Network Storage and Elastic File Systems

Andru Estes
Principal Author
andru-estes

What Is Amazon Elastic File System (EFS)?

Amazon EFS Introduction

Managed Network File System (NFS) that can be mounted by many EC2 instances

EFS works with EC2 instances across multiple Availability Zones

Highly available and scalable; however, it is much more expensive than EBS

Important Concepts

Uses standardized NFSv4.1 protocol

Currently only compatible with Linux-based AMIs

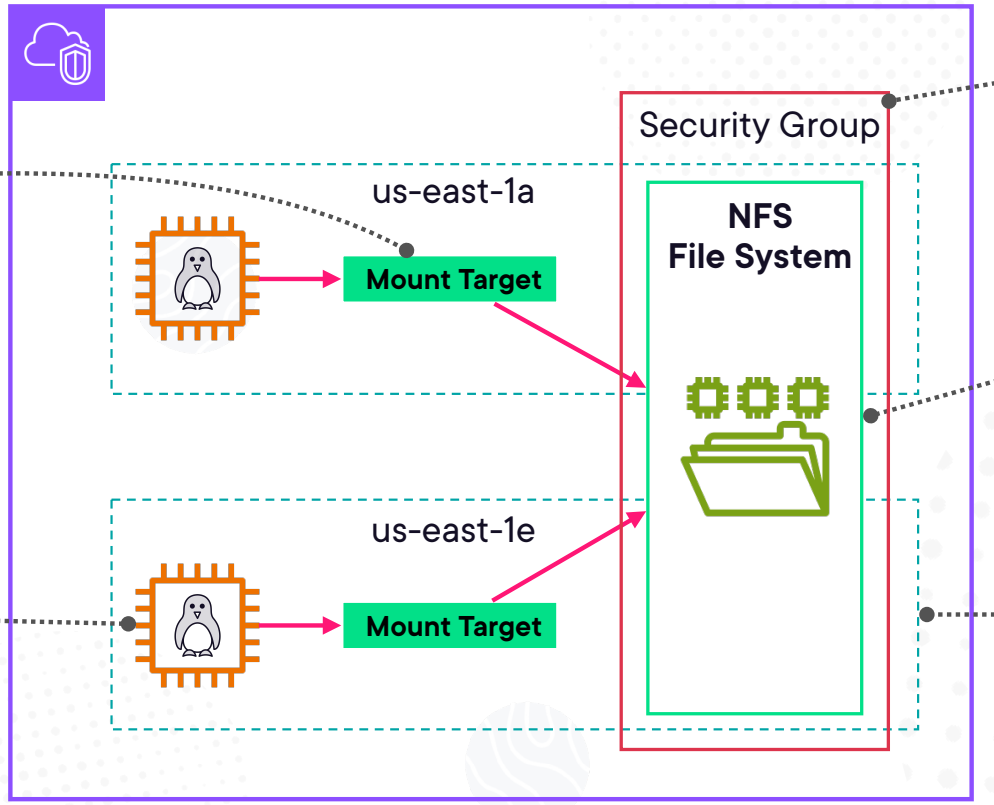
Capability to enable encryption at rest using KMS

Scales automatically; no capacity planning required

Pay per use and it is more expensive than using EBS

You use VPC Security Groups to control access to your EFS resources.

EFS Architecture Diagram



NFS clients mount the target for the shared file system

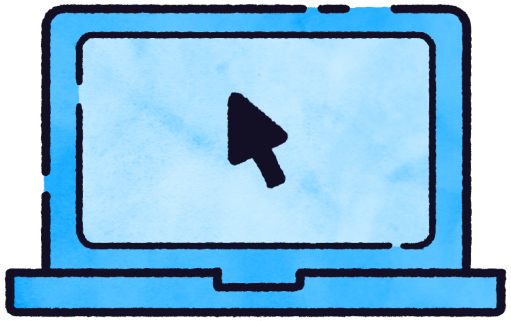
Only Linux-based AMIs are supported right now

Control network access via VPC security groups

POSIX file system supports standardized file API calls

EFS can span across multiple Availability Zones to support Highly-available architectures

Amazon EFS Use Cases



Content Management

Great fit for content management systems, as you can easily share content between EC2 instances

Web Servers

Also a great fit for simple web servers where you have a simple folder structure for your website

Exam Pro Tip: Question about needing a scalable filesystem or NFS?

Think Amazon EFS!

