



Hybrid Backup

SEP sesam

Verwendung von S3 als Sicherungsziel

Vorteile, Workflow und Benutzerszenarien

Index

Index	2
Einführung	3
SEP Si3 Deduplizierung	4
Backup, Restore, Replikation, Migration	5
Restore Optionen	6
Deduplizierungsworkflow	7
Disaster Recovery Szenarien	8-10
Kategorie 1: Backup Server nicht betroffen	8-9
Kategorie 2: Backup Server außer Betrieb	10
Performance	11
SEP sesam Zertifizierungen	12
Cloud Object	12
Storage	12

Einführung

Amazon Web Services (AWS) ist das öffentliche Cloud-Service-Angebot von Amazon. AWS hat eine eigene API für den Zugriff auf seinen öffentlichen Cloud-Objektspeicher entwickelt. Im Jahr 2006 wurde Amazon Simple Storage Service (S3) veröffentlicht, um den schnell wachsenden Bedarf an skalierbarer und hochverfügbarer Objektspeicherung direkt zu decken.



Die grundlegenden Speichereinheiten von Amazon S3 sind Objekte, die in Buckets organisiert sind. Jedes Objekt wird durch einen eindeutigen, vom Benutzer zugewiesenen Schlüssel identifiziert. Buckets können entweder über die von Amazon S3 bereitgestellte Konsole, programmgesteuert über das AWS SDK oder mit Hilfe der Amazon S3 REST Application Programming Interface (API).

Objekte können mit dem AWS SDK oder mit der Amazon S3 REST API verwaltet werden und können bis zu fünf Terabyte groß sein, mit zwei Kilobyte an Metadaten. Außerdem können Objekte über die HTTP-GET-Schnittstelle und das BitTorrent-Protokoll heruntergeladen werden.

S3 hat sich zu einem globalen Marktstandard für den Zugriff auf jede Art von Objektspeicher entwickelt - egal ob lokal oder in der Cloud. Für beide Kategorien gibt es viele Anbieter auf dem Markt. Die meisten öffentlichen Cloud-Anbieter haben ihre eigene API. Obwohl Microsoft seine eigene API mit Microsoft Azure BLOB-Storage bewirbt, bieten sie eine S3-Schnittstelle oder zumindest ein S3-Gateway an, um Marktkompatibilität zu gewährleisten.

Mehr Informationen finden Sie hier.

SEP Si3 Deduplizierung

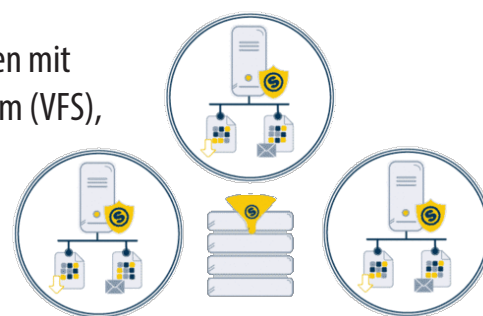
SEP Si3 ist der Name des patentierten und integrierten Deduplizierungsalgorithmus von SEP. Er arbeitet vollständig im Speicher und basiert auf variablen Blocklängen und bietet eine effektive Methode zur Datenreduzierung und Speicherplatzeinsparung durch die Schaffung optimaler Deduplizierungsverhältnisse.

Mit nur einem Klick aus einem Dropdown-Menü in der GUI können Sie die Deduplizierung für fast alle Arten von direkt angeschlossenem Speicher wie SAS/SATA, iSCSI, LUN oder SMB aktivieren.

Die große Auswahl an verschiedenen Konfigurationsmodi wie Zielseite, Quellseite und n:m Replikation auf Basis von Medienpools bietet echte globale Deduplizierung und minimiert die Menge an Backup-Daten in jeder Art von verteilter Umgebung drastisch.

Alle Arten von mount-basierten Wiederherstellungsoptionen werden mit Hilfe einer leistungsfähigen Schnittstelle, dem virtuellen Dateisystem (VFS), implementiert:

- Mount file Backup
- Mount Exchange EDB
- Mount virtual Disks von VMs
- Mount NDMP Backups



werden mit Hilfe einer leistungsfähigen Schnittstelle, dem virtuellen Dateisystem (VFS), durchgeführt. Wenn keine Listendateien verfügbar sind, wird während des Mount-Prozesses ein Index erstellt.

