CloudBerry Lab

#1 Cross-Platform Cloud Backup

A Complete Guide for Backup and DR on AWS

Amazon Web Services (AWS) cloud computing resources, networking facilities, and multi-user applications reduce the amount of financial and business resources spent maintaining your business IT infrastructure. Off-site cloud storage further protects your data against disaster. This whitepaper reviews best practices for backup and disaster recovery on the AWS Cloud.

About AWS

AWS has regions world-wide. Each region is a separate geographic area and is completely independent from other regions. This provides most customers with local region access, which is ideal for compliance and optimal speed.

Each region is composed of multiple, isolated locations known as Availability Zones which are connected to one another through a low-latency link. If your instances leverage multiple Availability Zones and one instance fails, you can design run your workloads from another Availability Zone. Same goes for storage.
If your storage uses multiple Availability Zones, you have 3 copies of your data for added resilience.

AWS uses a pay-as-you-go pricing model. You are only charged for the resources used. Because the AWS Cloud is highly scalable, you can increase workloads from simple storage to virtual servers, databases, and application workloads, as needed.

RPO and RTO

There are two terms that help determine the level of protection needed for backup and recovery: Recovery Point Objective (RPO) and Recovery Time Objective (RTO). RPO is the amount of data a company can afford to lose after an outage and resulting recovery. RTO defines the amount of time a company can afford to wait before the data is recovered.

Naturally, businesses want to get their information restored as quickly as possible and keep data loss to a minimum. To do this, you need reduce your RTO, possibly by finding ways to restore data more quickly, and you need to reduce your RPO to minimize data loss by running backups more frequently. Determining the ideal RPO and RTO is a discussion and negotiation that needs to be had with the business units that rely on the applications that use the data since there are costs for reducing both. Finding the right balance is key.

3-2-1 Backup Strategy

The simplest backup strategies involve making local backups to network storage. Local backups can be fast to restore because of the low-latency, fast local network speeds. But there is an inherent risk of a local storage server crash or disaster (fire, flood, earthquake) that can make restores impossible. It's never ideal to only store backups in the same location as the live data.

Some choose to move backups offsite. But offsite solutions like tape can be very slow to restore and require third-party services to pick up and drop off tapes daily.

For companies that need the security and protection of offsite cloud storage and the speed of local backups, a **3-2-1 backup strategy** is recommended:

- Have at least 3 copies of your data: The original data and two backups
- Keep two copies of your data on two different types of media: Protects from disaster as you can lose one backup (local or offsite) and still be able to restore from the other
- Store one copy of your data offsite: Protects from disaster in the local data center

Offsite storage should be:

- Reliable
- Able to store any kind of data
- Located away from your current location
- Able to be accessed when needed

Types of Backup

Companies may need a variety of backups performed on their data in order to meet recovery demands. Here are some common backup types and how best to utilize them.

File-Level Backup

File-level backup simply backs up a selection of folders / files. Think of file backup as a way to protect your user-generated content, like office documents, PDFs, music, video, and photos, whether that content is located on user desktop/laptops, servers, NAS, or SAN. Backups may keep multiple versions of file backups and allow for a restore of a previous version when going back to an old copy is needed.

The best solution for keeping your user data safe is file-level backup. But when you need to protect more than individual files, image-based backup may be a better option.

Image-Based Backup

Unlike file-level backup, image-based backup protects entire disk volumes (physical or virtual), and usually include backing up the operating system. Image-based backups are usually performed on servers, but can be done on desktop operating systems as well. In case of a server failure (hardware issue or otherwise), image-based backups can restore the entire server to another server (physical or virtual). By restoring the disks themselves, the operating system, applications, and user files are restored together, minimizing disaster recovery downtime. Image-level backups can also be used to restore individual folders / files when an entire disk volume is not needed for restore.

