

ÜBERLADENDE DATEIEN EFFEKTIV BEHANDELN

VON

MARC STAIMER

(PRESIDENT & CDS von DRAGON SLAYER CONSULTING)



Dragon Slayer Consulting



Dragon Slayer Consulting

Ein nachhaltiges Verfahren, um die durch überladene Dateien entstandenen, folgenschweren Ärgernisse zu beseitigen

Marc Staimer, President & CDS von Dragon Slayer Consulting

marcstaimer@mac.com +1 (503) 579-3763

Einleitung

Es gibt eine unauffällige Krankheit, die IT-Arbeitsplätze, Laptops und Dateiserver aller Unternehmen befallen hat.

Die Ursache dafür sind "aufgeblähte" Dateien. Viele IT-Experten haben noch nie von dieser digitalen Pandemie gehört und falls doch, spielen sie den Ernst der Lage gerne herunter.

Es wird oft angenommen, dass die Hardware-Technologie die Probleme überladener Dateien bezüglich Verarbeitung, Speicher, Input und Output sowie Bandbreite beseitigen oder entschärfen kann. Das ist falsch.

Erliegen Sie nicht dieser Täuschung, denn: Überladene Dateien sind ein dringliches IT-Problem mit schwerwiegenden Konsequenzen.

Was sind überladene Dateien?

Einfach ausgedrückt sind überladene Dateien, Dateien, die viel größer und überladener sind, als sie es eigentlich sein müssten. In viel zu vielen Fällen beträgt die Überlast dieser Dateien mehr als das 20-fache der Dateigröße.

Überladene Dateien weisen die Symptome einer biologischen Krankheit auf.

Die Probleme gründen sich auf leistungsschwache Software, veraltete Algorithmen oder unvollständige Annahmen von Seiten der Software-Entwickler darüber, wie ihre Software tatsächlich benutzt wird.

Dadurch kommt es oft zu einer unpassenden Wahl für das Bildformat, einer viel zu hohen Auflösung oder großen, für den normalen Ablauf völlig unnötigen Datenblöcken. Einige der schlimmsten Verursacher sind hier JPEG-, PowerPoint-, Excel-, und Word-Dateien. Die Tragweite wird besonders deutlich, wenn Dateien miteinander kombiniert oder eingebettet werden. Wird z.B. ein Bild im Format JPEG mit 300dpi in eine PowerPoint Datei eingefügt, wird im Grunde immer mit einer viel zu hohen Auflösung gearbeitet. Das Einfügen eines Teils einer Tabelle in eine Präsentation führt dazu, dass anstelle nur des sichtbaren und notwendigen Teils das gesamte Blatt eingefügt wird. Das Ergebnis für beide Beispiele: Erheblich überladene Dateien.

IT-Experten, die nichts von den Problemen und Hintergründen zu diesen überladenen Dateien wissen, müssen schon hinter dem Mond leben, wenn sie mit den hieraus resultierenden Konsequenzen nicht schon einmal konfrontiert worden sind.



Die düsteren Konsequenzen aus überladenen Dateien

Unbehandelt erzeugen überladene Dateien eine ganze Reihe ernster und sich gegenseitig hochschaukelnder Probleme. Größere Dateien führen zu einer verlangsamt Schreib- und verlängerten Öffnungszeiten. Die Anwendungen laufen langsamer. Die Produktivität der Benutzer sinkt. Speicherplatz wird immer schneller verbraucht, neuer Primärspeicher wird erforderlich. Weitere Speichersysteme werden unumgänglich. Mehr Speichernetzwerk-Ports, Schalter, Kabel, Regalplatz, Raum, Strom, Kühlung und Infrastruktur müssen beschafft werden.

All das eskaliert schnell zu einer großen Anzahl von zu verwaltenden Aufgaben (von denen die meisten manuell bearbeitet werden müssen). Dies wiederum steht in keinem Verhältnis zur benötigten Zeit durch den Administrator.

IT-Admins sind in ihrer Zeit und ihren Aufgabenbereichen eingeschränkt. Sie können nur eine bestimmte Anzahl von Problemen in einer vorgegebenen Zeit abarbeiten, unabhängig von den Produktivitätswerkzeugen, die Ihnen dazu zur Verfügung gestellt werden. Ein Zuviel an Aufgaben bedeutet, dass weitere IT-Administratoren eingestellt werden müssen. Diese Konsequenzen bedeuten schließlich einen sprunghaften Anstieg des Kapitaleinsatzes und Betriebsbudgets. Wow. Das sind offensichtlich miese Folgeerscheinungen.

Leider wird es noch schlimmer.

