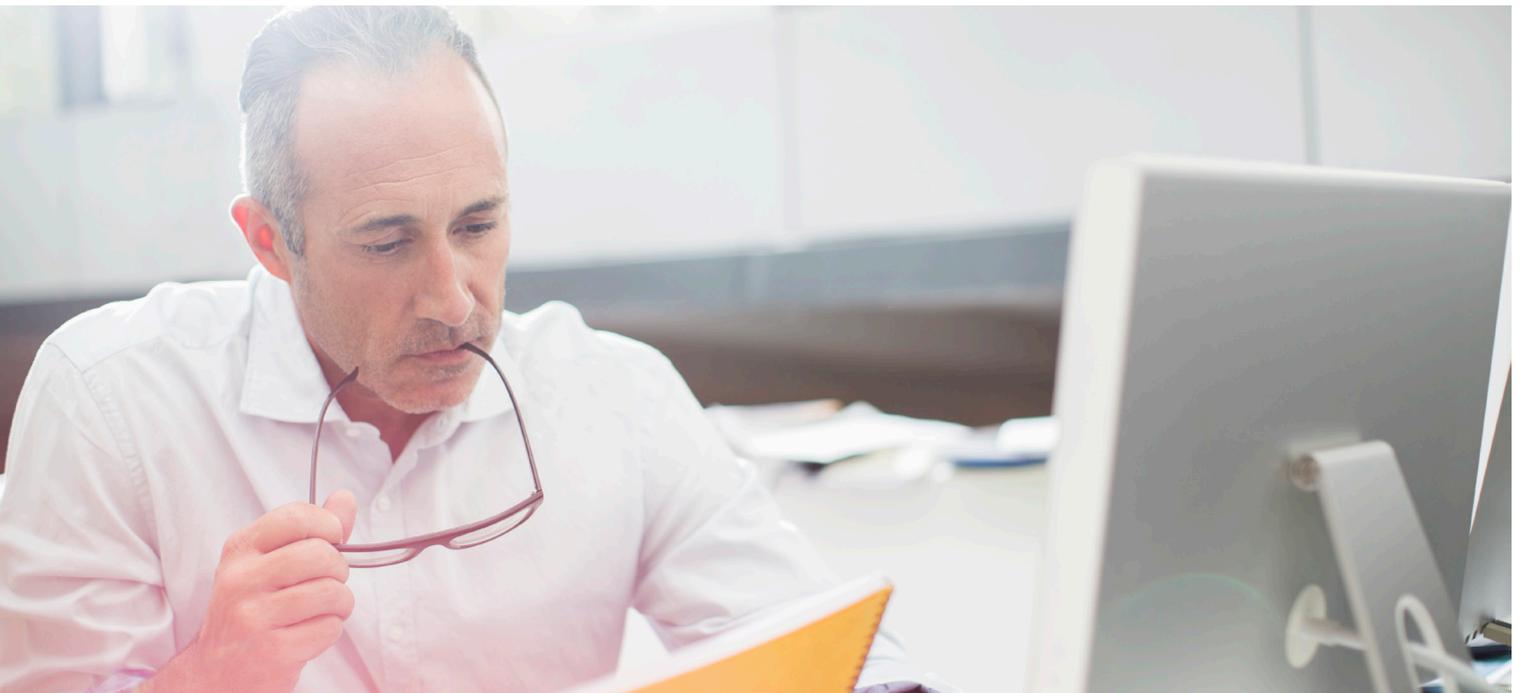


Deduplizierung: Die versteckte Wahrheit und was diese Sie kosten könnte

Nicht alle Deduplizierungstechnologien sind gleich. Erfahren Sie, wie Sie mit der Auswahl der richtigen Lösung Speicherplatz bis zu einem Faktor von 10 sparen können.

Von Adrian Moir, Senior Consultant, Produktmanagement, Quest Software



Datenwachstum ist etwas, mit dem wir alle zu kämpfen haben. Wir müssen immer mehr Daten für eine immer längere Zeit speichern, um geschäftliche oder gesetzliche Anforderungen zu erfüllen. Die weltweite Datensphäre wird bis 2025 wahrscheinlich auf 163 Zettabyte ansteigen. Zu diesem Zeitpunkt wird sich beinahe die Hälfte davon im Unternehmen befinden.¹

Der Schutz Ihrer Daten ist eine wesentliche Geschäftsfunktion. Dieser bringt jedoch auch seine ganz eigenen Herausforderungen mit sich. Rechnen Sie diese Herausforderungen zu den stetig wachsenden und sich verschiebenden Datensätzen hinzu, und Ihre Ressourcen werden sich sehr schnell verringern.