EFS Performance

EFS Performance Overview

Allows up to thousands of concurrent NFS client connections

Provides up to 20 Gbps of throughput for networking requirements

Automatically scales to petabyte sizes as soon as you need it

Performance Mode

General Purpose

Default option. Better for things like web servers and content management systems

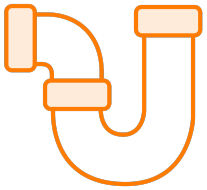
Max I/O

Previous generation. Originally meant for highly parallelized workloads that can handle higher latency

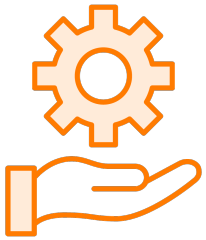
Max I/O performance mode has higher per-operation latencies than General Purpose performance mode.

Because of this, AWS recommends you always use General Purpose performance mode!

Throughput Modes



Elastic: Default mode for spiky or unpredictable workloads. Will automatically scale up or down as needed. Recommended mode!



Provisioned: Useful if you know the performance requirements of your workloads. Does not take in storage size for throughput.

Bursting: Useful for when you want throughput to scale with the amount of storage in EFS.

EFS Storage Classes

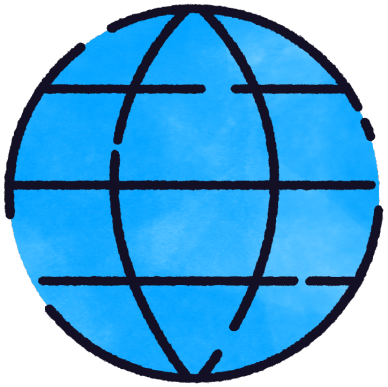
Storage Class Options

Standard: High-speed SSD storage to deliver sub-millisecond latency performance for your active application data. Frequently accessed data. Most expensive option.

Infrequent Access (EFS-IA): Cost-optimized storage classes offer a combination of low cost and high performance. Useful for files needed to satisfy audit requirements or performing historical analysis.

Archive: Cost-optimized for data that is accessed only a few times a year or less. Perfect for uses that do not need sub-millisecond latency. Prices can be up to 50% lower than EFS-IA.

File Systems Types



Regional

Spreads data across multiple Availability Zones for better durability and availability. Useful for important workloads.

One Zone

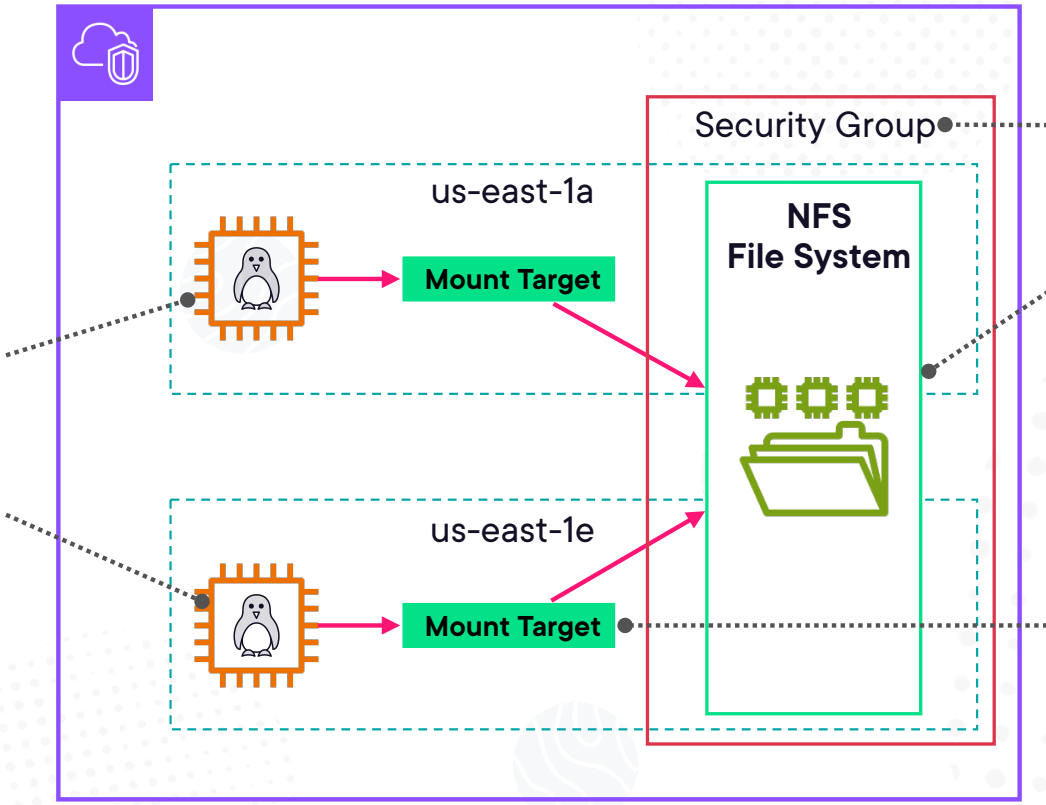
Data is only redundant within a single Availability Zone, with backups turned on by default. Useful for non-critical workloads.

