

Link: <https://www.computerwoche.de/a/backup-kosten-um-den-faktor-zehn-senken,2351494>

Daten-Deduplizierung

Backup-Kosten um den Faktor Zehn senken

Datum: 17.08.2010

Autor(en): Thomas Pelkmann

Unternehmen und Privatpersonen speichern immer mehr Daten auf Festplatten sowie optischen und Bandlaufwerken. Wer dabei richtig Geld sparen möchte, sollte sich Lösungen zur Daten-Deduplizierung anschauen.

Die Menge der auf Speichermedien abgelegten Daten wird sich einer Untersuchung von **IDC**¹ aus dem Jahr 2009 zwischen 2008 und 2012 zufolge verfünffachen: Waren es vor zwei Jahren noch 487 Exabytes, wird das Datenvolumen bis in zwei Jahren auf geschätzte 2.500 Exabyte ansteigen. Ein **Exabyte**² entspricht der schwindelerregenden Menge von einer Milliarde Gigabyte.

Ein guter Teil dieser explodierenden Datenmassen entfällt auf moderne Medien der Unterhaltungsindustrie: Filme, Musik, Digitalfotos, **Voice-over-IP**³-Sprachmitteilungen... Aber auch das Datenvolumen in den Unternehmen platzt aus allen Nähten, weil hier immer mehr Echtzeitdaten erhoben und verarbeitet werden wollen. So ist in den Unternehmen ein Großteil der gesamten schriftlichen Kommunikation bereits digitalisiert: E-Mails, Geschäftsbriefe, Angebote, Aufträge und Rechnungen. Zunehmend werden in der Zukunft auch Daten von Videokonferenzen oder VoIP-Telefonie dazu kommen.

Nur wenige Daten werden gelöscht

Governance⁴-Regeln und Geschäftsstrategien sorgen dafür, dass nur wenige dieser Daten nach absehbarer Zeit gelöscht werden. In Gegenteil: Die vorhandenen Speichermöglichkeiten bieten die Option, alle Geschäftsdaten quasi unbegrenzt zu speichern.

Und obwohl Speicherplatz immer verfügbarer und zudem, gerechnet etwa auf den Gigabyte-Preis, immer billiger wird, ist die Menge der Daten ein Thema für die Betreiber von firmeneigenen Rechenzentren. Auch preiswerter Speicherplatz kostet immerhin Geld: für die Anschaffung von Daten- oder Backup-Servern, für Speichermedien, für die Wartung und Instandhaltung. Auch preiswerte Speicher verbrauchen zudem Energie und verlangen nach Platz und Kühlung.

So wird aus einem exotischen Randthema für die IT ein Business-Case. Zumal es Lösungen gibt, die vorhandene Datenmenge drastisch zu reduzieren: Durch die sogenannte **Daten-Deduplizierung**⁵, so etwa IDC-Analyst Rüdiger Spies, lässt sich das anfallende Datenvolumen um mindestens den Faktor Zehn reduzieren.

Hersteller und Anbieter von Speziallösungen haben auch höhere Zahlen im Katalog, aber schon durch die zehnfache Verkleinerung der Datenmengen lässt sich bei flüchtigem Nachrechnen eine Menge Geld sparen - **ein gutes Argument für CIOs**⁶, die mit ihrem Management immer wieder über die Reduktion der IT-Kosten diskutieren.

Geld und Zeit sparen



Daten-Deduplizierung soll beim Backup nicht nur Geld, sondern auch Zeit sparen.

Die Rechnung geht auch auf, wenn man die geringere Zeit bedenkt, die für **Backup und Restore**⁷ deduplizierter Daten anfällt. **EMC**⁸, ein Anbieter von Deduplizierungslösungen, spricht von zehnmal schnelleren Backups sowie von einem raschen Single-Step-Recovery.

Bei der **Daten-Deduplizierung**⁹ geht es im Wesentlichen um das Backup von Daten: Das Verfahren sorgt dafür, dass ausschließlich einzigartige Daten gespeichert werden. Dabei orientiert sich Daten-Deduplizierung allerdings nicht an kompletten Dateien, etwa an Word- oder Excel-Dokumenten, sondern an kleinen, etwa vier KByte großen Einheiten, "Blocks" oder "Chunks" genannt.

