



Barracuda Backup Deduplizierung

---

White Paper

## Kurzbeschreibung

Datenschutztechnologien spielen in Unternehmen verschiedenster Größe eine wichtige Rolle. Es gibt jedoch verschiedene Herausforderungen zu bewältigen, um diese Technologien optimal einsetzen und nutzen zu können. Barracuda Backup bietet eine vollständig integrierte, kapazitätsoptimierte Lösung, die sich nicht nur problemlos bereitstellen lässt, sondern tatsächlich auch absolut zuverlässig und sicher ist. Mit der Barracuda Deduplizierung, einem festen Bestandteil sämtlicher Barracuda Backup-Produkte, können Unternehmen ihren Kapazitätsbedarf und ihre Backup-Kosten deutlich reduzieren. Unternehmen, die mehrere Standorte schützen müssen, können auf die globale Deduplizierungs- und Cloudspeichertechnologie von Barracuda vertrauen, mit der der Schutz der verteilten Netzwerke gewährleistet ist und zugleich der insgesamt benötigte Speicherplatz möglichst gering ausfällt.

## Herausforderungen für Backup-Speicher

Datenschutzbemühungen gibt es schon seit jeher, auch wenn das Volumen der zu schützenden Daten zunehmend gewachsen ist. Während die Herausforderungen, aufgrund des gewaltigen Wachstums der gespeicherten Datenmengen, nicht neu sind, hat sich die Wachstumsgeschwindigkeit selbst deutlich erhöht und viele Unternehmen haben versucht, ihre zunehmenden Speicher- und Aufbewahrungsanforderungen auf herkömmliche Weise zu lösen, indem sie die Daten auf externen Medien (z. B. Bandsicherung, SAN/NAS) sichern und speichern. Diese Verfahren sind häufig sehr komplex, zeitaufwendig und anfällig für menschliche Fehler oder Hardwarefehler. Die Verwendung herkömmlicher externer Medien zur Speicherung von Backups stellt keine effiziente Lösung des Speicherproblems dar und kostet die Unternehmen langfristig gesehen mehr, als von Anfang an in eine deutlich effizientere Lösung zu investieren. Die Kapazitätsoptimierung ist ein integraler Bestandteil jeder anspruchsvollen Backup-Plattform.

Neben Kapazitätsengpässen haben viele Unternehmen das Problem, dass ihre gesamte IT-Infrastruktur nicht zentralisiert ist und an entfernten Standorten entweder zu viele oder zu wenig IT-Mitarbeiter bereitstehen – oder eventuell sogar überhaupt kein Schutz gewährleistet ist. Bandsicherung und andere externe Medien sind standortübergreifend nicht effektiv anpassbar und erfordern zusätzliche Intervention von IT-Personal um konsistente Backups zu gewährleisten und Wachstum zu ermöglichen. Unternehmen, die standortübergreifende Konnektivität bereitstellen, sichern entfernte Standorte häufig über ein zentrales Rechenzentrum. Dies führt zu hoher Auslastung der verfügbaren Bandbreite und damit zu Verzögerungen bei Sicherung und Wiederherstellung. Durch eine Technologie wie Deduplizierung, die eine deutliche Reduktion der zu übertragenden Datenmenge ermöglicht, kann nicht nur die Überlastung unternehmensinterner Netzwerke vermieden, sondern auch der Zeitaufwand für Sicherung und Wiederherstellung reduziert werden.

