

# Signifikante Daten- reduktion im SAP BW

André Paul  
Business Intelligence

# Signifikante Datenreduktion im SAP-BW

## Unternehmens- einordnung

# Standorte und Mitarbeiter\*

•EDS ITELLIUM ist der führende Dienstleister für praxiserprobte und zukunftssichere IT-Lösungen im Handel – wir stehen als **Ihr** Partner für hohe Kompetenz, langjährige Erfahrung und nachgewiesene Business-Exzellenz.



<b>Essen</b>	<b>388</b>
<b>Frankfurt</b>	<b>17</b>
<b>Nürnberg</b>	<b>301</b>
<b>TOTAL*</b>	<b>706</b>

\*Stand: 1.7.2007

**Essen**  
Theodor-Althoff-Straße 2  
45133 Essen

**Frankfurt**  
Hanauer Landstraße 360  
60386 Frankfurt

**Nürnberg**  
Hundingstraße 11b  
90431 Nürnberg

# Portfolio Business Intelligence

## BI Consulting

1. Fachliches & technisches Consulting zur Finanz- und Warenwirtschaftlichen Planung
2. Fachliches & technisches Consulting zur Finanz- und Warenwirtschaftlichen Analyse und dem zugehörigen Reporting
3. Fachliches & technisches Consulting zum CRM
4. Fachliches & technisches Consulting bei der Tool-Evaluierung und Herstellerauswahl
5. Consulting bei technischen Themen
6. Erstellen Konzeptionen und Business Blueprints

## BI Implementierung

1. BI-Anforderungsanalyse
  - Zieldefinitionen und Definition KPI
  - BI-Konzeption
  - Kostenanalyse
2. Datenmodellierung
3. Design/Modellierung/ETL
4. Implementierung
5. Tuning/Optimierung
6. Einführung der BI-Komponenten inkl. Der relevanten Schulungen
7. Administration, Pflege, Weiterentwicklung

## BI Themen

1. Oberflächenanalyse
2. Auswahl Tools (Front End/Back End)
3. Query/Reporting und Analyse
4. Datenvisualisierung
5. Internet-/Intranet-Anbindung
6. Berücksichtigung bestehender BI-Lösungen für Finanzwirtschaft und Handel inkl. deren Integration bzw. Migration (ETL)
7. Data Mining
8. Planung (Finanz- und Warenwirtschaft)
9. Kampagnenmanagement

# Signifikante Datenreduktion im SAP-BW

## Ausgangslage

# Ausgangslage

- Eine EDW-Systemlandschaft hat immer mehrere Datenhaltungsschichten.
  - Data Warehouse Layer (PSA / DSO)
  - Reporting Layer (Cubes)
- Diese unterschiedlichen Datenhaltungsschichten werden unterschiedlich genutzt und unterliegen somit einem unterschiedlichem Zugriffsverhalten.
- Der Data Warehouse Layer stellt das historische Gedächtnis des Unternehmens dar, da die Quellsysteme in der Regel keine entsprechende Historie vorhalten.
- Das historische Gedächtnis wird in der Regel nicht bzw. nicht im operativen Reporting genutzt.
- Mit steigendem Datenvolumen steigen die Zeiten für das Backup, dem ggf. notwendigen Recovery und sowohl deren Kosten als auch für die Speicherung und Sicherung der Daten benötigten Datenhaltungsmedien.
- Verlängerte Recovery-Zeiten können zu erhöhten kalkulatorischen Kosten führen, da Systeme im Ausnahmefall nicht verfügbar sind.

