

AWS Cloud practitioner exam cheat sheet

Cloud computing :cloud computing is a remote virtual pool of on-demand shared resources offering compute, storage, database, and network services that can be rapidly deployed at scale.Cloud computing is based on virtualized technology. Typical cloud foundational resources fall under the categories of compute, storage, database, and network. And there are three main cloud deployment models, these being public, private, and hybrid. The main concepts of cloud computing are on-demand, scalability, economies of scale, flexibility, growth, utility-based metering, shared infrastructure, highly available, and security as shared responsibility. And there are three main cloud service models, infrastructure as a service, platform as a service, and software as a service.

Cloud key features:

- On demand resourcing
- Scalability
- Economy of scale
- Ability to Growth
- Utility based metering
- Shared infrastructure
- Highly available
- Security

Cloud service models :

Infrastructure as server (lowest level of customization) example :gmail

Platform as service (greater level of management and)managed by vendor

Software as service (highest level of customization)

There are some other service models but not needed to dig in for this course

Common use of cloud computing :

- a.Migrating of production services to public cloud
- b.Traffic bursting (heavy traffic on some times of year ,pay only for what you use)
- c.Back up /DR (access to unlimited storage)

- d.Web hosting (flexible infrastructure and easy to manage,customisable)
- e.Test/Dev environments(you can use instances when you need and shut them off when you are done)
- f.Proof of concept (show case)
- g.Big data /data manipulation

How Data Center architecture is reflected in the Cloud:

- A.location (regions all over the globe)
- B.physical security (managed by vendor ,end user has no access)
- C.mechanical and electrical infrastructure(generators .. vendors responsibility)
- D.network infrastructure (switches ..software level being replaced by vpc and others,you can only configure it to certain degree,vpc and so on in AWS)
- E.Service (instances or vms)
- F..Storage (unlimited in cloud)

AWS Storage fundamentals :

- A.S3
- b.Glacier
- C.Ec2 instance storage
- D.elastic block storage (EBS)
- E.elastic file system (EFS)
- F.cloudfront
- G.storage gateway
- H.Snowball

A.Amazon s3

- .fully managed objects based storage
- .highly available
- .highly durable

- .very cost effective
- .widely and easily accessible
- .unlimited storage capacity
- .highly salable

Smallest file size supported = 0 bytes
Largest file size supported = 5terabytes
Objects stored in S3 have a durability of 99.999999999%
S3 stores numerous copies of the same data in different AZs
Availability of S3 data objects is 99.99%

S3 Buckets

Objects are stored in S3 buckets
Bucket's name should be unique
Data can be uploaded into the buckets or folders within the buckets
Limitation of 100 buckets per AWS account
Objects have a unique object key identifying that object

