AWS Storage Extras

(1) AWS Snow family

Data migration → Snowcone
Snowball edge
Snowmobile

when network cost, limited low, connectivity shared BW so hence data migration

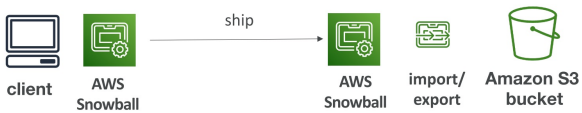
* snow family devices are offline devices.

Edge computing → Snowcone
Snowball edge

{ when it takes more than a week for data transfer use snow family ! }

- Direct upload to S3:

- With Snow Family:



limited or no connectivity or less computing powers.
ships, mines.
pre process, ML, transcode media

(1) Snowball edge

→ move TBs or PBs of data in & out of AWS

→ pay per data transfer → for block & S3 obj storage

snowball edge storage optimized → 80TB HDD, 40 vCPU, 210TB NVMe, 416 GiB RAM

snowball edge compute optimized → 42TB HDD, 104 vCPU, 28TB NVMe, 416 GiB RAM, Optional GPU, storage clustering

→ Large data cloud migrations, DR, DC decommission

② Snowcone & Snowcone SSD

→ few TBs of data. small, portable, drone transportable

Snowcone → 8TB of HDD
Snowcone SSD → 14 TB of SSD
} 2CPUs, 4 GiB mem. Power using USB-C

→ battery & cable powered

→ either send Offline or use "AWS Data sync" to send the data
- pre-installed on given graphical interface
- like a management console

③ Snowmobile { truck }

→ Transfer Exabytes of data { 1000 PBs = 1 EB TBs }

→ each snow mobile has 100 PB of capacity.

→ GPS, Temperature controlled.

Usage process

(1) Request one
(2) install snowball client / AWS OpsHub on your servers. (provides UI)
(3) connect snowball to your servers & copy files using client
(4) ship back to AWS
(5) Data loaded to S3
(6) snowball is wiped.

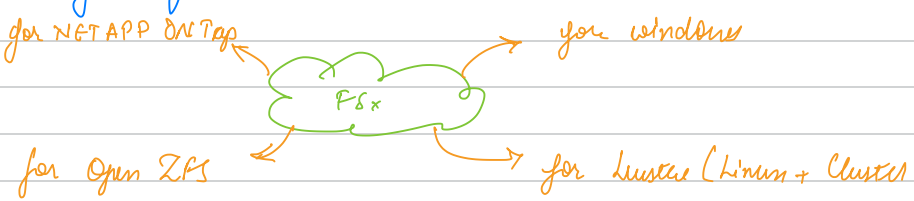
* long term deployment - 1 or 3 y for cost & ball good benifits. { for EC }

Solution Architecture: Snowball into Glacier

- Snowball cannot import to Glacier directly
- You must use Amazon S3 first, in combination with an S3 lifecycle policy

② Amazon FSx

→ AWS managed service to launch 3rd party high performance file systems on AWS.



Amazon FSx for Windows (File Server)

FSx

- FSx for Windows is a fully managed Windows file system share drive (server message Block)
- Supports SMB protocol & Windows NTFS
- Microsoft Active Directory integration, ACLs, user quotas
- Can be mounted on Linux EC2 instances
- Supports Microsoft's Distributed File System (DFS) Namespaces (group files across multiple FS)

- Scale up to 10s of GB/s, millions of IOPS, 100s PB of data
- Storage Options:
  - SSD – latency sensitive workloads (databases, media processing, data analytics, …)
  - HDD – broad spectrum of workloads (home directory, CMS, …)
- Can be accessed from your on-premises infrastructure (VPN or Direct Connect)
- Can be configured to be Multi-AZ (high availability)
- Data is backed-up daily to S3

Amazon FSx for Lustre

FSx

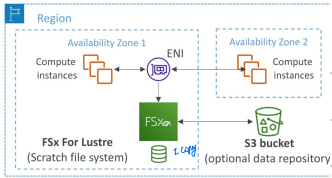
- Lustre is a type of parallel distributed file system, for large-scale computing
- The name Lustre is derived from "Linux" and "cluster"