# Ausgangslage

- Eine EDW-Systemlandschaft beinhaltet immer HOT and COLD DATA.
  - HOT Data werden operativ genutzt und „permanent“ zugegriffen. Diese müssen eine hohe Verfügbarkeit haben
    - ➔ Daten zum Reporting
  - COLD Data werden in der Regel nie oder nur äußerst selten operativ genutzt. Diese müssen zwar verfügbar sein; ein Zugriff kann aber auch eingeschränkt bzgl. der Zugriffszeit sein.
    - ➔ Daten, die für einen Neuaufbau bei geänderten / neuen Anforderungen genutzt werden

# Ausgangslage

- Für einen Großkunden von EDS Itellium wurde eine Evaluierung zur Archivierung der Daten des SAP-BI durchgeführt. Hierbei mussten die nachfolgenden Rahmenbedingungen eingehalten werden:
  - Möglichkeit des Online-Zugriffs auf die archivierten Daten (kein Restore bevor eine Nutzung möglich ist)
  - Optimierung der Plattennutzung → daher Komprimierung
  - Speicherung der archivierten Daten außerhalb der Datenbank, um Backup-Zeiten zu reduzieren
  - Nutzung ohne weitere HW → daher Nutzung auf identischer Systemplattform
  - Vollständige Integration in die bestehende Systemlandschaft