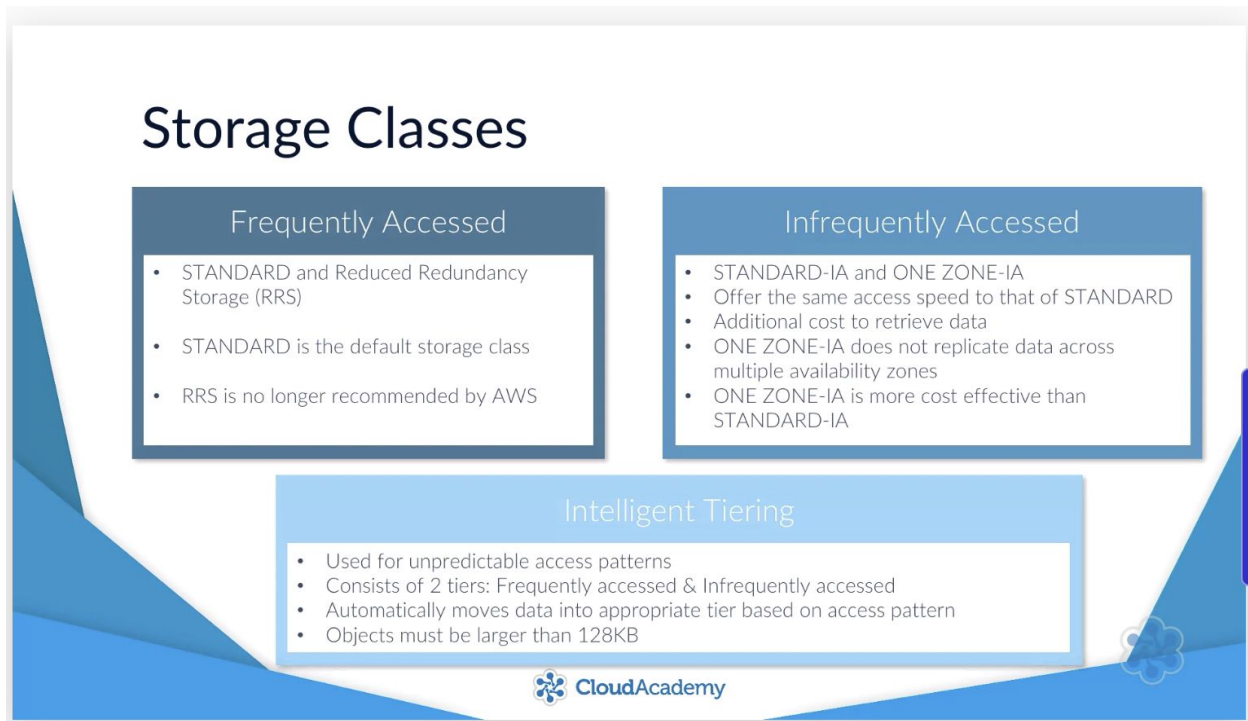
Storage classes in S3

Standard
Standard - IA (infrequent access)
Intelligent Tiering
One zone -IA (infrequent access)
Reduced Redundancy storage (RRS)

Storage Classes

Storage Class	Designed for	Durability (designed for)	Availability (designed for)	Availability Zones	Min storage duration	Min billable object size	Other Considerations
STANDARD	Frequently accessed data	99.999999999%	99.99%	>= 3	None	None	None
STANDARD_IA	Long-lived, infrequently accessed data	99.999999999%	99.9%	>= 3	30 days	128 KB	Per GB retrieval fees apply.
INTELLIGENT_TIERING	Long-lived data with changing or unknown access patterns	99.999999999%	99.9%	>= 3	30 days	None	Monitoring and automation fees per object apply. No retrieval fees.
ONEZONE_IA	Long-lived, infrequently accessed, non-critical data	99.999999999%	99.5%	1	30 days	128 KB	Per GB retrieval fees apply. Not resilient to the loss of the Availability Zone.
RRS (Not recommended)	Frequently accessed, non-critical data	99.99%	99.99%	>= 3	None	None	None

Main difference between classes is durability and availability that each class offers
We can divid storage classes into :
Data accessed frequently and data that accessed infrequently



Before selecting right class you should be asking these questions:

How often is the data likely to be accessed?

How critical is my data?

How reproducible is the data?

Can it be easily created again if need be?

And do I know the access patterns of my data?

Security :

Bucket policy :works as AMI and is written in JAVA

a. ACL:access control list for users outside of AWS account who has access to your bucket

b.Data encryption :server side and client side methods

Data management :

a.Versioning :allows for multiple versions of the same object to exist

Is is created by bucket upon creation (you cant disable versioning)

b.Lifecycle rules :provides an automatic method of managing the lifecycle of your data stored

Amazon S3 is good solution for these scenarios :

A.Data backup :cant be accessible from anywhere

B.static content and websites: S3 is perfect to store static data ,any object can have a URL

- 1..Aamcon cloud front interacts closely with Amazon S3
- 2.Entire static websites can be hosted on Amazon S3
- c.Large data sets
- D.service integration example: EBS can store snap shot on S3 or aws cloudtrail log files are automatically stored on preconfigured S3

Cost of storage depends on :

Region and storage class that you select also Request cost and data transfer cost
Data transfer to S3 is free ,however data transfer to another region can be costly

S3 anti patterns :

- A.data archiving for long term use
- B.dynamic and fast changing data
- C.file system requirements
- D.structured data that needs to be quired

