**Storage**

**Amazon S3**

Q: What is Amazon S3?

Amazon S3 is object storage built to store and retrieve any amount of data from anywhere on the Internet. It’s a simple storage service that offers an extremely durable, highly available, and infinitely scalable data storage infrastructure at very low costs.

Q: What can I do with Amazon S3?

Amazon S3 provides a simple web service interface that you can use to store and retrieve any amount of data, at any time, from anywhere on the web. Using this web service, you can easily build applications that make use of Internet storage. Since Amazon S3 is highly scalable and you only pay for what you use, you can start small and grow your application as you wish, with no compromise on performance or reliability.

Amazon S3 is also designed to be highly flexible. Store any type and amount of data that you want; read the same piece of data a million times or only for emergency disaster recovery; build a simple FTP application, or a sophisticated web application such as the Amazon.com retail web site. Amazon S3 frees developers to focus on innovation instead of figuring out how to store their data.

Q: What can developers do with Amazon S3 that they could not do with an on-premises solution?

Amazon S3 enables any developer to leverage Amazon’s own benefits of massive scale with no up-front investment or performance compromises. Developers are now free to innovate knowing that no matter how successful their businesses become, it will be inexpensive and simple to ensure their data is quickly accessible, always available, and secure.

Q: How much data can I store in Amazon S3?

The total volume of data and number of objects you can store are unlimited. Individual Amazon S3 objects can range in size from a minimum of 0 bytes to a maximum of 5 terabytes. The largest object that can be uploaded in a single PUT is 5 gigabytes. For objects larger than 100 megabytes, customers should consider using the [Multipart Upload](http://docs.amazonwebservices.com/AmazonS3/latest/dev/UploadingObjects.html) capability.

Q: What storage classes does Amazon S3 offer?

Amazon S3 offers a range of storage classes designed for different use cases. These include S3 Standard for general-purpose storage of frequently accessed data; S3 Intelligent-Tiering for data with unknown or changing access patterns; S3 Standard-Infrequent Access (S3 Standard-IA) and S3 One Zone-Infrequent Access (S3 One Zone-IA) for long-lived, but less frequently accessed data; and Amazon S3 Glacier (S3 Glacier) and Amazon S3 Glacier Deep Archive (S3 Glacier Deep Archive) for long-term archive and digital preservation.

Q: How is Amazon S3 data organized?

Amazon S3 is a simple key-based object store. When you store data, you assign a unique object key that can later be used to retrieve the data. Keys can be any string, and they can be constructed to mimic hierarchical attributes. Alternatively, you can use S3 Object Tagging to organize your data across all of your S3 buckets and/or prefixes.

Q: How do I interface with Amazon S3?

Amazon S3 provides a simple, standards-based REST web services interface that is designed to work with any Internet-development toolkit. The operations are intentionally made simple to make it easy to add new distribution protocols and functional layers.

**AWS Regions**

Q: Where is my data stored?

You specify an AWS Region when you create your Amazon S3 bucket. For S3 Standard, S3 Standard-IA, and S3 Glacier storage classes, your objects are automatically stored across multiple devices spanning a minimum of three Availability Zones, each separated by miles across an AWS Region. Objects stored in the S3 One Zone-IA storage class are stored redundantly within a single Availability Zone in the AWS Region you select.

Q: What is an AWS Region?

An AWS Region is a geographic location where AWS provides multiple, physically separated and isolated Availability Zones which are connected with low latency, high throughput, and highly redundant networking.

Q: What is an AWS Availability Zone (AZ)?

An AWS Availability Zone is an isolated location within an AWS Region. Within each AWS Region, S3 operates in a minimum of three AZs, each separated by miles to protect against local events like fires, floods, etc.

Amazon S3 Standard, S3 Standard-Infrequent Access, and S3 Glacier storage classes replicate data across a minimum of three AZs to protect against the loss of one entire AZ. This remains true in Regions where fewer than three AZs are publicly available. Objects stored in these storage classes are available for access from all of the AZs in an AWS Region.

The Amazon S3 One Zone-IA storage class replicates data within a single AZ. Data stored in this storage class is susceptible to loss in an AZ destruction event.

Q: How much does Amazon S3 cost?

With Amazon S3, you pay only for what you use. There is no minimum fee. You can estimate your monthly bill using the [AWS Simple Monthly Calculator](http://calculator.s3.amazonaws.com/calc5.html).

We charge less where our costs are less. Some prices vary across Amazon S3 Regions. Billing prices are based on the location of your bucket. There is no Data Transfer charge for data transferred within an Amazon S3 Region via a COPY request. Data transferred via a COPY request between AWS Regions is charged at rates specified in the pricing section of the Amazon S3 detail page. There is no Data Transfer charge for data transferred between Amazon EC2 and Amazon S3 within the same region, for example, data transferred within the US East (Northern Virginia) Region. However, data transferred between Amazon EC2 and Amazon S3 across all other regions is charged at rates specified on the [Amazon S3 pricing page](https://aws.amazon.com/s3/pricing/), for example, data transferred between Amazon EC2 US East (Northern Virginia) and Amazon S3 US West (Northern California).

Q: How secure is my data in Amazon S3?

Amazon S3 is secure by default. Upon creation, only the resource owners have access to Amazon S3 resources they create. Amazon S3 supports user authentication to control access to data. You can use access control mechanisms such as bucket policies and Access Control Lists (ACLs) to selectively grant permissions to users and groups of users. The Amazon S3 console highlights your publicly accessible buckets, indicates the source of public accessibility, and also warns you if changes to your bucket policies or bucket ACLs would make your bucket publicly accessible.