Natürlich ist der Schutz von Daten erforderlich, aber er sollte dennoch nicht zu einer Belastung werden. Wir alle suchen nach Lösungen, die Daten so schnell wie physikalisch möglich von A nach B bewegen. Wir erwarten Technologien, die unsere Erfahrung besser machen, schneller funktionieren und mehr speichern, während sie gleichzeitig immer die Kosten senken – eine wirkliche Herausforderung.

Wir haben die Entwicklung vieler Technologien verfolgt, die den Weg zur sekundären Datenspeicheroptimierung geebnet haben; mit der Absicht, die Ressourcenbelastungen durch Sicherung und Wiederherstellung zu senken. Bei der Deduplizierung

handelt es sich genau um eine solche Technologie. Zwar steht die Deduplizierung schon seit einiger Zeit auf dem Markt zur Verfügung, aber doch sind nicht alle Lösungen gleich. Daher ist es häufig hilfreich, genau zu verstehen, welche Auswirkungen jede Lösung wahrscheinlich auf Ihre verfügbaren Ressourcen und Budgets hat.

NICHT ALLE DINGE SIND GLEICH AUFGEBAUT

Und das gilt auch für Deduplizierungstechnologien. Deduplizierung ist ein Begriff für Datenreduktion geworden. Dennoch kann die Verwendung eines einzigen Worts zur Beschreibung unterschiedlicher Vorgehensweisen bestenfalls nur irreführend sein.

Lassen Sie uns einen Blick auf unterschiedliche Datenreduktionstechnologien werfen:

Komprimierung

Ein gutes Beispiel ist verlustfreie Datenkomprimierung. Diese basiert auf der Nutzung statistischer Redundanz ohne den Verlust von Daten, sodass Sie die Komprimierung „rückgängig“ machen und die Daten wiederherstellen können. Wir verwenden diese Methoden seit einigen Jahren, und sie gehören bereits bei den Technologien zum Standard. Ein GIF-Bild nutzt beispielsweise LZW-Komprimierung (Lempel-Ziv-Welch) zur Reduktion der

Fester Block

Originaler Datenstrom.

Nennt mich Ishmael. Vor einigen Jahren – wie lange es wirklich her ist, tut nichts zur Sache

Fester Block – Zweiter Datenstrom

Aufgrund der Datenänderung werden einzigartige Blöcke gespeichert.

Nennt mich Ish. Vor einigen Jahren – wie lange es wirklich her ist, tut nichts zur Sache

Fester Block – Dritter Datenstrom

Aufgrund der Datenänderung sind alle Blöcke einzigartig und müssen gespeichert werden.

Nennt mich Izzy. Vor einigen Jahren – wie lange es wirklich her ist, tut nichts zur Sache

Variabler Block

Vergleich mit dem originalen Datenstrom. Abgrenzungsänderungen senken die Anzahl an gespeicherten einzigartigen Blöcken.

Nennt mich Izzy. Vor einigen Jahren – wie lange es wirklich her ist, tut nichts zur Sache

Einzigartige Datenstücke, die gespeichert wurden.

Übereinstimmende Datenstücke, die referenziert, aber nicht gespeichert werden, und so die Speichernutzung senken.

Abbildung 1: Grenzen der Deduplizierung mit festen Blöcken.

Deduplizierung mit festen Blöcken ist ein guter Anfang, aber sie hat ihre Grenzen. Sie funktioniert nur gut bei einigen Datentypen, die direkt im Dateisystem gespeichert sind.

Dateigröße des Bildes, ohne dabei Informationen zu verlieren.

Single Instancing

Single Instancing ist eine sehr gute Beschreibung einer Speichermethodologie. Wenn ich dieselbe Datei zweimal speichere, behalte ich eine und referenziere sie für die andere. Dies funktioniert jedoch nur, wenn die Dateien identisch sind. Wird nur eine Kleinigkeit in der Datei geändert, dann wird die gesamte Datei erneut gespeichert. Dies eignet sich idealerweise für Systeme, deren Struktur mit denselben Inhalten eingepflegt werden, also möglicherweise für Produktionsspeicher, aber nicht so gut für die allgemeine Datensicherung.

