



Hybrid Backup

SEP sesam

Using S3 as a Backup Target

Benefits, Workflow and User Scenarios

Index

Index	2
Introduction	3
SEP Si3 Deduplication	4
Backup, Restore, Replication, Migration	5
Restore Options	6
Deduplication Workflow	7
Disaster Recovery Scenarios	8-10
Category 1: Backup Server not affected	8-9
Category 2: Backup Server out of order	10
Performance	11
SEP sesam Certifications	12
Cloud Object	12
Storage	12

Introduction

Amazon Web Services (AWS) is the public cloud service offering from Amazon. AWS has created their own API to access their public cloud object storage. In 2006, Amazon Simple Storage Service (S3) was released to directly address the fast-growing need for scalable and high available object storage.



The basic storage units of Amazon S3 are objects which are organized into buckets. Each object is identified by a unique, user-assigned key. Buckets can be managed using either the console provided by Amazon S3, programmatically using the AWS SDK, or with the Amazon S3 REST application programming interface (API).

Objects can be managed using the AWS SDK or with the Amazon S3 REST API and can be up to five terabytes in size with two kilobytes of metadata. Additionally, objects can be downloaded using the HTTP GET interface and the BitTorrent protocol.

S3 has become a global market standard for accessing any type of object storage – no matter if local or in the cloud. There are many providers active in market for both categories. Most public cloud providers have their own API. Even though Microsoft promotes their own API with Microsoft Azure BLOB storage they offer an S3 interface or at least an S3 Gateway is available to provide market compatibility.

For more information about S3 you can [find here](#)

SEP Si3 Deduplication

SEP Si3 is the name of SEP's patented and integrated deduplication algorithm. Working fully in-memory and based on variable block length it offers an effective method for data reduction and storage savings by creating optimal deduplication ratios.

With just one click from a dropdown menu in the GUI you can switch on deduplication for almost any kind of direct attached storage like SAS/SATA, iSCSI, LUN or SMB.

The vast variety of different configuration modes like target side, source side and n:m replication based on media pools offers real global deduplication and minimizes the amount of backup data dramatically in any kind of distributed environment.

All kind of mount-based restore options are implemented by means of a powerful layer called virtual file system (VFS) are available for:

- Mount file backup
- Mount Exchange EDB
- Mount virtual disks of VMs
- Mount NDMP backups



are implemented by means of a powerful layer called virtual file system (VFS).

If list-files are not available an index will be generated on the fly during the mount process.

The latest version SEP sesam v5.0 (Jaglioni), includes an improved version of its integrated deduplication engine called Si3-NG (new generation).

The Si3-NG internal deduplication algorithm has significantly improved performance, resource consumption, restrictions, and garbage collection. Write Performance has been increased but especially Read performance has been accelerated dramatically. Benchmarks have shown that the throughput of Si3-NG is now comparable to a backup to native plain Data Store, enabling new backup concepts based on Si3-NG as primary backup storage.

To transition from Si3 to Si3-NG a simple to configure replication can be applied.

In addition, a brand-new functionality has been released with this revision.

It is now possible to write backup data directly to S3 in deduplicated form without the need of a local Si3-NG Data Store. This is a major concept change to the Si3 S3 replication. Previously this process required a local Si3 Data Store and replicated and restored to and from S3 transparently in the background without any direct access to the data on S3 storage by the backup user. By utilizing Si3-NG deduplication the SEP S3 interface is optimized for being used over the WAN i.e. in cloud environments.

Backup, Restore, Replication, Migration

Si3-NG-S3 Data Store is treated and behaves exactly as a local Si3-NG Data Store with no difference from a GUI/WebUI usability perspective. The technical differences are completely internal and fully transparent to the backup user.

The backup, restore, replication and migration process require configuring a respective Data Store in the GUI, associated drives and media pools to specify the EOL for this backup target. The media pool can be used for backup, restore, replication and migration. For example, the different EOLs enable archiving a copy of your data to the cloud.

Utilizing a local Si3-NG Data Store, you can specify a replication in any combination of Si3-NG and Si3-NG-S3. Even a replication from S3 to S3 is possible and provides a perfect basis for multi cloud backup scenarios. This supports the growing demand to be independent of one single cloud provider or to migrate from one cloud to another. From a dataflow perspective, the S3-to-S3 replication works indirect via the SEP backup server or SEP Remote Device Server (RDS) and not directly from cloud to cloud.