## Was ist Deduplizierung?

Die Deduplizierung ist ein Prozess, bei dem Dateien und andere Daten in „Datenblöcke“ unterteilt werden und eine Tracking-Datenbank genutzt wird, um sicherzustellen, dass nur eine einzelne Kopie jedes Datenblocks innerhalb der gesamten Sicherungsdaten gespeichert wird. Bei nachfolgenden Backups werden eingehende Daten mit der Tracking-Datenbank abgeglichen, um zu prüfen welche Datenblöcke geschützt sind und nur einzigartige Datenblöcke zu übertragen und zu speichern. Wenn beispielsweise fünf unterschiedliche Server Daten auf eine Barracuda Appliance sichern und ein Datenblock ermittelt wird, der auf jeden dieser fünf Server vorhanden ist, wird nur eine einzelne Kopie dieses Datenblocks tatsächlich auf der Appliance gespeichert. Jeder Datenblock wird dabei mit Hinweisen versehen wie er, im Falle einer Wiederherstellung, über alle fünf Server aufbereitet (kompiliert) werden muss. Durch die Tracking-Datenbank wird sichergestellt, dass diese Datenblöcke aufbewahrt werden, bis alle Backups, die auf einen bestimmten Datenblock verweisen, gelöscht werden. Dass nur einzigartige Bestandteile von Daten auf dem Server gespeichert werden, führt zu einer immensen Reduktion der benötigten Kapazität. Bei einem Recovery wird eine Datei anhand der Informationen in der Tracking-Datenbank aufbereitet und dann an den Zielort zur Wiederherstellung gesendet.

## Barracuda Deduplizierung

Es gibt – je nach Hersteller – verschiedene Arten von Deduplizierungstechnologien und diese Unterschiede haben Einfluss auf das Ergebnis für IT-Administratoren. Bei einem älteren, einfachen Deduplizierungsverfahren mit der Bezeichnung „Post-Process“ muss bei Backups gewartet werden bis der Schreibvorgang auf die Festplatte abgeschlossen ist, bevor der Deduplizierungs- und Replizierungsvorgang gestartet werden kann. Dadurch verlängert sich die Zeitdauer bis zum vollständigen Schutz der Daten. Außerdem erhöht sich die Auslastung der lokalen Systeme, da die Daten auf der Festplatte drei Mal adressiert werden müssen, bevor sie repliziert werden können (schreiben als Backup-Daten, lesen für Deduplizierung, schreiben als deduplizierte Daten). Für die Post-Process Deduplizierung ist außerdem die doppelte Zielkapazität erforderlich – einerseits für den Backup-Datensatz und andererseits zur Speicherung der deduplizierten Daten. Daraus resultiert eine ineffiziente Nutzung der Speicherkapazität. In einigen Fällen müssen Anbieter, die auf Post-Process Deduplizierung vertrauen, Storage-Inhalte auf externe Medien wie Bandsicherungen oder Datenträger auslagern, um Datenaufbewahrungsanforderungen zu erfüllen.

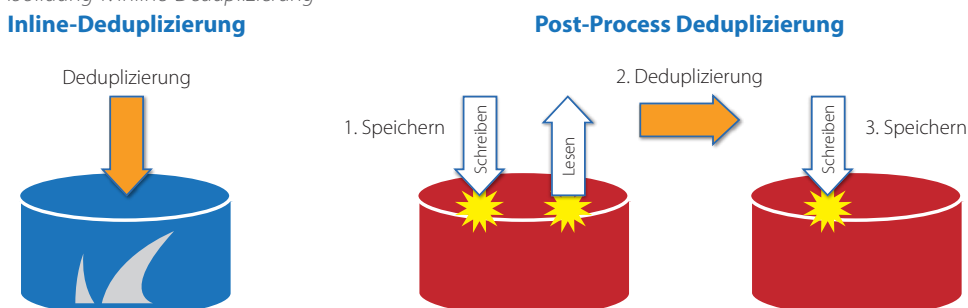
Anstelle von Post-Process entwickelte Barracuda seine Inline Deduplizierungstechnologie (Abbildung 1). Bei der Inline-Deduplizierung führt die Appliance die Deduplizierung in einem Schritt – bei der Datenaufnahme – aus. Damit erübrigt sich der Bedarf für überflüssig große Zielspeicherkapazität, wie sie für das langsamere Post-Process Verfahren mit seinen zwei Teilschritten nötig ist. Mit der Inline-Deduplizierung von Barracuda können Unternehmen Kosten sparen, da kein größeres Festplattensystem benötigt wird, um die aufgenommenen Daten zu speichern, bevor die Deduplizierung gestartet wird. Mit dieser Deduplizierungsmethode kann auch das Datenverlustrisiko reduziert werden: die Verarbeitungszeit bis zum Backup und die vollständige Replikation werden beschleunigt, da die Daten für die Replikation schon in die Warteschlange gestellt werden während der Backup-Job verarbeitet wird.