Each file system type is said to deliver 99.999999999% data durability within their respective AZs.

Exam Pro Tip: You can leverage lifecycle policies with EFS to move files between different storage classes!

Demo: Deploying an Elastic File System for EC2

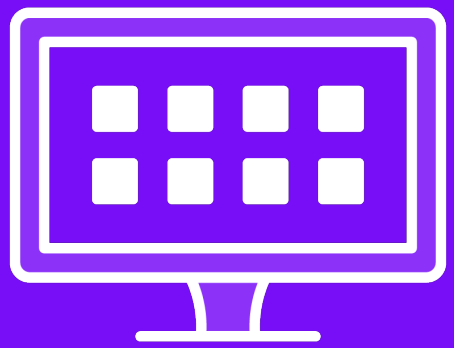
Deploy multiple Amazon Linux EC2 instances to multiple Availability Zones



Create a new EFS resource, and then create a new security group to only allow VPC traffic

Set up the EC2 instances to mount and use the new file system

Amazon FSx for Windows



Amazon FSx for Windows

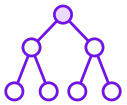
A fully managed native Microsoft Windows file system so you can easily move your Windows-based applications that require file storage to AWS

Amazon FSx for Windows Concepts

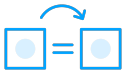
Fully managed Windows file system for easy Windows migrations