# Signifikante Datenreduktion im SAP-BW

Entscheidung

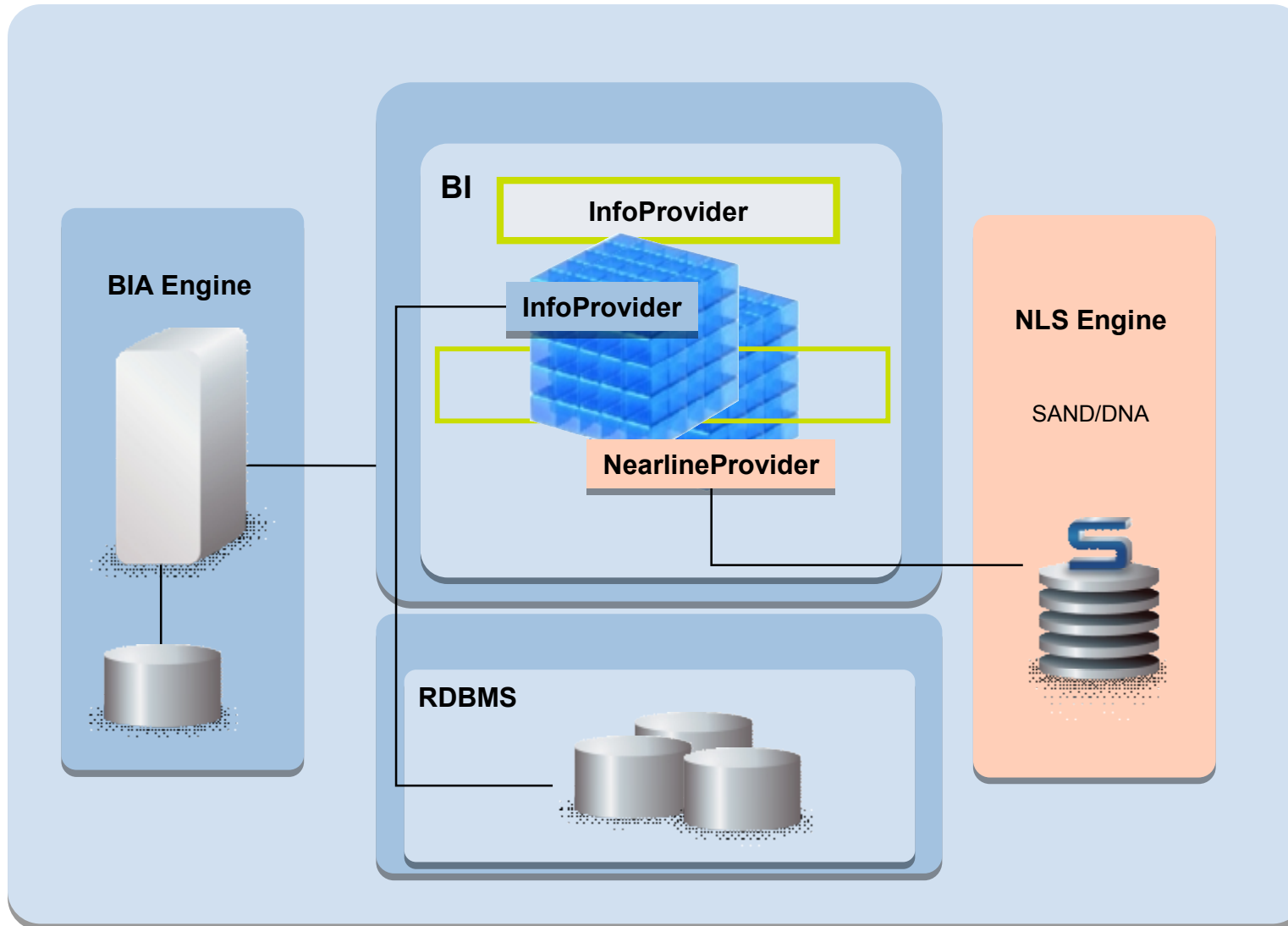
# Entscheidung

- Gemeinsam mit dem Großkunden von EDS Itellium wurde eine Entscheidung zu einer Nearline-Lösung zu Gunsten der Lösung SAND/DNA der Firma SAND getroffen.