Durch die Bereitstellung einer Barracuda Backup Appliance wird die DR-Bereitschaft deutlich verbessert, weil die benötigte Zeit für die Datenübertragung zu standortfernen Speicherorten durch Inline-Deduplizierung und Instant-Replikation deutlich reduziert wird. Durch das Inline-Deduplizieren stehen die Daten schneller zur Replikation bereit, als wenn sie erst nach Abschluss des Backups verarbeitet werden können. Und weil nicht die gesamte Datenmenge vor Beginn der Replikation aufgenommen werden muss, können die Daten bereits während das Backup und die Deduplizierung ausgeführt werden zu einem standortfernen Speicherort übertragen werden. Auf diese Weise wird schnellerer Offsite-Schutz gewährleistet.

Vergleich zwischen Post-Process-Deduplizierung und der Inline-Deduplizierung von Barracuda:

- **Kosten:** Weil die Post-Process Deduplizierung einen Zielspeicherort benötigt bevor die Daten in einen deduplizierten Zustand gebracht werden, ist, abhängig von der Größe des Datensatzes, oft eine größere und teurere Appliance nötig. Mit der Barracuda Inline Lösung ist dies nicht notwendig.
- **Zeit:** Bei der Post-Process Deduplizierung muss zuerst der Backup-Vorgang abgeschlossen sein, bevor der Deduplizierungsvorgang gestartet werden kann – und erst dann erfolgt die Datenreplikation. Obwohl dies den Backup-Vorgang zunächst zu beschleunigen scheint, sind die Daten noch nicht vollständig geschützt, da der Deduplizierungs- und der Replizierungsvorgang erst abgeschlossen sein müssen. Diese serielle Verarbeitung und der mehrstufige Vorgang an sich können – verglichen mit der Echtzeit-Inline-Deduplizierung mit gleichzeitiger Replikation – zu enormen Verzögerungen führen, bis der vollständige Datenschutz gewährleistet ist.
- **Risiko:** Bei der Post-Process Deduplizierung verlängert sich die Zeit bis zur DR-Bereitschaft und jeder Fehler in diesem dreistufigen Nachbearbeitungsvorgang kann zu einem beschädigten Datensatz führen. Falls Netzwerkprobleme auftreten oder es an einem Standort zu einem Ausfall kommt, können Daten verloren gehen – selbst wenn der Backup-Vorgang als abgeschlossen gilt.

Abbildung 1. Inline-Deduplizierung



## Deduplizierungsmethodik von Barracuda

Barracuda Backup nutzt ein dreistufiges Modell – Quellen-, Ziel- und Globale Deduplizierung. Dadurch kann die Bandbreite minimiert, die Speichernutzung optimiert und der Aufwand für die Datensicherung reduziert werden.

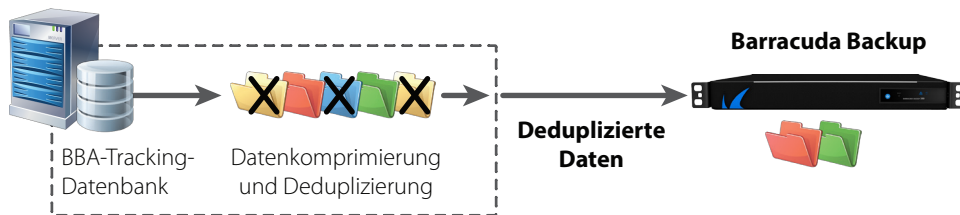
- **Quelendeduplizierung:** Lokale Daten werden an der Quelle dedupliziert und in deduplizierter Form an die Barracuda Backup Appliance gesendet, um dadurch die LAN-Bandbreite und die Daten, die an den lokalen Server gesendet werden, zu reduzieren. Dieser Vorgang wird auch als Client-seitige Deduplizierung bezeichnet.
- **Zieldeduplizierung:** Daten werden, über verschiedene Quellen hinweg direkt in der Backup Appliance dedupliziert. Dadurch reduziert sich die Datenmenge, die zwischengespeichert und repliziert werden muss.
- **Globale Deduplizierung:** Daten werden über alle lokalen Server hinweg dedupliziert, die auf eine zentrale Appliance oder in die Cloud repliziert wurden.