Der interne Deduplizierungsalgorithmus von Si3-NG hat die Leistungsfähigkeit, den Ressourcenverbrauch, die Einschränkungen und die Garbage Collection deutlich verbessert. Die Schreibperformance wurde gesteigert, aber vor allem die Leseperformance wurde dramatisch beschleunigt. Benchmarks haben gezeigt, dass der Durchsatz von Si3-NG jetzt vergleichbar ist mit einem Backup auf einen nativen, einfachen Datenspeicher, was neue Backup-Konzepte ermöglicht, die auf Si3-NG als primärem Backup-Speicher basieren.

Für den Übergang von Si3 zu Si3-NG kann eine einfach zu konfigurierende Replikation eingesetzt werden. Darüber hinaus wurde mit dieser Revision eine brandneue Funktionalität freigegeben.

Es ist nun möglich, Sicherungsdaten direkt in deduplizierter Form in S3 zu schreiben, ohne dass ein lokaler Si3-NG-Datenspeicher erforderlich ist. Dies ist eine wesentliche Änderung des Konzepts der Si3 S3-Replikation. Bisher war für diesen Prozess ein lokaler Si3-Datenspeicher erforderlich, und die Replikation und Wiederherstellung zu und von S3 erfolgte transparent im Hintergrund, ohne dass der Backup-Benutzer direkten Zugriff auf die Daten im S3-Speicher hatte. Durch den Einsatz von Si3-NG Deduplizierung wird die SEP S3 Schnittstelle für den Einsatz über das WAN, d.h. in Cloud-Umgebungen, optimiert.

Backup, Restore, Replikation, Migration

Der Si3-NG-S3-Datenspeicher wird genau wie ein lokaler Si3-NG-Datenspeicher behandelt und verhält sich auch so, ohne dass es aus Sicht der Benutzerfreundlichkeit der GUI/WebUI einen Unterschied gibt. Die technischen Unterschiede sind vollständig intern und für den Backup-Benutzer völlig transparent.

Der Sicherungs-, Wiederherstellungs-, Replikations- und Migrationsprozess erfordert die Konfiguration eines entsprechenden Data Stores in der GUI, der zugehörigen Laufwerke und Medienpools, um das EOL für dieses Sicherungsziel festzulegen. Der Medienpool kann für Backup, Restore, Replikation und Migration verwendet werden. Die verschiedenen EOLs ermöglichen zum Beispiel die Archivierung einer Kopie Ihrer Daten in der Cloud.

Unter Verwendung eines lokalen Si3-NG-Datenspeichers können Sie eine Replikation in einer beliebigen Kombination von Si3-NG und Si3-NG-S3 angeben. Sogar eine Replikation von S3 zu S3 ist möglich und bietet eine perfekte Basis für Multi-Cloud-Backup-Szenarien. Dies unterstützt die wachsende Nachfrage, von einem einzelnen Cloud-Anbieter unabhängig zu sein oder von einer Cloud in eine andere zu migrieren. Aus Sicht des Datenflusses funktioniert die S3-zu-S3-Replikation indirekt über den SEP Backup Server oder SEP Remote Device Server (RDS) und nicht direkt von Cloud zu Cloud.

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

Die Replikation mit SEP sesam ist nach wie vor ein asynchroner Prozess. Sie müssen einen Zeitplan angeben, um den Replikationsprozess über ein Replikationsereignis auszulösen. Der Replikationszeitplan kann so granular wie nötig sein. Dies hängt von der Änderungsrate Ihrer Daten, der Menge der zu übertragenden Daten und der verfügbaren Bandbreite ab.

