



Architect of an Open World™



Data Deduplication in Backup und Recovery

Josef Höckner

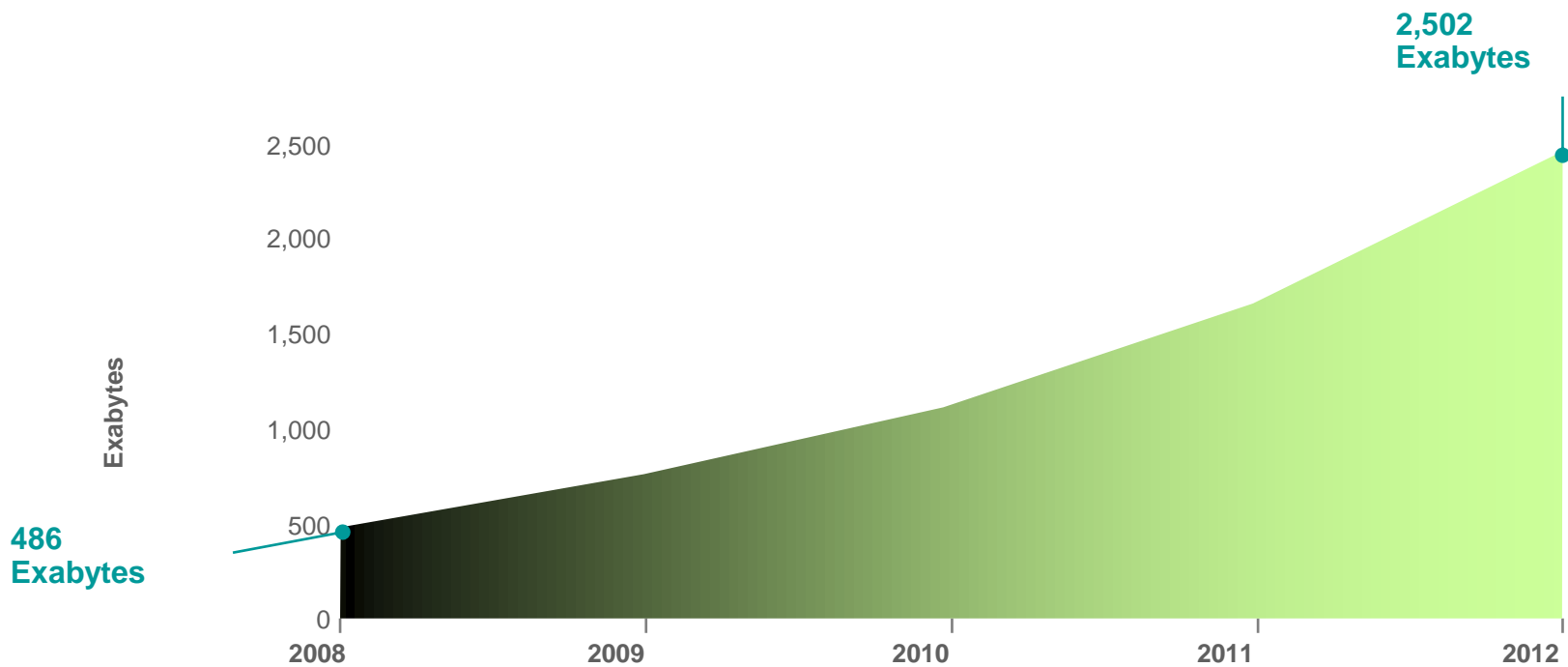
Storeway™ Solution Architect

....die alte Leier....

- **Bedarf an Speicherplatz steigt nach wie vor**
- **Zumindest gleichbleibende SLA's**
 - Backup-Zeiten außerhalb des Backup-Windows
 - B2D für schnelleres Recovery (Kosten!)
- **Ressourcen werden weniger (Personal und Budget)**



Datenwachstum



Source: IDC Digital Universe white paper, sponsored by EMC, May 2009



Wenn ein Sandkorn ein Byte wäre ...



1 Megabyte = 1 Million Bytes
Ein Esslöffel Sand



1 Gigabyte = 1 Milliarde Bytes
Ein Sandfleck – 0,25m², 30 cm tief



1 Terabyte = 1 Billion Bytes
Ein Sandkasten – 7m², 30 cm tief



1 Petabyte = 1.000 Terabytes
1.600m langer Strand – 30m breit, 30cm tief

Wenn ein Sandkorn ein Byte wäre ...