Quelle:

Die Quelendeduplizierung wird durch den Barracuda Backup Agent implementiert. Bei der Installation der Software wird auf dem Server eine kleine Datenbank erstellt, die Datenblöcke nachverfolgt, damit nur einzigartige Daten, die der Agent ermittelt, komprimiert und zur Verarbeitung an die Appliance gesendet werden. Dadurch wird der Netzwerkverkehr reduziert und das Zeitfenster für Backup verkleinert (Abbildung 2).

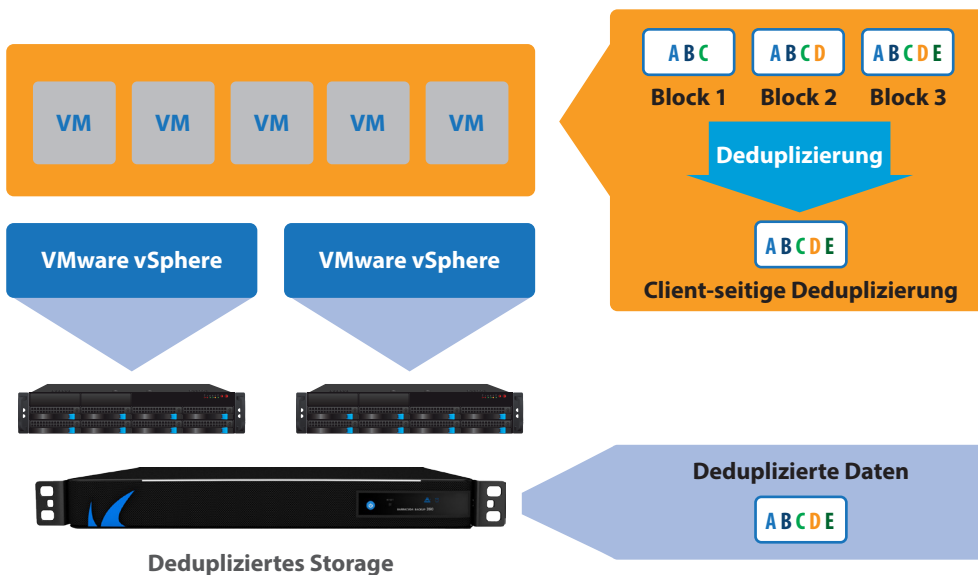
Abbildung 2. Quelendeduplizierung

### Clientcomputer



Bei VMware-Backups nutzt Barracuda zur Sicherung der virtuellen Datenträger die vStorage APIs for Data Protection (VADP) von VMware. Mit VADP kann Barracuda das Changed Block Tracking-Verfahren (CBT, d. h. Erfassung der veränderten Datenblöcke) nutzen, damit nur einzigartige Datenblöcke an die Barracuda Backup Appliance gesendet werden (Abbildung 3).