You can securely upload/download your data to Amazon S3 via SSL endpoints using the HTTPS protocol. If you need extra security you can use the Server-Side Encryption (SSE) option to encrypt data stored at rest. You can configure your Amazon S3 buckets to automatically encrypt objects before storing them if the incoming storage requests do not have any encryption information. Alternatively, you can use your own encryption libraries to encrypt data before storing it in Amazon S3.

Q: How can I control access to my data stored on Amazon S3?

Customers may use four mechanisms for controlling access to Amazon S3 resources: Identity and Access Management (IAM) policies, bucket policies, Access Control Lists (ACLs), and Query String Authentication. IAM enables organizations with multiple employees to create and manage multiple users under a single AWS account. With IAM policies, customers can grant IAM users fine-grained control to their Amazon S3 bucket or objects while also retaining full control over everything the users do. With bucket policies, customers can define rules which apply broadly across all requests to their Amazon S3 resources, such as granting write privileges to a subset of Amazon S3 resources. Customers can also restrict access based on an aspect of the request, such as HTTP referrer and IP address. With ACLs, customers can grant specific permissions (i.e. READ, WRITE, FULL\_CONTROL) to specific users for an individual bucket or object. With Query String Authentication, customers can create a URL to an Amazon S3 object which is only valid for a limited time.

Q: What options do I have for encrypting data stored on Amazon S3?

You can choose to encrypt data using SSE-S3, SSE-C, SSE-KMS, or a client library such as the [Amazon S3 Encryption Client](http://docs.amazonwebservices.com/AWSJavaSDK/latest/javadoc/com/amazonaws/services/s3/AmazonS3EncryptionClient.html). All four enable you to store sensitive data encrypted at rest in Amazon S3.

SSE-S3 provides an integrated solution where Amazon handles key management and key protection using multiple layers of security. You should choose SSE-S3 if you prefer to have Amazon manage your keys.

SSE-C enables you to leverage Amazon S3 to perform the encryption and decryption of your objects while retaining control of the keys used to encrypt objects. With SSE-C, you don’t need to implement or use a client-side library to perform the encryption and decryption of objects you store in Amazon S3, but you do need to manage the keys that you send to Amazon S3 to encrypt and decrypt objects. Use SSE-C if you want to maintain your own encryption keys, but don’t want to implement or leverage a client-side encryption library.

SSE-KMS enables you to use [AWS Key Management Service](https://aws.amazon.com/kms/) (AWS KMS) to manage your encryption keys. Using AWS KMS to manage your keys provides several additional benefits. With AWS KMS, there are separate permissions for the use of the master key, providing an additional layer of control as well as protection against unauthorized access to your objects stored in Amazon S3. AWS KMS provides an audit trail so you can see who used your key to access which object and when, as well as view failed attempts to access data from users without permission to decrypt the data.

Q: What is an Amazon VPC Endpoint for Amazon S3?

An Amazon VPC Endpoint for Amazon S3 is a logical entity within a VPC that allows connectivity only to S3. The VPC Endpoint routes requests to S3 and routes responses back to the VPC.

Q: What is Amazon Macie?

Amazon Macie is an [AI-powered security service](https://aws.amazon.com/macie/) that helps you prevent data loss by automatically discovering, classifying, and protecting sensitive data stored in Amazon S3. Amazon Macie uses machine learning to recognize sensitive data such as personally identifiable information (PII) or intellectual property, assigns a business value, and provides visibility into where this data is stored and how it is being used in your organization. Amazon Macie continuously monitors data access activity for anomalies, and delivers alerts when it detects risk of unauthorized access or inadvertent data leaks.

Q: What can I do with Amazon Macie?

You can use Amazon Macie to protect against security threats by continuously monitoring your data and account credentials. Amazon Macie gives you an automated and low touch way to discover and classify your business data. It provides controls via templated Lambda functions to revoke access or trigger password reset policies upon the discovery of suspicious behavior or unauthorized data access to entities or third-party applications. When alerts are generated, you can use Amazon Macie for incident response, using Amazon CloudWatch Events to swiftly take action to protect your data.

Q: How reliable is Amazon S3?

Amazon S3 gives any developer access to the same highly scalable, highly available, fast, inexpensive data storage infrastructure that Amazon uses to run its own global network of web sites. The S3 Standard storage class is designed for 99.99% availability, the S3 Standard-IA storage class is designed for 99.9% availability, and the S3 One Zone-IA storage class is designed for 99.5% availability.

Q: How durable is Amazon S3?

Amazon S3 Standard, S3 Standard–IA and S3 Glacier are all designed to provide 99.999999999% durability of objects over a given year. For S3 One Zone-IA durability is 99.999999999% within an Availability Zone. This durability level corresponds to an average annual expected loss of 0.000000001% of objects. For example, if you store 10,000,000 objects with Amazon S3, you can on average expect to incur a loss of a single object once every 10,000 years. In addition, Amazon S3 Standard, S3 Standard-IA, and S3 Glacier are all designed to sustain data in the event of an entire S3 Availability Zone loss.

As with any environment, the best practice is to have a backup and to put in place safeguards against malicious or accidental deletion. For S3 data, that best practice includes secure access permissions, Cross-Region Replication, versioning, and a functioning, regularly tested backup.

Q: How are Amazon S3 and Amazon S3 Glacier designed to achieve 99.999999999% durability?

Amazon S3 Standard, S3 Standard-IA, and S3 Glacier storage classes redundantly store your objects on multiple devices across a minimum of three Availability Zones (AZs) in an Amazon S3 Region before returning SUCCESS. The S3 One Zone-IA storage class stores data redundantly across multiple devices within a single AZ. These services are designed to sustain concurrent device failures by quickly detecting and repairing any lost redundancy, and they also regularly verify the integrity of your data using checksums.