Amazon Glacier

It's an extremely low cost, long-term, durable storage solution which is often referred to as cold storage, ideally suited for long-term backup and archival requirements. It's capable of storing the same data types as Amazon S3 effectively any object. However, it does not provide instant access to your data. In addition to this, there are other fundamental differences which makes this service fit for purpose for other use cases. The service itself again has eleven 9's of durability,it does replicate data within multiple availability zones
Retrieval data can take up to several hours so it is not fast at all
it is similar to S3 and interacts with S3 lifecycle rules but it is much cheaper than S3
It doesn't provide you with instance access to data

Data are stored in Vaults and Archives in glacier

Vaults are container for Glacier archives ,vaults are regional
Archives can be any object similarity to S3 ,you can have unlimited Glacier vaults

Glacier dashboard only allows to create vaults
Any operational process to upload or retrieved has to be using code
API or AWS SDKs

MOving data to glacier is a two step process :
First create your vault by glacier console
Second you have to move your data to vault via API/SDK or by S3 lifecycle rules
For data retrieval you have to use some form of code too

Data retrieval methods from glacier :

Expedited :used for urgent access ,less than 250 MB and it takes 1-5 mins ,cost is per GB and per request

Standard

Use to retrieve any data regardless of size .3-5 hours cost is per GB and per request

Bulk

Used to retrieve petabytes of data ,5-12 hours .cost is per GB and per request

Glacier security :

- a.Data is encrypted by default and
- b.Vault access policies
- c.vault lock policies

Glacier Pricing :

Glacier offers a single storage regardless of amount cost but varies between region

Data going in is free but going out has a different cost from region to region

Amazon Elastic Block storage (EBS)

- a.EBS also provides block level storage to your EC2 instances
- b.offers persistent and durable data storage
- c.Greater flexibility than that of instance store volume
- d.EBS volumes can be attached to EC2 Instances for rapidly changing data
- E.used to retain valuable data due to it's persistent qualities
- F.independent than ec2 ,ebs volumes act as network attached storage devices but each volume can only be attached to one ec2 instance ,on the other hand multiple volumes can be attached to a single ec2 instance
- G.data is retained if the ec2 instance is stopped.restated or terminated

EBs offers the ability to provide back ups known as snap shots

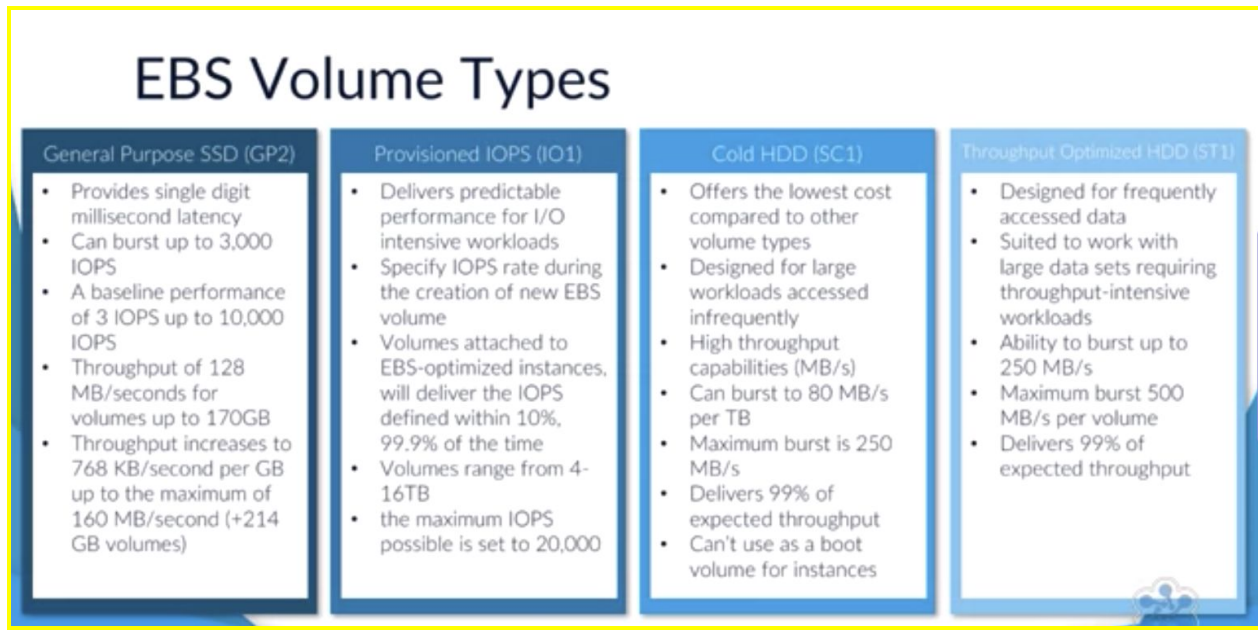
Snapshots store automatically in s3 so they are reliable.