Abbildung 3. Changed Block Tracking

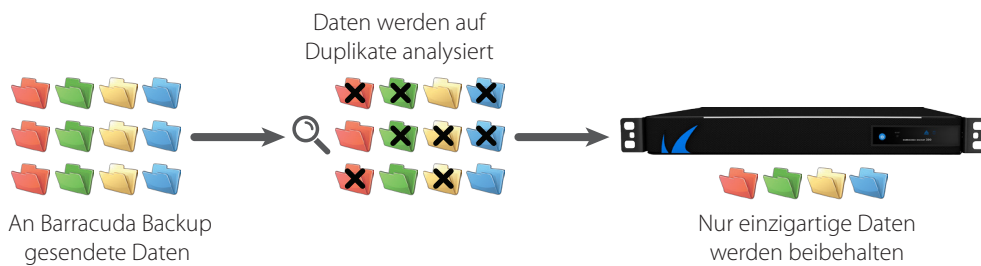


Bei Microsoft Hyper-V-Backups verkürzt der Barracuda Backup Agent das Backup-Zeitfenster durch die Deduplizierung der VHD-Dateien auf dem Host-Server, um so die Datenmenge zu reduzieren, die an die Backup Appliance gesendet wird.

*Ziel:*

Zieldeduplizierung erfolgt auf der Barracuda Backup Appliance um eine Redundanz über alle lokale Agents auszuschliessen und eine Minimierung des lokalen Cache und der für Backups benötigten Kapazität in der Cloud zu erreichen. Unternehmen die SAN- oder NAS-Dateiserver, VMware und File-Shares sichern, können häufig keine quellenbasierte Deduplizierung verwenden, wodurch Zieldeduplizierung zur primären Methodik wird (Abbildung 4).

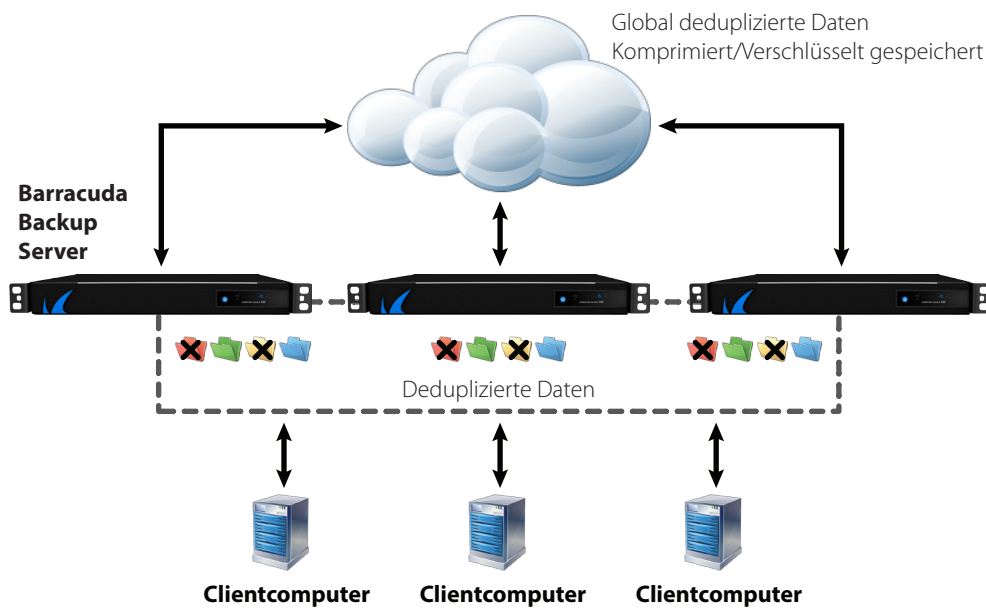
Abbildung 4. Zieldeduplizierung.



*Global:*

Die globale Deduplizierung wird entweder auf einer Appliance implementiert, die als zentrales Replikationsziel genutzt wird, oder in der Cloud. Auf diese Weise entfällt die Redundanz zwischen Backup-Appliances innerhalb der weltweiten Infrastruktur und Unternehmen erhalten die Möglichkeit, die Kapazität, die zum Speichern von Backup Daten in komprimierter und verschlüsselter Form notwendig ist, zu reduzieren (Abbildung 5).