Große, überladene Dateien haben ernst zu nehmende Auswirkungen auf die Datensicherung, die Systemwiederherstellung nach einem Ausfall und das Notfallmanagement. Backups, Speicherauszüge, Replikationen, Spiegel, etc. benötigen insgesamt deutlich länger. Oft sind überladene Dateien auch der Grund für verpasste Backup-Fenster. Wiederherstellungen (a.k.a. recovery time objective oder RTO) benötigen ebenfalls länger. Und dabei gilt zu bedenken, dass Wiederherstellungen immer dringend sind. Es kommt nicht darauf an, ob es um eine Datei, ein System oder eine Seite geht, der Zeitfaktor hat schlicht immer oberste Priorität.

Da aber Datenschutztechnologien dazu tendieren, multiple Kopien einer Datei zu erzeugen, ist der Speicherverbrauch auch vier bis zehnmal größer als beim Primärspeicher. Unabhängig davon, ob der Datenschutz auf Band oder Disk erfolgt, wird durch die überladenen Dateien immer mehr Platz in Anspruch genommen. Mehr Speicher führt wieder zu weit mehr Speichersystemen, Speichernetzwerk-Ports, Schaltern, Kabel, Regalplatz, Raum, Strom, Kühlung und benötigter IT-Adminzeit als notwendig. All das kostet Zeit und Geld. Die Konsequenzen aus überladenen Dateien sind alarmierend. Leider gibt es sogar noch weitere Probleme, wenn es um E-Mails geht. Überladene Dateien werden an E-Mails angehängt, um sie intern oder extern an Kollegen, Klienten, Kunden, Partner und Zulieferer weiterzugeben. Größere Anhänge führen zu langsameren E-Mail-Servern, größeren Datenbanken, umfangreicheren Postfächern für die Angestellten, einem komplizierteren E-Mail-Datenschutz, aufwändigeren Systemwiederherstellung nach einem Ausfall sowie einem erweiterten Notfallmanagement. Die wenigsten IT-Organisationen wollen mehr Zeit als nötig auf einen zusammengebrochenen E-Mail-Server verwenden (und E-Mail-Server brechen irgendwann zwangsläufig zusammen), geschweige denn, dass sie zwei, drei, vier oder sogar bis zu zehn Mal mehr Zeit als wirklich unbedingt notwendig darauf verschwenden wollen. Diese Leistungsfragen werden oft durch den Einsatz weiterer E-Mail-Server - einer Verkomplizierung des E-Mail-Managements, Backups, der Datenwiederherstellung etc. beantwortet. Überladene Dateien werden auch mittels SharePoint, NFS, CIFS, SSH, und FTP ausgetauscht. Voluminösere Dateien erfordern eine immer größere Bandbreite. Eine erweiterte Bandbreite kostet mehr Geld und löst meist nicht das Problem überladener Dateien - insbesondere auch nicht für das Weitferntz (WAN). Die Geräte zur WAN-Optimierung bedeuten weiteres Equipment, Regalplatz, Raum, Strom, Kühlung, Verwaltung und Kosten.

Überladene Dateien sind somit das perfekte Beispiel für die Chaostheorie. Etwas, das im Grunde nicht danach aussieht überhaupt irgendeine Konsequenz zu haben, erzeugt de facto massive, sich verstärkende Probleme. Aber IT-Unternehmen legen die Hände nicht in den Schoß und hoffen, dass sich die Probleme von selbst lösen. Viele versuchen die durch überladene Dateien entstehenden Symptome mithilfe von Provisorien zu umgehen.



Provisorien für überladene Dateien

Der häufigste Weg, der eingeschlagen wird, um überladene Dateien zu vermeiden ist die Deduplizierung und die verlustfreie Komprimierung.

Beide Ansätze mildern einige der Symptome, aber keines davon packt das Übel bei der Wurzel. Das Verständnis über die Funktionsweise dieser Technologien verdeutlicht zwar, was sie tatsächlich bewirken, aber auch in welchen Punkten sie die Probleme unter Umständen sogar noch verschlimmern.

o Deduplizierung und Überladene Dateien

Die Deduplizierung arbeitet im Grunde am besten bei Duplikaten von Dateien, Blöcken oder Blockteilen. Einfach ausgedrückt entfernt eine Deduplizierung Kopien. Es wird eine Datenbank verwendet, die jede auf ihr oder durch sie gespeicherte Dateneinheit im Auge behält. Wird ein Duplikat entdeckt, wird es entfernt und stattdessen ein Hinweis zum Speicherort des Originals abgelegt.