Snapshots will copy the changes only if different from the past snapshot

You can recover a snapshot and attach it to the desired EC2

Is it possible to move snapshot from one region to other

EBS Volume types



EBS encryption

a.EBS offers encryption at rest and in transit (you dont have to worry about encryption keys to perform data encytopin process)

B.Encryption is managed by the EBS service itself

C.itcan be enabled with a checkbox



Any snapshot or volume created will be encrypted. encryption are available on certain instance types .

Two ways of creating new EBS volumes from within the management console :

- 1, during the creation of EC2 instance
2. as stand alone standing EBS volume

Changing size of EBS volume :

- A. volumes are elastically scalable
- B. increase the size via AWS console
- C. after the increase you must extend the file system on the EC2 instance
- D. it is possible to resize volume by creating a new volume from a snapshot .

EBS is not recommended for temporary storage of multi-instance storage access ,very high durability and availability

EBS Pricing :

Pricing



The image shows a pricing table for Amazon EBS. On the left, there is a red icon of a money bag with a white dollar sign inside a circle, and the text 'Amazon EBS' below it. The table has four rows, each representing a different EBS volume type. Each row has a red box containing the volume type name and a light red box containing the pricing information.

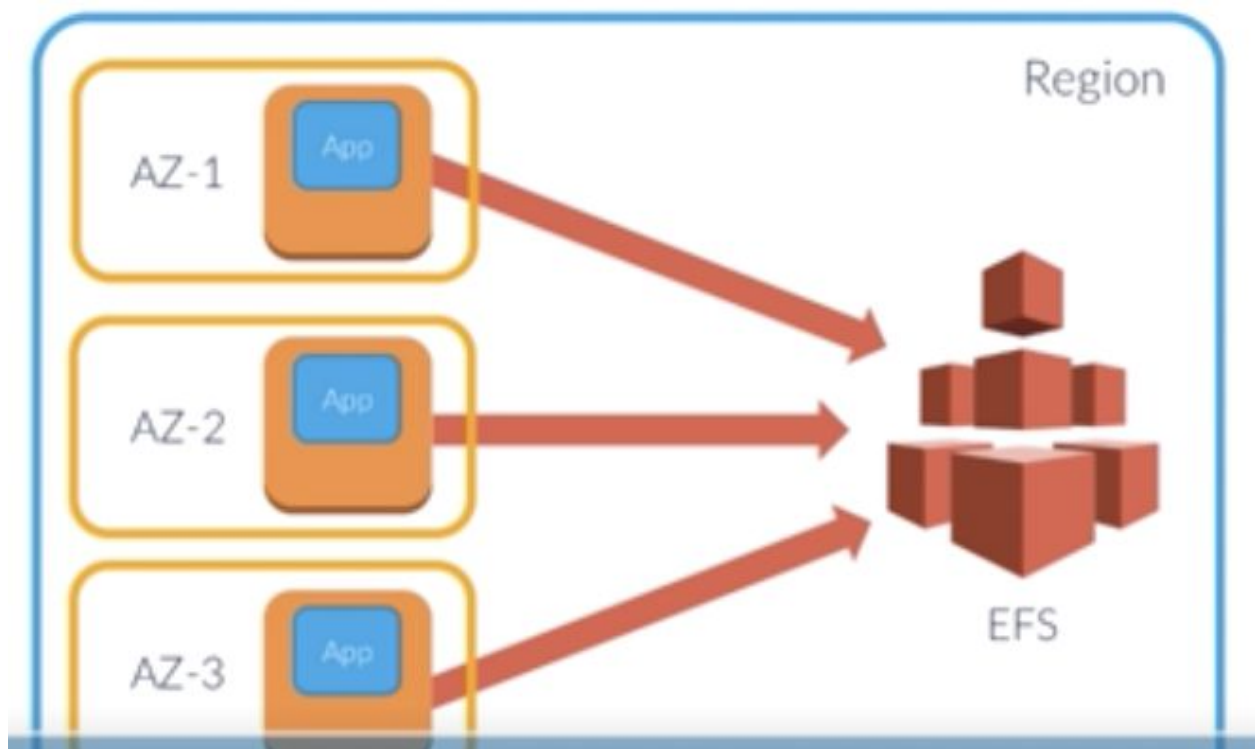
General Purpose SSD (GP2)	\$0.116 per GB-month of provisioned storage
Provisioned IOPS SSD (IO1)	\$0.145 per GB-month of provisioned storage \$0.076 per provisioned IOPS-month
Throughput Optimized HDD (ST1)	\$0.053 per GB-month of provisioned storage
Cold HDD (SC1)	\$0.029 per GB-month of provisioned storage