Abbildung 5. Globale Deduplizierung



Barracuda Backup nutzt die dreistufige Deduplizierung für mehrere Datenquellen wie in der folgenden Tabelle beschrieben:

Datenquellentyp	Backup-Methode	Deduplizierungsmethode
Microsoft Exchange Server Microsoft SQL Server Microsoft Hyper-V Microsoft Windows Microsoft Active Directory Microsoft SharePoint	Barracuda Windows Backup Agent	Quelle, Ziel und Global
Lotus Domino Server	Windows Volume Shadow Copy Service (VSS)	Quelle, Ziel und Global
Linux-Systeme	Barracuda Linux Backup Agent	Quelle, Ziel und Global
Novell Open Enterprise Server 2 SP2.0+	Barracuda Linux Backup Agent	Quelle, Ziel und Global
Mac OS X	Barracuda Macintosh Backup Agent	Quelle, Ziel und Global
Unix	CIFS/SSHFS	Ziel und Global
Network Addressable Storage	CIFS/Access Control Lists (ACL)	Ziel und Global
VMware Server und Guests	VDAP mit Changed Block Tracking (CBT)	Quelle, Ziel und Global

## Implementierung der Barracuda Deduplizierung

Barracuda Backup bietet maximale Datenreduktion und minimalen Speicherkapazitätsbedarf durch anwendungsorientierte Inline Deduplizierung auf Blockebene mit variabler Länge. Dadurch wird der insgesamt benötigte Speicherplatz, die Anforderungen an die Bandbreite und die Zeitdauer für Backup und Wiederherstellung reduziert. Im Laufe der Zeit können Benutzer, durch den dreistufigen Deduplizierungsprozess, eine Datenreduktion um das 20- bis 50-fache (Durchschnittswert) für die Sicherung von Standarddatensätzen feststellen.

Die Länge der gesicherten Datenblöcke, die für die Deduplizierung genutzt werden, ist abhängig von Dateityp und -größe. Jeder Datenblock erhält drei eindeutige Hash-Werte (digital fingerprints), MD5sum, SHA1 und Größe der Datei. Jeder Hash-Wert ist für jeden einzelnen Datenblock individuell und wird vom Barracuda Backup Agent in einer Datenbank gespeichert. Der Barracuda Backup Agent wird dabei auf dem lokalen Server ausgeführt, gemeinsam mit einer Datenbank auf der lokalen Appliance. Bei der Ausführung des Backups werden die einzelnen berechneten Hash-Werte mit den Werten der Datenblöcke abgeglichen, die bereits verarbeitet wurden. Sofern es sich um einen einzigartigen Wert handelt, wird der Datenblock an die Appliance übertragen. Ist ein Hash-Wert identisch mit einem bereits gespeicherten, wird lediglich ein Hinweis an die Appliance übermittelt. Wenn die Daten zur lokalen Barracuda Backup Appliance hinzugefügt werden, erfolgt erneut ein Abgleich der Hash-Werte über alle Agents. Wenn doppelte Einträge gefunden werden, speichert die Appliance eine einzelne Kopie der Daten auf der lokalen Appliance und macht einen Vermerk, dass diese Kopie gesichert wurde und auf jedem Server wiederhergestellt werden kann, der den entsprechenden Hash anfordert.

Die folgenden Informationen sollten im Hinblick auf eine optimale Performance des Barracuda Backup Agent beachtet werden:

### Anpassung der Größe der Deduplizierungsdatenbank:

Die lokale Deduplizierungsdatenbank macht nur einen kleinen Anteil der gesamten Dateisystemgröße aus. Die Größe der Datenbank nimmt linear zu den insgesamt gespeicherten deduplizierten Daten zu. Planen Sie bei der Anpassung der Datenbankgröße circa 1-3 GB Datenbankgröße pro TB an gespeicherten, deduplizierten Daten, d. h. bei 2 TB deduplizierten Daten entspricht dies einer Datenbankgröße von 4 GB.

## Prozessorauslastung

Durch den Barracuda Backup Agent kann sich die Prozessorauslastung während eines Backups erhöhen, da der Barracuda Backup Agent für den Quellenduplizierungsvorgang und die Komprimierung die CPU des Clientcomputers nutzt. Es erfolgt keine Einschränkung von Seiten des lokalen Barracuda Backup Agent, um die Prozessorressourcen des Clientcomputers während der Datensicherung und –wiederherstellung einzugrenzen.