Weil es weit mehr Kopien in Backups, Speicherausügen, replizierten (?), CDP oder gespiegelten Daten gibt, erzielt die Deduplizierung in Datenschutzsystemen die besten Ergebnisse. Hier wird am ehesten eine hohe Deduplizierungsquote beobachtet. Das ist beim Primärspeicher aber nicht der Fall. Denn hier gibt es nicht diese Menge an Duplikaten wie im Backup, in Images oder im repliziertem Speicher. Zudem ist der größere Teil der unstrukturierten Primärdaten komprimiert oder zusätzlich verschlüsselt.

Die Deduplizierung auf Block oder Blockteilniveau (und damit die häufigste) arbeitet nicht gut mit komprimierten (Blöcke werden verändert) oder verschlüsselten Daten (die Blöcke sind immer anders). Diese Faktoren reduzieren die Effektivität der Deduplizierung beim Primärspeicher drastisch.

Wie verarbeitet das Deduplizieren überladene Dateien?

Das Deduplizieren behandelt alle Dateien gleich. Es reduziert überladene Dateien nicht im geringsten. Diese Dateien führen sogar dazu, dass der Prozess länger dauert, da mehr Daten verarbeitet werden müssen. Deduplizierung während der Übertragung vergrößert die Schreib- und Lesezeiten (zum Öffnen muss die Datei vollständig wiederhergestellt werden). Bei der nachträglichen Deduplizierung vergrößert es das Volumen des benötigten Disk-Caches und somit die Zeit, die benötigt wird, um die überladenen Dateien zu deduplizieren und die Zeit, um diese bei Bedarf wiederherzustellen.

Zusammengefasst: Die Deduplizierungsquoten für Backups, Speicherausüge, replizierte (?), CDP und gespiegelte Daten im geschützten Speicher sind relativ hoch. Für den Primärspeicher ist dieser Prozentsatz deutlich geringer, aber trotzdem noch annehmbar. Im Grunde befasst sich die Deduplizierung jedoch nicht mit der Ursache überladener Dateien, obwohl sie einige Symptome behandelt. Überladene Dateien verschlimmern manche Unzulänglichkeiten des Deduplizierens. Bei der Deduplizierung während der Übertragung vergrößern sich die Lese- und Schreibzeiten, die sowieso schon länger sind als bei nicht dedupliziertem Speicher. Bei der nachträglichen Deduplizierung vergrößert sich benötigter Disk-Cache, die Zeit für den Prozess an sich und die Dauer, um die Dateien bei Bedarf wiederherzustellen (was erneut merklich länger dauert).





o *Verlustfreie Komprimierung und Überladene Dateien*

Die beiden grundlegenden Typen einer verlustfreien Komprimierung sind LZ und inhaltsbasiertes LZ sowie Lempel Ziv (a.k.a. LZS oder LZO). Sie sind eigentlich der Deduplizierung relativ ähnlich.

LZ reduziert die Dateigröße und den Speicherverbrauch, indem doppelte Muster wie Dateien, Wörter oder Blöcke entfernt werden. Es verwendet eine Datenbank oder ein Lexikon mit Verweisen oder Übersetzungen. Ähnlich dem Deduplizieren hält es konstant Ausschau und lernt, wenn die Muster sich verändern.



Obwohl es nicht ganz so effektiv arbeitet, wenn es darum geht Duplikate zu entfernen und den Speicher wieder freizusetzen, ist es immer noch sehr wirkungsvoll und hat ähnliche Effekte auf den sekundären Speicher sowie ähnliche Probleme mit dem Primärspeicher. Insofern hat die LZ-Komprimierung auch die gleichen Defizite, denn sie unternimmt nichts gegen überladene Dateien. Einige der Symptome, die Speicher verbrauchen, werden behandelt, wobei merkliche Verzögerungen im Dateizugriff entstehen. Diese nehmen natürlich mit größeren Dateien auch noch weiter zu.

Die inhaltsbasierte Komprimierung entfernt alle zusätzlichen Leerzeichen und setzt ein Zeichen an Stellen, an denen ein identischer String wiederholt wird. Kleinere Strings, die sich regelmäßig wiederholen, werden ebenfalls ersetzt. Damit ist diese Form der Komprimierung nicht so effektiv wie die LZ-Methode und auch sie unternimmt nichts gegen überladene Dateien. Da diese wiederum um einiges länger zur Komprimierung und Dekomprimierung brauchen (es müssen mehr Daten bearbeitet werden), sind die Reaktionszeiten zum Schreiben und Lesen ebenfalls länger.