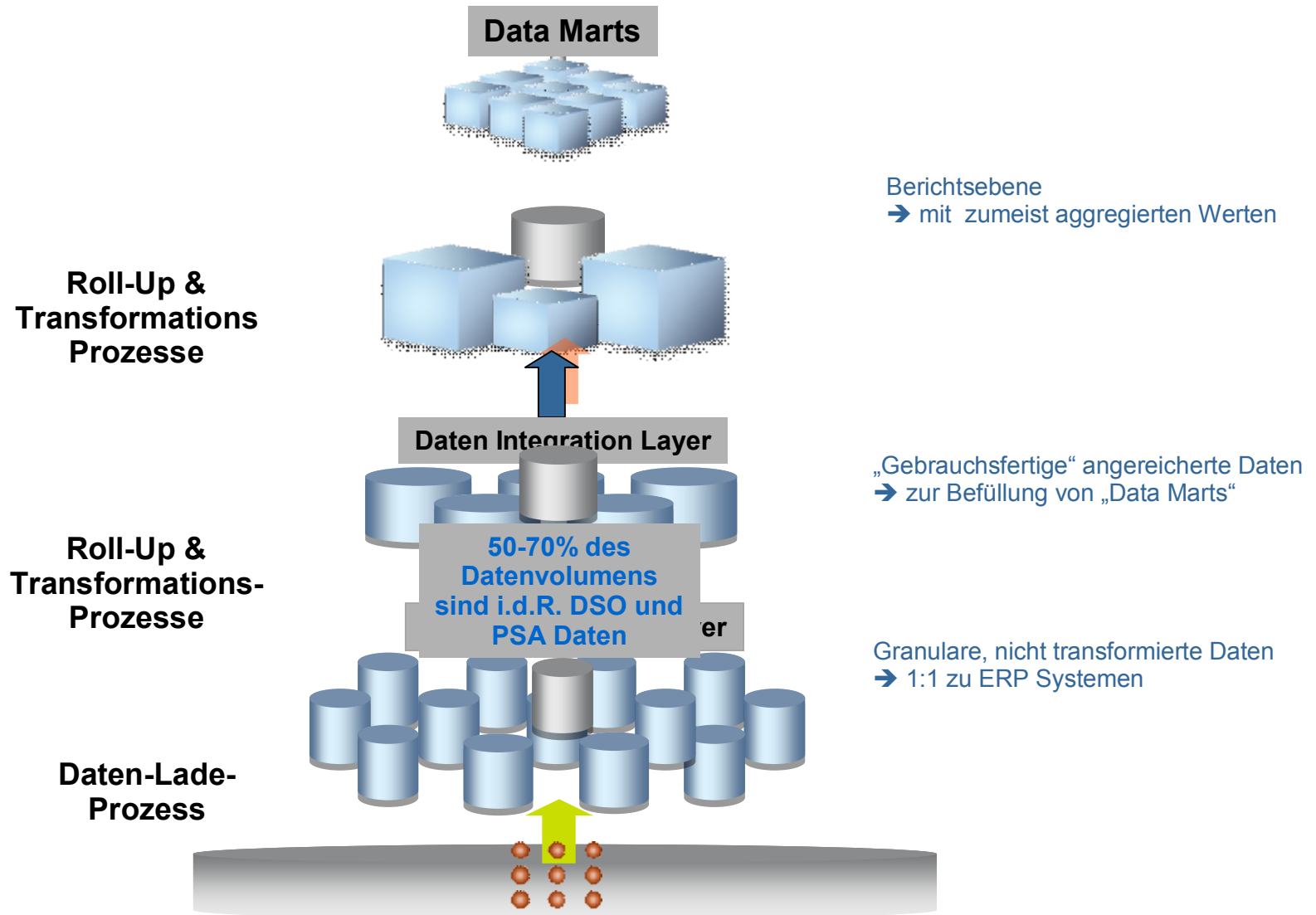
# Signifikante Datenreduktion im SAP-BW

## Architektur und