New Gen:	Si3-NG	↔	Si3-NG-S3
Mixed:	Si3	↔	Si3-NG-S3
S3 to S3:	Si3-NG-S3	↔	Si3-NG-S3

Replication with SEP sesam still is an asynchronous process. You must specify a schedule to trigger the replication process via a replication event. The replication schedule can be as granular as needed. This depends on your data change rates, quantity of data to be transmitted and your available bandwidth.

A remote device server or backup server can have only one deduplication Data Store configured. In order to have a copy of or archiving your data to the cloud, you will need to configure two Si3-NG:

- A local Si3-NG as source
- A Si3-NG-S3 as target

By default, you would need two RDS' for this configuration. However, there is a key in the global GUI settings available to overwrite this default restriction if necessary.

When configuring the Si3-NG-S3 Data Store in Sesam GUI, you must specify a size for this Data Store even if any S3 target shows unlimited storage. This is relevant for size calculation and licensing.

The configuration of a Si3/Si3-NG Data Store including Si3-NG-S3 can be automated via our command [line interface](#).

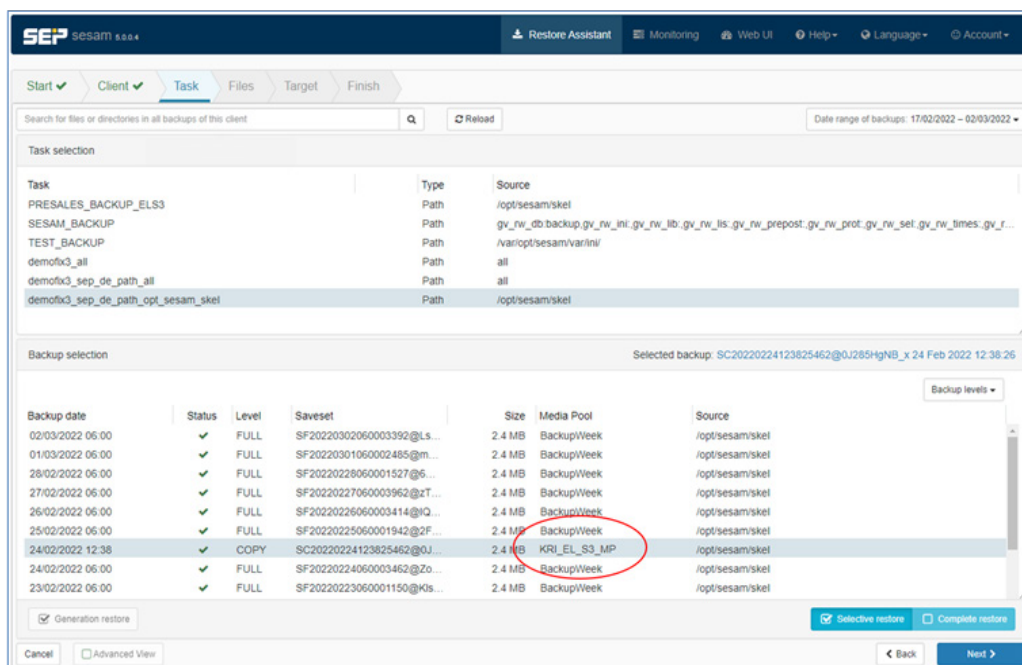
Best practices and the complete configuration process of a Si3-NG-S3 Data Store is described in detail on our [technical wiki](#).

Restore Options

As an Si3-NG-S3 DataStore behaves identical to a local Si3-NG Data Store it also offers the same vast restore functionalities including:

- Simple restore via GUI or WebUI Restore Assistant of file data, DB, VMs, etc.
- Restoring locally or to any other location by specifying a target server
- Mounting of backups e.g. Path backups for cross platform restore
- NDMP for Single File Restore
- Exchange (EDB) for browsing with SEP Exchange Recovery Pro and Single Object Restore e.g. mails, calendar entries, etc.
- VMs (virtual disks) for Single File Restore
- VMware VMs with Instant Recovery
- DR scenarios described in one a later chapter

Example from the WebUI Restore Assistant to show the easy selection of the backup on the right media pool for restore from an S3 backup storage:



Deduplication Workflow

When using Si3-NG-S3 the normal deduplication algorithm will be applied to the back data stream in memory of the RDS. The mathematical shifting window process generates chunks of variable block size with a maximum of 128kB per chunk.