Auf einem Remote Device Server oder Backup Server kann nur ein Deduplizierungs-Datenspeicher konfiguriert sein. Um eine Kopie oder Archivierung Ihrer Daten in der Cloud zu haben, müssen Sie zwei Si3-NG konfigurieren:

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

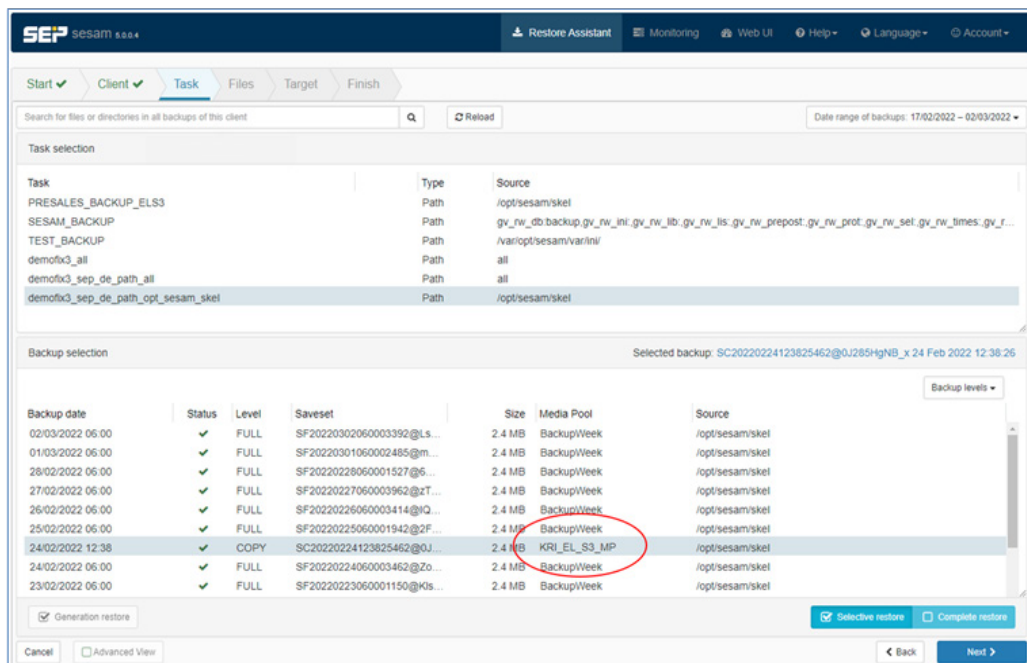
Standardmäßig würden Sie für diese Konfiguration zwei RDS' benötigen. Es gibt jedoch einen Schlüssel in den globalen GUI-Einstellungen, mit dem diese Standardbeschränkung bei Bedarf überschrieben werden kann. Bei der Konfiguration des Si3-NG-S3-Datenspeichers in der Sesam-GUI müssen Sie eine Größe für diesen Datenspeicher angeben, auch wenn ein S3-Ziel unbegrenzten Speicherplatz anzeigt. Dies ist für die Größenberechnung und Lizenzierung relevant. Die Konfiguration eines Si3/Si3-NG Data Stores inklusive Si3-NG-S3 kann über unsere Kommandozeilenschnittstelle automatisiert werden. Best Practices und der komplette Konfigurationsprozess eines Si3-NG-S3 Data Stores sind in unserem technischen Wiki detailliert beschrieben.

Restore Optionen

Da sich ein Si3-NG-S3 DataStore identisch verhält wie ein lokaler Si3-NG DataStore, bietet er auch die gleichen umfangreichen Wiederherstellungsfunktionalitäten, einschließlich:

- Einfache Wiederherstellung über GUI oder WebUI Wiederherstellungsassistent von Dateidaten, DB, VMs, etc.
- Wiederherstellung lokal oder an jedem anderen Ort durch Angabe eines Zielserver
- Einbinden von Backups, z.B. Path-Backups für plattformübergreifende Wiederherstellung
- NDMP für die Wiederherstellung einzelner Dateien
- Exchange (EDB) zum Durchsuchen mit SEP Exchange Recovery Pro und Single Object Restore z.B. Mails, Kalendereinträge, etc.
- VMs (virtuelle Festplatten) für Single File Restore
- VMware VMs mit Instant Recovery
- DR-Szenarien, die in einem späteren Kapitel beschrieben werden

Beispiel aus dem WebUI-Wiederherstellungsassistenten, um die einfache Auswahl des Backups auf dem rechten Medienpool für die Wiederherstellung aus einem S3-Sicherungsspeicher zu zeigen:



Deduplizierungsworkflow

Bei der Verwendung von Si3-NG-S3 wird der normale Deduplizierungsalgorithmus auf den Rückdatenstrom im Speicher des RDS angewendet. Der mathematische Shifting-Window-Prozess generiert Chunks mit variabler Blockgröße mit maximal 128kB pro Chunk.

