More Storage Options

SnowFamily

These are secure, portable devices that allows you to college and process data at edge location, AND migrate data in and out of AWS. Two use cases.

It is used to perform offline data migration since if you are going to transfer data using the network it is going to take too much time. If it takes you more than a week to transfer over the network, then you should use snowball devices. You basically get a physical device shipped it to you and then you would upload your data, then ship it back. Then AWS will take your device and deposit the data importing it into S3 bucket.

There are three different device family

Snowball Edge

Allows you to move TBs or PBs of data in or out of AWS. You will be paying per data transfer job.

There are two kinds, snowball edge storage optimized where you get 80 TB of HDD capacity. The other is snowball edge compute optimized where you get 42 TB of HDD. The difference is in the compute power that you get from this machine which we will discuss later, we only focusing on transferring data for now.

Snowcone and Snowcone SSD

Meant for environment where you will only collect little bit of data. Can withstand harsh weather.

Snowcone: 8TB

Snowcone SSD: 14 TB.

You must provide your own battery / cables. You can sent back data by shipping it or connect to using online option via AWS DataSync. Online option is only available for snowcone family

Snowmobile

An actual truck where you can upload your data into to. It can hold exabytes of data! It is a lot of freaking data.

There is GPS, 24/7 video surveillance. Use snowmobile if you need to transfer more than 10 PB.

Edge computing

There are two devices that can also be used to do offline computing in edge location. What's an edge location? Where you aren't able to access the internet or the cloud and have no compute power. Mining station, ship at sea, or truck in road.

Snowcone and Snowball edge are the two devices that can also do edge computing. They can preprocess data, do machine learning, and if you need to transfer the data back to AWS you can ship it right back.

This is good for if you need to work on the data close to the site that you are collecting from.

Snowcone & Snowcone SSD: 2 CPU, 4GB of memory, can have wired and wireless access

Snowball edge compute optimized: 52 CPU! 208 GB of RAM! Even add GPU, but only 42 TB of storage

Snowball edge storage optimized: 40 CPU! 80 GB of RAM.

These two devices can run EC2 instance and AWS lambda using AWS OpsHub (A software you install on your computer and laptop to manage your snow family devices). Use it to spin up EC2 or lambdas.

Pricing is marketed using long-term deployment.

Data to glacier

What if you want to transfer data to glacier, because Snowfamily can only do it into a S3 bucket, you will have to set up lifecycle policy to transfer those data to glacier

Amazon FSx

Managed file system service.

FSx for Windows file server

Fully managed windows file system. Can be mounted even by Linux EC2 instances.

You can access it from on-premise infrastructure by VPN or direct connect. You can also configure to be multi-AZ, backed-up daily to S3.

Storage option: SSD for low latency and HDD.

FSx for Lustre

Distributed file system for large-scale computing. **Used for machine learning and high performance computing!**

FSx for NetApp ONTAP

Use this if you already use ONTAP to move it to cloud.

This works with lots of platform, Linux, Windows, MacOS, very broad compatibility.

Storage shrinks or grow.

FSx for OpenZFS

Compatible only for ZFS.

Move workload already on ZFS.

Also broad compatibility.

Storage gateway

Gives your on-premise data center the access to unlimited cloud storage.

There are four different storage types. If you don't have additional virtualization capability to run these gateway you can order it from Amazon and install it into your server.

S3 file gateway

You will connect your on-premise data server to the S3 bucket. The on-premise server connect to the S3 via S3 file gateway. On-premise will access it using NFS or SMB protocol, but underlying it will be translated to HTTPS request.

The S3 file gateway will also cache mostly used data.

FSx file gateway

You deploy a Window File Server on FSx. You can add a FSx file gateway for caching frequently used data. Give you low latency, the main reason why to add storage gateway.

Volume gateway

On-premise data volume will be transported to S3 bucket that is then backed up into EBS volume snapshots.

Cached volumes: low latency access to most recent data. Data is stored into S3 for low-latency access.

Stored volumes: Entire dataset is on premise, scheduled backup to S3. For disaster recovery.

Tape gateway

Some companies have backup process using physical tapes. Then tape gateway use the same process but back it up to the cloud.

AWS transfer family

Sent file in and out of Amazon S3 or EFS using FTP protocol.

Support FTP, FTPS (adds security on FTP), SFTP (Uses ssh)

Use it to share files, public datasets. Can add authentication for users.

It is a managed infrastructure!

User can access the protocol using the endpoint provided by the family or use your own DNS name using Route 53.

DataSync

Move large amount of data to and from. On-premise / other cloud locations to AWS. An agent is required to move the data.

You can move data from AWS to AWS different storage services. An agent is not needed!

S3, EFS, FSx are all supported. The replication tasks can be scheduled hourly, daily, weekly. Not going to be synchronized immediately.

File permissions and metadata are preserved.

Snowcone have agent pre-installed for transferring the data after you ship it back to AWS.

Revision #1
Created 2023-02-20 20:08:14 UTC by Tamarine
Updated 2023-02-20 22:02:07 UTC by Tamarine