Each chunk yields to a hash value which will be mapped to the hash directory. If this hash is already listed only a pointer will be set and no data chunk must be transmitted.

This meta data will be kept locally in a directory which will be specified during the SEP Si3-NG-S3 Data Store configuration. In this directory, the chunks will be collected for transmission via S3 in dedicated container files until one of several predefined thresholds will trigger the data transmission. All chunks are compressed and the amount of intermediate local data will not exceed 130 MB. Rarely, it could be possible that this limit will be exceeded when having many parallel streams during the initial seed phase. Best practices recommends a storage buffer in these cases. The amount of meta data being in the range of some 100kB is not relevant.

Purging of an Si3-NG-S3 works similar to Si3-NG. After a certain amount of chunks are orphaned and marked for deletion, the chunks are reorganized and obsolete container files will be deleted via the respective S3 API command.

We only use a few basic S3 API calls (upload, download, delete) to ensure compatibility with all S3 targets. The connection to S3 is based on https and the connection is securely encrypted. Encryption of the data itself is in development with a scheduled release later this year.

Backup or replication to a S3 cloud via the WAN can be difficult to define. They are highly dependent on the individual configuration and preconditions e.g.

- Amount of backup data before deduplication (-> Frontend data)
- Amount of data after deduplication (-> dedupe ratio)
- Amount of logical data (-> how often backup the same data)
- Change rate (-> how many new chunks)
- Type of data (-> File, DB, compressed, encrypted etc.)
- Recovery Time Objective (RTO) for restore

When using the Si3-NG deduplication algorithm the same requirement must be followed. These are described and kept up to date in our [Wiki](#)

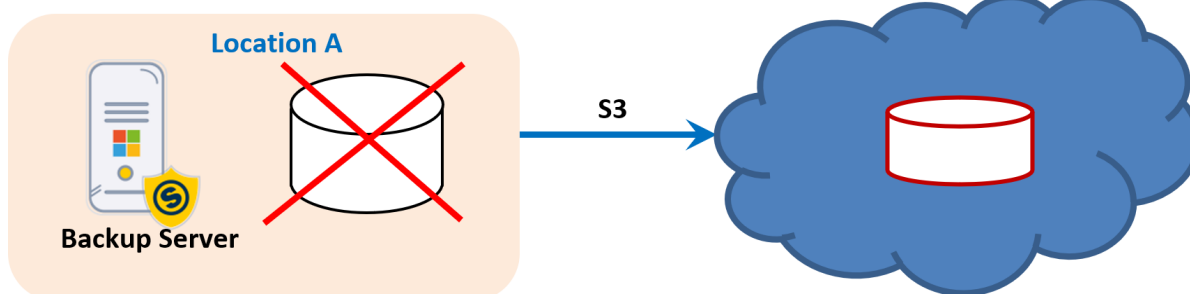
Disaster Recovery Scenarios

The Disaster Recovery (DR) scenarios described in this chapter are generic because the S3 backup target could always be replaced by any location with a standard S3-NG replication. The advantage of an S3 target is that you can transfer your data to a different location (cloud) while having a local infrastructure. Here you can find the Wiki article. When talking about DR, best practices dictates that you must consider two user scenarios:

⇒ With and without outage of your Backup Server

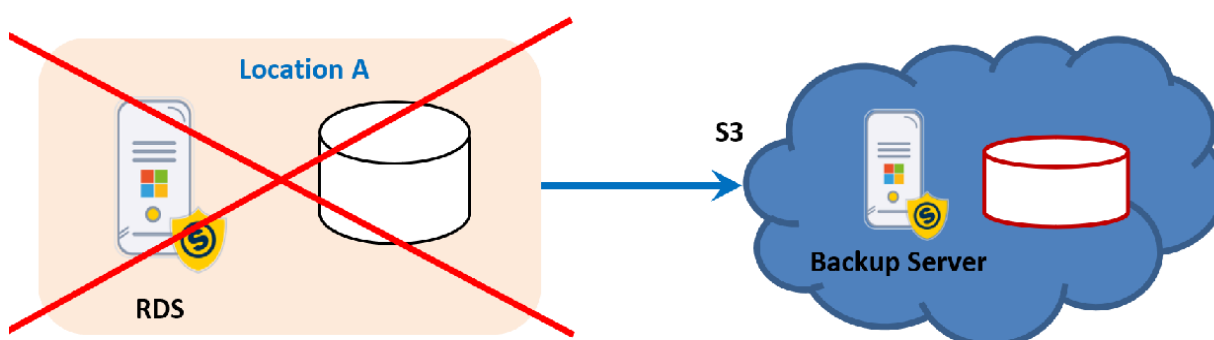
User Scenario 1: Backup Server not affected