## Speicherauslastung

Während des Backups verursacht der Barracuda Backup Agent eine erhöhte Speichernutzung des Systems. Der Agent nutzt den Arbeitsspeicher des Clientcomputers für die Speicherung der Backup-Daten, bevor diese (in komprimierter Form) ebenso wie die Hashes an die Barracuda Backup Appliance gesendet werden. Der Agent nutzt beim Backup-Vorgang bis zu 512 MB Speicher für die Speicherung der Datenblöcke. Auf diese Weise ist ein schnelles Absuchen des Dateisystems möglich.

## Barracuda-Lösung im Vergleich zu anderen Deduplizierungsansätzen

Bei Produkten ohne Änderungserkennung und Deduplizierung werden immer wieder dieselben Daten gesichert. Diese Vorgehensweise verursacht lange Backup-Zeiten und massive Speicher- und Bandbreitenauslastungen. Durch das Deduplizierungsverfahren von Barracuda Backup kann die für Backups benötigte Zeit reduziert werden. Kunden erhalten die Möglichkeit, nächtliche Backups auszuführen. Barracuda nutzt drei unterschiedliche Deduplizierungsoptionen, die eine zuverlässige und zugleich einfach anwendbare Lösung bereitstellen. In diesem Abschnitt sollen die spezifischen Charakteristika näher untersucht werden, die der Barracuda-Technologie zugrunde liegen.

### Blöcke mit fester bzw. variabler Länge

Die Deduplizierung von Blöcken mit fester Länge ist die einfachste Deduplizierungsmethode. Bei diesem Verfahren werden spezifische Datenblöcke, mit fester Größe, des zu sichernden Datensatzes überprüft. Da sich die Blockgröße nicht ändert, benötigt die Deduplizierungsmethode mit festen Blöcken geringere CPU- und Festplattenressourcen. Die möglichen Reduktionen, verglichen mit anspruchsvolleren Deduplizierungsverfahren, sind jedoch begrenzt, da aufgrund des vordefinierten Datenblocks Duplikate mancher Datensätze nicht gefunden werden.

Die anwendungsorientierte Deduplizierung mit variablen Blöcken ist ein fortschrittliches Verfahren, das die zu sichernden Daten/Anwendungen analysiert und die Blockgröße für optimale Ergebnisse entsprechend vergrößert oder reduziert. Da sich die Blockgröße je nach Beschaffenheit der zu sichernden Daten ändern kann, sind zusätzliche CPU- und Festplattenressourcen erforderlich, um die Deduplizierung auszuführen. Es wird jedoch eine maximale Datenreduzierung erreicht. Barracuda Backup verwendet die Deduplizierung mit variablen Blöcken für alle drei Stufen. Bei der erweiterten Deduplizierung mit variablen Blöcken von Barracuda wird der Datentyp und die Blockgröße analysiert und eine Blockgröße festgelegt, mit der eine optimale Deduplizierung erzielt werden kann, ohne dadurch zusätzliche CPU- und Plattenressourcen zu beanspruchen.

Barracuda Backup ist eine Hardware-Appliance und kann daher die Deduplizierung mit variablen Blöcken ausführen, ohne CPU- und Festplattenressourcen zu beanspruchen. Die zugrunde liegende Hardware und Software in der Backup Appliance wird für diese Blockmethode optimiert und ermöglicht eine maximale Datenübernahmerate.

