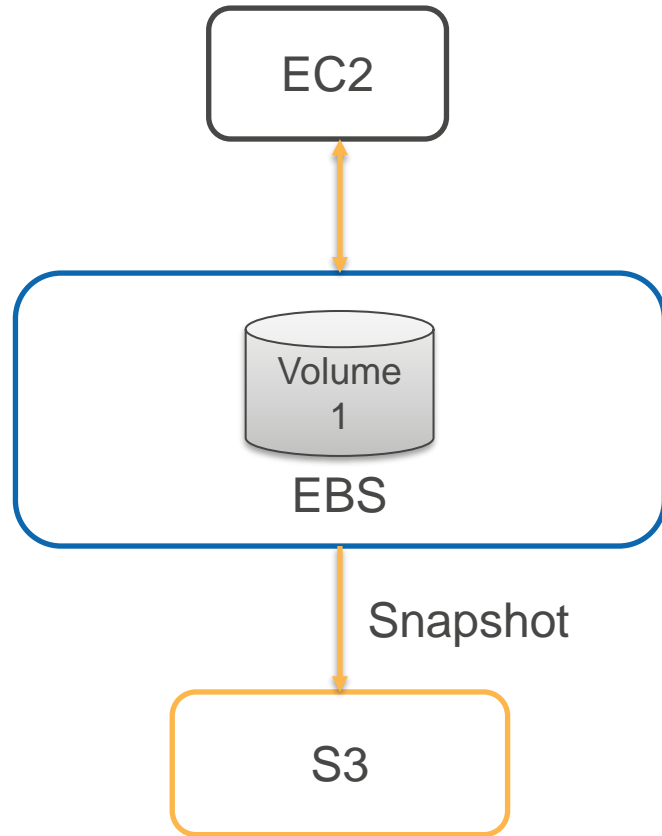
Relational Databases

NoSQL Databases

Big data analytics

Media Workflows

# Snapshot



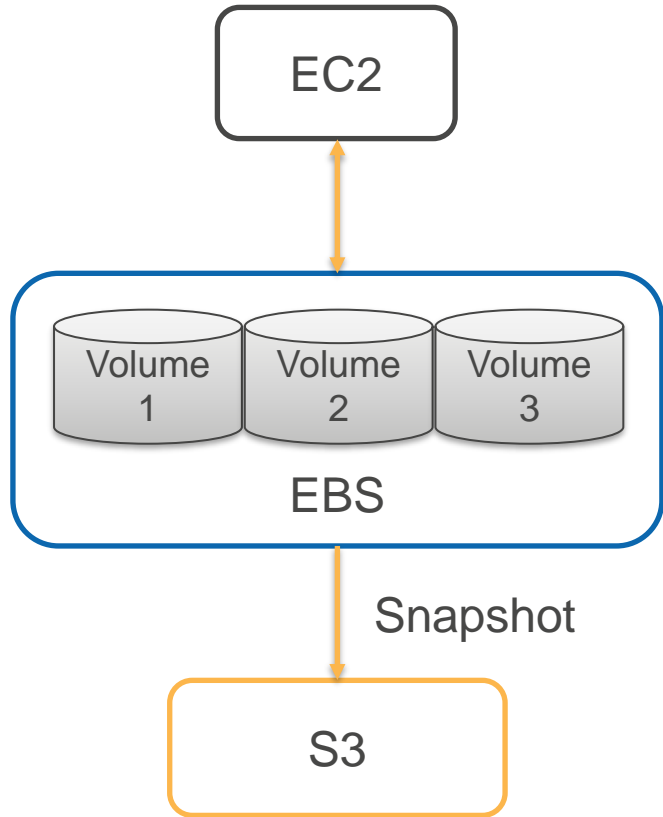
Snapshot – Incremental Backup of EBS Volume to S3

Point-in-time – Volume needs to be in consistent state (all application cache needs to be flushed and no traffic) when snapshot command is issued