Flexible restore options allow restores to a variety of targets:

- Restore as a virtual machine in the cloud (Amazon EC2)
- Restore with a USB flash drive directly from the cloud
- Restore to the same or dissimilar hardware
- Restore to a Microsoft Hyper-V or VMware virtual machine

Application and Database Backup

Your applications and databases may have specific backup mechanisms (APIs, programs, etc.) to ensure backups are performed in a reliable way. As an example, relational databases like Microsoft SQL Server include a set of backup commands that execute a variety of backup types to address RPOs.

- **Full Backup:** Creates a complete backup of the database. Full backups are the most time-consuming backups to run, but they are necessary and provide a base for future differential and transaction log restored. When restoring your database, you'll always start with the full backup
- **Differential Backup:** Backs up all database changes since the last full backup. Differential backups are generally smaller than their full equivalents and tend to run faster. They can also be used as a base for transaction log restores. When restoring your database, you can restore a differential backup after the full backup is restored to get the database closer to the desired restore point.
- **Transaction Log Backup:** Backs up all database changes (transactions) since the last transaction log backup (think of these as incremental backups). Transaction log backups tend to be the fastest to execute since they only contain transactions run on the database since the last transaction log backup. Transaction log backups are restored in sequence after a full or after a full and a differential backup is restored to get the database to a specific point-in-time in order to minimize data loss.

Most database administrators employ a combination of Full + Transaction Log or Full + Differential + Transaction Log backups to meet RPOs and keep backups within available maintenance windows. As an example, full backups may be run once a week with differential backups run daily, in-between. Transaction log backups are run based on RPO needs (e.g. every 15 minutes).

Initial Backup Seeding

Those initial full backups can be time-consuming. Backing up even large data sets locally can run be fast because of the low-latency and speed of a local network and storage. However, sending large data sets offsite, may require significantly more time because of limitations on internet connection and target cloud speed. When these initial data sets are too large, AWS has solutions to help.

- **Amazon S3 Transfer Acceleration:** Speeds up data transfer into Amazon S3. This is especially helpful when you have customers all over the world backing up to a non-local region, you transfer large amounts of data across continents, or you want to be able to better utilize your available internet bandwidth. Amazon S3 Transfer Acceleration takes advantage of Amazon CloudFront's globally distributed edge locations to securely transfer files over long distances. There is a charge for this service.
- **AWS Snowball / Snowball Edge / Snowmobile:** These are secure, hardware transport solutions that allows large amounts of data to be backed up locally, sent to AWS, and then securely transferred into your Amazon S3 account. You order the appliance from Amazon, connect it directly to your local

network, run your backups, and then send the hardware appliance back to Amazon where it is loaded into your account. This is a great solution for initial seeding of very large backups or when you regularly generate very large data sets and cannot upload to Amazon S3 via your available internet connection. There is a charge for this service.

Whether or not you need this type of advanced data transfer will be based on many factors, including:

- Size of initial data set for backup
- How much new / changed data is generated daily
- Speed of internet connection
- How long you can wait for the initial seed to complete
- Budget

Storage Classes

Amazon Web Services offers different storage classes to accommodate all your data needs. Storage classes can reduce storage costs for medium and long-term archival storage needs.

Amazon S3 Standard

Amazon S3 Standard is designed for optimal performance with frequently-accessed data.

- Low latency and high throughput performance
- Data is stored in a minimum of three Availability Zones (AZs) for durability
- 99.999999999% durability (0.000000001% chance of losing data)
- 99.99% availability over a given year (one hour of unavailability for every ten thousand hours)
- Billing per GB for data storage
- Some cost to retrieve data
- Covered by the Amazon S3 Service Level Agreement for Availability

Amazon S3 Standard is billed based on average use for the month. Objects in S3 Standard can be deleted as needed as there are no retention requirements. It is a good solution for any backups that you expect to access or update / delete in the first 30 days.

Amazon S3 Standard-Infrequent Access

Amazon Standard Infrequent Access (S3-IA) is designed for data which has less frequent access than S3 Standard, but requires high performance when needed.