Deduplizierung mit festen Blöcken

Dies ist unser erster Vorstoß in den Bereich, den man als richtige Deduplizierungsmethodologie beschreiben würde. Die Deduplizierung mit festen Blöcken nimmt einen Datenstrom und unterteilt diesen in Blöcke einer festgelegten Größe. Wir nennen diese „Datenstücke“.

Diese Datenstücke werden dann mithilfe verschiedener Methoden verglichen. Werden sie als identisch erachtet, wird nur ein einziges „Datenstück“ und eine Referenz für jede nachfolgende Übereinstimmung gespeichert. Kommt Ihnen das bekannt vor? Betrachten Sie dies als Single Instancing, aber auf

Unterdatei-Ebene: also das Betrachten der Blöcke, aus denen eine Datei besteht.

Diese Methodologie klingt zwar besser, aber sie hat auch ihre Grenzen. Sie funktioniert gut bei einigen Datentypen, die direkt im Dateisystem gespeichert werden, da sie byte-ausgerichtet sind und ihre Dateisysteme wie beispielsweise bei virtuellen Maschinen in 4K-, 8K- oder 32K-Blöcken geschrieben sind.

In diesem Fall können Lösungen mit festen Blöcken sehr effektiv sein. Dies funktioniert jedoch nicht gut bei einer Mischung von Daten, bei denen diese Grenzen nicht einheitlich sind oder durch unterschiedliche Softwarearten gesichert sind, wodurch sich die Ausrichtung ändert. Wir wissen, dass Daten alles andere als einheitlich sind und abhängig von ihrem Typ einen unterschiedlichen Aufbau, eine unterschiedliche Blockgröße und Byte-Ausrichtung sowie unterschiedliche Inhalte aufweisen.

Deduplizierung mit variablen Blöcken

Wir blicken nun erneut auf eine Technologie, welche die Probleme bei der Deduplizierung mit festen Blöcken lösen und in der Lage sein soll, verschiedene Datentypen zu handhaben, während wir gleichzeitig die Datenreduktion erhalten, nach der wir suchen. Und hier kommen die Mathematiker ins Spiel!

Wie bestimmen Sie, welche Daten in einer Datei mit den von Ihnen gespeicherten

Single-Instance	Fester Block	Variabler Block
<ul style="list-style-type: none"> • Erfordert genaue Übereinstimmungen • Gut für Anwendungen • Nicht gut für Sicherungsdaten • < 5 % Ersparnis 	<ul style="list-style-type: none"> • Übereinstimmung auf Unterdatei-Ebene • Bessere Deduplizierung als Single-Instance • Funktioniert bei Sicherungsdaten • < 30 % Ersparnis 	<ul style="list-style-type: none"> • Übereinstimmung auf Byte-Ebene • Beste Deduplizierung • Am besten für Sicherungsdaten geeignet • Bis zu 93 % Ersparnis

— Auswirkungen auf die Festplattenkapazität abhängig von Technologie.* —



* Festplatten nur zu Erläuterungszwecken dargestellt, nicht als direktes Verhältnis.

Abbildung 2: Deduplizierung mit variablen Blöcken führt zum besten Speicherplatz und den größten Kostenersparnissen.

Daten übereinstimmen, ohne dass Sie jede Kleinigkeit überprüfen müssen? Die Notwendigkeit, eine sich verändernde Datenblockgröße an ein neues „Datenstück“ anzupassen, das wiederum an einen anderen Datentyp angepasst werden kann, stellt eine große Herausforderung dar. Um diese zu bewältigen, wird eine Deduplizierung variabler Datenstücke mit gleitendem Fenster und Rabin-Fingerabdrücken verwendet.

Zum Finden von Duplikaten durchlaufen die Daten einen Rabin-Fingerabdruck-Algorithmus und es wird ein Datenstück erstellt, wenn eine einzigartige Reihe von Bytes gefunden wird. Da die Daten variabel sind und über ein gleitendes Fenster berechnet werden, kann dieselbe Reihe von Bytes der Daten (unterschiedliche Datenstücke) immer wieder identifiziert werden, unabhängig davon, wo sie sich im Datenstrom befindet. So müssen die Daten nicht mehr genau angeordnet sein, um Duplikate zu erfassen, wie es bei den Deduplizierungssystemen mit festen Blöcken der Fall ist.