Snapshot is async and works in the background

System can start using the volume – no need to wait for backup to complete

# Snapshot – Multiple Volumes



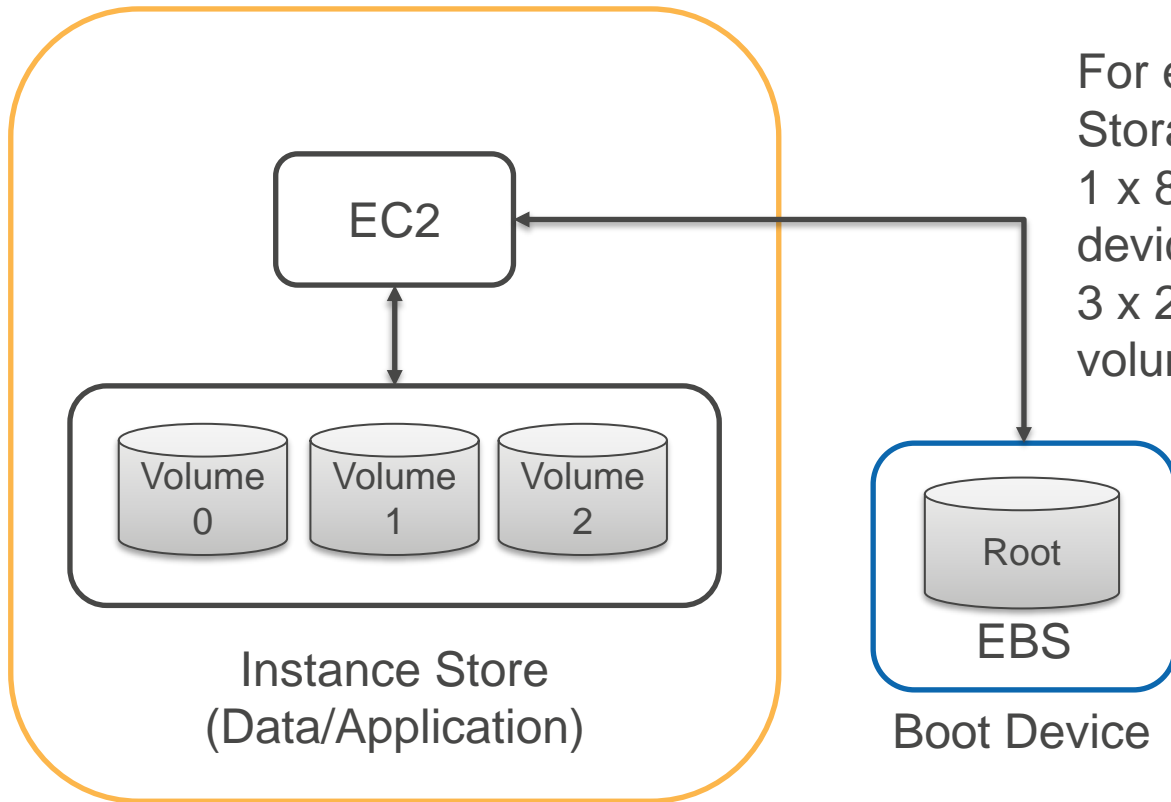
In this example, EC2 instance is using three volumes. You can use any of the following options:

1. Stop the system and issue snapshot for each volume and restart
2. Take backup at application level (for example, using database software)
3. Use the **\*\*New\*\*** [multi-volume consistent snapshot](#)

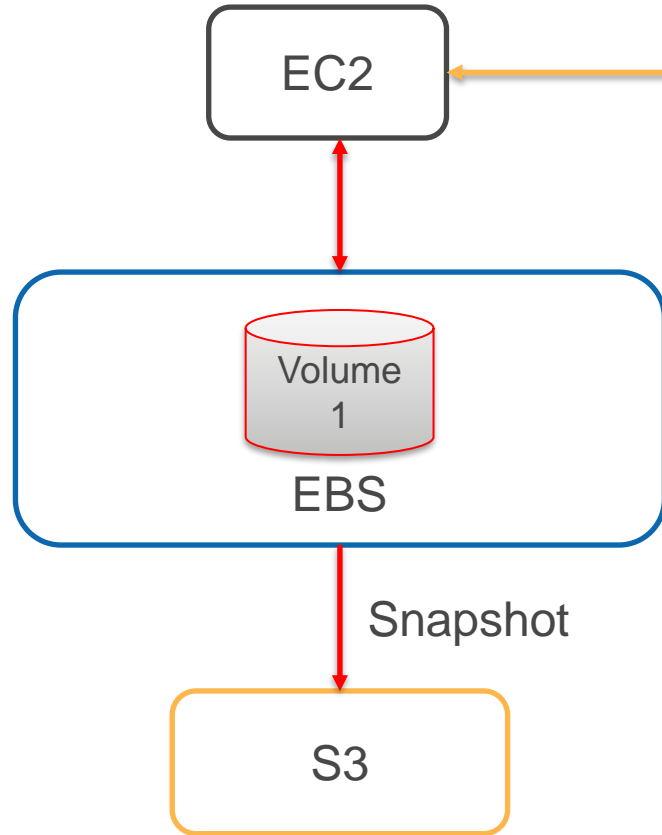
# Mix-and-Match

Mix and match volumes of different types

For example,  
Storage Optimized (d2.xlarge)  
1 x 8 GiB EBS volume as boot device and  
3 x 2 TiB HDD Instance Store volumes for application use



