

AWS FSx

AWS FSx : Les systèmes de fichiers partagés hautes performances

Le besoin global

Amazon EFS est excellent, mais il ne gère que le protocole POSIX (Linux). Amazon S3 n'est pas un système de fichiers (c'est du stockage objet).

- **Le besoin** : Tu as des applications existantes (legacy) qui ont besoin d'un "vrai" réseau de fichiers partagés (un lecteur réseau `Z:\` sous Windows ou un point de montage spécifique sous Linux) avec des performances extrêmes, sans réécrire le code de l'application.

Est-ce que AWS FSx permet de délocaliser mon système de fichier vers S3 ?

Non ! Quand tu crées un système AWS FSx (que ce soit pour Windows, Lustre, ONTAP ou OpenZFS), AWS provisionne des serveurs de stockage dédiés (généralement des disques SSD ou HDD ultra-rapides) cachés derrière le service.

- **Ce n'est pas du S3.** S3 est un stockage "Objet" (via des clés/valeurs et des requêtes HTTP). FSx est un stockage "Fichiers" traditionnel (avec des dossiers, des sous-dossiers et des protocoles réseau comme SMB ou NFS).
- **Le cas de la VM** : Si tu as une machine virtuelle (EC2), tu détaches ton stockage local ou tu crées simplement un partage réseau, et tu le connectes à FSx. Ta VM écrit directement sur les disques de FSx, pas sur S3.

La confusion vient souvent de **FSx for Lustre**, qui possède une fonctionnalité unique de liaison avec S3 :

- **Le besoin** : Tu as des téraoctets de données brutes stockées à bas coût dans un bucket S3 (par exemple, des images satellites). Tu veux faire un calcul intensif dessus (du Machine Learning).
- **Le fonctionnement** : Tu crées un FSx for Lustre et tu lui dis de se "lier" à ton bucket S3. FSx va alors indexer le bucket S3. Pour tes machines virtuelles, le système FSx semblera contenir tous les fichiers. Dès qu'une VM demandera un fichier, FSx for Lustre ira le chercher de manière transparente dans S3, le copiera sur ses disques SSD ultra-rapides pour que la VM puisse travailler dessus à la vitesse de l'éclair. Une fois le calcul fini, FSx peut renvoyer les résultats dans S3.

A) FSx for Windows File Server

- **Le contexte** : Tu migres une application d'entreprise basée sur Windows (comme un CRM maison, des profils utilisateurs).
- **Points clés SAA-C03** :
 - **NTFS, SMB, AD** : Utilise le système Windows natif (NTFS), le protocole SMB, et s'intègre obligatoirement avec **Microsoft Active Directory** pour la gestion des permissions.
 - **Montage sur Linux EC2 : Oui, c'est possible !** Bien que conçu pour Windows, Linux gère le protocole SMB. Si l'examen te parle d'un stockage partagé entre instances Windows et Linux, FSx for Windows est une option.
 - **Microsoft DFS (Distributed File System)** : Supporté. Permet de regrouper plusieurs serveurs de fichiers en un seul chemin logique et de faire de la réplication.
 - **SSD / HDD** : SSD pour les bases de données et la haute performance (IOPS élevés), HDD pour les partages de fichiers classiques ou l'archivage économique.
 - **Accès depuis l'On-Premise** : Accessible depuis ton centre de données local via un VPN AWS ou AWS Direct Connect.

B) FSx for Lustre

- **Le besoin** : Le High Performance Computing (HPC). Tu as besoin de traiter des volumes gigantesques de données à une vitesse phénoménale (des millions d'IOPS).
- **Le contexte** : Machine Learning (entraînement de modèles), modélisation financière, rendu 3D, analyse génomique.
- **Points clés SAA-C03** :
 - **Intégration transparente avec S3** : C'est sa force. Il peut se lier à un bucket S3. FSx for Lustre va "lire" les données de S3 comme s'il s'agissait de fichiers locaux, faire le calcul ultra-rapide, et réécrire le résultat directement dans S3.
 - **On-Premises** : Utilisable depuis le local via Direct Connect pour du calcul hybride.

Quand tu crées du FSx for Lustre, tu dois choisir entre ces deux modes :

Caractéristique	Scratch File System (Stockage temporaire)	Persistent File System (Stockage long terme)
Le Besoin	Coût minimal pour du calcul brut à court terme.	Données hautement disponibles qui doivent durer dans le temps.
Le Contexte	Tu lances un calcul de 4 heures pour du Machine Learning. Si un serveur crash, tu relances le job, ce n'est pas grave.	Tu as un pipeline de données sensible qui tourne en continu sur plusieurs semaines ou mois.
Comportement	Pas de réplication. Si le matériel lâche, les données non sauvegardées sur S3 sont perdues .	Réplication automatique au sein de la même Zone de Disponibilité (AZ). Remplacement automatique en cas de panne.