Es spielt keine Rolle, ob die Daten des Datenstroms vorher oder nachher hinzugefügt wurden. Ist ein Datenstück identifiziert, wird ein SHA-1-Hash generiert und im Deduplizierungsverzeichnis gespeichert. Jegliches zukünftige Auftreten der Daten wird erkannt, da dasselbe Datenstück identifiziert wird. Die SHA-1-Hashs stimmen dabei überein und die Daten werden dedupliziert.

Die Hashs werden dann in einem Bloom²-Filter überprüft. Dies dient zu einer schnelleren Prüfung, um herauszufinden,

ob dieser Hash bereits bekannt ist und ob dieses „Datenstück“ gespeichert oder referenziert werden muss. Fortlaufend übereinstimmende Hashs ermöglichen die Erstellung von „Datenstücken“, wenn keine Übereinstimmungen gefunden werden, und die Erstellung von Referenzen, wenn Übereinstimmungen gefunden werden.

Insgesamt stellt das gleitende Fenster mehr Übereinstimmungen bereit und reduziert so die einzigartigen Daten, die gespeichert werden müssen. Damit wird gegenüber anderen Deduplizierungstechnologien eine erhebliche Menge an Speicherplatz eingespart.

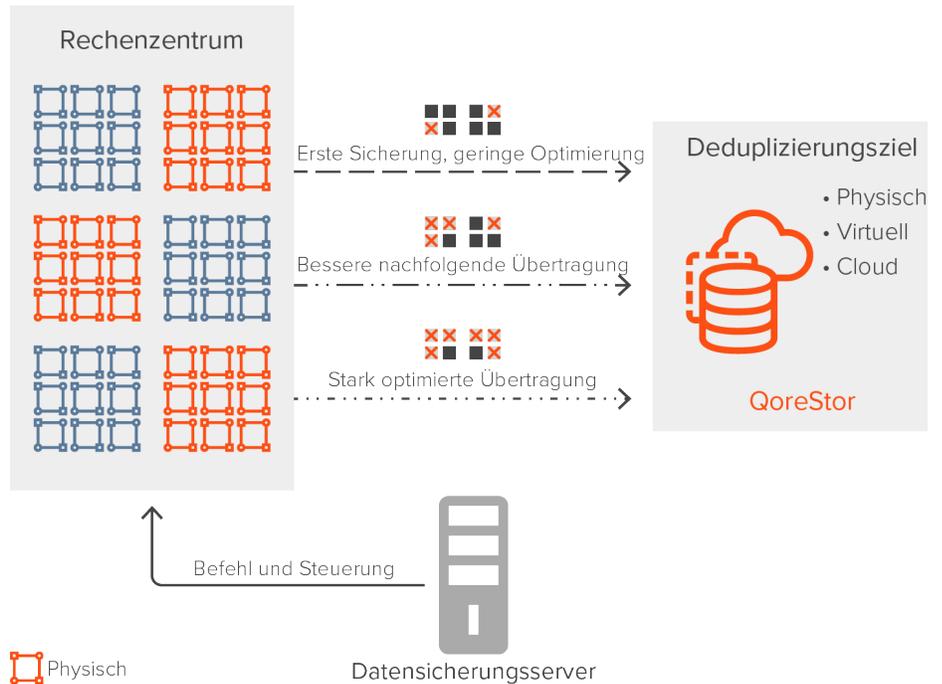
Inhaltsorientierte Deduplizierung mit variablen Blöcken

Es gibt einen nächsten Schritt, der noch weitere Einsparungen bieten kann: die Orientierung am Inhalt des Datenstroms selbst. Quest® QoreStor™ fügt beispielsweise einen inhaltsorientierten Algorithmus zu seiner Unterteilung in unterschiedlich große Datenstücke hinzu. Der Algorithmus identifiziert Muster innerhalb der Daten – ungeachtet der Verschiebung, die durch das Hinzufügen oder Löschen von Daten im Datenstrom entsteht – und richtet dann die Start- und Endpunkte der Blöcke so aus, dass sie duplizierte Datenstücke ergeben. Gleichzeitig identifiziert er nur die veränderten Datenstücke als einzigartig.

ALLE GUTEN DINGE BEGINNEN AN DER QUELLE

Wie wir bereits gesehen haben, gibt es Deduplizierungstechnologien in unterschiedlichen Formen. Eine

Insgesamt bietet die Deduplizierung mit variablen Blöcken mehr Übereinstimmungen und reduziert so die Menge an einzigartigen Daten, die gespeichert werden muss.



