

White Paper

Datendeduplizierung

Basierend auf Active Backup for Business 2.3.0



Inhaltsverzeichnis

Einleitung	3
Technologie zur Reduktion der Datenmenge	4
Deduplizierungsprozesse	
Changed Block Tracking (CBT)	
Deduplizierungsrate	
Fazit	9

Ihre Informationen finden

Synology veröffentlicht eine breite Palette von unterstützenden Dokumenten.

Im **Knowledge Center** finden Sie nützliche Hilfe- und FAQ-Artikel sowie Schritt-für-Schritt-Videoanleitungen. Ebenso finden Sie dort Benutzerhandbücher, Lösungshinweise, Broschüren und White Paper. Erfahrene Anwender und Administratoren finden Antworten und Anleitungen in technischen Administrator-Handbüchern und Entwicklerhandbüchern.

Sie haben ein Problem und können die Lösung in unserer offiziellen Dokumentation nicht finden? Durchsuchen Sie hunderte Antworten von Anwendern und Support-Mitarbeitern in der **Synology Community** oder wenden Sie sich mittels Webformular, per E-Mail oder telefonisch an den **Synology Support**.



Einleitung

Datenverlust oder Systemausfälle können fatale Folgen haben, egal, welches Speichergerät oder System Sie verwenden. Ihre wichtigsten Geräte, wie Ihre Arbeitsrechner oder Server, brauchen eine konkrete und effektive Datensicherungsstrategie, mit der Sie im Fall von Naturkatastrophen oder technischen Problemen kostspielige Systemausfälle vermeiden können.

Mit **Active Backup for Business (ABB)** können Sie die Sicherung aller Ihrer Geräte mit VMware, Hyper-V, Windows oder Linux und Ihrer Dateiserver über eine zentrale Konsole auf dem Synology NAS verwalten. Mit sofortiger Wiederherstellung von Windows-Geräten, virtuellen Maschinen und einzelnen Dateien können Sie selbst in einem Notfall wichtige Dienste funktionsfähig halten. Darüber hinaus nutzt Active Backup for Business einige der neuesten Technologien, wie **Dateneduplizierung** und **inkrementelle Sicherung**, um die Leistung der Datensicherung zu erhöhen und den benötigten Speicherplatz zu minimieren. Da diese Ansätze in verschiedenen Sicherungsdiensten häufig genutzt werden, haben unsere Benutzer öfter Fragen zur Technologie und Messung der Deduplizierungsrate in ABB gestellt.

Dieses Dokument stellt einige verwendete Deduplizierungstechnologien vor und definiert die in Synology Active Backup for Business implementierte Deduplizierungsrate.

Technologie zur Reduktion der Datenmenge

Durch die Reduktion der Datenmenge verbraucht die Datensicherung weniger Speicherplatz und durch diese optimierte Speichernutzung können so viele Geräte wie möglich gesichert werden. Active Backup for Business verwendet eine fortschrittliche Technologie, um die Datenmenge zu reduzieren und die Leistung der Datensicherung insgesamt zu verbessern.

Deduplizierungsprozesse

Deduplizierung ist ein effektiver Weg, durch das Entfernen redundanter Daten die Speichernutzung zu optimieren. Dies kann auf zwei grundlegende Arten erfolgen: Vorab oder im Nachhinein. Bei der **Deduplizierung vorab** werden redundante Daten entfernt, **bevor** sie in den Speicher geschrieben werden. Bei der **Deduplizierung im Nachhinein** werden Daten dagegen erst **nach** Abschluss der Sicherung entfernt.

Die Deduplizierung vorab kann je nach Sicherungslösung entweder im Sicherungsagent oder im Sicherungsziel erfolgen. Bei Deduplizierung über den Agent finden Datensuche, Deduplizierung und Übertragung jedoch gleichzeitig statt, was die Sicherung verlängert und die Leistung im Sicherungsziel beeinträchtigt.



Synology **Active Backup for Business** wurde mit **Deduplizierung vorab** entwickelt. Sie findet jedoch erst im Ziel statt, bevor die Daten in den Speicher geschrieben werden. Das verhindert, dass die Sicherung auf der Sicherungsquelle länger dauert und CPU-intensiver ist, wie es bei quellenseitiger Deduplizierung in der Regel der Fall ist. Außerdem können Benutzer mit Active Backup for Business verschiedene Systeme und Geräte in einem einzigen Ziel sichern, das möglicherweise redundante Daten enthält.