1. Outage of your primary data storage or backup storage



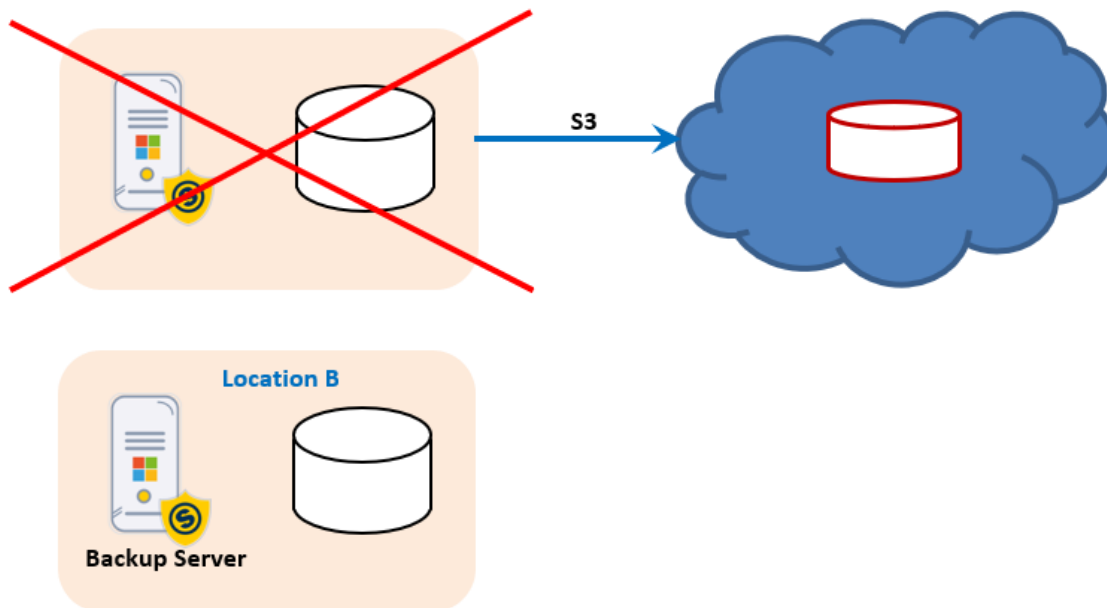
Simply restore or replicate your data from the cloud.

2. Outage of location A with Backup Server in the cloud



This scenario is not very complicated as the RDS does not have a significant amount of metadata stored locally, only for the S3 Data Store. Simply install the SEP package on the new RDS hardware, configure the same IP address and restore the meta data of the S3 Data Store. Then trigger the back-replication from the Backup Server which is recoverable from the infrastructure in the cloud.

3. Outage of location A with RDS

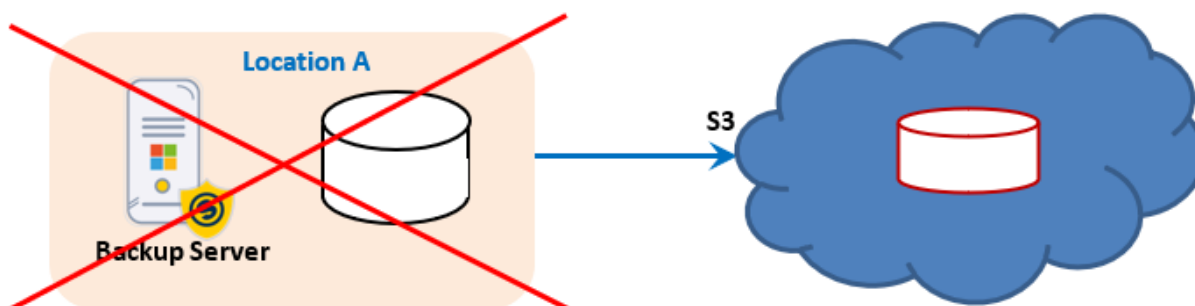


This scenario is not very complicated as the RDS does not have a significant amount of metadata stored locally, only for the S3 Data Store. Simply install the SEP package on the new RDS hardware, configure the same IP address and restore the meta data of the S3 Data Store. Then trigger the back-replication from the Backup Server which is recoverable from the infrastructure in the cloud.