# Encryption



EBS supports volume encryption

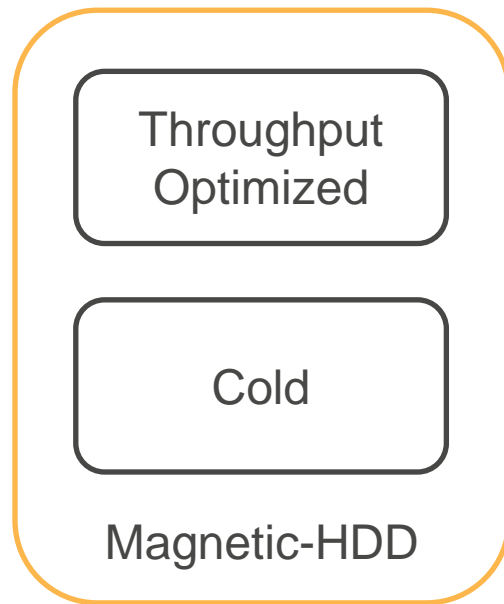
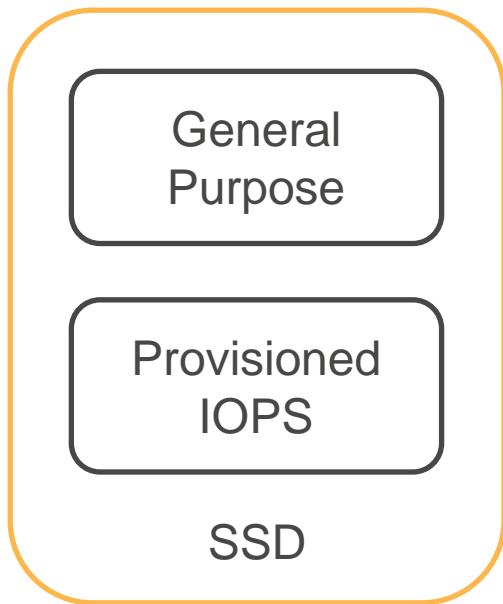
Data in transit from host computer and at rest in EBS volume is encrypted

Snapshots are encrypted

Restored volumes are also encrypted

Encryption Key – EBS managed Key or Customer Master Key (CMK). CMK gives you fine grained access control to encrypted data.

# EBS Volume Types



## EBS – General Purpose SSD (gp2)

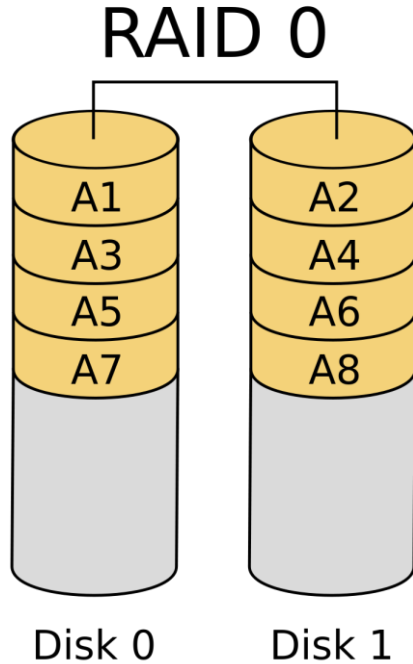
“Balances price-performance for a wide variety of workloads”

Feature	Value
Baseline IOPS per GiB	3
Burst IOPS	3,000
<b>Max IOPS/Volume</b>	<b>16,000 (at 5.3 TiB storage or larger)</b>
Max Throughput (MiB/s)	250
Size	1 GiB – 16 TiB