## Software-Deduplizierung im Vergleich zu Hardware-Deduplizierung

### Software-Deduplizierung

Bei Software-Deduplizierung handelt es sich üblicherweise um Add-Ons/Plug-Ins, die von Softwareherstellern angeboten werden, mit dem Ziel den Speicherplatzbedarf zu reduzieren, der für die Sicherung von Backup-Daten auf Datenträger benötigt wird. Die Software-Deduplizierung wird häufig in Ergänzung zur Komprimierung eingesetzt, bevor die Daten auf ein Medium geschrieben werden. Bei der Software-Deduplizierung müssen die Unternehmen in Bezug auf ihre Bereitstellung strenge Hardwareanforderungen einhalten, da durch den Deduplizierungsvorgang zusätzliche Ressourcen gebunden werden. Die Software-Deduplizierung ist häufig kostengünstiger oder in die Backup-Software integriert. Die Software stellt jedoch nur einen Bruchteil der tatsächlichen Backup-Bereitstellung dar: Unternehmen übersehen dabei häufig die komplexen Anforderungen des schwerfälligen Software-Deduplizierungsmodells.

### Hardware-Deduplizierung

Die Hardware-Deduplizierung erfolgt mit einem spezifischen Storage-Gerät, das als Ziel für die Backup-Software fungiert bzw. über eine eigene integrierte Softwarelösung verfügt. Diese Lösung hat sich aufgrund ihrer Effizienz und Performancevorteile gegenüber der Software-Deduplizierung in den meisten mittleren und großen Unternehmen durchgesetzt. Die Hardware-Deduplizierung wird dabei verwendet, um die zusätzlichen Ressourcen, die zur Dateneduplizierung nötig sind auszulagern. Die Hardware-Appliance übernimmt CPU- und speicherintensive Prozesse für die Deduplizierung und eliminiert dadurch den Bedarf für mehrere Geräte zum Schutz der Netzwerkumgebung.

## Schlussfolgerung

Durch die Deduplizierung von Barracuda Backup wird der Datenschutz vereinfacht und Fixkosten, Medienkosten und Netzwerkkosten reduziert. Die dreistufige Inline-Deduplizierungslösung mit variabler Blocklänge von Barracuda ermöglicht eine effiziente langfristige Speicherung der geschützten Server. Darüber hinaus verkürzt sich die für Backups benötigte Zeit. Mit der Barracuda-Deduplizierung erhalten Unternehmen die Möglichkeit, entfernte Niederlassungen mit überschaubarem Aufwand von Ressourcen von einem zentralen Standort aus zu schützen. Unternehmen, die das Ziel verfolgen, virtuelle Umgebungen zu schützen, können ihren gewünschten Wiederherstellungszeitpunkt (Recovery Point Objective, RPO) sicherstellen und zugleich den insgesamt benötigten Speicherplatz reduzieren, indem sie für VMware und Hyper-V jeweils die CBT-Methode nutzen und dadurch nur die erforderlichen Änderungen innerhalb des Datensatzes speichern.

### Über Barracuda Networks, Inc. (NYSE: CUDA)

Barracuda Networks bietet Cloud-vernetzte Security- und Storage-Lösungen, die die IT-Infrastruktur vereinfachen. Bereits über 150.000 Kunden weltweit vertrauen den leistungsstarken, benutzerfreundlichen und kostengünstigen Appliances. Sie stehen als virtuelle, Cloud-basierte oder hybride Lösung zur Verfügung. Das kundenorientierte Geschäftsmodell von Barracuda konzentriert sich auf hochwertige und Abonnement-basierte IT-Lösungen, die das Netzwerk und die Daten der Kunden schützen.

Das Unternehmen ist an der New York Stock Exchange (NYSE) unter dem Symbol CUDA gelistet. Der internationale Hauptsitz liegt in Campbell, Kalifornien (USA), der Hauptsitz für Kontinentaleuropa in Innsbruck (Österreich). Weitere Informationen stehen auf [barracuda.com](http://barracuda.com) zur Verfügung.

Barracuda Networks und das Barracuda Networks-Logo sind eingetragene Marken von Barracuda Networks, Inc. in den Vereinigten Staaten. Alle anderen Namen sind Eigentum der jeweiligen Rechteinhaber.



### Barracuda Networks AG

Eduard-Bodem-Gasse 1  
6020 Innsbruck  
Österreich

**t:** +43 508 100  
**e:** [office@barracuda.com](mailto:office@barracuda.com)  
**w:** [barracuda.com](http://barracuda.com)