Q: What is Versioning?

Versioning allows you to preserve, retrieve, and restore every version of every object stored in an Amazon S3 bucket. Once you enable Versioning for a bucket, Amazon S3 preserves existing objects anytime you perform a PUT, POST, COPY, or DELETE operation on them. By default, GET requests will retrieve the most recently written version. Older versions of an overwritten or deleted object can be retrieved by specifying a version in the request.

Q: Why should I use Versioning?

Amazon S3 provides customers with a highly durable storage infrastructure. Versioning offers an additional level of protection by providing a means of recovery when customers accidentally overwrite or delete objects. This allows you to easily recover from unintended user actions and application failures. You can also use Versioning for data retention and archiving.

Q: Can I setup a trash, recycle bin, or rollback window on my Amazon S3 objects to recover from deletes and overwrites?

You can use [Lifecycle rules](http://docs.aws.amazon.com/AmazonS3/latest/dev/object-lifecycle-mgmt.html) along with [Versioning](http://docs.aws.amazon.com/AmazonS3/latest/dev/ObjectVersioning.html) to implement a rollback window for your Amazon S3 objects. For example, with your versioning-enabled bucket, you can set up a rule that archives all of your previous versions to the lower-cost Glacier storage class and deletes them after 100 days, giving you a 100-day window to roll back any changes on your data while lowering your storage costs.

Q: How can I ensure maximum protection of my preserved versions?

Versioning’s [Multi-Factor Authentication (MFA)](https://aws.amazon.com/mfa/) Delete capability can be used to provide an additional layer of security. By default, all requests to your Amazon S3 bucket require your AWS account credentials. If you enable Versioning with MFA Delete on your Amazon S3 bucket, two forms of authentication are required to permanently delete a version of an object: your AWS account credentials and a valid six-digit code and serial number from an authentication device in your physical possession.

**S3 Intelligent-Tiering**

Q: What is S3 Intelligent-Tiering?

Amazon S3 Intelligent-Tiering (S3 Intelligent-Tiering) is an S3 storage class for data with unknown access patterns or changing access patterns that are difficult to learn. It is the first cloud storage class that delivers automatic cost savings by moving objects between two access tiers when access patterns change. One tier is optimized for frequent access and the other lower-cost tier is designed for infrequent access.

Objects uploaded or transitioned to S3 Intelligent-Tiering are automatically stored in the frequent access tier. S3 Intelligent-Tiering works by monitoring access patterns and then moving the objects that have not been accessed in 30 consecutive days to the infrequent access tier. If the objects are accessed later, S3 Intelligent-Tiering moves the object back to the frequent access tier. This means all objects stored in S3 Intelligent-Tiering are always available when needed. There are no retrieval fees, so you won’t see unexpected increases in storage bills when access patterns change.

Q: What performance does S3 Intelligent-Tiering offer?

S3 Intelligent-Tiering provides the same performance as S3 Standard storage. S3 Intelligent-Tiering is designed for the same 99.999999999% durability as S3 Standard. S3 Intelligent-Tiering is designed for 99.9% availability.

Q: How do I get my data into S3 Intelligent-Tiering?

There are two ways to get data into S3 Intelligent-Tiering. You can directly PUT into S3 Intelligent-Tiering by specifying INTELLIGENT\_TIERING in the x-amz-storage-class header or set lifecycle policies to transition objects from S3 Standard or S3 Standard-IA to S3 INTELLIGENT\_TIERING.

Q: Is there a minimum duration for S3 Intelligent-Tiering?

S3 Intelligent-Tiering has a minimum storage duration of 30 days, which means that data that is deleted, overwritten, or transitioned to a different S3 Storage Class before 30 days will incur the normal usage charge plus a pro-rated charge for the remainder of the 30-day minimum.

Q: Can I have a bucket that has different objects in different storage classes?

Yes, you can have a bucket that has different objects stored in S3 Standard, S3 Intelligent-Tiering, S3 Standard-IA, and S3 One Zone-IA.

**S3 Standard-IA**

Q: What is S3 Standard-Infrequent Access?

Amazon S3 Standard-Infrequent Access (S3 Standard-IA) is an Amazon S3 storage class for data that is accessed less frequently but requires rapid access when needed. S3 Standard-IA offers the high durability, throughput, and low latency of the Amazon S3 Standard storage class, with a low per-GB storage price and per-GB retrieval fee. This combination of low cost and high performance make S3 Standard-IA ideal for long-term storage, backups, and as a data store for disaster recovery.

Q: Why would I choose to use S3 Standard-IA?

S3 Standard-IA is ideal for data that is accessed less frequently, but requires rapid access when needed. S3 Standard-IA is ideally suited for long-term file storage, older sync and share storage, and other aging data.

Q: How do I get my data into S3 Standard-IA?

There are two ways to get data into S3 Standard-IA. You can directly PUT into S3 Standard-IA by specifying STANDARD\_IA in the x-amz-storage-class header. You can also set Lifecycle policies to transition objects from the S3 Standard to the S3 Standard-IA storage class.

Q: Is there a minimum storage duration charge for S3 Standard-IA?

S3 Standard-IA is designed for long-lived but infrequently accessed data that is retained for months or years. Data that is deleted from S3 Standard-IA within 30 days will be charged for a full 30 days.

**S3 One Zone-IA**

Q: What is S3 One Zone-IA storage class?

S3 One Zone-IA storage class is an Amazon S3 storage class that customers can choose to store objects in a single availability zone. S3 One Zone-IA storage redundantly stores data within that single Availability Zone to deliver storage at 20% less cost than geographically redundant S3 Standard-IA storage, which stores data redundantly across multiple geographically separate Availability Zones.