Amazon S3-IA differs from S3 Standard in the following ways:

- Lower cost per GB for storage
- 99.9% availability over a given year
- Additional cost to retrieve data
- 30 days minimum storage duration charge
- 128 KB minimum capacity charge per object

The minimum storage period charge is for 30 days. Meaning, data can be deleted before 30 days, but you are billed as though it was present for the first 30 days. Amazon S3-IA is best used for data you expect to keep a minimum of 30 days and that you do not expect to access in that time, but need the restore performance if you do need access. There is an added cost to restore from S3-IA, but this can be easily offset by the lower storage costs. Because of the 128 KB minimum object size for billing, Amazon S3-IA may not be the best choice for a very large number of very small files. S3 Standard-Infrequent Access is suitable for all backup types and can be used instead of S3 Standard if the data is kept long enough and you only expect very limited access.

Amazon S3 One Zone-Infrequent Access

Amazon S3 One Zone-IA is less expensive and less durable version of S3 Standard-Infrequent Access. Amazon S3 One Zone-IA differs from S3-IA in the following ways:

- Storage is 20% less expensive than Amazon S3 Standard-Infrequent Access
- Data is stored in a single Availability Zones (AZ) and will be lost of that AZ is destroyed
- 99.5% availability over a given year

S3 One Zone-IA is good choice for infrequently accessed data that does not require the availability and resilience of S3 Standard or S3 Standard-IA, for storing secondary backup copies of on-premises data, or for data that is easily re-creatable.

Amazon Glacier

Amazon Glacier is used for long-term, archival storage of data that that is not accessed often, if at all, and does not require immediate access when needed. Amazon Glacier is extremely low-cost storage and differs from Amazon S3 Standard storage classes as follows:

- Extremely low cost
- Not covered by the Amazon S3 Service Level Agreement for Availability
- 90 days minimum storage duration charge
- 128 KB minimum capacity charge per object
- Additional cost to retrieve data
- Latency to first byte in minutes or hours

The service is optimized for archived data, with a standard retrieval time of several hours. It's a good option for any long-term storage needed for compliance, legal hold, and data that will not be accessed frequently, if at all. It is recommended that backups are not performed directly to Glacier, but instead to one of the Amazon S3 storage tiers, and then transitioned to Glacier for archive storage. This can be performed using an Object Lifecycle Policy.

Object Lifecycle Policies

Objects stored can be easily transitioned between storage tiers using an Object Lifecycle Policy. While you can skip storage classes, you can only move data in the following direction:

1. Amazon S3 Standard
2. Amazon S3 Standard-Infrequent Access
3. Amazon S3 One Zone-Infrequent Access
4. Amazon Glacier

This makes it easy, for example, to backup to Amazon S3 Standard and then automatically transition data to S3 Standard-Infrequent Access after 30 days and then to Glacier after 60 days. Or you could back up to Amazon S3 Standard-Infrequent Access and then transition to Amazon Glacier after 90 days. Once data is transitioned to a higher tier, there is no way to move it back without making a copy of the data (which would incur additional storage charges).

CloudBerry and AWS for Backup and Disaster Recovery

CloudBerry products make it easy to implement backup and disaster recovery on AWS. With two solutions available: A SaaS solution that runs on AWS for companies and IT service providers and our individually licensed products for very small business and consumers, CloudBerry allows everyone to utilize AWS features with ease.

File-Level Recovery with CloudBerry

CloudBerry makes it easy to protect and recover your important user-generated files from AWS. Create as many jobs as needed to protect your office docs, employee data, SQL Server or Exchange databases, family photos and videos, or your personal music collection. CloudBerry easily manages multiple file versions and can automatically transition backups between AWS storage tiers. Compression is applied to reduce data transfer size and storage costs, and end-to-end-encryption encryption is utilized to protect your data in transit and in storage.

To recover, simply tell CloudBerry which files or folders you need back, whether you want the most recent versions or older file versions, and whether you need your NTFS permissions restored to simplify post-restore security and administration. It's that easy.

Image-Level Recovery with CloudBerry

Just like file backups, CloudBerry can easily restore an entire server or desktop; operating system and all. This can be done to the same or dissimilar hardware using CloudBerry Bare-Metal Restore (BMR) technology. You can also restore to a VMware or Hyper-V virtual machine, and as a benefit for AWS users, you can restore directly to Amazon EC2 as a virtual machine (or AMI). This makes it easy to recovery entire systems locally or in the cloud. When disaster strikes, it's important to know you can recover an entire system when needed.