- EBS Volumes are billed on a per-second basis
- Remember that EBS snapshots are stored on Amazon S3 which will incur standard S3 storage costs

Amazon Elastic file system (EFS)

- a.EFS provides a file level storage service
- b.it is fully managed
- c.Highly available and durable
- d.Ability to create shared file system
- e.Highly scalable
- F.concurrent access by 1000s of instances
- G.limitless storage

EFS is regional



It can be created via console

- A.EFS is only compatible with NFS V4.0 and v4.1
- b.EFS does not support the windows OS

AMAZON RDS

AWS Relational Database Service is a managed database service that lets you focus on building your application storage by taking away the administrative components, such as

backups, patches, and replication. It supports a variety of different relational database builders and it offers a reliable infrastructure for running your own database in multiple availability zones.

RDS also makes Read Replicas possible. Amazon RDS will keep your databases up to date with the latest patches. You can exert optional control over when your instance is patched. Another benefit is Database Event Notifications. So RDS databases can notify you via email or SMS of database events through Amazon SNS, the Simple Notification Service.

So you can use the AWS Management Console or the Amazon RDS APIs to subscribe to over 40 different database events associated with your database instances. Another key benefit is the availability and durability that RDS provides. You get automated backups turned on by default. The automated backup feature of RDS enables point-in-time recovery for your database instances. Amazon RDS will backup your database in transaction logs and store both for a user specified retention period. So this allows you to restore your database instance to any second during your retention period up to the last five minutes. Your automated backup retention period can be configured up to 35 days.

Database snapshots are another benefit. So snapshots are user-initiated backup of your instances stored and they are kept until you explicitly delete them. You can create a new instance from a database snapshot or load database snapshots serve operationally as full backups. They're only built for the incremental storage use.

The other great benefit is Multi AZ Deployments. So Amazon RDS Multi AZ Deployments provide availability and durability for database instances. When you provision a Multi AZ Database Instance, Amazon RDS synchronously replicates the data to a standby instance in a different availability zone. Another benefit is you get automatic host replacement. So Amazon RDS will automatically replace the compute instance powering your deployment in the event of a hardware failure. All very, very useful for highly available fault tolerant solutions.

Another great benefit is the resource-level permissions. So RDS is integrated with AWS Identity and Access Management and provides you with the ability to control the actions that your AWS IAM users and groups can take on specific Amazon RDS resources.

Another key thing to remember about RDS is the Maintenance Window. So RDS performs maintenance on RDS resources for you. It's a managed service.

Dynamo DB

Dynamo DB is a NoSQL key-value store. Table scanning is made possible using secondary indexes based on your application search parameters. You can update streams which allow you to hook them into item label changes. In other use cases, consider Dynamo DB when your application model is schemaless and nonrelational. It can also serve as a persistent session storage mechanism for applications to applications and take away service state.