User Scenario 2: Backup Server out of order

There are a couple of possible configurations under this user scenario. Both scenarios require a DR copy of the SEP Sesam backup server directory. This is a well-defined but manual process a simple configuration needed. Recovery of the backup server process is clearly documented in our [Wiki](#):

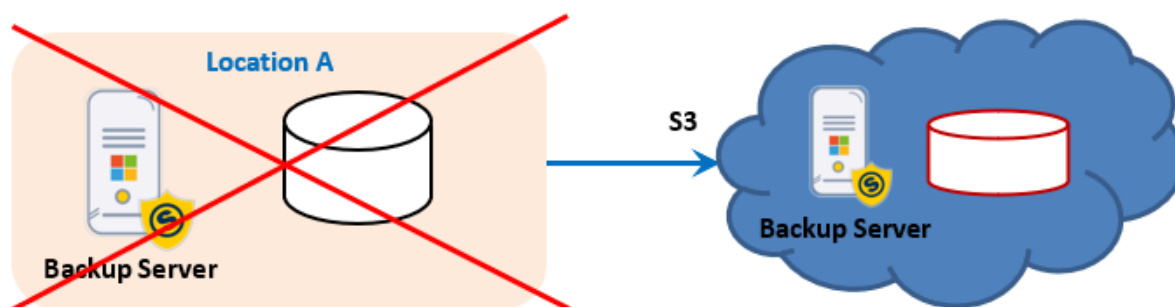
4. Outage of location A with physical machine recovery from the cloud.



You have to setup your complete infrastructure from scratch.

- Build and configure the needed physical infrastructure
- Recover the SEP sesam backup server directory
- Restore the data from the cloud

5. Outage of location A with virtual machine recovery from the cloud.



For a faster replacement of location A and to provide a temporary working environment in the cloud it is possible to

- Setup your Backup Server via DR in the cloud infrastructure
- Ensure access to the backup data on S3
- Recovery of all VMs in the cloud
- In case of VMware do an Instant Recovery of all VMs

Performance

During a recovery from the cloud, the backup data is transferred from S3 to local storage in deduplicated form and rehydrated locally. This ensures that performance for backup and restore will not differ much and that bandwidth will not act as an increasing bottleneck for recovery.

As Si3-NG-S3 is based on the Si3-NG algorithm, essentially the same performance values are applicable.

Benchmarks have shown that Si3-NG performance is comparable to the performance of a normal Data Store. Therefore, with S3 we have reached the same results when using local S3 object storage.

The network is the most limiting role. By providing more hardware resources like CPU and RAM can increase the performance. However, when using S3 public cloud via the WAN, bandwidth will in most cases be the limiting factor. To estimate your performance, you need a deeper understanding of how deduplication works. As an example, when there are no new data chunks that need to be transmitted this means that only meta data has to be changed which is saved locally and backup performance will be increased.

Here are the averaged figures from our benchmark tests below:

- system 'dd' as a baseline for the performance capabilities of your system
- -10% to write to a plain DataStore (block device)
- -20% to write to an Si3-NG DataStore or Si3-NG-S3 local object storage
- -30%+ to write to a Si3-NG-S3 cloud storage

These figures could vary significantly according to:

- Action type:
 - o backup
 - o recovery
- Network connection:
 - o disk storage
 - o bandwidth
- IOPS of disk storage
- Deduplication ratio
- Number of parallel backups
- Workload of the Backup Server/RDS

SEP sesam Certifications

The vendors listed below highlights the current certifications and validated cloud and object storage providers. This list will only grow over time. Validation and certification is planned for all vendors listed below. However, any vendor that utilizes the AWS S3 should work accordingly with SEP Sesam replication.

Cloud Storage Providers:

 **wasabi**[®] Wasabi (certified)

 **EL** | storage
powered by  wasabi Ebertlang S3 Cloud Storage (certified)

 **aws** AWS (validated)

 **IBM Cloud** IBM Cloud (validated)

Object Storage Provides

 **DATA CORE** DataCore Swarm S3 (validated)

 **Objects** Nutanix Objects (validated)


 **PURE STORAGE**[®] PureStorage FlashBlade (validated)

 **MINIO** Min.io (validated)

Contact

 +49 8024 463310

 sales@sep.de

 SEP AG, Konrad-Zuse-Straße 5
83607 Holzkirchen, Germany