Mit quellseitiger Übereinstimmung gibt es weniger Daten, die über die Leitung gesendet werden. Dies ermöglicht es Ihnen, mehr Sicherungen zeitgleich auszuführen.

Abbildung 3: Technologien mit direkter Übertragung zum Ziel bieten bessere quellseitige Deduplizierung.

weiterer entscheidender Punkt ist es, zu berücksichtigen, wo die Deduplizierung stattfindet und welche Auswirkungen diese haben könnte. Wir haben nur die Technologien betrachtet, die üblicherweise bei Zielspeicherlösungen verwendet werden.

Lassen Sie uns nun die Probleme der zeitgleichen Sicherung mehrerer Systeme, Anwendungen und Daten betrachten.

Wir alle möchten so viele Daten wie möglich mit den geringsten Auswirkungen für unsere Benutzer übertragen. Daher definieren wir ein Zeitfenster, in dem dies stattfinden kann. Dies wird allgemein als „Backup-Fenster“ bezeichnet. Häufig liegt dies außerhalb der Geschäftszeiten, um die Auswirkungen durch die Übertragung großer Datenmengen zwischen Netzwerken zu minimieren – dieselben Netzwerke, die Benutzer zum Zugriff auf Anwendungen nutzen.

Heutzutage wird die Datenübertragung zu einem Problem, da immer mehr Unternehmen rund um die Uhr agieren und es begrenzen Raum für die Auswirkungen gibt, die durch diese Sicherungen verursacht werden. Dabei gibt es eigentlich mehr als ein Problem. Das erste Problem liegt beim Anwendungsserver in der Auswirkung auf die Sättigung seiner Netzwerkverbindung. Das zweite

Problem ist das Zielgerät, dessen Netzwerkverbindung mit allen Daten von unterschiedlichen Maschinen gesättigt wird.

Da Zeit endlich ist, kommen Sie zu einem Punkt, an dem Sie bei dem, was Sie in Ihrem festgelegten Backup-Fenster erreichen können, eingeschränkt sind.

Es können Verbesserungen vorgenommen werden, indem einige Ressourcen bei jedem Anwendungsquellserver verwendet werden, der gesichert wird. Mit Blick auf das erste Problem müssen durch übereinstimmende Daten-Hashs im Backup-Datenstrom beim Quellserver nur einzigartige Daten über das Netzwerk übertragen werden. Zwar bedeutet dies einen kleinen Mehraufwand für die Anwendungsquellserver, aber dies kommt nicht annähernd an die Auswirkungen einer überlasteten Netzwerkverbindung heran.

Die Verwendung der quellseitigen Übereinstimmungsmethode hat auch Einfluss auf das zweite Problem. Da weniger Daten über die Leitung gesendet werden, können mehr Sicherungen zeitgleich ausgeführt werden. Nachdem dies erfolgt ist, kann entweder das Backup-Fenster reduziert oder mehr Sicherungen im selben Zeitrahmen durchgeführt werden. Dadurch wird die Datendurchsatzleistung deutlich verbessert. Diese Technologie wird quellseitige Deduplizierung genannt.

DIE KOSTEN DES EINGEHENS VON KOMPROMISSEN BEI TECHNOLOGIE

Die Auswirkungen der Wahl der falschen Deduplizierungstechnologie können am besten anhand der Größe und den damit verbundenen Kosten erläutert werden. Lassen Sie uns von der Größe eines Datensatzes von 100 TB ausgehen. Zur Vereinfachung nehmen wir an, dass Sie wöchentliche Sicherungen haben, die Sie auf 12 Wochen begrenzen möchten. Jede Woche würden 100 TB zum Zielspeicher hinzukommen. Ohne Datenreduktionstechnologie würden nach drei Monaten 1,2 PB Speicher verbraucht, bevor die anfänglichen Sicherungen beginnen, zu verfallen.

Betrachten wir eine Deduplizierung mit festen Blöcken, die Ihnen eine Reduktion um 30 % ermöglichen würde. Somit würde der Speicherbedarf auf 840 TB gesenkt. Das ist zwar eine gute Ersparnis, aber immer noch beträchtlich und wahrscheinlich eine kostspielige Lösung.