Keep in mind that disk volumes may be large and restore times may vary. If you have RTOs in place, test your restores to ensure you can meet those objectives. If you can't restore your images from Amazon S3 to a local machine in the required time, then consider using Hybrid Backups (that make two backup copies to local storage and Amazon S3) or restoring to a running Amazon EC2 instance so the data stays within a region within AWS.

Large Data Transfers out of Amazon S3 and Glacier

If you have the bandwidth (and time) you can move data from Amazon S3 and Glacier directly, but if you have time constraints because of an extremely large restore, then you can leverage the previously mentioned AWS features to speed up the data transfer out of AWS.

- **Amazon S3 Transfer Acceleration:** Speeds up data transfer out of Amazon S3. This is especially helpful when your restore location and AWS storage region are far apart. S3 Transfer Acceleration can help you more fully utilize your available bandwidth by minimizing the effect of distance on throughput. Amazon provides a site to help you understand the estimated benefits based on your location: https://s3-accelerate-speedtest.s3-accelerate.amazonaws.com/en/accelerate-speed-comparsion.html
- **AWS Snowball / Snowball Edge / Snowmobile:** These secure, hardware transport solutions allow large amounts of data to be securely moved from AWS to the appliance, sent to you, and then restored to your local network. This is a great solution when very large amounts of data need to be restored locally.

Total estimated cost of retrieving 50 TB of data Comparison of S3 Transfer Acceleration and AWS Snowball		
	S3 Acceleration	S3 Snowball
Cost	$7,100 (est)	$1,700 (est)
Time	2.5 days (2 Gbps connection) 10 Days (500 Mbps connection) 50 Days (100 Mbps connection)	< 7 days (includes shipping and data transfer time)
Additional Fees	None	Shipping fees
Other	Speed depends on available bandwidth	Availability of service depends on region and physical location

CloudBerry Software Licensing

CloudBerry is available in two versions:

- **CloudBerry Managed Backup Service:** A subscription, Backup-as-a-Service (BaaS) offering that provides centralized agent, job, reporting, monitoring, and administration support. A perfect solution for companies, education, government, and IT service providers. Prices start at $49.99 / year per computer (with quantity discounts available)
- **CloudBerry Backup:** A stand-alone, perpetually licensed solution for very small business and consumers. Prices start at $49.99 for Desktop Backup and $119.99 for Server Backup (these are one-time cost licenses)

AWS Storage Cost

How much data you need to keep in storage depends on many factors, including:

- How much source data you are protecting
- Whether you are running file or image backups
- How many versions of backups you are keeping
- How long you are keeping backups in storage (retention)
- Whether you are using an object lifecycle policy to move longer-term data to less-costly storage options

We can illustrate estimate storage costs for two options as follows:

- 10 TB of data stored in AWS
- 2 scenarios (Amazon S3 Standard only, Amazon S3 Standard → S3-IA → Glacier)
- 200 GB (or about 2%) of data restored on average per month

AWS Storage Costs (US East - N. Virginia Region)		
	Amazon S3 Standard (no storage tier transitions)	Amazon S3 (20%) → S3-IA (20%) → Glacier (60%)
Monthly	**$252 (est)**	**$111 (est)**
Annual	**$3,024 (est)**	**$1,332 (est)**

Please refer to AWS Calculator to better estimate your actual storage costs:
https://calculator.s3.amazonaws.com/index.html

Conclusion

Your backup strategy hinges on proper planning and product selection. Combining the right backup product with the best public cloud storage option provides an ideal combination for disaster recovery and long-term archival storage. A combined CloudBerry and AWS solution gives customers all the needed features to manage backups across your IT environment; whether that environment is business data across 10,000 endpoints and servers or a single laptop which contains those valuable family photos.

For a free trial and to get more information about CloudBerry solutions, please visit us at: www.cloudberrylab.com.