Trifft die Deduplizierungs-Software auf einen bereits gespeicherten Block, referenziert sie beim neuen Backup-Vorgang lediglich auf diesen Block, ohne ihn erneut zu speichern. Dieser Verweis (Pointer) verbraucht nach Angaben von EMC lediglich 20 Bytes, ist also deutlich kleiner als der Originalblock.

Daten-Deduplizierung - zwei Verfahren

Bei der Deduplizierung gibt es zwei wesentliche Verfahren: "at source", also an der Quelle, und "at target", also am Ziel. Bei der ersten Variante werden Duplikate bereits an der Quelle erkannt und nur solche Blöcke ans Backup geschickt, die sich von den bereits gespeicherten unterscheiden. Das Source-Verfahren schont die Bandbreite im Netzwerk, während die Quell-Version besser zur Storage-Hardware passt und zudem eine größere Unabhängigkeit von Übertragungsprotokollen bietet.

Interessant ist die **Daten-Deduplizierung**¹⁰ nicht nur für das Backup von Unternehmensdaten. Auch bei der Virtualisierung spielt das Verfahren eine wichtige Rolle. Eine IT-Infrastruktur verwaltet zwischen mehreren hundert bis Tausende von Desktop-PCs, die bis auf die individuellen Benutzereinstellungen über eine identische Ausstattung mit Betriebssystem und Anwendungen verfügen. Wenn diese Systeme über virtuelle Images verwaltet werden, lässt sich durch Daten-Deduplizierung spielend ein Großteil dieser für die Images anfallenden Datenmengen reduzieren.

Insofern ist die Deduplizierung auch ein interessantes Thema für Outsourcing-Dienstleister und **Cloud**¹¹-Anbieter: Je mehr externe Daten im Rechenzentrum verwaltet und gesichert werden müssen, desto wichtiger ist es auch hier, Speicherplatz und die damit verbundenen Kosten reduzieren zu können.

Links im Artikel:

¹ <http://www.emc.com/collateral/demos/microsites/idc-digital-universe/iview.htm>

² <http://de.wikipedia.org/wiki/Byte>

³ <https://www.computerwoche.de/netzwerke/tk-netze/1927928/>

⁴ <https://www.computerwoche.de/schwerpunkt/i/IT-Governance.html>

⁵ <http://whitepaper.computerwoche.de/index.cfm?cid=38&pkdownloads=3865&source=subnetbox>

⁶ <http://whitepaper.computerwoche.de/index.cfm?cid=38&pkdownloads=3867&source=subnetbox>

⁷ <https://www.computerwoche.de/hardware/storage/1929449/>

⁸ <http://germany.emc.com/>

⁹ [http://whitepaper.computerwoche.de/index.cfm?](http://whitepaper.computerwoche.de/index.cfm?cid=38&pkdownloads=3862&source=subnetbox)

[cid=38&pkdownloads=3862&source=subnetbox](http://whitepaper.computerwoche.de/index.cfm?cid=38&pkdownloads=3862&source=subnetbox)

¹⁰ <https://www.computerwoche.de/hardware/storage/1888507/>

¹¹ <https://www.computerwoche.de/hardware/storage/1894517/>

IDG Tech Media GmbH

Alle Rechte vorbehalten. Jegliche Vervielfältigung oder Weiterverbreitung in jedem Medium in Teilen oder als Ganzes bedarf der schriftlichen Zustimmung der IDG Tech Media GmbH. dpa-Texte und Bilder sind urheberrechtlich geschützt und dürfen weder reproduziert noch wiederverwendet oder für gewerbliche Zwecke verwendet werden. Für den Fall, dass auf dieser Webseite unzutreffende Informationen veröffentlicht oder in Programmen oder Datenbanken Fehler enthalten sein sollten, kommt eine Haftung nur bei grober Fahrlässigkeit des Verlages oder seiner Mitarbeiter in Betracht. Die Redaktion übernimmt keine Haftung für unverlangt eingesandte Manuskripte, Fotos und Illustrationen. Für Inhalte externer Seiten, auf die von dieser Webseite aus gelinkt wird, übernimmt die IDG Tech Media GmbH keine Verantwortung.