Verfahrensweisen

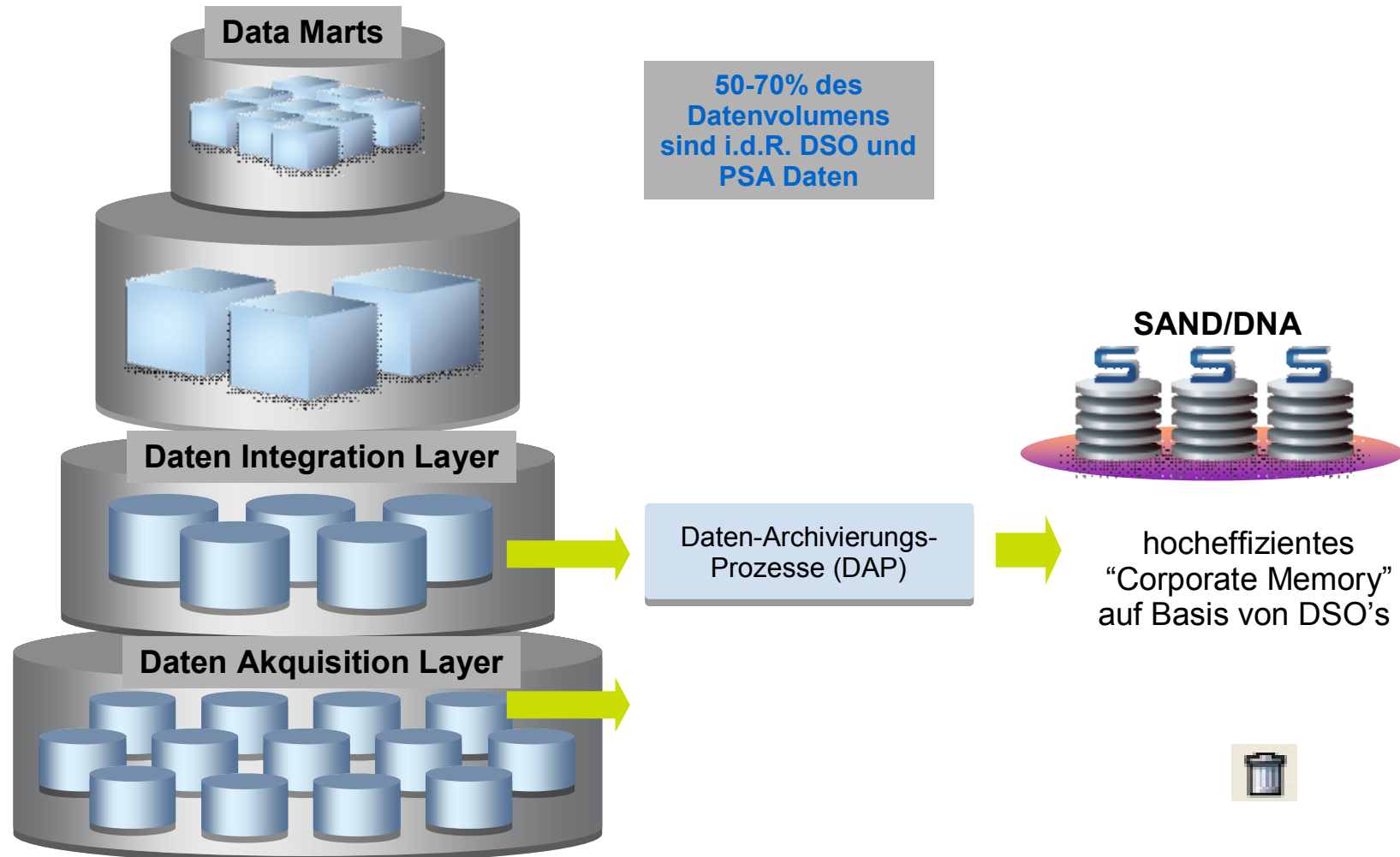
# System-Integration der Nearline-Lösung