Q: What use cases are best suited for S3 One Zone-IA storage class?

Customers can use S3 One Zone-IA for infrequently-accessed storage, like backup copies, disaster recovery copies, or other easily re-creatable data.

**Amazon S3 Glacier**

Q: Why is Amazon Glacier now called Amazon S3 Glacier?

Customers have long thought of Amazon Glacier, our backup and archival storage service, as a storage class of Amazon S3. In fact, a very high percentage of the data stored in Amazon Glacier today comes directly from customers using S3 Lifecycle policies to move cooler data into Amazon Glacier.

Q: Does Amazon S3 provide capabilities for archiving objects to lower cost storage classes?

Yes, Amazon S3 enables you to utilize Amazon S3 Glacier’s extremely [low-cost storage service for data archival](https://aws.amazon.com/glacier/). Amazon S3 Glacier stores data for as little as $0.004 per gigabyte per month. To keep costs low yet suitable for varying retrieval needs, Amazon S3 Glacier provides three options for access to archives, ranging from a few minutes to several hours. Some examples of archive uses cases include digital media archives, financial and healthcare records, raw genomic sequence data, long-term database backups, and data that must be retained for regulatory compliance.

Q: How can I store my data using the Amazon S3 Glacier storage class?

If you have storage which should be immediately archived without delay, or if you make custom business decisions about when to transition objects to S3 Glacier that can’t be expressed through an Amazon S3 Lifecycle policy, S3 PUT to Glacier allows you to use S3 APIs to upload to the S3 Glacier storage class on an object-by-object basis. There are no transition delays and you control the timing. This is also a good option if you want your application to make storage class decisions without having to set a bucket-level policy.

You can use [Lifecycle rules](http://docs.aws.amazon.com/AmazonS3/latest/dev/object-lifecycle-mgmt.html) to automatically archive sets of Amazon S3 objects to S3 Glacier based on object age. Use the Amazon S3 Management Console, the AWS SDKs, or the Amazon S3 APIs to define rules for archival. Rules specify a prefix and time period. The prefix (e.g. “logs/”) identifies the object(s) subject to the rule. The time period specifies either the number of days from object creation date (e.g. 180 days) or the specified date after which the object(s) should be archived. Any S3 Standard, S3 Standard-IA, or S3 One Zone-IA objects which have names beginning with the specified prefix and which have aged past the specified time period are archived to S3 Glacier. To retrieve Amazon S3 data stored in S3 Glacier, initiate a retrieval job via the Amazon S3 APIs or Management Console. Once the retrieval job is complete, you can access your data through an Amazon S3 GET object request.

Q: How can I retrieve my objects that are archived in Amazon S3 Glacier and will I be notified when the object is restored?

To retrieve Amazon S3 data stored in the S3 Glacier storage class, initiate a retrieval request using the Amazon S3 APIs or the Amazon S3 Management Console. The retrieval request creates a temporary copy of your data in the S3 RRS or S3 Standard-IA storage class while leaving the archived data intact in S3 Glacier. You can specify the amount of time in days for which the temporary copy is stored in S3. You can then access your temporary copy from S3 through an Amazon S3 GET request on the archived object.

With restore notifications, you can now be notified with an [S3 Event Notification](https://docs.aws.amazon.com/AmazonS3/latest/dev/NotificationHowTo.html) when an object has successfully restored from S3 Glacier and the temporary copy is made available to you. The bucket owner (or others, as permitted by an [IAM](https://aws.amazon.com/iam/) policy) can arrange for notifications to be issued to [Amazon Simple Queue Service (SQS)](https://aws.amazon.com/sqs/) or [Amazon Simple Notification Service (SNS)](https://aws.amazon.com/sns/). Notifications can also be delivered to [AWS Lambda](https://aws.amazon.com/lambda/) for processing by a Lambda function.

Q: What am I charged for archiving objects in Amazon S3 Glacier?

Amazon S3 Glacier storage class is priced based on monthly storage capacity and the number of Lifecycle transition requests into Amazon S3 Glacier. Objects that are archived to Amazon S3 Glacier have a minimum of 90 days of storage, and objects deleted before 90 days incur a pro-rated charge equal to the storage charge for the remaining days.

**S3 Glacier Deep Archive**

Q: What is S3 Glacier Deep Archive?

S3 Glacier Deep Archive is a new [Amazon S3 storage class](https://aws.amazon.com/s3/storage-classes/) that provides secure and durable object storage for long-term retention of data that is accessed once or twice in a year. From just $0.00099 per GB-month (less than one-tenth of one cent, or about $1 per TB-month), S3 Glacier Deep Archive offers the lowest cost storage in the cloud, at prices significantly lower than storing and maintaining data in on-premises magnetic tape libraries or archiving data off-site.

Q: What use cases are best suited for S3 Glacier Deep Archive?

S3 Glacier Deep Archive is an ideal storage class to provide offline protection of your company’s most important data assets, or when long-term data retention is required for corporate policy, contractual, or regulatory compliance requirements. Customers find S3 Glacier Deep Archive to be a compelling choice to protect core intellectual property, financial and medical records, research results, legal documents, seismic exploration studies, and long-term backups, especially in highly regulated industries, such as Financial Services, Healthcare, Oil & Gas, and Public Sectors. In addition, there are organizations, such as media and entertainment companies, that want to keep a backup copy of core intellectual property. Frequently, customers using S3 Glacier Deep Archive are able to reduce or discontinue the use of on-premises magnetic tape libraries and off-premises tape archival services.

Q: How does S3 Glacier Deep Archive differ from S3 Glacier?