Jeder Chunk ergibt einen Hash-Wert, der im Hash-Verzeichnis abgebildet wird. Ist dieser Hash bereits eingetragen, wird nur ein Zeiger gesetzt und es muss kein Datenchunk übertragen werden.

Diese Metadaten werden lokal in einem Verzeichnis gehalten, das bei der Konfiguration des SEP Si3-NG-S3 Data Store angegeben wird. In diesem Verzeichnis werden die Chunks für die Übertragung via S3 in dedizierten Containerdateien gesammelt, bis einer von mehreren vordefinierten Schwellenwerten die Datenübertragung auslöst. Alle Chunks sind komprimiert, und die Menge der lokalen Zwischendaten wird 130 MB nicht überschreiten. In seltenen Fällen kann es vorkommen, dass diese Grenze überschritten wird, wenn während der anfänglichen Seed-Phase viele parallele Streams vorhanden sind. Für diese Fälle wird ein Speicherpuffer empfohlen. Die Menge der Metadaten, die im Bereich von einigen 100kB liegt, ist nicht relevant.

Die Bereinigung eines Si3-NG-S3 funktioniert ähnlich wie bei Si3-NG. Nachdem eine bestimmte Anzahl von Chunks verwaist und zur Löschung markiert ist, werden die Chunks reorganisiert und veraltete Containerdateien über den entsprechenden S3-API-Befehl gelöscht.

Wir verwenden nur wenige grundlegende S3-API-Aufrufe (Upload, Download, Delete), um die Kompatibilität mit allen S3-Zielen zu gewährleisten. Die Verbindung zu S3 basiert auf https und die Verbindung ist sicher verschlüsselt. Die Verschlüsselung der Daten selbst befindet sich in der Entwicklung und soll noch in diesem Jahr veröffentlicht werden.

Die Sicherung oder Replikation zu einer S3-Wolke über das WAN kann schwierig zu definieren sein. Sie sind stark von der individuellen Konfiguration und den Voraussetzungen abhängig, z. B.

- Menge der Sicherungsdaten vor der Deduplizierung (-> Frontend-Daten)
- Datenmenge nach Deduplizierung (-> Deduplizierungsquote)
- Menge der logischen Daten (-> wie oft werden die gleichen Daten gesichert)
- Änderungsrate (-> wie viele neue Chunks)
- Art der Daten (-> Datei, DB, komprimiert, verschlüsselt etc.)
- Recovery Time Objective (RTO) für die Wiederherstellung

Bei der Verwendung des Si3-NG Deduplizierungsalgorithmus müssen die gleichen Anforderungen beachtet werden. Diese werden in unserem Wiki beschrieben und auf dem neuesten Stand gehalten.

Disaster Recovery Szenarien

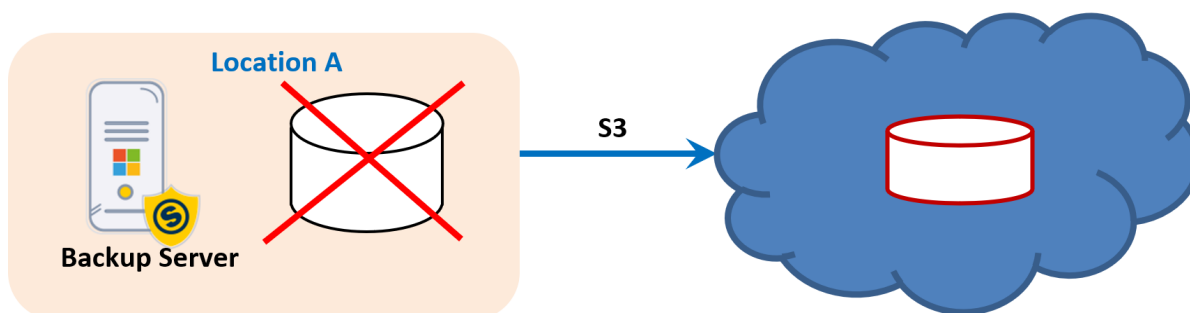
Die in diesem Kapitel beschriebenen Disaster Recovery (DR)-Szenarien sind generisch, da das S3-Sicherungsziel jederzeit durch einen beliebigen Standort mit einer Standard-Si3-NG-Replikation ersetzt werden kann. Der Vorteil eines S3-Ziels besteht darin, dass Sie Ihre Daten an einen anderen Standort (Cloud) übertragen können, während Sie eine lokale Infrastruktur haben. Hier geht's zum Wiki Artikel.