1 Exabyte = 1.000 Petabytes
Der gleiche Strand – 2x die Landesgrenze Österreich / Deutschland



1 Zetabyte = 1.000 Exabytes
Der gleiche Strand – entlang ganz Europas



1 Yottabyte = 1.000 Zetabytes
Genug um die gesamte Fläche Europas mit 100m Sand zu bedecken



Was ist Deduplizierung?

Deduplizierung ist eine Technologie die es ermöglicht große Mengen an Kapazität einzusparen.

In der Regel werden durch diese Methode redundante (mehrfach vorhandene) Daten eliminiert und durch Platzhalterverweise auf den Ursprungsdatensatz ersetzt.

Dies geschieht völlig transparent für Anwendungen und Anwender.



Wie funktioniert Deduplizierung?

Die Art und Weise wie die Erkennung von Redundanzen vor sich geht, lässt sich unterteilen in **objektbasiert** und **blockbasiert**.

Bei **objektbasierter** Erkennung (Single Instanz) genügt schon das Ändern eines einzigen Buchstabens in einem langen Dokument um eine Ungleichheit hervorzurufen und die Deduplizierung unwirksam zu machen.

Bei der **blockbasierten** Deduplizierung erfolgt die Redundanzerkennung auf Datenblockebene, wodurch auch bei Änderungen an Dateien noch große Teile als redundant erkannt und erfolgreich reduziert werden können.



Arten der Deduplizierung

Datei-Ebene (Object-Level)

Nur eine Kopie pro File wird gespeichert



HITACHI
DATA SYSTEMS

BULL

Arten der Deduplizierung

Beispiel Blockbasiert (fixe Blocklängen)

Original

```
000110100101101000110010110001010100101
100101000010100110100110001101001010110
010100101011011000100011101001010001010
010100010110101001101010010100100101001
1010010101010100101010010110101010010
100010010111100101100101001001000110011
000110100101101000110010110001010100101
100101000010100110100110001101001010110
010100101011011000100011101001010001010
010100010110101001101010010100100101001
1010010101010100101010010110101010010
100010010111100101100101001001000110011
```

Variante

```
000110100101101000110010110001010100101
100101000010100110100110001101001010110
010100101011011000100011101001010001010
010100010110101001101010010100100101001
101001010101001010010010010110101010010
100010010111100101100101001001000110011
000110100101101000110010110001010100101
100101000010100110100110001101001010110
010100101011011000100011101001010001010
010100010110101001101010010100100101001
1010010101010100101010010110101010010
100010010111100101100101001001000110011
```

=> Übereinstimmung 93%



Arten der Deduplizierung

Beispiel Blockbasiert (variable Blocklängen)

Original

```
000110100101101000110010110001010100101
100101000010100110100110001101001010110
010100101011011000100011101001010001010
010100010110101001101010010100100101001
101001010101010010101010010110101010010
100010010111100101100101001001000110011
000110100101101000110010110001010100101
100101000010100110100110001101001010110
010100101011011000100011101001010001010
010100010110101001101010010100100101001
101001010101010010101010010110101010010
100010010111100101100101001001000110011
```