S3 Glacier Deep Archive expands our data archiving offerings, enabling you to select the optimal storage class based on storage and retrieval costs, and retrieval times. Choose S3 Glacier when some of your archived data is needed in as little as 1-5 minutes using Expedited retrievals. S3 Glacier Deep Archive, in contrast, is designed for colder data that is very unlikely to be accessed, but still requires long-term, durable storage. S3 Glacier Deep Archive is up to 75% less expensive than S3 Glacier and provides retrieval within 12 hours using the Standard retrieval speed. You may also reduce retrieval costs by selecting Bulk retrieval, which will return data within 48 hours.

Q: How durable and available is S3 Glacier Deep Archive?

S3 Glacier Deep Archive is designed for the same 99.999999999% durability as the S3 Standard and S3 Glacier storage classes. S3 Glacier Deep Archive is designed for 99.9% availability

Q: Are there minimum storage duration and minimum object storage charges for S3 Glacier Deep Archive?

S3 Glacier Deep Archive is designed for long-lived but rarely accessed data that is retained for 7-10 years or more. Objects that are archived to S3 Glacier Deep Archive have a minimum of 180 days of storage, and objects deleted before 180 days incur a pro-rated charge equal to the storage charge for the remaining days.

S3 Glacier Deep Archive has a minimum billable object storage size of 40KB. Objects smaller than 40KB in size may be stored but will be charged for 40KB of storage.

Q: What is Amazon Athena?

Amazon Athena is an interactive query service that makes it easy to [analyze data in Amazon S3 using standard SQL queries](https://aws.amazon.com/athena/). Athena is serverless, so there is no infrastructure to setup or manage, and you can start analyzing data immediately. You don’t even need to load your data into Athena, it works directly with data stored in any S3 storage class.

Q: What can I do with Amazon S3 event notifications?

Amazon S3 event notifications enable you to run workflows, send alerts, or perform other actions in response to changes in your objects stored in S3. You can use S3 event notifications to set up triggers to perform actions including transcoding media files when they are uploaded, processing data files when they become available, and synchronizing S3 objects with other data stores. You can also set up event notifications based on object name prefixes and suffixes. For example, you can choose to receive notifications on object names that start with “images/."

**Amazon S3 Transfer Acceleration**

Q: What is S3 Transfer Acceleration?

Amazon S3 Transfer Acceleration enables fast, easy, and secure transfers of files over long distances between your client and your Amazon S3 bucket. S3 Transfer Acceleration leverages Amazon CloudFront’s globally distributed AWS Edge Locations. As data arrives at an AWS Edge Location, data is routed to your Amazon S3 bucket over an optimized network path.

Q: How should I choose between S3 Transfer Acceleration and AWS Snow Family (Snowball, Snowball Edge, and Snowmobile)?

The AWS Snow Family is ideal for customers moving large batches of data at once. The AWS Snowball has a typical 5-7 days turnaround time. As a rule of thumb, S3 Transfer Acceleration over a fully-utilized 1 Gbps line can transfer up to 75 TBs in the same time period. In general, if it will take more than a week to transfer over the Internet, or there are recurring transfer jobs and there is more than 25Mbps of available bandwidth, S3 Transfer Acceleration is a good option. Another option is to use both: perform initial heavy lift moves with an AWS Snowball (or series of AWS Snowballs) and then transfer incremental ongoing changes with S3 Transfer Acceleration.

Q: Can S3 Transfer Acceleration complement AWS Direct Connect?

AWS Direct Connect is a good choice for customers who have a private networking requirement or who have access to AWS Direct Connect exchanges. S3 Transfer Acceleration is best for submitting data from distributed client locations over the public Internet, or where variable network conditions make throughput poor. Some AWS Direct Connect customers use S3 Transfer Acceleration to help with remote office transfers, where they may suffer from poor Internet performance.

**S3 object tags**

Q: What are S3 object tags?

S3 object tags are key-value pairs applied to S3 objects which can be created, updated or deleted at any time during the lifetime of the object. With these, you’ll have the ability to create Identity and Access Management (IAM) policies, setup S3 Lifecycle policies, and customize storage metrics. These object-level tags can then manage transitions between storage classes and expire objects in the background.

Q: Why should I use object tags?

Object tags are a tool you can use to enable simple management of your S3 storage. With the ability to create, update, and delete tags at any time during the lifetime of your object, your storage can adapt to the needs of your business. These tags allow you to control access to objects tagged with specific key-value pairs, allowing you to further secure confidential data for only a select group or user. Object tags can also be used to label objects that belong to a specific project or business unit, which could be used in conjunction with S3 Lifecycle policies to manage transitions to other storage classes (S3 Standard-IA, S3 One Zone-IA, and S3 Glacier) or with S3 Cross-Region Replication to selectively replicate data between AWS Regions.

**S3 Lifecycle management**

Q: What is S3 Lifecycle management?

S3 Lifecycle management provides the ability to define the lifecycle of your object with a predefined policy and reduce your cost of storage. You can set a lifecycle transition policy to automatically migrate objects stored in the S3 Standard storage class to the S3 Standard-IA, S3 One Zone-IA, and/or S3 Glacier storage classes based on the age of the data. You can also set lifecycle expiration policies to automatically remove objects based on the age of the object. You can set a policy for multipart upload expiration, which expires incomplete multipart uploads based on the age of the upload.

Q: How can I use Amazon S3 Lifecycle management to help lower my Amazon S3 storage costs?