Um sämtliche redundanten Daten verschiedener Ziele zu entfernen, nutzt Active Backup for Business auch **globale Deduplizierung**. Damit kann Active Backup for Business Daten über viele Sicherungsaufgaben in einem einzigen gemeinsamen Ordner hinweg deduplizieren und so den tatsächlichen Speicherverbrauch optimieren, ohne dass die Leistung der Sicherung darunter leidet.

Einleitung

Für die Deduplizierung nutzt Active Backup for Business eine einzigartige ID zur Klassifizierung aller gesicherten Dateien und Dateiblöcke. Dabei unterteilt die Deduplizierung von Active Backup for Business gesicherte Daten in einzelne Blöcke von **4 KB** und weist diesen jeweils einen einzigartigen Fingerabdruck zu. Dieser wird von einem Deduplizierungsmodul berechnet, das die Blöcke mit dem **SHA-256-Algorithmus** verschlüsselt.

Nachdem ein Fingerabdruck für einen Dateiblock generiert wurde, erstellt das System einen Index der Fingerabdrücke für alle gespeicherten Blöcke. Damit werden eingehende Blöcke mit ihren gespeicherten Gegenübern verglichen. Werden redundante Daten erkannt, werden diese von der Deduplizierung sofort entfernt.

Changed Block Tracking (CBT)

Changed Block Tracking (CBT) ist eine gängige Form der inkrementellen Datensicherung in einer Reihe verschiedener Sicherungslösungen. **Inkrementelle Sicherung** ist eine Art der Datensicherung, bei der nur seit der letzten Sicherungsversion geänderte oder neu erstellte Daten gesichert werden. CBT kann helfen, bei der inkrementellen Sicherung Zeit und Ressourcen zu sparen, indem nur geänderte Datenblöcke gesichert werden, anstatt jedes Mal eine vollständige Sicherung auszuführen. Deshalb wird eine geringere Menge von Daten übertragen. Das heißt, die **Größe der übertragenen Sicherungsdaten** ist geringer als die ursprüngliche Datengröße in der Sicherungsquelle.

Active Backup for Business integriert diese Funktion in seine Sicherung von PCs, physischen Servern und virtuellen Maschinen. ABB führt für Sicherungsaufgaben eine Forever-Incremental-Datensicherung aus, was die Anzahl verfügbarer Sicherungsversionen maximiert und den Speicherverbrauch minimiert, da nach der ersten vollständigen Datensicherung nur mehr geänderte Daten gesichert werden.



Deduplizierungsrate

Definition

Die **Deduplizierungsrate** misst die Effektivität der von einer Sicherungslösung ausgeführten Dateneduplizierung. Sie ist oft ein entscheidender Faktor für die Wahl einer bestimmten Lösung, da damit Funktionen und Leistung verschiedener Anbieter verglichen werden können. Die Deduplizierungsrate kann jedoch je nach verwendeter Technologie und allgemeiner Charakteristik der Daten variieren.

Eine **Deduplizierungsrate** oder **Komprimierungsrate** wird definiert als Verhältnis der gesicherten Daten und der Menge an Speicher, die die Sicherungsdaten im Ziel nach dem Entfernen redundanter Daten tatsächlich benötigen.

Beispiel: Angenommen, Sie haben 100 GB an Daten von Ihrem System gesichert und auf dem Sicherungsserver wurden 20 GB belegt. Indem Sie nun die 100 GB an gespeicherten Daten durch die auf dem Server belegten 20 GB dividieren, kommen Sie auf eine Deduplizierungsrate von 5:1, oder anders ausgedrückt 80 %.

Die Deduplizierungsrate wird auch häufig im Marketing verwendet, um die Leistungsfähigkeit einer Sicherungslösung zu bewerben. Beispielsweise kann angegeben werden, ein Sicherungsprodukt hätte eine 20x höhere Deduplizierungsrate als andere Lösungen auf dem Markt. Allerdings variiert die Deduplizierungsrate je nach verwendeter Technologie, allgemeinen Charakteristiken der Daten, verwendeter Sicherungsmethode und Berechnungsmethode.

Daher ist sie vielleicht nicht immer das beste Mittel, um verschiedene Produkte zu vergleichen. Sie sollten also auch berücksichtigen, wie die Deduplizierungsrate jeder einzelnen Lösung berechnet wird und wie viel Speicherplatz im Sicherungsziel tatsächlich belegt wird.

