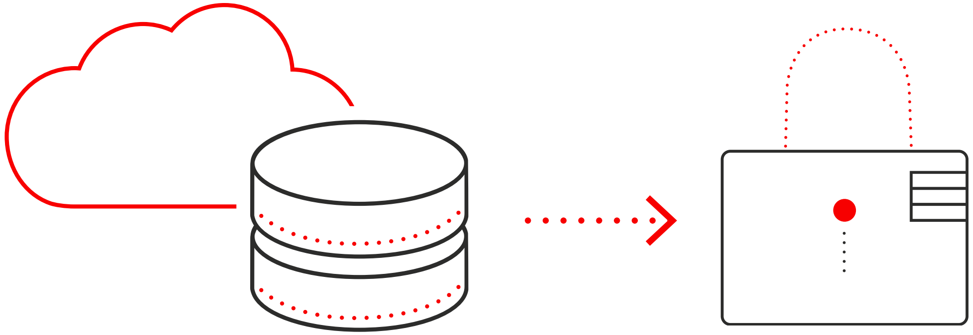
PoINT
software & systems

Backup for Object Storage

Criteria and Recommended Actions



Backup for Object Storage

Criteria and Recommended Actions

Object storage systems are increasingly being used for storing productive data. They offer decisive advantages for this:

- The **S3 API** is very easy to use and significantly simplifies application development.
- Object storage is also characterized by high **scalability** and **high availability**.

Systems achieve high availability through **redundant components**, **erasure coding** and **replication**.

> However, high availability must not be confused with data protection!
>
> Object storage is also exposed to **data loss risks** that cannot be eliminated by high-availability functions.
>
> A **backup of object storage** is therefore indispensable.

The most important criteria for such a backup are:

- that the backup is object-based,
- that it allows direct access
- and that it uses economical storage media – such as tape.

Scenarios for data loss in object storage

Data loss in object storage can occur in a variety of scenarios. These scenarios can be divided into two categories:

- The first category includes disk, node, rack, and site failures. In these cases, data loss can be prevented by proprietary object storage functions such as erasure coding and replication.
- The second category includes software errors, accidental or malicious deletion, ransomware and the human factor. The risk of data loss in these cases can only be minimized by a suitable backup.

	Scenarios for data loss	Solutions
Category 1	• Hard Drive failure • Node failure • Rack failure • Site failure	Erasure Coding Replication, CRR
Category 2	• Software errors • Accidental or malicious deletion • Ransomware • „The Human Factor“	**Backup**

CATEGORY 1: DISK, NODE, RACK, AND SITE FAILURES

Depending on the configuration, a large number of hard disks are installed in object storage systems. The **failure of hard disks** is therefore part of the daily routine. **Erasure coding methods** are used to protect against data loss. This means that recovery times are significantly shorter than with a classic RAID. The number of hard disks that may fail at one time depends on the set erasure coding rate and is ultimately paid for by a higher overhead.

Also in the event of a **node failure**, erasure coding or **replication** help to ensure that operation can continue.

Depending on the object storage, the data center architecture can be set up so that the **failure of a rack** does not affect operations: **separate fire compartments** can reduce the risk of a complete site outage. Replication or erasure coding procedures also come into play here.

Many companies have **two data centers** that they operate in an active/active or active/passive setup. In this way IT operations can be maintained even in the event of a complete site failure. In the case of two sites, replication ensures redundant storage and prevents data loss. For three or more sites erasure coding can play to its strengths, to prevent data loss in combination with low storage overhead.

CATEGORY 2: SOFTWARE ERRORS, ACCIDENTAL OR MALICIOUS DELETION, RANSOMWARE, HUMAN FACTOR

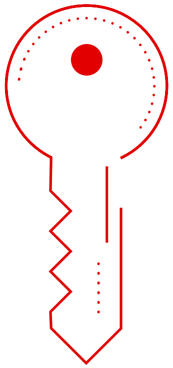
Despite maximum availability, **software errors** in the object storage can result in data loss. Erasure coding is of no use in this case.

If the permissions are set inappropriately or have been stolen, for example, a simple **S3 operation** is sufficient to **delete** thousands of objects in one go. Cross-site mirroring or erasure coding are of no help in this case either. Versioning, if supported by the object storage at all, only helps to a limited extent.

Even object storage is not immune to **ransomware**. Rhino Security Labs, a provider of penetration tests, has shown what an attack on an S3 bucket can look like. The result: even within very short periods of time, the attacker can succeed in encrypting large amounts of data. If only 10 minutes pass before the attack is stopped, more than 500GB can already be lost. [Source: Rhino Security Labs, Technical Blog, „S3 Ransomware Part 1: Attack Vector", https://rhinosecuritylabs.com/aws/s3-ransomware-part-1-attack-vector/]

Finally, the **„human factor“** is also a potential risk for data losses. Many data breaches and data loss have already occurred due to **misconfigured S3 buckets**. Encryption is also a risk. It does create security – but **if you lose the key**, you also lose access to the data. This is equivalent to data loss. Errors can also occur quickly during **maintenance work** and lead to data loss; for example, when applying patches.