Lassen Sie uns nun eine variable, inhaltsorientierte Deduplizierungstechnologie in Betracht ziehen. Mit einer Reduktion um 93 % würde dies Ihren Speicherbedarf auf nur 84 TB senken. Wie Sie sehen, kann mit der Auswahl der richtigen Deduplizierungstechnologie Speicherplatz bis zu einem Faktor von 10 eingespart werden.

Bedenken Sie nun den Unterschied der Speicherkosten zwischen 840 TB und 84 TB. Unter Berücksichtigung der Kosten der Kosten für Leistung und eventuell Rack-Fläche bei Nutzung eines externen Rechenzentrums werden die möglichen Einsparung sehr schnell deutlich. Die Wahl der falschen Technologie hat viele Auswirkungen – nicht nur auf den Speicherplatz. Kosten werden immer gedrückt und müssen sorgfältig erwogen werden. Wählen Sie die richtige Technologie und senken Sie so Ihre Kosten sowie Ihren Datenfußabdruck.

EINES DIESER DINGE IST NICHT WIE DIE ANDEREN

Recherchieren Sie bei der Erwägung einer optimierten Speicherlösung für Ihre Sicherheitsbedürfnisse sorgfältig, damit Sie ein umfassendes Verständnis davon haben, welche Technologien auf ihre Umgebung angewendet werden können. Im Laufe der Jahre wurde sich sehr stark mit Deduplizierungsanwendungen, fest installierter Hardware und Software befasst, die dazu entwickelt wurden, eine Lösung bereitzustellen.

Dies war zwar ein guter Ansatz und er hat gute Dienste geleistet, aber nun verändert sich die Natur der Infrastruktur. Wir befinden uns jetzt in einer Welt, in der von uns mehr Agilität hinsichtlich Bereitstellung, Infrastruktur und Standort gefordert wird.

Auch die Auswirkungen in Bezug auf die Kosten rücken immer mehr in den Vordergrund. Den Speicher alle drei bis fünf Jahre aktualisieren und für die Softwareelemente einer Anwendung erneut bezahlen zu müssen, ist nicht ideal. Dies ist einer der Bereiche, in denen Quest QoreStor anders ist.

QoreStor ist eine wirkliche softwaredefinierte Sekundärspeichertechnologie. Es handelt sich dabei um Software, die direkt auf der physischen Serverhardware, einer virtuellen Maschine oder in einer Public oder Private Cloud-Umgebung installiert werden kann – öffentlich oder privat. Da QoreStor nicht an eine bestimmte Hardwareplattform gebunden ist (wie Deduplizierungsanwendungen), können Sie die Sicherungssoftware auswählen, die Sie mögen – von Ihrem bevorzugten Anbieter, zu Ihren üblichen Nachlässen. Wenn es Zeit für die Aktualisierung der Hardware wird, tun Sie einfach genau das: Zahlen Sie nur für die neue Hardware und behalten Sie die bestehende QoreStor Softwarelizenz.

Softwaredefiniert zu sein, bedeutet nicht, dass QoreStor selbst eingeschränkt ist. Es nutzt die Deduplizierungstechnologie mit variablen Blöcken und bietet sowohl Inline- als auch quellseitige Deduplizierungsfähigkeiten. Darüber hinaus komprimiert es die deduplizierten „Datenstücke“ und verschlüsselt diese zur Sicherheit.

Ferner bietet QoreStor Datentrennung mithilfe seiner „Datengruppen“-Technologie. Dadurch ermöglicht es unterschiedliche Komprimierungs- und Verschlüsselungseinstellung für verschiedene Datensätze oder Benutzer. All dies ist in der QoreStor Lizenz enthalten, zusammen mit der Fähigkeit der Replikation von einem QoreStor Vorgang zum anderen unter Einsatz sicherer und optimierter Datenübertragung.

Wenn Sie Ihren Sekundärspeicher für Sicherung und Wiederherstellung optimieren, mehrere optimierte Kopien von Daten an unterschiedlichen Orten, einschließlich der Cloud, auf kostengünstige und sichere Weise aufbewahren möchten, dann treffen Sie mit Quest QoreStor genau die richtige Wahl.

Quest QoreStor kann direkt auf der Hardware, in einer VM oder einer Cloudumgebung installiert werden. Es arbeitet mit Ihrer bereits vorhandenen Sicherungssoftware.

¹ Reinsel, David; Gantz, John; Rydning, John; „Data Age 2025: The Evolution of Data to Life-Critical“, IDC, April 2017.