Elasticache

Elasticache is a managed in-memory cache service for fast, reliable data access. The underlying engines behind ElastiCache are Memcached and Redis. With the Redis engine you can take advantage of multiple availability zones for high availability and scaling to read replicas. ElastiCache will automatically detect failed nodes and replace them without manual intervention. A typical use case for ElastiCache is low latency access of frequently retrieved data. So think cache database results of data within frequent changes for use in a heavily utilized web application for example. It can serve as temporary storage for compute-intensive workloads or when storing the results from IO intense queries or calculations

RedShift

RedShift is a fully managed petabyte-scale data warehouse optimized for fast delivery performance with large data sets. So if you're using HSMs, CloudHSM, and AWSQ management services, you can encrypt your data at rest. RedShift is fully compliant with a variety of compliance standards, including SOC 1, SOC 2, SOC 3, and PCI DSS Level 1. And you can query your data using standard SQL commands through ODBC or JDBC connections. And RedShift integrates with other services, including AWS data pipeline and Kinesis. You can use RedShift to archive large amounts of infrequently used data. When you want to execute analytical queries on large data sets, then RedShift is a really good service.

Amazon Cloudfront

- a.Acts as a content delivery network (CDN)
- b.Distributions data requested through web traffic closer to the end user via edge locations
- C.as the data is cached ,durability of the data is not possible
- D.origin data can be S3

Edge locations

- a.AWS edge locations are sites deployed near highly populated areas across the globe
- B.edge locations are not used to deploy infrastructure (EC2/EBS/etc)
- c.they are used by AWS services such as AWS CloudFront to cache data and reduce latency for end user access

Storage gateway

Storage Gateway allows you to provide a gateway between your own data center's storage systems, such as your SAN, NAS, or DAS, and Amazon S3 in Glacier on AWS.

The storage gateway itself is a software appliance that can be installed within your own data center, which allows integration between your on-premise storage and that of AWS

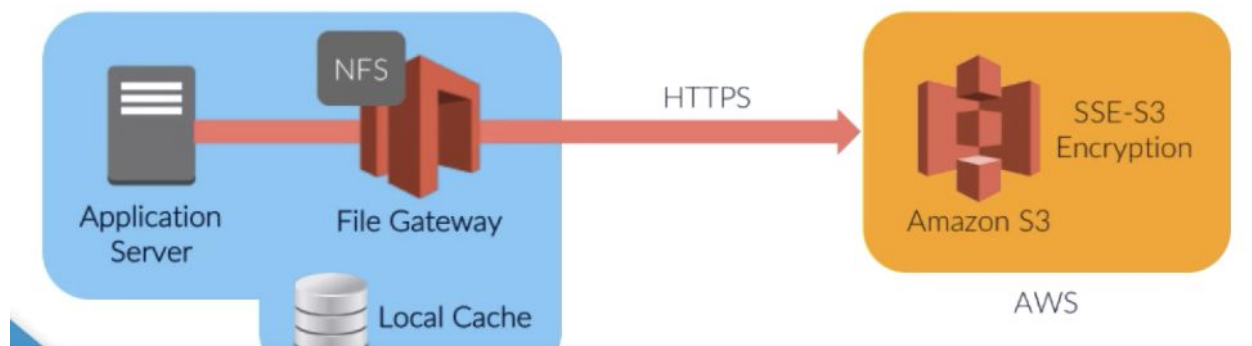
File gateways allow you to securely store your files as objects within S3. Using it as a type of file share which allows you to mount or map drives to an S3 Bucket as if the share was held locally on your own corporate network.

The second option we have as a gateway configuration are volume gateways. And these can be figured in one of two different ways, stored volume gateways and cached volume gateways. Let me explain stored volume gateways first.

Cached volume gateways is data storage provided by AMAZON S3 and they are local data storage that is used for buffering and a local cache for recently accessed data (cached data)