Your data is the key, so don‘t lose it.



Criteria for Object Storage Backup

INDEPENDENT STORAGE MEDIUM

A regular backup of the object storage to an independent storage medium creates protection. If the object storage fails, a backup makes the data available quickly and in its entirety. The backup should definitely be **saved on a different storage medium, e.g. tape**, and **stored spatially separated**. The high level of protection for the data outweighs the additional costs for an extra storage medium. In addition, backing up to tape media is very sustainable and environmentally friendly, as tape storage systems are very energy-efficient.

ON-PREMISES BACKUP SYSTEM

Another crucial criterion is the local installation of the backup system with the backed-up data: The data should be stored **in the company's own data center** and not with a cloud provider. In the event of a cyberattack, for security reasons, Internet access is always interrupted for an unspecified period of time. This means that a backup stored with a cloud provider cannot be accessed. A restore is then not possible. With increasing data volumes, the transfer times also increase , and access costs rise. This also speaks against an external backup with a cloud provider.

OBJECT BASED BACKUP

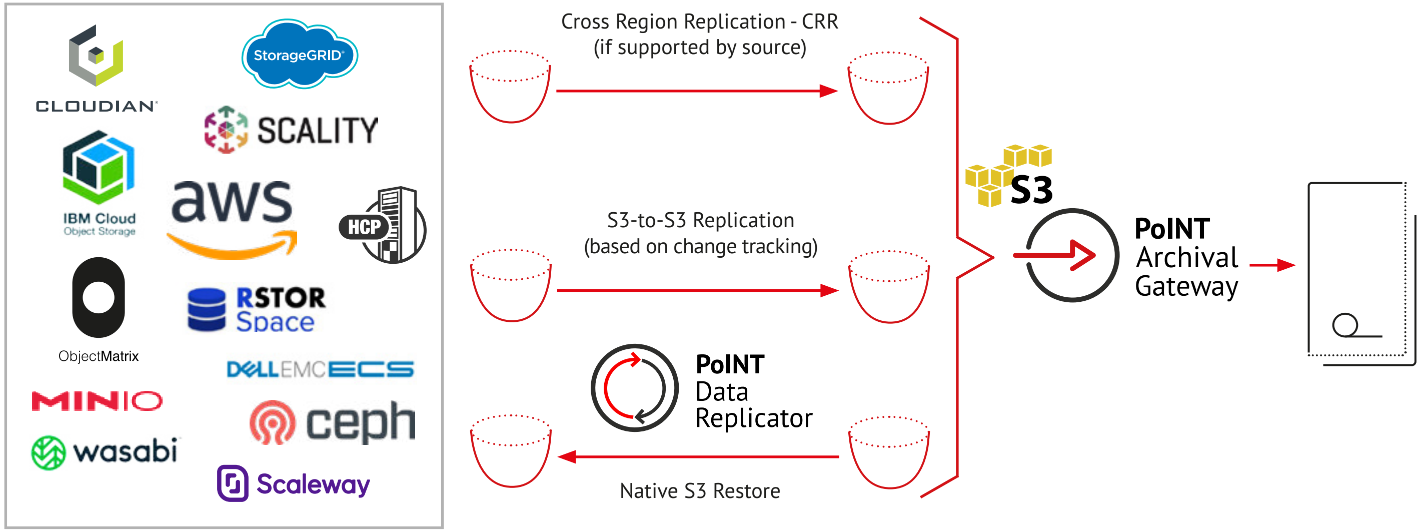
Like the source system, the backup system must be object-based, i.e. it must **save the objects in their native format as objects** including the metadata. In the event of an emergency, direct access to the backed-up data on the object-based storage system is possible, and a time-consuming restore process does not have to be performed first. Work can continue while the restore process is running at the same time.

Backup Methods

Basically, two backup methods are possible for object-based storage systems. Both methods allow direct access to the backed-up data if the primary storage system fails. This protects not only the valuable data, but also the work processes that rely on the data.

(1) BACKUP WITH "CROSS REGION REPLICATION"

If the source system supports the function **„Cross Region Replication" (CRR)**, this function can be used for direct backup to an S3-capable backup target system. For example, the **tape-based object storage PoINT Archival Gateway** can serve as such a system. Here, the backup system consists of the S3 object storage „PoINT Archival Gateway" and a tape library. This method allows a direct connection of the source system to this backup system.