NOTE: Review the above link for up-to-date Max IOPS data



# RAID using EBS volumes



One way to increase IOPS, Throughput, Volume is by using RAID 0

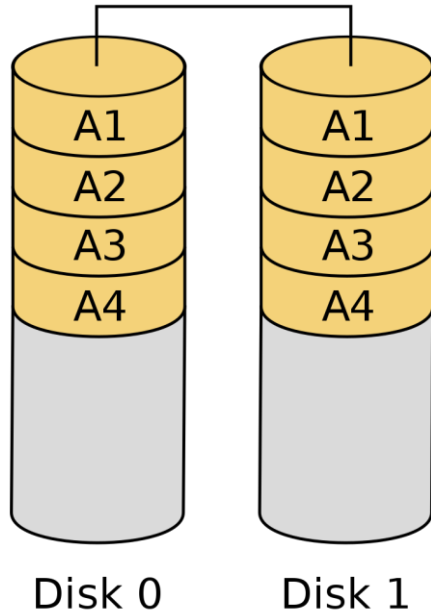
RAID 0 stripes the blocks across two EBS volumes

Each volume in-turn is automatically replicated by EBS service for high availability and durability

RAID 0 doubles the IOPS, Throughput and Size

RAID requires additional software components that you need to manage

# RAID 1



RAID 1 mirrors the blocks

EBS volumes are already replicated – so there is no benefit in using RAID 1

RAID 1 is generally not used with EBS volumes

Image Credit: Cburnett, <https://commons.wikimedia.org/w/index.php?curid=1509082>  
[https://en.wikipedia.org/wiki/Standard\\_RAID\\_levels](https://en.wikipedia.org/wiki/Standard_RAID_levels)

# Uses - General Purpose SSD

Boot Volumes

Small to Medium databases

Development and test environment

## EBS – Provisioned IOPS SSD (io1)

“Highest performance volume for latency sensitive transactional workloads”

Feature	Value
Max IOPS/Volume	64,000
Max Provisioned Performance to Volume Size (GB)	50:1 i.e., you can Provision IOPS up to 50 times the volume size in GiB
Max Throughput (MiB/s)	1,000
Size	4 GiB – 16 TiB
Provisioned Performance Guarantee	99.9% of the time

NOTE: Review the above link for up-to-date Max IOPS data

# Uses - Provisioned IOPS SSD (io1)

Boot Volumes

Critical Business applications

Large databases – Cassandra, MongoDB, SQL Server, Oracle, PostgreSQL, MySQL

## EBS – Throughput Optimized HDD (st1)

“Low cost volume designed for frequently accessed, throughput intensive workloads”

Feature	Value
Max IOPS/Volume	500
Max Throughput (MiB/s)	500
Size	500 GiB – 16 TiB

NOTE: Review the above link for up-to-date data

# Uses - Throughput Optimized HDD (st1)

Big data

Data warehouse

Log processing

## EBS – Cold HDD (sc1)

“Low cost volume designed for infrequently accessed workloads”

Feature	Value
Max IOPS/Volume	250
Max Throughput (MiB/s)	250
Size	500 GiB – 16 TiB

NOTE: Review the above link for up-to-date data



## Uses - Throughput Optimized HDD (st1)

Inexpensive storage

Ideal for infrequently accessed sequential workloads

# Pricing Example

Volume Type	Price per GB-month	IOPS Cost per month	Size (GB)	IOPS - Provisioned	Storage Cost	IOPS Cost	Total Monthly Cost
General Purpose SSD	0.100	-	500.00	Baseline 1500 IOPS (500x3) and can burst up to 3000	50.00		50.00
Provisioned IOPS SSD	0.125	0.065	500.00	3,000.00	62.50	195.00	257.50
Provisioned IOPS SSD	0.125	0.065	500.00	5,000.00	62.50	325.00	387.50
Throughput Optimized HDD	0.045		500.00		22.50		22.50
Cold HDD	0.025		500.00		12.50		12.50