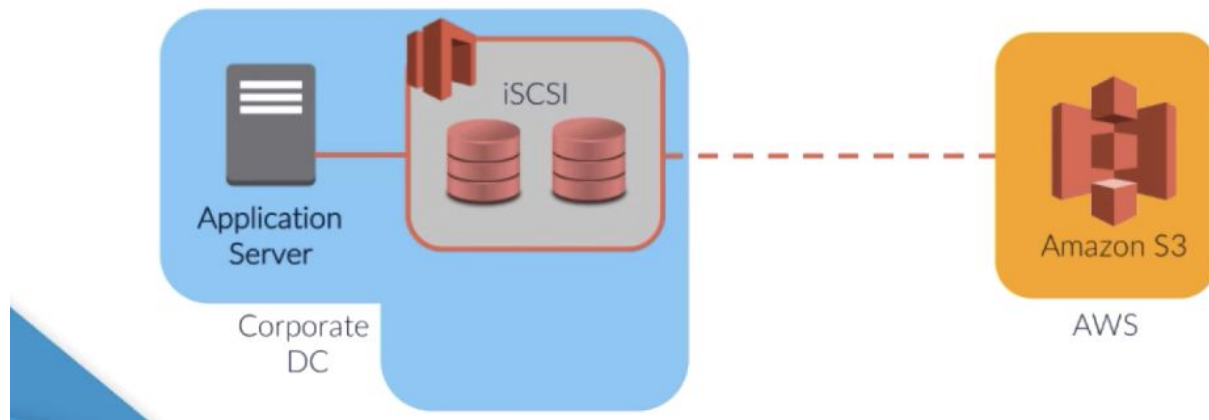
File Gateways

- Allow you to securely store your files as objects within S3
- Ability to mount or map drives to an S3 Bucket as if it was a share held locally



Stored Volume Gateways

- Used to backup your local storage volumes to Amazon S3
- Your entire local data set remains on-premise ensuring low latency data access



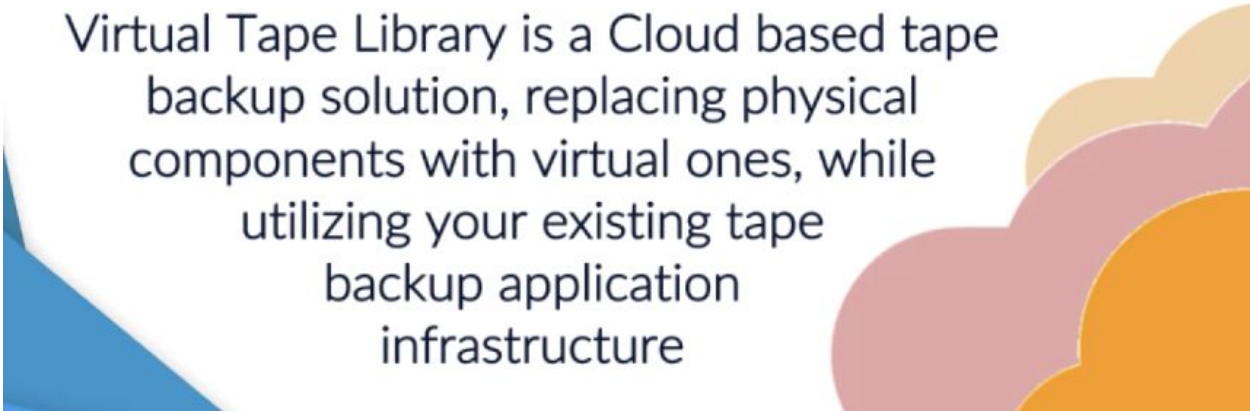
Stored Volume Gateways

- Volumes can be between 1GiB – 16TiB
- Each Storage Gateway can hold 32 volumes
- Maximum storage of 512TiB
- A storage buffer using on-premise storage is used as a staging point for data
- Data is uploaded across an SSL channel and stored in an encrypted form
- Snapshots can be taken of volumes at any point and stored as EBS snapshots on S3

Gateway-Virtual Tape Library

- Allows you to backup data to S3 from your on-premise data center
- Leverage Amazon Glacier for data archiving

Virtual Tape Library is a Cloud based tape backup solution, replacing physical components with virtual ones, while utilizing your existing tape backup application infrastructure



Cached Volume Gateways

- Each Volume can be up to 32TiB
- The Storage Gateway can support 32 Cached Volume Gateways
- Total storage capacity of 1024TiB per Cached Volume Gateway
- Snapshots of volumes stored on S3 as EBS snapshots