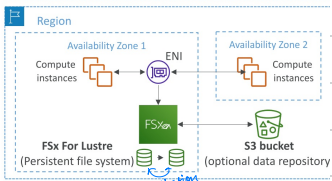
- Machine Learning, High Performance Computing (HPC)
- Video Processing, Financial Modeling, Electronic Design Automation
- Scales up to 100s GB/s, millions of IOPS, sub-ms latencies
- Storage Options:
  - SSD – low-latency, IOPS intensive workloads, small & random file operations
  - HDD – throughput-intensive workloads, large & sequential file operations
- Seamless integration with S3
  - Can "read S3" as a file system (through FSx)
  - Can write the output of the computations back to S3 (through FSx)
- Can be used from on-premises servers (VPN or Direct Connect)

FSx Lustre - File System Deployment Options

- Scratch File System
  - Temporary storage
  - Data is not replicated (doesn't persist if file server fails)
  - High burst (6x faster, 200MBps per TiB)
  - Usage: short-term processing, optimize costs
- Persistent File System
  - Long-term storage
  - Data is replicated within same AZ
  - Replace failed files within minutes
  - Usage: long-term processing, sensitive data

Storage



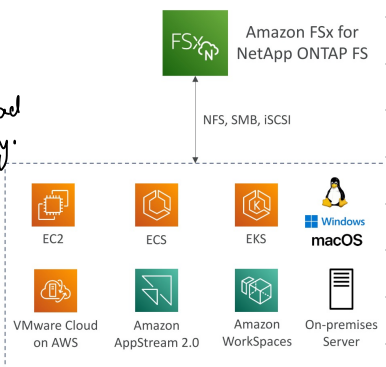


Amazon FSx for NetApp ONTAP

- Managed NetApp ONTAP on AWS
- File System compatible with NFS, SMB, iSCSI protocol
- Move workloads running on ONTAP or NAS to AWS
- Works with:
  - Linux
  - Windows
  - MacOS
  - VMware Cloud on AWS
  - Amazon Workspaces & AppStream 2.0
  - Amazon EC2, ECS and EKS
- Storage shrinks or grows automatically
- Snapshots, replication, low-cost, compression and data de-duplication
- Point-in-time instantaneous cloning (helpful for testing new workloads)

to stage away from production.

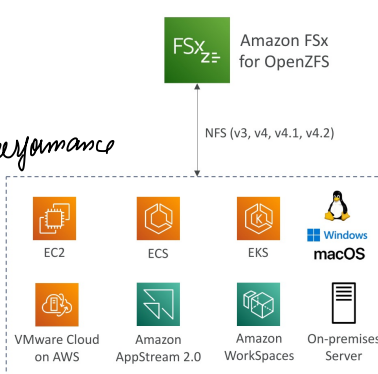
many! broad compatibility.



Amazon FSx for OpenZFS

- Managed OpenZFS file system on AWS
- File System compatible with NFS (v3, v4, v4.1, v4.2)
- Move workloads running on ZFS to AWS
- Works with:
  - Linux
  - Windows
  - MacOS
  - VMware Cloud on AWS
  - Amazon Workspaces & AppStream 2.0
  - Amazon EC2, ECS and EKS
- Up to 1,000,000 IOPS with < 0.5ms latency
- Snapshots, compression and low-cost
- Point-in-time instantaneous cloning (helpful for testing new workloads)

good performance



(3) AWS Storage Gateway { A bridge between on premises & data storage on AWS }

AWS Storage Gateway

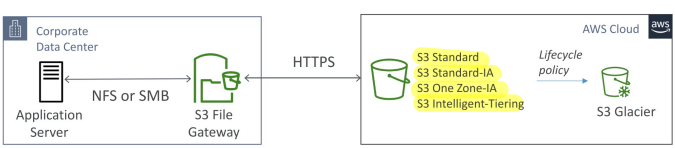
- Bridge between on-premises data and cloud data
- Use cases:
  - disaster recovery
  - backup & restore
  - tiered storage
  - on-premises cache & low-latency files access
- Types of Storage Gateway:
  - S3 File Gateway
  - FSx File Gateway
  - Volume Gateway
  - Tape Gateway

} 4 type of gateway

Amazon S3 File Gateway

- Configured S3 buckets are accessible using the NFS and SMB protocol
- Most recently used data is cached in the file gateway
- Supports S3 Standard, S3 Standard IA, S3 One Zone A, S3 Intelligent Tiering
- Transition to S3 Glacier using a Lifecycle Policy
- Bucket access using IAM roles for each File Gateway
- SMB Protocol has integration with Active Directory (AD) for user authentication

NOT Glacier!