AWS Simple Monthly Calculator - <https://calculator.s3.amazonaws.com/index.html>

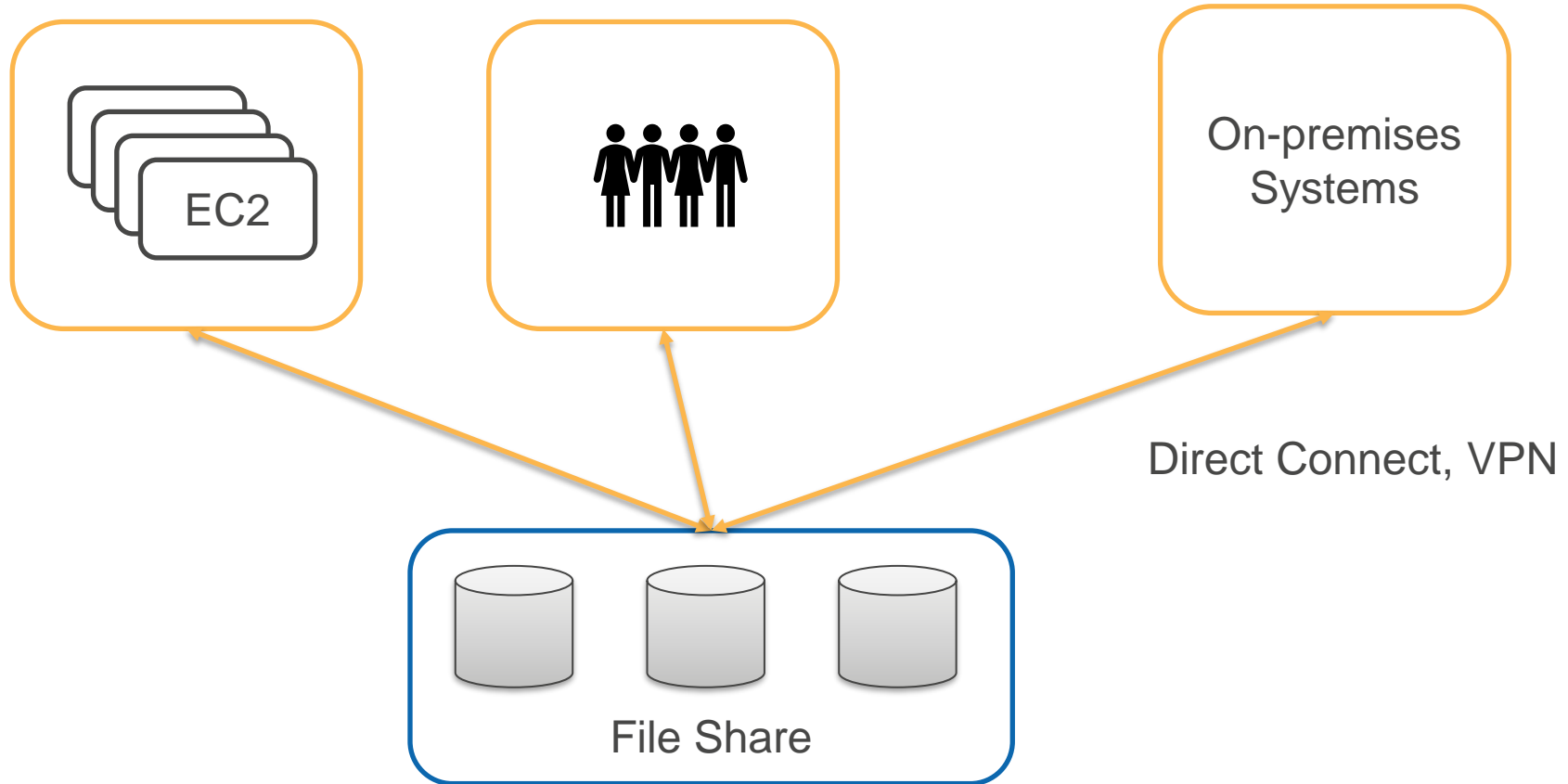
# File Share

Elastic File System (EFS)

FSx for Windows

FSx for Lustre

# AWS File Share Solution



# AWS Managed File Shares - Uses

- User Home Directories
- Lift-Shift Enterprise Applications to Cloud
- Analytics – Data Storage
- Web Content – Common share for all webserver
- Media Workflows – Video Editing and Production
- Database backups – easily mount to database servers and take backup

# File Storage Services

Service	Purpose
Elastic File System (EFS)	File share for Linux EC2 instances on AWS
FSx for Windows	File share for Windows EC2 instances on AWS
FSx for Lustre	File share Optimized for High Performance Computing – Linux EC2 instances Access S3 as a file share (like Storage Gateway)

- Fully managed – automatically grow and shrink
- High Durability and Availability – Replicated across multiple availability zones in a region
- All these file shares can be used from On-Premises using AWS Direct Connect or VPN

# Elastic File System (EFS)

File Share for Linux EC2 instances on AWS

NFS Filesystem and traditional file permissions model

Standard and Infrequent Access Storage Tiers (Similar to S3)

Lifecycle Management – to move files to lower cost storage

# FSx for Windows

File share for Windows Systems

NTFS Filesystem using SMB Protocol

Integrated with Active Directory



# FSx for Lustre

High performance file share for Linux Systems

Optimized for high performance computing and fast processing

Two modes:

- Standalone File share or

- Link to S3 Bucket and access S3 as a file share



Chandra Lingam

57,000+ Students



For AWS self-paced video courses, visit:

<https://www.cloudwavetraining.com/>