Supports Windows NTFS and Server Message Block (SMB) protocols

Designed for Windows and, accordingly, Windows applications

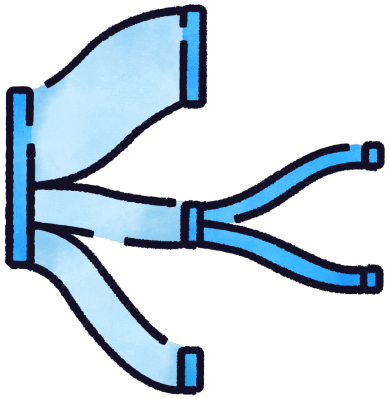


Supports AD users, access control lists, groups, and security policies



Also supports Distributed File System (DFS) namespaces and replication

Amazon FSx for Windows Performance and Durability



Allows for throughput of up to 10+ GB per second

Can store hundreds of Petabytes of data

Multi-AZ capable and backs up to S3 daily

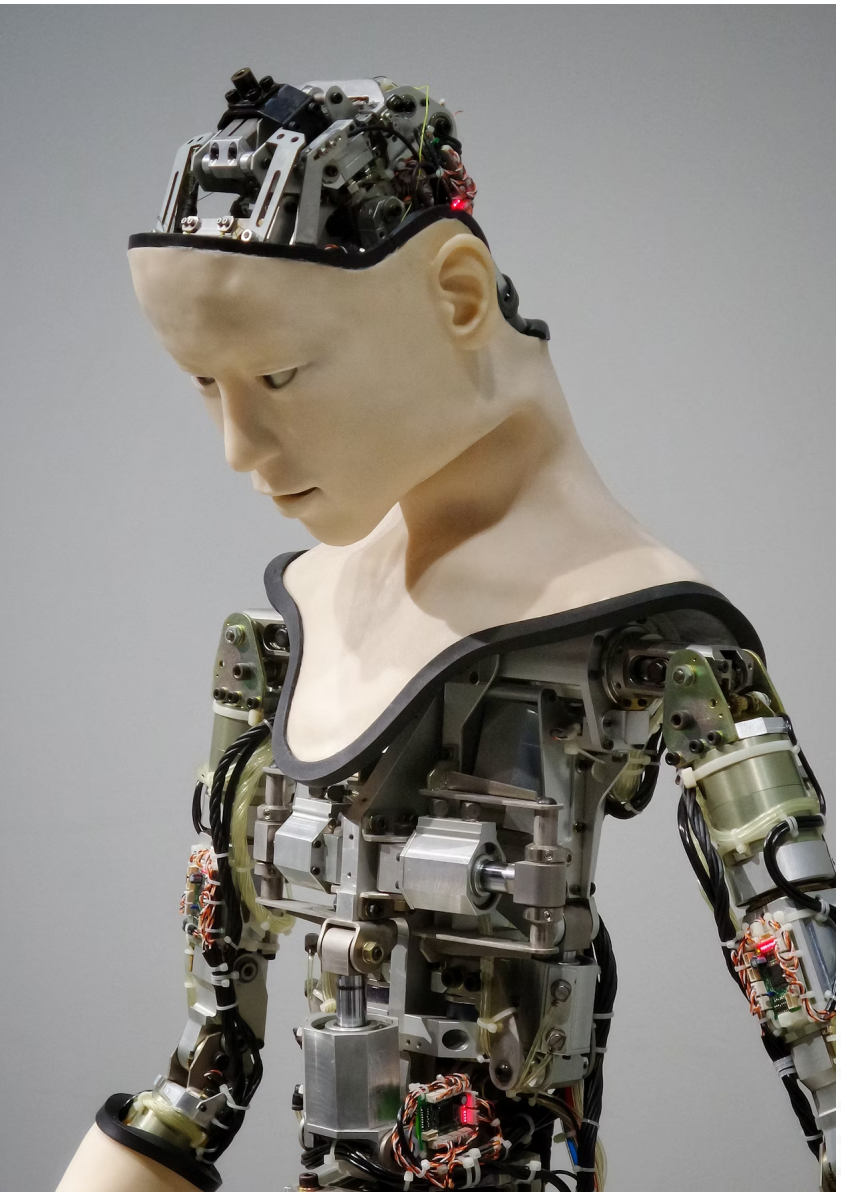
Think Amazon FSx for Windows when you need centralized storage for Windows-based applications.

These can be SharePoint, Microsoft SQL Server, Workspaces, IIS Web Server, or any other native Microsoft application.

Amazon FSx for Lustre

Amazon FSx for Lustre

A fully managed, parallel, distributed file system that is optimized for compute-intensive workloads



Amazon FSx for Lustre Concepts

Geared towards Linux-based Operating Systems

Useful for HPC and Machine Learning workloads

Easily supports millions of IOPS and sub-millisecond latencies

Process massive datasets at up to hundreds of gigabytes per second

Can leverage Amazon S3 as a file system

Examples:

- Video processing
- Electronic design automation

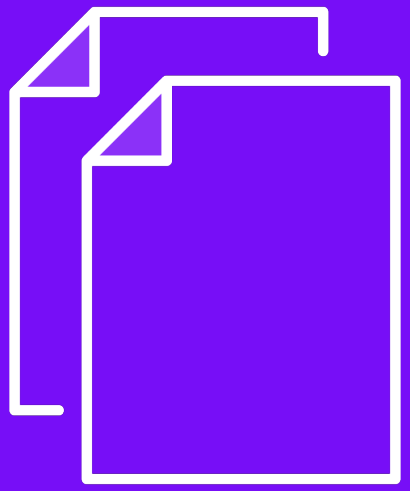
Image Source: https://unsplash.com/

Think Amazon FSx for Lustre when you need high-speed, high-capacity distributed storage for applications.

These include high performance computing (HPC), financial modeling, etc.

Remember that FSx for Lustre can store data directly on S3!

Amazon FSx for NetApp ONTAP



Amazon FSx for NetApp ONTAP

Fully managed shared storage in the AWS Cloud with the popular data access and management capabilities of ONTAP

Amazon FSx for NetApp ONTAP Concepts

AWS-managed NetApp ONTAP service

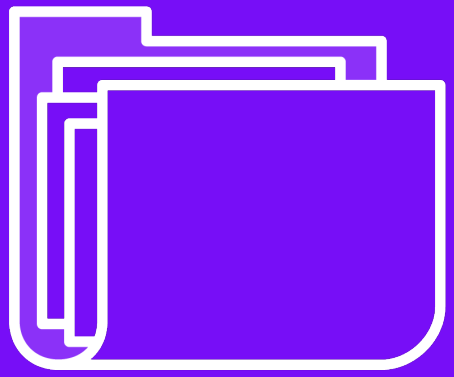
Works with a majority of major Operating Systems

Your storage can shrink and grow as you need

Supports industry-standard NFS, SMB, iSCSI, and NVMe-over-TCP

Think Amazon FSx for NetApp ONTAP when you need to migrate or use NetApp ONTAP file systems within AWS!