² Ein Bloom-Filter ist eine platzsparende probabilistische Datenstruktur, die von Burton Howard Bloom im Jahre 1970 entwickelt wurde. Sie wird dazu verwendet, zu testen, ob ein Element in einem Datensatz schon einmal aufgetreten ist.

ÜBER QUEST

Bei Quest versuchen wir, komplexe Herausforderungen mit einfachen Lösungen zu bewältigen. Dies gelingt uns dank unserer speziellen Unternehmensphilosophie, bei der hervorragender Service und unser allgemeines Ziel – ein unkomplizierter Geschäftspartner zu sein – im Vordergrund stehen. Unsere Vision besteht darin, Technologien bereitzustellen, bei denen Sie sich nicht zwischen Effizienz und Effektivität entscheiden müssen. Dadurch müssen Sie und Ihre Organisation sich weniger um die IT-Verwaltung kümmern und haben mehr Zeit für Unternehmensinnovation.

© 2018 Quest Software Inc. Alle Rechte vorbehalten.

Dieses Handbuch enthält urheberrechtlich geschützte Informationen. Die in diesem Handbuch beschriebene Software wird im Rahmen einer Softwarelizenz- oder Vertraulichkeitsvereinbarung bereitgestellt. Diese Software darf nur gemäß den Bestimmungen der entsprechenden Vereinbarung genutzt oder kopiert werden. Dieses Handbuch darf ohne schriftliche Genehmigung von Quest Software Inc. – außer zur persönlichen Nutzung durch den Käufer – weder ganz noch in Teilen in irgendeiner Form oder Weise (elektronisch, mechanisch, zum Beispiel durch Fotokopiertechnik oder Aufzeichnung) reproduziert oder an Dritte weitergegeben werden.

Die Informationen in diesem Dokument beziehen sich auf Quest Software Produkte. Dieses Dokument sowie der Verkauf von Quest Software Produkten gewähren weder durch Rechtsverwirkung noch auf andere Weise ausdrückliche oder implizite Lizenzen auf geistige Eigentumsrechte. Es gelten ausschließlich die in der Lizenzvereinbarung für dieses Produkt festgelegten Geschäftsbedingungen. Quest Software übernimmt keinerlei Haftung und lehnt jegliche ausdrückliche oder implizierte oder gesetzliche Gewährleistung in Bezug auf die Produkte von Quest Software ab, einschließlich, jedoch nicht beschränkt auf, stillschweigende Gewährleistung der handelsüblichen Qualität, Eignung für einen bestimmten Zweck und Nichtverletzung der Rechte Dritter. In keinem Fall haftet Quest Software für direkte oder indirekte Schäden, Folgeschäden, Schäden aus Bußgeldern, konkrete Schäden oder beiläufig entstandene Schäden, die durch die Nutzung oder die Unfähigkeit zur Nutzung dieses Dokuments entstehen können (einschließlich, jedoch nicht beschränkt auf, entgangene Gewinne, Geschäftsunterbrechungen oder Datenverlust), selbst wenn Quest Software auf die Möglichkeit derartiger Schäden hingewiesen wurde. Quest Software gibt keinerlei Zusicherungen oder Gewährleistungen hinsichtlich der Richtigkeit oder Vollständigkeit der Informationen in diesem Dokument und behält sich das Recht vor, die Spezifikationen und Produktbeschreibungen jederzeit ohne Benachrichtigung zu ändern. Quest Software verpflichtet sich nicht dazu, die Informationen in diesem Dokument zu aktualisieren.

Patente

Wir von Quest Software sind stolz auf unsere fortschrittliche Technologie. Dieses Produkt ist möglicherweise durch Patente oder Patentanmeldungen geschützt. Aktuelle Informationen zu den für dieses Produkt geltenden Patenten finden Sie auf unserer Website unter www.quest.com/legal.

Marken

Quest, QoreStore[®] und das Quest Logo sind Marken und eingetragene Marken von Quest Software Inc. Eine vollständige Liste aller Quest Marken finden Sie unter www.quest.com/legal/trademark-information.aspx. Alle anderen Marken sind Eigentum der jeweiligen Markeninhaber.

Sollten Sie Fragen hinsichtlich der potenziellen Nutzung des Materials haben, wenden Sie sich bitte an:

Quest Software Inc.

Attn: LEGAL Dept
4 Polaris Way
Aliso Viejo, CA 92656

Auf unserer Website (Quest.com/de) finden Sie Informationen zu unseren regionalen oder internationalen Büros.