Berechnung

Mit den folgenden Werten können Sie die Deduplizierungsrate berechnen:

- Gesamtkapazität der gesicherten Daten vor dem Entfernen von Duplikaten
- Größe der übertragenen Sicherungsdaten
- Im Speicherziel tatsächlich belegte Kapazität

Allgemein ist es eine gute Berechnungsmethode, die **Größe der übertragenen Sicherungsdaten** durch die **im Speicherziel tatsächlich belegte Kapazität** zu dividieren. So macht es auch Synology.

Einleitung

Manche Anbieter dividieren jedoch die **Gesamtkapazität der gesicherten Daten vor dem Entfernen von Duplikaten** durch die **im Speicherziel tatsächlich belegte Kapazität**, was nicht zwangsläufig eine zutreffende Deduplizierungsrate ergibt.

Denken Sie beispielsweise an folgendes Szenario:

- Sicherungslösung: A
- Gesamtkapazität der gesicherten Daten vor dem Entfernen von Duplikaten: 6 TB
- Größe der übertragenen Sicherungsdaten: 1,2 TB
- Im Speicherziel tatsächlich belegte Kapazität: 0,83 TB
- Beworbene Deduplizierungsrate: 7,2:1

Hier wird Sicherungslösung A mit einer Deduplizierungsrate von **7,2:1** (6 TB / 0,83 TB) beworben. Diese Berechnung ist jedoch nicht vollkommen zutreffend, da nur **1,2 TB an Sicherungsdaten** übertragen wurden. Die Menge an redundanten Daten wurde dann durch CBT reduziert und die Daten wurden zur Speicherung im Speicherziel weiter auf **0,83 TB** komprimiert.

Sicherungslösung A hat die **Gesamtkapazität der gesicherten Daten** durch die **im Speicherziel tatsächlich belegte Kapazität** dividiert, um eine höhere Deduplizierungsrate zu erreichen. Tatsächlich zeigt jedoch nur die **Größe der übertragenen Sicherungsdaten**, wie hoch die deduplizierte Datenmenge tatsächlich ist.

Die **Größe der übertragenen Sicherungsdaten** durch die **Im Speicherziel tatsächlich belegte Kapazität** zu dividieren, würde zu einem noch akkurateren Ergebnis führen. Die Berechnung der Deduplizierung kann daher angepasst werden:

Deduplizierungsrate

Größe der übertragenen Sicherungsdaten	÷	Im Speicherziel tatsächlich belegte Kapazität	=	Dateneduplizierungsrate
1,2 TB		0,83 TB		1,44:1

Deduplizierungsrate

(Größe der übertragenen Sicherungsdaten	-	Im Speicherziel tatsächlich belegte Kapazität)	÷	Größe der übertragenen Sicherungsdaten	x 100 %	=	Deduplizierungsrate
	1,2 TB		0,83 TB			1,2 TB			30,8 %

Sicherungsmethoden

Ein weiterer Aspekt, der die Deduplizierungsrate beeinflussen kann, ist die **Sicherungsmethode** – besonders, ob eine **vollständige** oder eine **inkrementelle Datensicherung** durchgeführt wird. Lösungen mit ausschließlich vollständiger Datensicherung können vermeintlich eine höhere Deduplizierungsrate haben als jene mit inkrementeller Sicherung. Dies liegt daran, dass bei der inkrementellen Sicherung nur geänderte Datenblöcke übertragen und redundante Daten bereits vorab entfernt werden, wodurch im Sicherungsziel weniger Speicher belegt wird. Vollständige Sicherungen enthalten im Vergleich dazu große Mengen an redundanten Daten, die sich seit der letzten Sicherungsversion nicht geändert haben.

Daher benötigen inkrementelle Datensicherungslösungen weniger tatsächlichen Speicherplatz im Ziel, auch wenn sie vermeintlich eine schlechtere Deduplizierungsrate haben als Lösungen mit vollständiger Sicherung.