Das Beste, was über Komprimierung und überladene Dateien gesagt werden kann, ist, dass beide Typen der Komprimierung hardwareseitig ausgeführt werden können, was die Zeit zum Komprimieren und Dekomprimieren an sich beschleunigt. Trotzdem behandelt es nicht die Ursache des Problems. Es sollte einen besseren Weg zur Behandlung dieses Übels geben und glücklicherweise ist das auch so.



Neuxpowers Lösung für überladene Dateien: NXPowerLite

NXPowerLite ist eine Dateioptimierungslösung, die auf eine verlustbehaftete Optimierung ausgelegt ist - sie entfernt die Überlast der Dateien. Die Technologie kann die Größe von Microsoft Office- und JPEG-Dateien um bis zu 95% reduzieren ¹.



Die NXPowerLite Software ist sowohl für den Desktop als auch für den Dateiserver erhältlich.

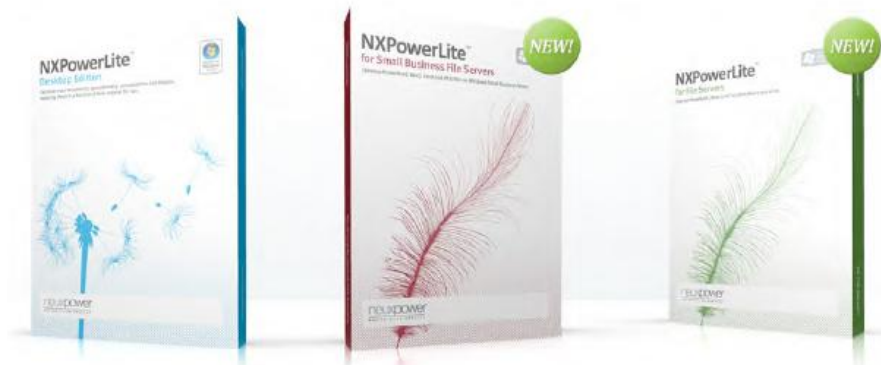
NXPowerLite greift überladene Dateien direkt an. Es entfernt unnötigen Ballast und überflüssige Daten, Grafiken werden in das am besten geeignete Format und die passende Auflösung formatiert.

Übrig bleibt eine kleinere und effizientere Datei, die genauso behandelbar ist und genauso aussieht wie das Original, aber eben um einiges schlanker geworden ist.

Die optimierten Dateien behalten ihr ursprüngliches Format mit all ihren Attributen und der Integrität der visuellen Inhalte. Damit sind das Aussehen und die Funktionalität gemeint, die in jeder Hinsicht identisch mit der zuvor überladenen Original bleibt.

Anders als deduplizierte oder komprimierte Dateien müssen die optimierten Dateien nicht wiederhergestellt oder dekomprimiert werden, um gelesen werden zu können. Das bedeutet viel schnellere Antwortzeiten. Zusätzlich entfällt die Notwendigkeit einer bestimmten Anzeigesoftware, um die mit NXPowerLite optimierten Dateien zu lesen. Wie die ursprüngliche, überladene Datei können diese von jedem geöffnet und bearbeitet werden. Das Beste aber ist, dass die mit NXPowerLite optimierten Dateien zusammen mit einer Deduplizierung und/oder Komprimierungstechnologie verwendet werden können. Durch das Entfernen der unnötigen Bestandteile einer Datei beschleunigt NXPowerLite eine Deduplizierung und die Komprimierung und macht sie zudem deutlich effektiver.

Mit mehr als einer Million Implementierungen bis heute, einschließlich vieler in militärisch relevanten Bereichen, ist NXPowerLite eine bewährte Lösung gegen überladene Dateien.



Fazit

Überladene Dateien sind ein bitteres, zeitraubendes und teures Problem. Der Schlüssel zur Lösung liegt in einer verlustbehafteten Dateioptimierung. NXPowerLite für Desktops und Dateiserver bietet genau diesen Lösungsansatz. Für eine Testversion oder weitere Informationen über Neuxpower und NXPowerLite, besuchen Sie www.nxpowerlite.de

Über den Autor: Marc Staimer ist der Gründer, Senior Analyst und CDS von Dragon Slayer Consulting in Beaverton, OR. Er hat 12 Jahre Beratungspraxis mit Schwerpunkten in den Bereichen strategische Planung, Produktentwicklung und Marktentwicklung. Mit mehr als 30 Jahren im Marketing, Vertrieb und Unternehmenserfahrung in den Bereichen Infrastruktur, Speicher, Server, Software und Virtualisierung gilt er als einer der führenden Industrieexperten. Marc kann unter folgender E-Mail-Adresse kontaktiert werden:

marcstaimer@mac.com.

¹ Source: NXPowerLite Trident Warrior Experimentation and Results, FORCEnet