You could create a rule that archives into S3 Glacier all objects with the common prefix “logs/” 30 days from creation and expires these objects after 365 days from creation. You can also create a separate rule that only expires all objects with the prefix “backups/” 90 days from creation. S3 Lifecycle policies apply to both existing and new S3 objects, helping you optimize storage and maximize cost savings for all current data and any new data placed in S3 without time-consuming manual data review and migration. Within a lifecycle rule, the prefix field identifies the objects subject to the rule. To apply the rule to an individual object, specify the key name. To apply the rule to a set of objects, specify their common prefix (e.g. “logs/”). You can specify a transition action to have your objects archived and an expiration action to have your objects removed. For time period, provide the creation date (e.g. January 31, 2015) or the number of days from creation date (e.g. 30 days) after which you want your objects to be archived or removed. You may create multiple rules for different prefixes.

Q: Why would I use an S3 Lifecycle policy to expire incomplete multipart uploads?

The S3 Lifecycle policy that expires incomplete multipart uploads allows you to save on costs by limiting the time non-completed multipart uploads are stored. For example, if your application uploads several multipart object parts, but never commits them, you will still be charged for that storage. This policy can lower your S3 storage bill by automatically removing incomplete multipart uploads and the associated storage after a predefined number of days.

**S3 Cross-Region Replication (CRR)**

Q: What is Amazon S3 Cross-Region Replication (CRR)?

CRR is an Amazon S3 feature that automatically replicates data between AWS Regions. With CRR, you can set up replication at a bucket level, a shared prefix level, or an object level using S3 object tags. You can use CRR to provide lower-latency data access in different geographic regions. CRR can also help if you have a compliance requirement to store copies of data hundreds of miles apart.

Q: How can I use CRR Replication rules?

Now, you can establish S3 Cross-Region Replication rules to make direct copies of data into the S3 Glacier storage class in a different region for backup or other purposes, without having to manage data lifecycle policies.

**IPv6**

Q: What is IPv6?

Every server and device connected to the Internet must have a unique address. Internet Protocol Version 4 (IPv4) was the original 32-bit addressing scheme. However, the continued growth of the Internet means that all available IPv4 addresses will be utilized over time. Internet Protocol Version 6 (IPv6) is the new addressing mechanism designed to overcome the global address limitation on IPv4.

Q: What can I do with IPv6?

Using IPv6 support for Amazon S3, applications can connect to Amazon S3 without the need for any IPv6 to IPv4 translation software or systems. You can meet compliance requirements, more easily integrate with existing IPv6-based on-premises applications, and remove the need for expensive networking equipment to handle the address translation. You can also now utilize the existing source address filtering features in IAM policies and bucket policies with IPv6 addresses, expanding your options to secure applications interacting with Amazon S3.

**AWS Snowball**

Q: What is AWS Snowball?

[AWS Snowball](https://docs.aws.amazon.com/snowball/latest/ug/whatissnowball.html) is a data transport solution that accelerates moving terabytes to petabytes of data into and out of AWS using storage devices designed to be secure for physical transport. Using Snowball helps to eliminate challenges that can be encountered with large-scale data transfers including high network costs, long transfer times, and security concerns.

Q: How does Snowball work?

AWS Snowball uses devices designed to be secure and the Snowball Client to accelerate petabyte-scale data transfers into and out of AWS. You start by using the AWS Management Console to create one or more jobs to request one or multiple Snowball devices (depending on how much data you need to transfer), and download and install the Snowball Client ("Client"). Once the device arrives, connect it to your local network, set the IP address either manually or with DHCP, and use the Client to identify the directories you want to copy. The Client will automatically encrypt and copy the data to the device and notify you when the transfer job is complete. When the transfer is complete and the device is ready to be returned, the E Ink shipping label will automatically update to indicate the correct AWS facility to ship to, and you can track the job status by using [Amazon Simple Notification Service (SNS)](https://aws.amazon.com/sns/), text messages, or directly in the console.

Q: Who should use Snowball?

Snowball is the right data transfer choice if you need to more securely and quickly transfer terabytes to many petabytes of data to AWS. Snowball can also be the right choice if you don’t want to make expensive upgrades to your network infrastructure, if you frequently experience large backlogs of data, if you're located in a physically isolated environment, or if you're in an area where high-bandwidth Internet connections are not available or cost-prohibitive.

Q: How much data can I transfer using Snowball?

You can transfer virtually any amount of data with Snowball using multiple devices in parallel or serially one after another. For example, move 150 TB at one time with two 80TB devices, or order a single device to move 80TB and then order a second device for the remaining 70TB.

Q: When should I consider using Snowball instead of AWS Direct Connect?

[AWS Direct Connect](https://aws.amazon.com/directconnect/) provides you with dedicated, fast connections from your premises to the AWS network. If you need to transfer large quantities of data to AWS on an ongoing basis, AWS Direct Connect might be the right choice.

Q: Can I use Snowball to migrate data from one AWS region to another AWS region?

No. Snowball is intended to serve as a data transport solution for moving high volumes of data into and out of a designated AWS region. For use cases that require data transfer between AWS regions, we recommend using [S3 Cross-Region Replication](https://docs.aws.amazon.com/AmazonS3/latest/dev/crr.html) as an alternative.

**Snapshots**

Q: Will I be able to access my snapshots using the regular Amazon S3 API?

No, snapshots are only available through the Amazon EC2 API.

Q: Do volumes need to be un-mounted to take a snapshot?

No, snapshots can be done in real time while the volume is attached and in use. However, snapshots only capture data that has been written to your Amazon EBS volume, which might exclude any data that has been locally cached by your application or OS. To ensure consistent snapshots on volumes attached to an instance, we recommend detaching the volume cleanly, issuing the snapshot command, and then reattaching the volume. For Amazon EBS volumes that serve as root devices, we recommend shutting down the machine to take a clean snapshot.

