

Link: <https://www.computerwoche.de/a/deduplizierung-ist-rechtskonform-und-spart-geld,2358424>

Chunks statt Redundanz

## Deduplizierung ist rechtskonform und spart Geld

Datum: 23.11.2010

Autor(en): Thomas Pelkmann

**Wer mit explodierendem Datenaufkommen seine Storage-Infrastruktur nicht hoffnungslos überfordern möchte, sollte sich über eine Reduktion der Daten Gedanken machen. Daten-Deduplizierung spart nicht nur Bandbreite und Speicherplatz; in der Summe lässt sich damit bares Geld sparen.**

Ein Großteil der in Unternehmensnetzen gespeicherten Daten ist Schrott, hat jüngst der Datenintegrationsspezialist **Informatica in einer Umfrage herausgefunden**<sup>1</sup>: Acht von zehn Managern haben dort eingeräumt, dass in ihren Firmennetzen ungenutzte Daten und Applikationen schlummern. Zudem seien viele Informationen redundant oder inkonsistent vorhanden. Der Datenschrott kostet die Unternehmen richtig Geld; in Einzelfällen berichteten die IT-Verantwortlichen in der Informatica-Umfrage von jährlichen Kosten in Millionenhöhe.

### IDC-Studie

#### ROI-Daten im Überblick

Die Analysten von IDC haben neun EMC-Kunden nach dem ROI ihrer Deduplizierungsprojekte gefragt. Dabei gaben die Kunden die folgenden Daten zu Protokoll:

- Kosteneinsparungen pro Jahr von über 1.000.000 US-Dollar.
- ROI über drei Jahre von 264 Prozent und eine Amortisation innerhalb von 6,6 Monaten.
- Kosteneinsparungen im ersten Jahr nach der Bereitstellung von 200.000 US-Dollar für nicht geleistete Anschaffungskosten für Bandlaufwerke-/bibliotheken.
- Jährliche Einsparung von rund 156.000 Dollar im Bereich der Wartung von vorhandener Hardware für Bandlaufwerke-/bibliotheken.
- Vermiedene Kosten in Höhe von 80.000 US-Dollar für neue Server.
- Verkürzte Wiederherstellungszeit von 4,6 Stunden auf etwa 35 Minuten.
- Einsparung von 1,6 Vollzeitäquivalenten für IT-Mitarbeiter über drei Jahre.

Aber gibt es überhaupt überflüssige Daten in den Unternehmen? Die befragten CIOs erzählen von Daten, die aus **Rechtsgründen**<sup>2</sup> aufbewahrt werden müssen; in Deutschland gilt zum Beispiel für alle Dokumente, die zu Streitigkeiten führen könnten, eine Aufbewahrungsfrist von sechs Jahren; für Buchungssätze und Jahresabschlüsse gelten sogar zehn Jahre. Einzelne Patientenakten in Krankenhäusern und Arztpraxen müssen gar bis zu 30 Jahren vorgehalten werden. Kein Wunder, dass es bei den Lebenszyklen moderner Datenbankanwendungen schnell zu vermeintlich unbrauchbaren und unlesbaren Datenmüllbergen kommt. Eine Alternative zur Aufbewahrung existiert dennoch nicht.

Durch das Reduzieren redundanter Daten gibt es aber zumindest eine Möglichkeit, solche Alt-, aber auch aktuelle Daten platzsparend abzulegen. **Dateneduplizierung**<sup>3</sup> heißt das Verfahren der Wahl: Hier werden bei jedem Speichervorgang nicht einfach alle Daten auf die Laufwerke gespielt, sondern nur solche, die sich seit dem letzten Speichervorgang geändert haben.

Um das möglichst effizient zu erledigen, zerlegen Deduplizierungsalgorithmen die zu speichernden Daten in kleinste Teile (**Chunks**<sup>4</sup>) und vergleichen bei späteren Speichervorgängen die neuen Schnipsel mit den bereits abgelegten. Treffen Deduplizierungsprogramme auf schon bekannte Byte-Folgen, wird lediglich ein Verweis darauf abgelegt, nicht aber die Folge selbst.

## **Deduplizierung ist rechtssicher**

Die Methode klingt bestechend, zumal diese platzsparende Art der Datenablage für die Benutzer eher unauffällig stattfindet: Sie können wie gewohnt auf die Daten zugreifen. Dennoch gibt es Diskussionen darum, ob sie auch zulässig ist. So beschäftigt sich etwa Rechtsanwalt Wilfried Reiners **in einem Fachbeitrag**<sup>5</sup> mit der Frage, ob Deduplizierung mit steuerrechtlichen Aufbewahrungspflichten zusammenpasse. Er antwortet mit einem klaren "Ja". So schreibe der Gesetzgeber zwar vor, dass beispielsweise von versendeten Handelsbriefen eine "mit der Urschrift übereinstimmende Wiedergabe" aufzubewahren sei. Dies könne, so der Anwalt, eine Kopie, ein Abdruck, eine Abschrift oder sonstige Wiedergabe des Wortlauts auf einem Schrift-, Bild- oder anderen Datenträger sein. Das schließe elektronische Kopien ein: "Damit spricht nichts gegen Deduplizierung."

Im Gegenteil, meint der Anwalt, spricht sich das **Bundesdatenschutzgesetz**<sup>6</sup> sogar dafür aus, Daten platzsparend aufzubewahren: "Die Erhebung, Verarbeitung und Nutzung personenbezogener Daten und die Auswahl und Gestaltung von Datenverarbeitungssystemen sind an dem Ziel auszurichten, so wenig personenbezogene Daten wie möglich zu erheben, zu verarbeiten oder zu nutzen."