Wenn Sie über DR sprechen, müssen Sie laut Best Practices zwei Benutzerszenarien berücksichtigen:

⇒ With and without outage of your Backup Server

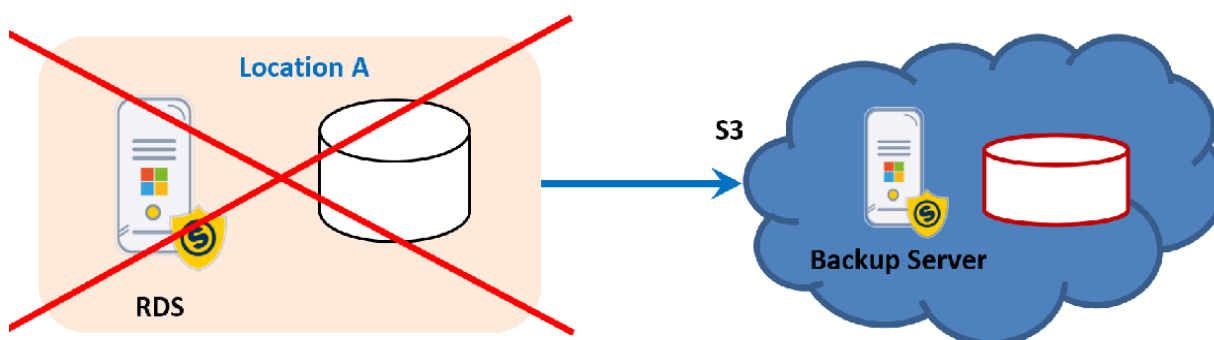
Benutzerszenario 1: Backup Server nicht betroffen

1. Ausfall Ihres primären Datenspeichers oder Backup-Speichers



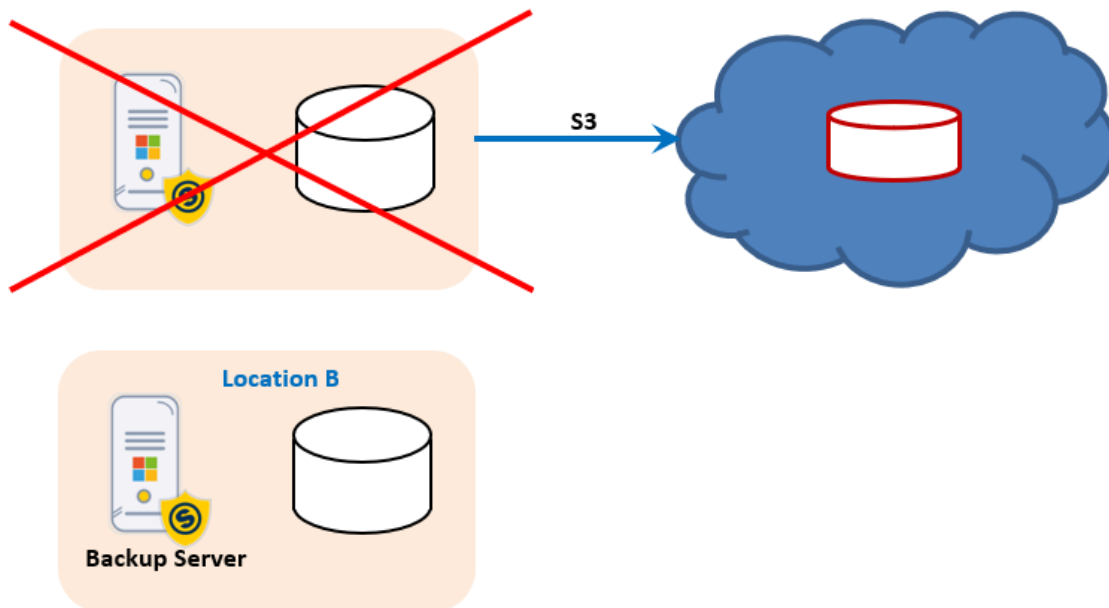
Stellen Sie Ihre Daten einfach wieder her oder replizieren Sie sie aus der Cloud.