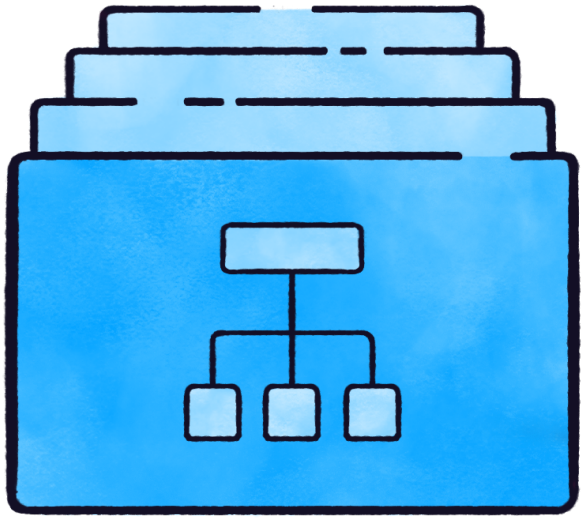
Amazon FSx for OpenZFS



Amazon FSx for OpenZFS

Fully managed file storage service that makes it easy to move data to AWS from on-premises ZFS or other Linux-based file servers

Amazon FSx for OpenZFS Concepts

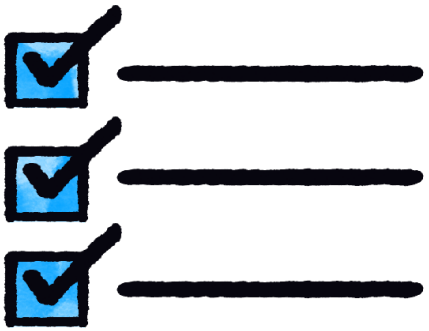


- Useable by multiple Operating Systems, including macOS, Windows, and Linux
- Supports the industry-standard NFS protocol (*v3, v4.0, v4.1, and v4.2*)
- Offers Multi-AZ (HA), Single-AZ (HA), and Single-AZ (non-HA)
- Fully managed file system backups stored on S3
- Near-instant point-in-time OpenZFS snapshots

Think Amazon FSx for OpenZFS when you need support for open-source OpenZFS file systems.

Module Summary and Exam Tips

Amazon EFS Exam Tips

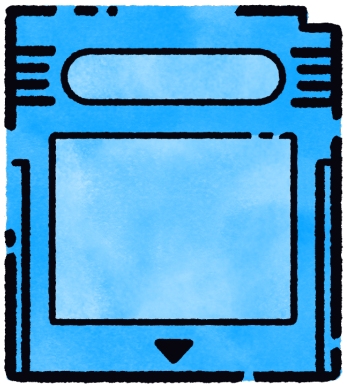


What to remember about EFS:

- Supports the Network File System version 4.1 (NFSv4.1) protocol
- Only pay for the storage you use (*no pre-provisioning required*)
- Can scale up to petabytes
- Can support thousands of concurrent NFS connections
- Data is stored across multiple AZs within a Region
- Read-after-write consistency

If you have a scenario-based question around highly scalable shared storage using NFS, think EFS.

Amazon EFS Exam Tips



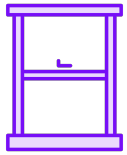
Storage Classes

Remember you can leverage different storage classes, including EFS-IA, Standard, and Archive

Lifecycles

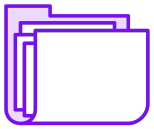
EFS supports lifecycle policies to move data between the different storage classes

Amazon FSx Exam Tips

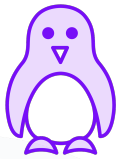


Amazon FSx for Windows: Centralized storage for Windows-based applications (*SharePoint, Microsoft SQL Server, IIS Web Server*)

Amazon FSx for Lustre: High-speed, high-capacity distributed storage. HPC applications, financial modeling, etc. Can store data directly on S3!



Amazon FSx for NetApp ONTAP: For when you need to migrate or use NetApp ONTAP file systems within AWS



Amazon FSx for OpenZFS: When you need support for open-source OpenZFS file systems