Wenn nun schon nichts gegen **Dateneduplizierung**<sup>7</sup> spricht, um Speicherplatz zu sparen - was spricht dann dafür? Fakt ist, dass Speicher immer billiger wird: Nach Einschätzung der Marktforscher von Experton sinken die Preise für **Storage**<sup>8</sup>-Hardware im Durchschnitt um satte 30 Prozent pro Jahr.

Auf der anderen Seite wächst der Speicherbedarf rasant an. Experten sprechen von einem jährlichen Mehrbedarf zwischen 30 und 60 Prozent. Die zunehmende Digitalisierung der Geschäftsprozesse und die steil wachsenden Datenmengen im Internet durch Twitter oder Facebook sorgen für eine den Preisverfall neutralisierende Nachfrage nach zusätzlichem Speicher.

## **Hohe Reduktionsraten sprechen für Deduplizierung**

Genau hier liegt der Business Case von Dateneduplizierung, wenngleich verlässliche Zahlen über das Sparpotenzial nur schwer zu ermitteln sind. So spricht **EMC**<sup>9</sup> davon, dass sich durch die nicht-redundante Ablage von Daten bis zu 98 Prozent Speicherplatz einsparen ließe. Backup- und Restore-Vorgänge würden sich zudem um bis zu 90 Prozent beschleunigen lassen. **Data Domain**<sup>10</sup>, seit 2009 Mitglied der EMC-Familie, rechnet mit einem um Faktoren zwischen 10 und 30 reduzierten Speicherbedarf. Im günstigsten Falle bedeutet das, dass für die Datenablage statt einer 30-GByte-Platte auch ein Stick mit einem GByte reichen würde.

In der Praxis schwanken die Reduktionsraten von Anbieter zu Anbieter und hängen auch von der konkreten Infrastruktur vor Ort ab. Insofern lassen sich darüber keine verbindlichen, dafür aber exemplarische Aussagen treffen. So berichtet das Versicherungsunternehmen Signal Iduna davon, dass sich durch Deduplizierung der gesicherte Datenbestand von 283 Terabyte "problemlos" auf ganze 16 Terabyte zusammenfassen ließ. Das entspricht einer Reduktion um den Faktor 17.

Bei der Berechnung der Wirtschaftlichkeit spielt auch die Frage eine Rolle, ob die Deduplizierungs-Algorithmen "at the source" (an der Quelle) oder "at the target" (am Ziel) eingesetzt werden. Die erste Methode fordert an Desktop oder Server höhere Prozessorleistung, schont aber die Bandbreiten im Netzwerk. Da nur neue oder geänderte Daten an die Speichermedien geschickt werden, lassen sich tägliche Backups sehr viel schneller erledigen. Bei der zweiten Methode wird erst unmittelbar vor der Speicherung der Daten dedupliziert, was mit einem erheblich höheren Bedarf an Bandbreite für die Übertragung der vollständigen und redundanten Daten einher geht, dafür aber die Prozessoren schont. Die Ziel-Version passt in der Regel besser zur Storage-Hardware und bietet zudem eine größere Unabhängigkeit von Übertragungsprotokollen.

Besonders Unternehmen, die ihre gesamte Desktop-IT virtualisieren möchten, können mit **Deduplizierungsverfahren**<sup>11</sup> massiv sparen. So muss man nicht für jeden einzelnen PC einer mitunter mehrere tausend Geräte umfassenden Infrastruktur virtuelle Maschinen vorhalten. Stattdessen lässt sich in einer deduplizierten Umgebung ein Großteil der für die einzelnen Images nötigen Datenmenge einsparen.

Schließlich gibt es mittelbare Spareffekte durch Deduplizierung, die Unternehmen bei ihren Wirtschaftlichkeitsrechnungen berücksichtigen sollten. Dazu heißt es in einem **Whitepaper von IDC**<sup>12</sup>: "Die Deduplizierung ermöglicht Ressourceneffizienz und Kosteneinsparungen durch die Verringerung des Energiebedarfs, der Kühlungs- und Stellflächenanforderungen im Rechenzentrum sowie der Speicherkapazität, Netzwerkbandbreite und der Personaleffizienz."

### Links im Artikel:

<sup>1</sup> <https://www.cio.de/knowledgecenter/bi/2249968/>

<sup>2</sup> <https://www.computerwoche.de/schwerpunkt/c/Compliance.html>

<sup>3</sup> <https://www.computerwoche.de/hardware/storage/2351494/>

<sup>4</sup> <http://de.wikipedia.org/wiki/Chunking>

<sup>5</sup> <http://www.it-business.de/news/recht/vorschriften-regulierungen/compliance/articles/262597/>

<sup>6</sup> <http://de.wikipedia.org/wiki/Bundesdatenschutzgesetz>

<sup>7</sup> <http://whitepaper.computerwoche.de/index.cfm?cid=38&pkdownloads=4208&source=subnetbox>

<sup>8</sup> <https://www.computerwoche.de/subnet/emc/1938383/>

<sup>9</sup> <http://germany.emc.com/index.htm>

<sup>10</sup> <http://whitepaper.computerwoche.de/index.cfm?cid=38&pkdownloads=3866&source=subnetbox>

<sup>11</sup> <http://whitepaper.computerwoche.de/index.cfm?cid=38&pkdownloads=4124&source=subnetEMC>

<sup>12</sup> <http://germany.emc.com/collateral/analyst-reports/idc-20090519-data-deduplication.pdf>

---

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.