Q: Does it take longer to snapshot an entire 16 TB volume as compared to an entire 1 TB volume?

By design, an EBS Snapshot of an entire 16 TB volume should take no longer than the time it takes to snapshot an entire 1 TB volume. However, the actual time taken to create a snapshot depends on several factors including the amount of data that has changed since the last snapshot of the EBS volume.

**Amazon EBS encryption**

Q: What is Amazon EBS encryption?

Amazon EBS encryption offers seamless encryption of EBS data volumes, boot volumes and snapshots, eliminating the need to build and maintain a secure key management infrastructure. EBS encryption enables data at rest security by encrypting your data using Amazon-managed keys, or keys you create and manage using the [AWS Key Management Service](https://aws.amazon.com/kms/) (KMS). The encryption occurs on the servers that host EC2 instances, providing encryption of data as it moves between EC2 instances and EBS storage.

Q: What is the AWS Key Management Service (KMS)?

AWS KMS is a managed service that makes it easy for you to create and control the encryption keys used to encrypt your data. AWS Key Management Service is integrated with other AWS services including Amazon EBS, Amazon S3, and Amazon Redshift, to make it simple to encrypt your data with encryption keys that you manage. AWS Key Management Service is also integrated with AWS CloudTrail to provide you with logs of all key usage to help meet your regulatory and compliance needs.

**AWS Storage Gateway**

Q: What is the AWS Storage Gateway service?

The AWS Storage Gateway service enables hybrid storage between on-premises environments and the AWS Cloud. It seamlessly integrates on-premises enterprise applications and workflows with Amazon’s block and object cloud storage services through industry standard storage protocols. It provides low-latency performance by caching frequently accessed data on premises, while storing data securely and durably in Amazon cloud storage services. It provides an optimized data transfer mechanism and bandwidth management, which tolerates unreliable networks and minimizes the amount of data being transferred. It brings the security, manageability, durability, and scalability of AWS to existing enterprise environments through native integration with AWS encryption, identity management, monitoring, and storage services. Typical use cases include backup and archiving, disaster recovery, moving data to S3 for in-cloud workloads, and tiered storage.

AWS Storage Gateway supports three storage interfaces: file, volume, and tape. Each gateway you have can provide one type of interface.

The [*file gateway*](https://aws.amazon.com/storagegateway/file/) *enables you to store and retrieve objects in Amazon S3 using file protocols, such as NFS. Objects written through file gateway can be directly accessed in S3.*

The [*volume gateway*](https://aws.amazon.com/storagegateway/features/#volume-gateway) *provides block storage to your applications using the iSCSI protocol. Data on the volumes is stored in Amazon S3. To access your iSCSI volumes in AWS, you can take EBS snapshots which can be used to create EBS volumes.*

The [*tape gateway*](https://aws.amazon.com/storagegateway/vtl/) *provides your backup application with an iSCSI virtual tape library (VTL) interface, consisting of a virtual media changer, virtual tape drives, and virtual tapes. Virtual tape data is stored in Amazon S3 or can be archived to Amazon Glacier.*

Q: How do I use the AWS Storage Gateway service?

You can have two touchpoints to use the service: the AWS Management Console and a gateway that is available as a virtual machine (VM) or as a physical hardware appliance.

You use the AWS Management Console to download the virtual appliance gateway or purchase the hardware appliance, configure storage, and manage and monitor the service. The gateway connects your applications to AWS storage by providing standard storage interfaces. It provides transparent caching, efficient data transfer, and integration with AWS monitoring and security services.

To get started, [sign up for an AWS account](https://console.aws.amazon.com/storagegateway/home) and visit [the AWS Storage Gateway Management Console](https://console.aws.amazon.com/storagegateway/home) to download a gateway VM appliance, or purchase the hardware appliance. Once you’ve installed your gateway, you associate it with your AWS Account through our activation process. After activation, you configure the gateway to connect to the appropriate storage type. For file gateway, you configure file shares that are mapped to selected S3 buckets, using IAM roles. For volume gateway, you create and mount volumes as iSCSI devices. For tape gateway, you connect your backup application to create and manage tapes. Once configured, you start using the gateway to write and read data to and from AWS storage. You can monitor the status of your data transfer and your storage interfaces through the AWS Management Console. Additionally, you can use the API or SDK to programmatically manage your application’s interaction with the gateway.

Q: What is file gateway?

File gateway presents a file-based interface to Amazon S3, which appears as a network file share. It enables you to store and retrieve Amazon S3 objects through standard file storage protocols. File gateway allows your existing file-based applications or devices to use secure and durable cloud storage without needing to be modified. With file gateway, your configured S3 buckets will be available as Network File System (NFS) mount points or Server Message Block (SMB) file shares. Your applications read and write files and directories over NFS or SMB, interfacing to the gateway as a file server. In turn, the gateway translates these file operations into object requests on your S3 buckets. Your most recently used data is cached on the gateway for low-latency access, and data transfer between your data center and AWS is fully managed and optimized by the gateway. Once in S3, you can access the objects directly or manage them using features such as S3 Lifecycle Policies, object versioning, and cross-region replication. You can run file gateway on-premises or in EC2.

Q: What is volume gateway?

Volume gateway provides an iSCSI target, which enables you to create block storage volumes and mount them as iSCSI devices from your on-premises or EC2 application servers. The volume gateway runs in either a cached or stored mode.

* In the cached mode, your primary data is written to S3, while retaining your frequently accessed data locally in a cache for low-latency access.
* In the stored mode, your primary data is stored locally and your entire dataset is available for low-latency access while asynchronously backed up to AWS.