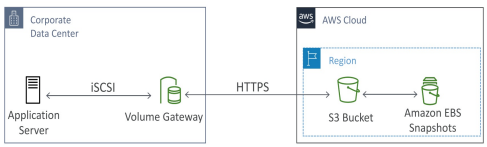
Amazon FSx File Gateway

why use gateway for FSx क्योंकि वो तो ऐसे ही हैं जाएगा !

- Native access to Amazon FSx for Windows File Server
- Local cache for frequently accessed data
- Windows native compatibility (SMB, NTFS, Active Directory...)
- Useful for group file shares and home directories

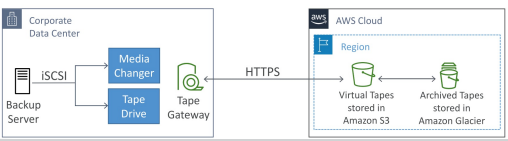
Volume Gateway

- Block storage using iSCSI protocol backed by S3
- Backed by EBS snapshots which can help restore on-premises volumes!
- **Cached volumes:** low latency access to most recent data
- **Stored volumes:** entire dataset is on premise, scheduled backups to S3



Tape Gateway

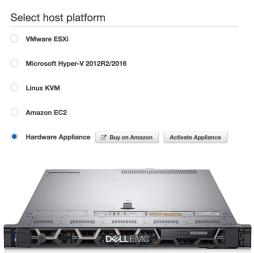
- Some companies have backup processes using physical tapes (!)
- With Tape Gateway, companies use the same processes but, in the cloud
- Virtual Tape Library (VTL) backed by Amazon S3 and Glacier
- Back up data using existing tape-based processes (and iSCSI interface)
- Works with leading backup software vendors



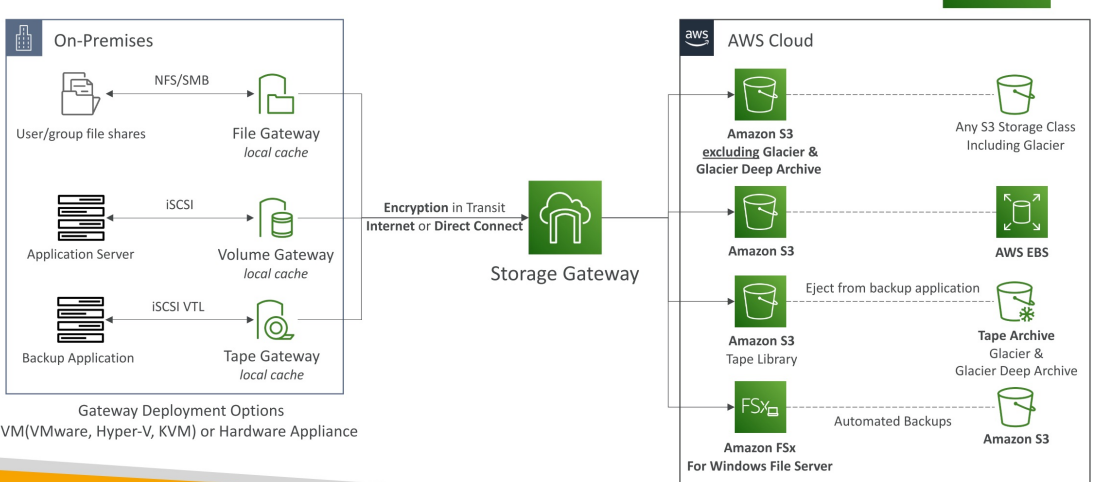
Storage Gateway – Hardware appliance

- Using Storage Gateway means you need on-premises virtualization
- Otherwise, you can use a **Storage Gateway Hardware Appliance**
- You can buy it on amazon.com

- Works with File Gateway, Volume Gateway, Tape Gateway
- Has the required CPU, memory, network, SSD cache resources
- Helpful for daily NFS backups in small data centers