C) FSx for NetApp ONTAP

- **Le besoin** : Tu utilises déjà la technologie de stockage **NetApp** dans ton propre centre de données et tu veux migrer vers AWS sans changer tes habitudes ni tes scripts d'administration.

- **Le contexte** : Environnements d'entreprise hautement spécialisés et multi-protocoles.
- **Points clés SAA-C03** :
 - **Multi-protocole** : Compatible simultanément avec NFS (Linux), SMB (Windows) et iSCSI (stockage en bloc pour les VM).
 - **Fonctionnalités avancées** : Supporte la déduplication et la compression (low cost), les **Snapshots** instantanés, et le **Hot Cloning** (cloner un volume de plusieurs To en 2 secondes pour les environnements de dev ou de gaming sans dupliquer l'espace disque).

D) FSx for OpenZFS

- **Le besoin** : Tu migres des serveurs de fichiers basés sur le système de fichiers open-source ZFS vers un service managé.
 - **Le contexte** : Souvent utilisé pour remplacer des serveurs de fichiers Linux customisés gourmands en bande passante.
 - **Points clés SAA-C03** :
 - Compatible **NFS** (optimisé pour Linux/Unix).
 - Offre des performances de type "milliseconde" (très faible latence).
 - Permet le **Point-in-time cloning** (pratique pour copier instantanément des données pour tester un patch d'application sans impacter la prod).
-

Exemples

Scénario 1 : Le Cloud 100% Natif (Partage de fichiers entre VMs)

C'est le cas d'usage le plus classique. Tu as migré tes applications sur AWS, elles tournent sur des instances EC2, et elles ont besoin d'accéder *en même temps* aux mêmes fichiers.

- **Le besoin** : Tu as une application web Windows déployée sur 10 instances EC2 différentes (pour la haute disponibilité). Ces instances doivent partager un dossier unique pour stocker les images de profil des utilisateurs.
- **La solution FSx** : Tu crées un **FSx for Windows File Server**. Les 10 instances EC2 connectent ce lecteur réseau (ex: le lecteur `Z:\`). Quand l'instance n°1 écrit une image, les 9 autres la voient instantanément.
- **Le contexte** : Cloud natif. Pas de serveur On-Premise ici.

Scénario 2 : Le "Lift-and-Shift" (Migration d'applications d'entreprise)

Tu as une application lourde dans ton centre de données local et ton patron te dit : "On déménage tout dans le cloud ce week-end, mais on n'a pas le temps de réécrire le code".

- **Le besoin** : Ton application utilise des protocoles spécifiques (comme SMB pour Windows ou des bases de données SQL Server qui ont besoin de stockage partagé, ou des configurations complexes NetApp ONTAP).
- **La solution FSx** : Au lieu de recréer des serveurs de fichiers complexes sur AWS (ce qui demanderait de gérer des VMs, des disques, de la réplication), tu crées un service **FSx** correspondant. Tu migres tes données du local vers FSx (avec DataSync par exemple), et tu branches ton application dessus. L'application ne voit aucune différence, elle pense être toujours sur son stockage d'origine.
- **Le contexte** : Migration de l'On-Premise vers le Cloud.

Scénario 3 : L'Hybride (Extension du On-Premise)

C'est le scénario où tes serveurs restent en local, mais tu utilises la puissance d'AWS.

- **Le besoin** : Tu as une équipe de chercheurs dans ton laboratoire local. Ils génèrent des données mais tes serveurs locaux n'ont pas la puissance de calcul pour les analyser.
- **La solution FSx (Lustre ou ONTAP)** : Tu crées un système FSx sur AWS connecté à ton centre de données via un câble dédié (**AWS Direct Connect**). Tes chercheurs peuvent écrire directement dans FSx à travers le réseau. Ensuite, des dizaines d'instances EC2 sur AWS récupèrent ces fichiers sur FSx, lancent des calculs massifs d'Intelligence Artificielle, et renvoient le résultat.
- **Le contexte** : Hybride (On-Prem + Cloud).

Revision #3

Created 2026-06-04 07:54:46 UTC by Victor

Updated 2026-06-04 15:49:23 UTC by Victor