2. Ausfall von Standort A mit Backup Server in der Cloud



Dieses Szenario ist nicht sehr kompliziert, da der RDS keine große Menge an Metadaten lokal gespeichert hat, nur für den S3-Datenspeicher. Installieren Sie einfach das SEP-Paket auf der neuen RDS-Hardware, konfigurieren Sie die gleiche IP-Adresse und stellen Sie die Metadaten des S3-Datenspeichers wieder her. Dann lösen Sie die Rückreplikation vom Backup Server aus, der von der Infrastruktur in der Cloud wiederherstellbar ist.

3. Ausfall von Standort A mit RDS

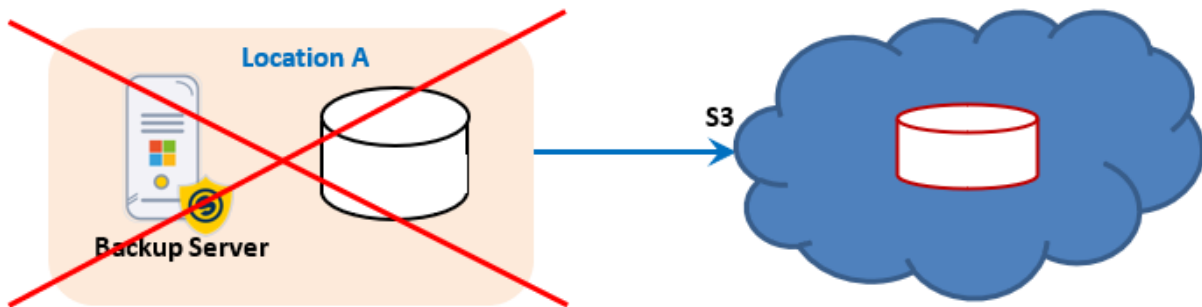


Dieses Szenario ist nicht sehr kompliziert, da der RDS keine große Menge an Metadaten lokal gespeichert hat, nur für den S3-Datenspeicher. Installieren Sie einfach das SEP-Paket auf der neuen RDS-Hardware, konfigurieren Sie die gleiche IP-Adresse und stellen Sie die Metadaten des S3-Datenspeichers wieder her. Dann lösen Sie die Rückreplikation vom Backup Server aus, der von der Infrastruktur in der Cloud wiederherstellbar ist.

Benutzerszenario 2: Backup Server außer Betrieb

Für beide Szenarien ist eine DR-Kopie des SEP Sesam Backup Server Verzeichnisses erforderlich. Dies ist ein wohldefinierter aber manueller Prozess, der eine einfache Konfiguration benötigt. Die Wiederherstellung des Backup Server Prozesses ist in unserem Wiki klar dokumentiert:

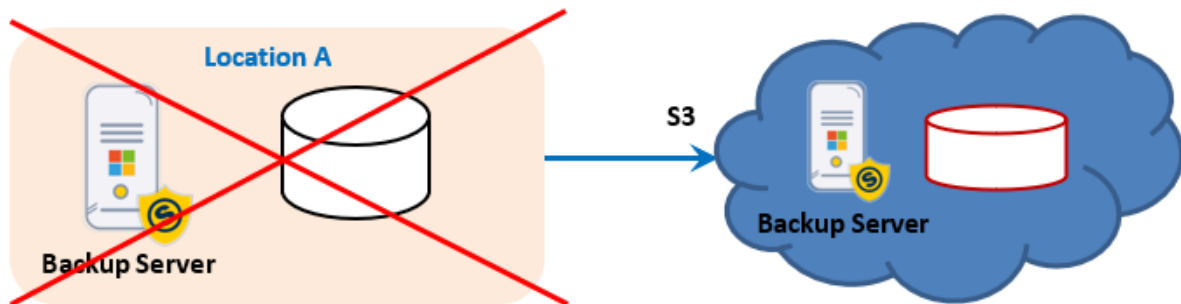
4. Ausfall von Standort A mit Wiederherstellung der physischen Maschine aus der Cloud.



Sie müssen Ihre gesamte Infrastruktur von Grund auf neu aufbauen.