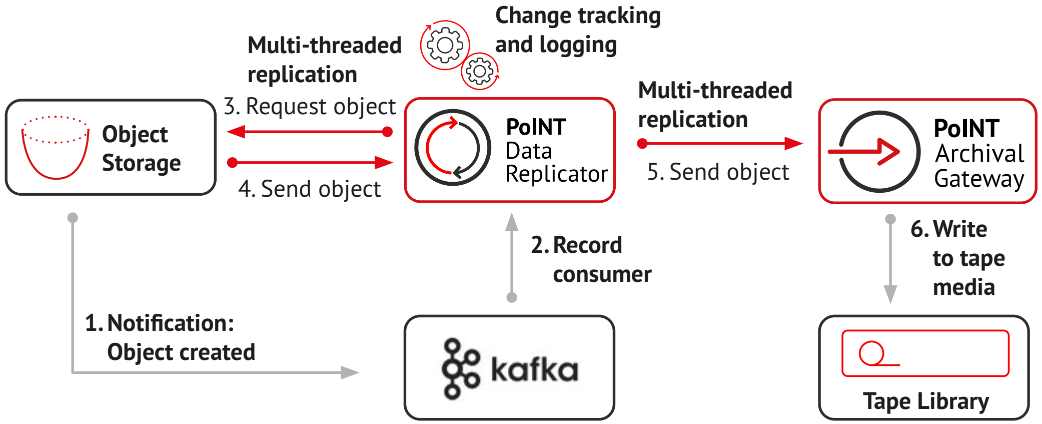


(2) BACKUP WITH REPLICATION

If the source system does not support „Cross Region Replication", a **replication software** can be used. With **PoINT Data Replicator**, for example, the data to be backed up can be replicated in native format via the S3 interface of the source system to the S3-capable backup target system – for example to PoINT Archival Gateway.

By means of individual filters, the replication process can be limited to specific data. To guarantee an efficient process, a database supports the operation. Thus, only new objects are replicated during a new run.

For continuous backup PoINT Data Replicator also supports queuing services like Apache Kafka and AWS SQS. This avoids long scan times on the source system. The following graphic illustrates the workflow using the Kafka service.

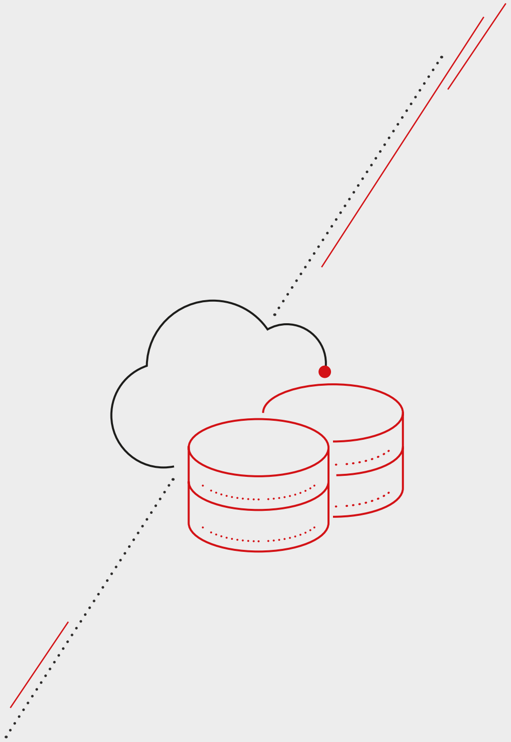


Conclusion

Backing up your data is and remains your responsibility. Therefore, it is careless if object storage systems are not backed up. It is important to realize that data availability is not synonymous with data protection. Therefore, in addition to high data availability, it is mandatory to back up the data of an object storage system.

- It is essential that the backup be performed to an **independent offline-capable medium**, such as tape, so that a media break and „air gap" are achieved. A backup to tape is not without reason referred to as the „last line of defense."
- The backup software itself should be object-based and not use proprietary formats. Objects must be backed up **in their native format**.
- The backup system should provide **direct access** to the backed-up data.

PoINT Software & Systems



PoINT Software & Systems is specialized in the development of software solutions for storage and management of data using all available mass storage technologies like cloud and object storage, hard disk, magnetic tape and optical. Close collaboration with leading hardware manufacturers enables an early support of innovative storage technologies. Besides complete solutions PoINT also offers its know-how as Toolkits, which can be easily integrated in other applications by the programming interface. Furthermore we project entire storage solutions and provide consultancy with our long-term and versatile experience.

PoINT products are distributed in more than 25 countries world-wide and have been installed successfully in more than two million installations. Our customers range from end users expecting a compact and secure solution to large corporations, which comply with our solutions their complex demands by providing the necessary reliability and perfection.

PoINT Software & Systems GmbH
Eiserfelder Straße 316
57080 Siegen, Germany

P +49 271 3841-0
M info@point.de
W www.point.de