In either mode, you can take point-in-time snapshots of your volumes, which are stored as Amazon EBS Snapshots in AWS, enabling you to make space-efficient versioned copies of your volumes for data protection, recovery, migration and various other copy data needs.

Q: What is tape gateway?

Tape gateway is a cloud-based Virtual Tape Library (VTL). It presents your backup application with a VTL interface, consisting of a media changer and tape drives. You can create virtual tapes in your virtual tape library using the AWS Management Console. Your backup application can read data from or write data to virtual tapes by mounting them to virtual tape drives using the virtual media changer. Virtual tapes are discovered by your backup application using its standard media inventory procedure. Virtual tapes are available for immediate access and are backed by Amazon S3. You can also archive tapes. Archived tapes are stored in Amazon S3 Glacier or Amazon S3 Glacier Deep Archive.

Q: What benefits does AWS Storage Gateway provide?

AWS Storage Gateway provides a set of features that enable you to effectively leverage AWS storage within your existing applications and workflows. It provides a standard set of protocols such as iSCSI, SMB and NFS, which allow you to use your existing applications without any changes. Through its local cache, the gateway provides low-latency access to recently used data. The gateway optimizes data transfer to AWS storage, such as optimization of transfer through intelligent buffering, upload management to address network variations, and bandwidth management. The gateway provides you an effective mechanism to store data in AWS across the range of storage services most suitable for your use cases. The gateway is easy to deploy and can use your existing virtual infrastructure and hypervisor investments, or can be installed in your data center or remote offices as a hardware appliance. The gateway software running as a VM or on the hardware appliance is stateless, allowing you to easily create and manage new instances of your gateway as your storage needs evolve. Finally, the service integrates natively into AWS management services such as Amazon CloudWatch, AWS CloudTrail, AWS Key Management Service (KMS), and AWS Identity and Access Management (IAM).

Q: What performance can I expect?

The AWS Storage Gateway sits between your applications and Amazon storage services. The performance you experience depends on what host platform (hardware appliance, virtual machine, Amazon EC2 instance) you are using to run Storage Gateway software and a number of other factors. These include the network bandwidth between your iSCSI initiator or NFS client and gateway, the speed and configuration of your underlying local disks, the configuration of your VM, the amount of local storage allocated to your gateway, and the bandwidth between your gateway and Amazon storage.

**File gateway**

Q: What is file gateway?

[File Gateway](https://aws.amazon.com/storagegateway/file/) is a configuration of the AWS Storage Gateway service that provides your applications a file interface to seamlessly store files as objects in Amazon S3, and access them using industry standard file protocols.

Q: What can I do with file gateway?

Use cases for file gateway include: (a) migrating on-premises file data to Amazon S3, while maintaining fast local access to recently accessed data, (b) Backing up on-premises file data as objects in Amazon S3 (including Microsoft SQL Server and Oracle databases and logs), with the ability to use S3 capabilities such as lifecycle management, versioning and cross region replication, and, (c) Hybrid cloud workflows using data generated by on-premises applications for processing by AWS services such as machine learning, big data analytics or serverless functions.

Q: What are the benefits of using file gateway to store data in S3?

File gateway enables your existing file-based applications, devices, and workflows to use Amazon S3, without modification. File gateway securely and durably stores both file contents and metadata as objects, while providing your on-premises applications low-latency access to cached data.

**Volume gateway**

Q: How much volume data can I manage per gateway? What is the maximum size of a volume?

Each *volume gateway* can support up to 32 volumes. In *cached mode*, each volume can be up to 32 TB for a maximum of 1 PB of data per gateway (32 volumes, each 32 TB in size). In *stored mode*, each volume can be up to 16 TB for a maximum of 512 TB of data per gateway (32 volumes, each 16 TB in size). For more information, please refer to our [documentation on Storage Gateway limits](https://docs.aws.amazon.com/storagegateway/latest/userguide/resource-gateway-limits.html).

Volume gateways compress data before that data is transferred to AWS and while stored in AWS. This compression can reduce both data transfer and storage charges. Volume storage is not pre-provisioned; you will be billed for only the amount of data stored on the volume, not the size of the volume you create.

Q: When I look in Amazon S3 why can’t I see my volume data?

Your volumes are stored in an Amazon S3 bucket maintained by the AWS Storage Gateway service. Your volumes are accessible for I/O operations through AWS Storage Gateway. You cannot directly access them using Amazon S3 API actions. You can take point-in-time snapshots of gateway volumes that are made available in the form of Amazon EBS snapshots, which can be turned into either Storage Gateway Volumes or EBS Volumes. Use the file gateway to work with your data natively in S3.

**Tape gateway**

Q: What are the benefits of storing virtual tapes in AWS compared to warehousing tapes offsite?

You get 11 9s of data durability, fixity checks by AWS on a regular basis, data encryption, right data when you restore, and cost savings, when storing virtual tapes in AWS using Tape Gateway with S3 Glacier Deep Archive compared to warehousing physical tapes offsite. First, all virtual tapes stored in S3 Glacier Deep Archive are replicated and stored across at least three geographically-dispersed Availability Zones, protected by 11 9s of durability. Second, AWS performs fixity checks on a regular basis to confirm your data can be read and no errors have been introduced. Third, all tapes stored in S3 Glacier Deep Archive are protected by S3 Server Side Encryption using default keys or your KMS keys. In addition, you also avoid physical security risk associated with tape portability. Fourth, compared to the experience of warehousing tapes offsite where you may receive an incorrect or broken tape during restore, with Tape Gateway, you always get correct data. Finally, you can save in monthly storage costs when storing your data in S3 Glacier Deep Archive compared to warehousing tapes offsite.