- Aufbau und Konfiguration der benötigten physischen Infrastruktur
- Das Verzeichnis des SEP sesam Backup Servers wiederherstellen
- Wiederherstellung der Daten aus der Cloud

5. Ausfall von Standort A mit Wiederherstellung der virtuellen Maschine aus der Cloud.



Für einen schnelleren Ersatz von Standort A und zur Bereitstellung einer vorübergehenden Arbeitsumgebung in der Cloud ist es möglich

- Einrichten Ihres Backup-Servers über DR in der Cloud-Infrastruktur
- Sicherstellung des Zugriffs auf die Sicherungsdaten auf S3
- Wiederherstellung aller VMs in der Cloud
- Im Falle von VMware eine sofortige Wiederherstellung aller VMs durchführen

Performance

Bei einer Wiederherstellung aus der Cloud werden die Sicherungsdaten in deduplizierter Form von S3 auf den lokalen Speicher übertragen und lokal rehydriert. Dadurch wird sichergestellt, dass sich die Leistung bei der Sicherung und Wiederherstellung nicht wesentlich unterscheidet und die Bandbreite nicht zu einem zunehmenden Engpass bei der Wiederherstellung wird.

Da Si3-NG-S3 auf dem Si3-NG-Algorithmus basiert, gelten im Wesentlichen die gleichen Leistungswerte.

Benchmarks haben gezeigt, dass die Leistung von Si3-NG mit der Leistung eines normalen Datenspeichers vergleichbar ist. Daher haben wir mit S3 die gleichen Ergebnisse erzielt, wenn wir einen lokalen S3-Objektspeicher verwenden.

Das Netzwerk ist der größte Hemmschuh. Durch die Bereitstellung von mehr Hardware-Ressourcen wie CPU und RAM kann die Leistung gesteigert werden. Bei der Nutzung der öffentlichen S3-Cloud über das WAN ist jedoch in den meisten Fällen die Bandbreite der begrenzende Faktor. Um die Leistung abschätzen zu können, müssen Sie die Funktionsweise der Deduplizierung besser verstehen. Wenn beispielsweise keine neuen Datenchunks übertragen werden müssen, bedeutet dies, dass nur Metadaten geändert werden müssen, die lokal gespeichert werden, und die Backup-Leistung wird erhöht.

Nachfolgend finden Sie die gemittelten Zahlen aus unseren Benchmark-Tests:

- System 'dd' als Basis für die Leistungsfähigkeit Ihres Systems
- -10% für das Speichern in einen einfachen DataStore (Blockgerät)
- -20% für das Speichern in einen Si3-NG DataStore oder Si3-NG-S3 lokalen Objektspeicher
- -30 %+ für das Speichern in einem Si3-NG-S3 Cloud-Speicher

Diese Zahlen können erheblich variieren, je nach

- Art der Aktion:
 - o Backup
 - o Recovery
- Netzwerkverbindung:
 - o Disk Storage
 - o Bandwidth
- IOPS des Plattenspeichers
- Deduplizierungsrate
- Anzahl der parallelen Backups
- Arbeitslast des Backup-Servers/RDS

SEP sesam Zertifizierungen

Die unten aufgeführten Anbieter zeigen die aktuellen Zertifizierungen und validierten Cloud- und Objektspeicheranbieter. Diese Liste wird im Laufe der Zeit noch erweitert. Eine Validierung und Zertifizierung ist für alle unten aufgeführten Anbieter geplant. Allerdings sollte jeder Anbieter, der AWS S3 nutzt, entsprechend mit SEP Sesam Replikation arbeiten.

Cloud Storage Providers

 wasabi	Wasabi (certified)
---	--------------------

	Ebertlang S3 Cloud Storage (certified)
---	--

	AWS (validated)
---	-----------------

 IBM Cloud	IBM Cloud (validated)
---	-----------------------

Object Storage Provides

 DATACORE	DataCore Swarm S3 (validated)
---	-------------------------------

 Objects	Nutanix Objects (validated)
--	-----------------------------


 PURESTORAGE	PureStorage FlashBlade (validated)
--	------------------------------------

 MINIO	Min.io (validated)
--	--------------------

Kontakt

 +49 8024 463310

 sales@sep.de

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