Gateway-Virtual Tape Library

- Your applications and backup software can mount tape drives using the media changer as an iSCSI device
- You can create Virtual Tapes as and when you need them to backup your data, which remember are stored on Amazon S3
- Archiving tapes moves data from Amazon S3 to Amazon Glacier

Gateway-Virtual Tape Library

- Your applications and backup software can mount tape drives using the media changer as an iSCSI device
- You can create Virtual Tapes as and when you need them to backup your data, which remember are stored on Amazon S3
- Archiving tapes moves data from Amazon S3 to Amazon Glacier

VTL Components

- **Storage Gateway:** This is configured as a Tape-Gateway acting as a VTL with a capacity of 1500 Virtual Tapes
- **Virtual Tapes:** Virtual equivalent to a physical tape cartridge with capacity of 100GiB – 2.5TiB. Data stored on VT's are backed by Amazon S3 and visible in the Virtual Tape Library
- **Virtual Tape Library (VTL):** Virtual equivalent to a Tape Library containing Virtual Tapes
- **Tape Drives:** Each VTL comes with 10 Tape Drives, presented as iSCSI devices to your backup applications
- **Media Changer:** A virtual device presented as an iSCSI device to backup applications that manages tapes between your Tape Drive and VTL
- **Archive:** Equivalent to an off-site storage facility, giving you the ability to archive tapes from your VTL to Amazon Glacier

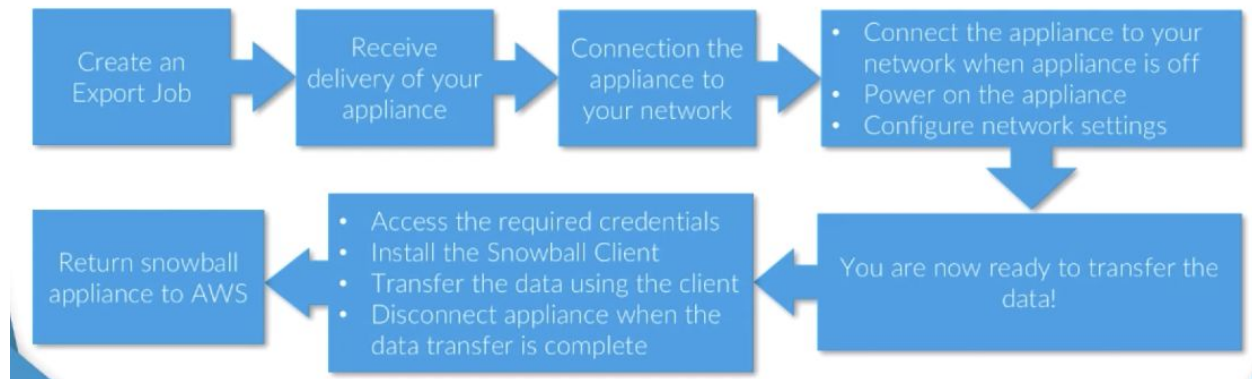
AWS Snowball



- Used to securely transfer large amounts of data in and out of AWS (Petabyte scale)
- Either from your on-premise data center to Amazon S3, or from Amazon S3 back to your data center using a physical appliance, known as a snowball
- The snowball appliance comes as either a 50TB or 80TB device
- The snowball appliance is dust, water and tamper resistant
- Built for high speed data transfer:
 - RJ45 (Cat6)
 - SFP+ Copper
 - SFP+ Optical

- End to end tracking using an E Ink shipping label
- Assists with the delivery to the correct AWS premises
- The Snowball appliance can be tracked with SNS (Simple Notification Service) text messages or via the AWS Management Console
- AWS Snowball is also HIPAA compliant allowing the transfer of protected health data into and out of S3
- Data removal from the appliance is the responsibility of AWS, conforming to NIST standards

AWS Snowball Process



Route 53

Route53 is a highly available domain name service offered by AWS. It supports geographical routing based on an end user's location. It also provides a DNS failover feature which can redirect users to an alternate location if there's an outage. It accomplishes this with health checks that monitor endpoints. Route53 makes it easy to perform a graceful application failover from a dynamic site accessible using an elastic load balancer pointing to a static S3 base site and that's a common use case. Companies can also run multi-region systems using the location-based routing functionality to send users to a region closes to them.