Deduplizierungsrate von Active Backup for Business

Wie erwähnt kann die Berechnung der Deduplizierungsrate anhand der **Gesamtkapazität der gesicherten Daten vor dem Entfernen von Duplikaten** eine höhere Deduplizierungsrate ergeben, spiegelt möglicherweise aber nicht die tatsächlich deduplizierte Menge an Daten wider. Active Backup for Business verwendet daher zur Berechnung stattdessen die **Größe der übertragenen Sicherungsdaten**, da dies die Menge an Daten ist, die tatsächlich dedupliziert wird. Wie in der folgenden Tabelle ersichtlich, führt die Deduplizierung in Active Backup for Business dazu, dass **im Speicherziel weniger tatsächliche Kapazität belegt wird**. Sie spart also Benutzern insgesamt mehr Speicherplatz.

Zusammengefasst sollten sich Benutzer statt auf die Deduplizierungsrate mehr auf den tatsächlich belegten Speicher im Verhältnis zur Größe der übertragenen Sicherungsdaten konzentrieren, um genauer zu sehen, wie sehr eine Lösung ihnen Speicherplatz sparen kann.

Die folgende Tabelle zeigt die tatsächliche Deduplizierungsrate von **Active Backup for Business** und vergleicht sie mit jener anderer Anbieter.

Anbieter	Größe der übertragenen Sicherungsdaten	Im Speicherziel tatsächlich belegte Kapazität	Deduplizierungsrate	Deduplizierungsrate
Synology Active Backup for Business	1,2 TB	637,2 GB	1,2 TB/0,62 TB = 1,93:1	48,18 %
Anbieter A	1,2 TB	819,8 GB	1,2 TB/0,8 TB = 1,5:1	33,33 %
Anbieter B	1,2 TB	847,7 GB	1,2 TB/0,82 TB = 1,45:1	31,08 %

Sicherungsquellen: 9 Windows-PCs, 1 Windows Server auf VM



Fazit

Die Dateneduplizierung ist ein komplexer Prozess, mit dem Benutzer die Menge an redundanten Daten reduzieren und so ihre Datensicherungen effizienter verwalten und die Speichernutzung optimieren können. Da die Deduplizierung in einer Reihe verschiedener Sicherungslösungen zum Einsatz kommt, vergleichen viele Benutzer Produkte anhand ihrer Deduplizierungsrate. Angesichts einer Reihe möglicher Messmethoden und verschiedener Faktoren, die die Deduplizierungsrate beeinflussen können, sind die angegebenen Raten nicht immer zu 100 Prozent zuverlässig. Beim Vergleich der Deduplizierungsraten verschiedener Lösungen sollten Sie daher also mehr darauf achten, wie viel physischer Speicherplatz nach der Deduplizierung tatsächlich belegt wird.

Synology Active Backup for Business liefert nicht nur eine umfassende Sicherungslösung für verschiedene Plattformen, sondern verfügt auch über schnelle, zuverlässige und effektive Dateneduplizierung, damit Sie Zeit sparen und Ihren Speicherplatz optimal nutzen können.



SYNOLOGY INC.

9F, No. 1, Yuan Dong Rd.
Banqiao, New Taipei 220545
Taiwan
Tel.: +886 2 2955 1814

**SYNOLOGY
AMERICA CORP.**

3535 Factoria Blvd SE, Suite #200,
Bellevue, WA 98006, USA
Tel.: +1 425 818 1587

**SYNOLOGY UK
LTD.**

Unit 5 Danbury Court, Linford Wood,
Milton Keynes, MK14 6PL, United
Kingdom
Tel.: +44 (0)1908048029

**Synology France
SARL**

102 Terrasse Boieldieu (TOUR W)
92800 Puteaux France
Tel.: +33 147 176288

**SYNOLOGY
GMBH**

Grafenberger Allee
29540237 Düsseldorf
Deutschland
Tel.: +49 211 9666 9666

**SYNOLOGY
SHANGHAI**

200070, Room 201, No.
511 Tianmu W. Rd.,
Jingan Dist., Shanghai,
China

**SYNOLOGY JAPAN
CO., LTD.**

4F, No. 3-1-2, Higashikanda
Chiyoda-ku, Tokio, 101-0031
Japan

Synology®



synology.com

Synology kann ohne vorherige Ankündigung jederzeit Änderungen an den technischen Daten und Produktbeschreibungen vornehmen. Copyright © 2022 Synology Inc. Alle Rechte vorbehalten. Synology und Namen anderer Synology-Produkte sind geschützte Marken oder eingetragene Warenzeichen von Synology Inc. Weitere hier genannte Produkte und Firmennamen sind Warenzeichen der entsprechenden Eigentümer.