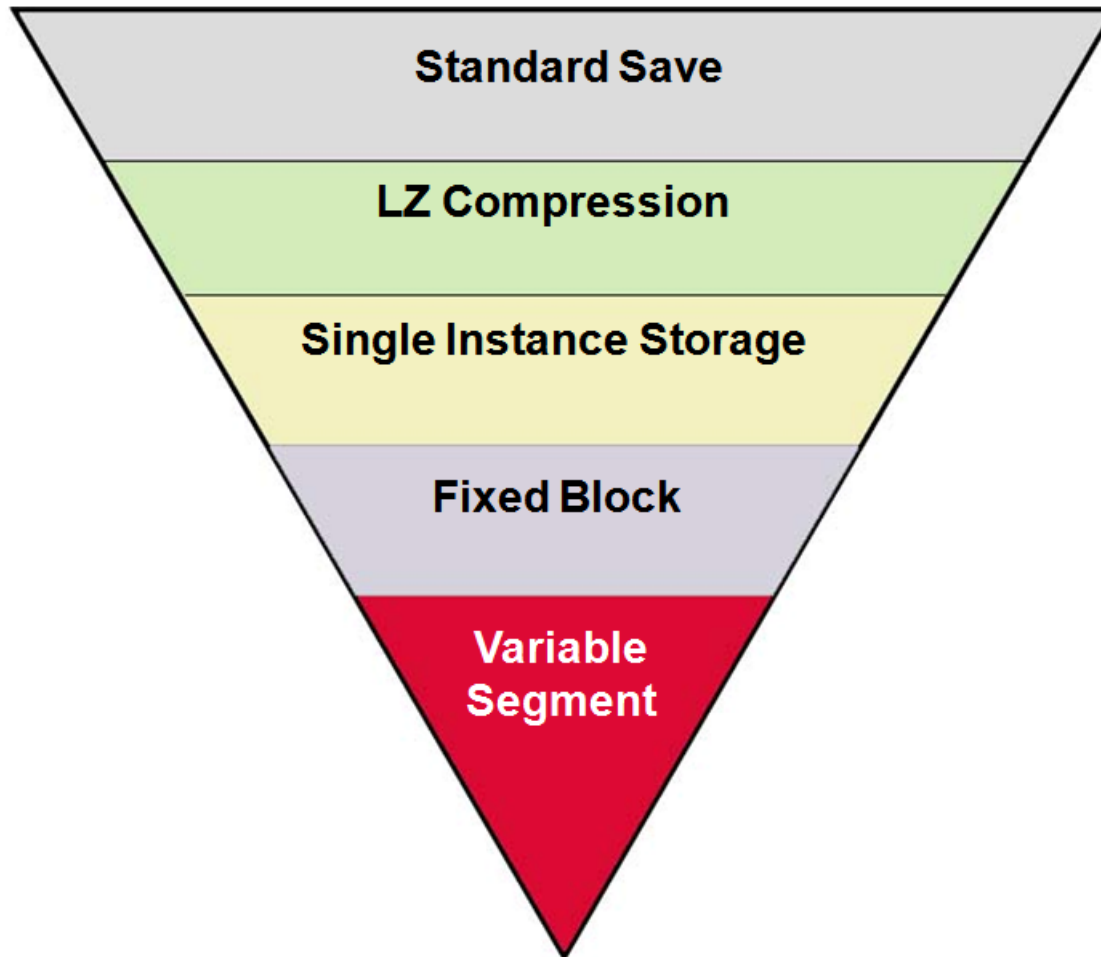
Variante

```
000110100101101000110010110001010100101
100101000010100110100110001101001010110
010100101011011000100011101001010001010
010100010110101001101010010100100101001
10100101010101001010100010110101010010
100010010111100101100101001001000110011
000110100101101000110010110001010100101
100101000010100110100110001101001010110
010100101011011000100011101001010001010
010100010110101001101010010100100101001
101001010101010010101010010110101010010
100010010111100101100101001001000110011
```

=> Übereinstimmung 99%



Arten der Deduplizierung



Orte der Deduplizierung

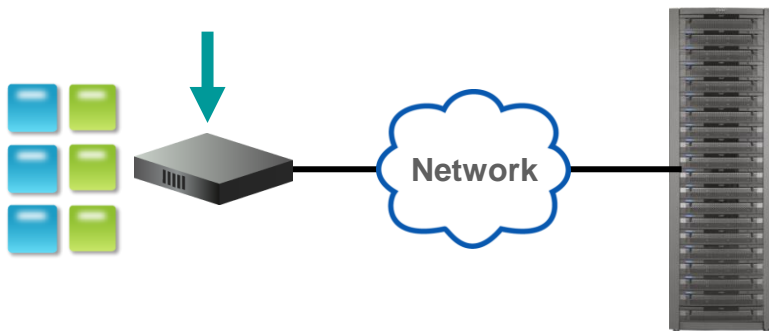
Source

- Client software agents identify repeated sub-file data segments at the source
- Only new, unique segments are transferred across the network and stored to disk
- Shorter backup window reduces daily impact on physical/virtual infrastructure

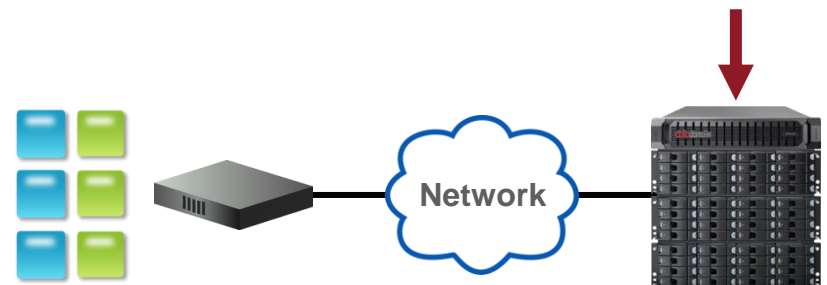
Target

- Backup application sends native data to a target storage device
- Data is deduplicated once it reaches the target—during or after the backup
- Found in virtual tape libraries (VTLs) or LAN backup-to-disk appliances
- Transparency to backup application offers users a “plug-and-play” experience