AWS Storage Gateway

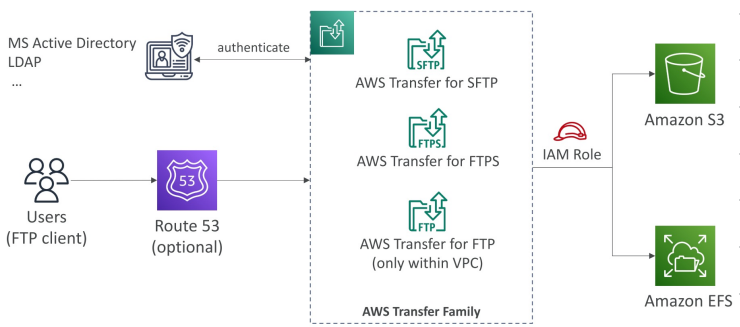


(4) AWS Transfer Family

for having FTP to transfer files

AWS Transfer Family

- A fully-managed service for file transfers into and out of Amazon S3 or Amazon EFS using the FTP protocol
- Supported Protocols
  - AWS Transfer for FTP (File Transfer Protocol (FTP))
  - AWS Transfer for FTPS (File Transfer Protocol over SSL (FTPS))
  - AWS Transfer for SFTP (Secure File Transfer Protocol (SFTP))
- Managed infrastructure, Scalable, Reliable, Highly Available (multi-AZ)
- Pay per provisioned endpoint per hour + data transfers in GB
- Store and manage users' credentials within the service
- Integrate with existing authentication systems (Microsoft Active Directory, LDAP, Okta, Amazon Cognito, custom)
- Usage: sharing files, public datasets, CRM, ERP, …

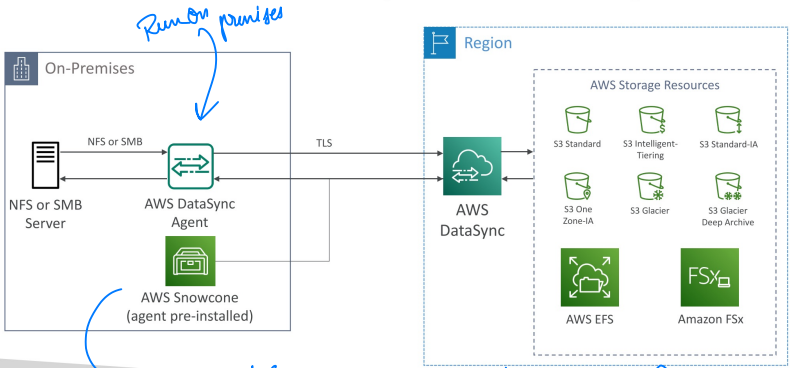


AWS DataSync

- Move large amount of data to and from
  - On-premises / other cloud to AWS (NFS, SMB, HDFS, S3 API…) – needs agent
  - AWS to AWS (different storage services) – no agent needed
- Can synchronize to:
  - Amazon S3 (any storage classes – including Glacier)
  - Amazon EFS
  - Amazon FSx (Windows, Lustre, NetApp, OpenZFS...)
- Replication tasks can be scheduled hourly, daily, weekly
- File permissions and metadata are preserved (NFS POSIX, SMB…)
- One agent task can use 10 Gbps, can setup a bandwidth limit

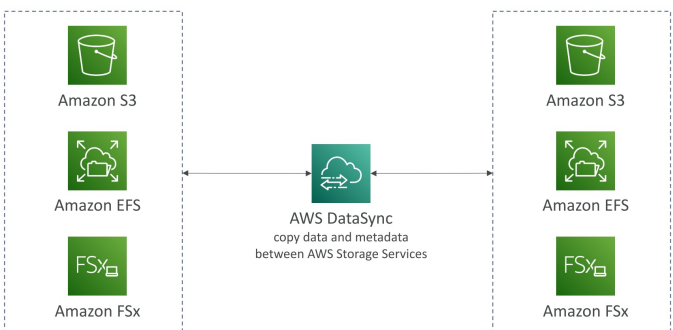
not continuous
(regularly!)

AWS DataSync
NFS / SMB to AWS (S3, EFS, FSx…)



We don't have network capacity to use Data sync so we snowcone

AWS DataSync
Transfer between AWS storage services



Storage Comparison

- **S3:** Object Storage
- **S3 Glacier:** Object Archival
- **EBS volumes:** Network storage for one EC2 instance at a time
- **Instance Storage:** Physical storage for your EC2 instance (high IOPS)
- **EFS:** Network File System for Linux instances, POSIX filesystem
- **FSx for Windows:** Network File System for Windows servers
- **FSx for Lustre:** High Performance Computing Linux file system
- **FSx for NetApp ONTAP:** High OS Compatibility
- **FSx for OpenZFS:** Managed ZFS file system
- **Storage Gateway:** S3 & FSx File Gateway, Volume Gateway (cache & stored), Tape Gateway
- **Transfer Family:** FTP, FTPS, SFTP interface on top of Amazon S3 or Amazon EFS
- **DataSync:** Schedule data sync from on-premises to AWS, or AWS to AWS
- **Snowcone / Snowball / Snowmobile:** to move large amount of data to the cloud, physically
- **Database:** for specific workloads, usually with indexing and querying

hybrid