Deduplication at Source



Deduplication at Target



Was bringt Deduplizierung

Durch das Beseitigen redundanter Datenobjekte und die ausschließliche Referenzierung auf das Ursprungsobjekt ergibt sich ein unmittelbarer Nutzen in punkto Storage-Effizienz.

Kostenvorteil: Geringere Storage-Anschaffungskosten oder längere Abstände bei der Erweiterung der Storage-Kapazität.

Managementvorteil: Die Fähigkeit, „mehr“ Daten pro Storage-Einheit zu speichern oder Online-Daten über längere Zeiträume vorzuhalten.

Bandbreiteneinsparung: Durch Deduplizierung wird auch bei der Bandbreite (speziell im WAN-Bereich wichtig) eingespart oder überhaupt erst ein zentrales Backup ermöglicht



Was bringt Deduplizierung



Unterstützung mit Bull Advisory Services

Understanding of operational and business challenges, transversal technical expertise, best practices...

Data Protection

- Servers, file servers, desktop
- Applications & non-structured information
- Data restore, RPO / RTO

Archiving

- Email, content management, files systems, databases
- Governance & compliance
- Retention periods
- Search and discovery

Consulting

- Financial analysis
- Service improvements
- Service alignment
- ROI analysis
- Impact analysis
- Process evaluation



Virtualisation

- Services and storage
- Resource activation and reconfiguration
- Capacity optimisation
- Application and data mobility
- Management of physical and virtual environments

Disaster Recovery

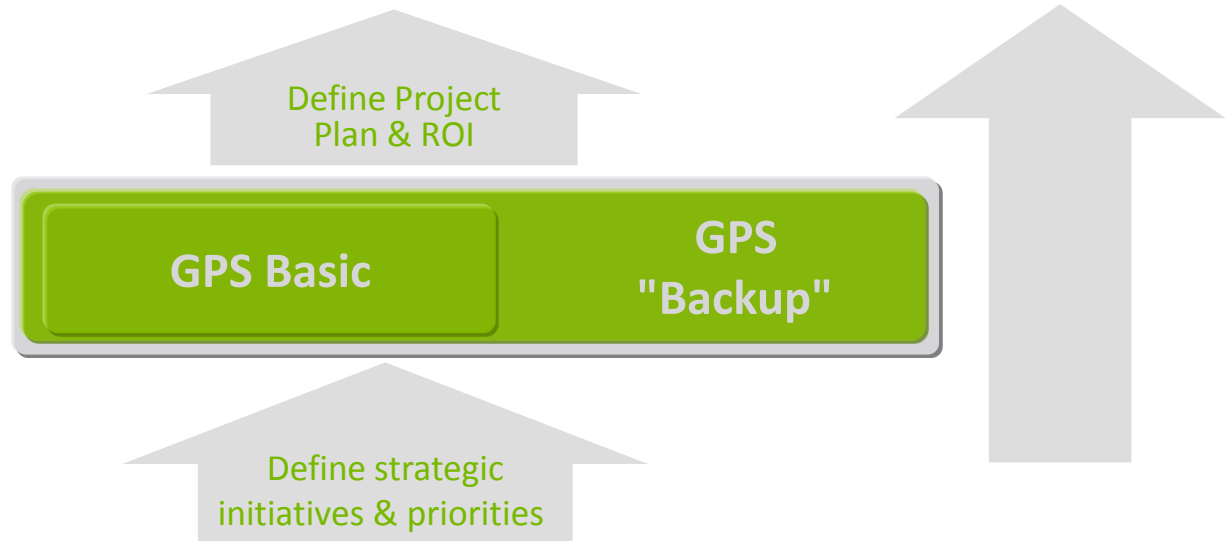
- Risk management
- Local, regional and international backup sites
- Telecommunication services
- Business continuity planning

Unterstützung mit Bull Advisory Services

COMPASS
Packages



GPS
Packages



Powered by  **GLASSHOUSE**



Bull StoreWay™