# EDW: Datenbewirtschaftung & Transformation



# Aktueller Einsatz bei EDS-Itellium

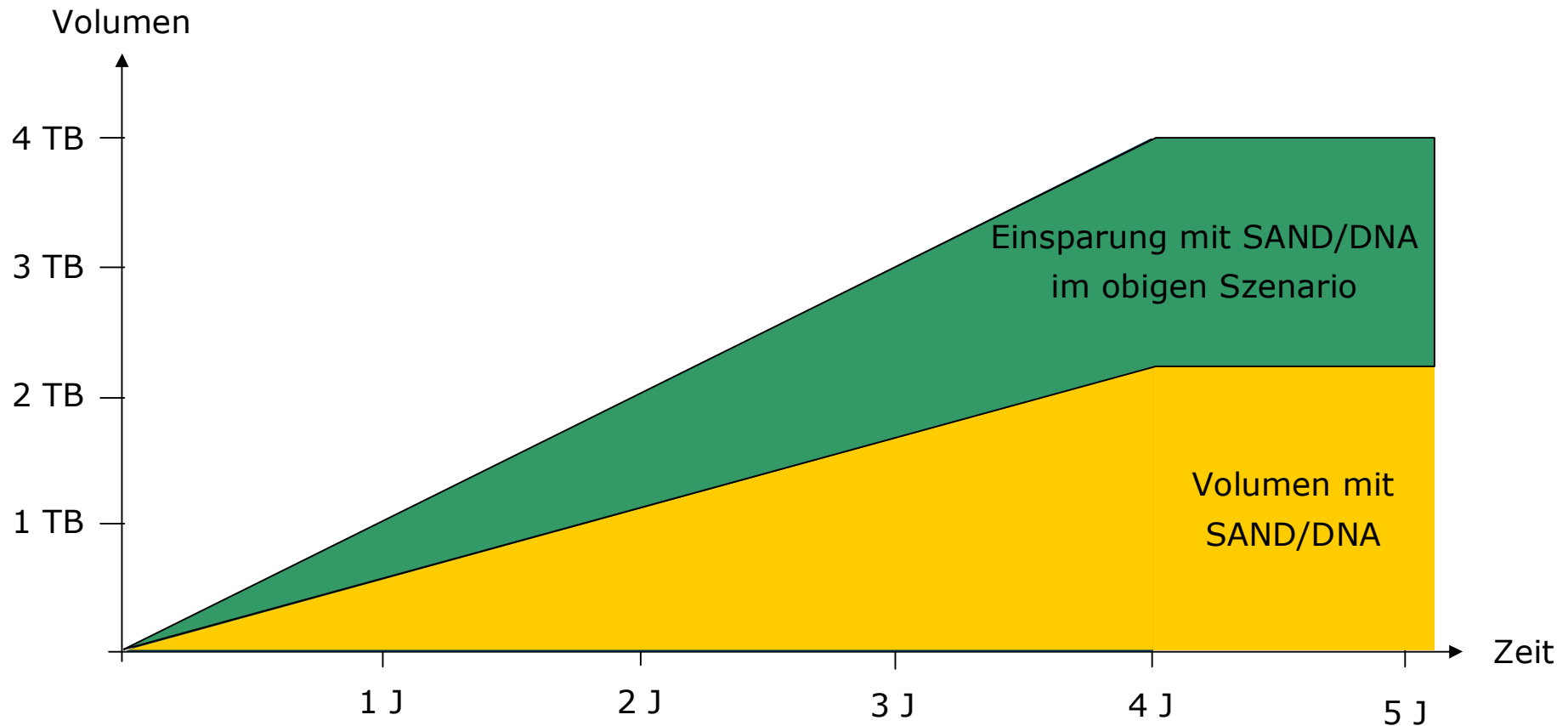


# Signifikante Datenreduktion im SAP-BW

Einsparpotentiale  
Disk

# Mengenwachstum (Beispielszenario)

- Bei einem angenommenen linearen Wachstum von 1 TB im Jahr über 4 Jahre (ab dem fünften Jahr wird soviel gelöscht wie neu hinzukommt) und der Möglichkeit 50% der Datenmenge über SAND/DNA auszulagern, ergibt sich die nachfolgende Betrachtung:





# Mengenwachstum (Beispielszenario)

- Bei einem angenommenen linearen Wachstum von 1 TB im Jahr über 4 Jahre (ab dem fünften Jahr wird soviel gelöscht wie neu hinzukommt) und der Möglichkeit 50% der Datenmenge über SAND/DNA auszulagern, ergibt sich die nachfolgende Betrachtung:

Szenario	Jahr 1	Jahr2	Jahr 3	Jahr 4	Jahr5 ff.
Ohne SAND/DNA	1 TB	2 TB	3 TB	4 TB	4 TB
Mit SAND/DNA	0,55 TB	1,1 TB	1,65 TB	2,2 TB	2,2 TB
Details mit SAND/DNA	0,5 TB BW 0,05 TB SAND	1 TB BW 0,1 TB SAND	1,5 TB BW 0,15 TB SAND	2 TB BW 0,2 TB SAND	2 TB BW 0,2 TB SAND

# Testergebnisse

- Überleitung aus einem ODS in die Storage-Objekte des SAND/DNA

Selektion	Anzahl Datensätze	Größe (KB)	Laufzeit (s) Archivierung + Verifikation	Laufzeit (s) Daten löschen	Gesamt - Laufzeit (s)	Größe Archivfile (KB)
01.03. bis 03.03.2005	1.683.246	1.346.597	2.445	869	3.314	131.681
04.03. bis 06.03.2005	1.688.452	1.350.762			2.528	129.472

Eine Komprimierung im Verhältnis 1:10 konnte bei der Nutzung von Oracle als DBMS nachgewiesen werden.

Die Verarbeitung in 2 Schritten (Archivierung und Verifikation und Löschung in separaten Jobs) ist nachweislich langsamer als eine Verarbeitung in einem Schritt.

# Signifikante Datenreduktion im SAP-BW

## Laufzeiten

# Laufzeiten

- Fortschreibung aus den Storage-Objekten des SAND/DNA in ein ODS

Im ersten Fall wurde ein ODS in ein zweites fortgeschrieben.

Im zweiten Fall wurde aus SAND/DNA in das zweite ODS fortgeschrieben

	ODS zu ODS	SAND zu ODS
Fortschreibung 1.1.05 - 31.1.05:	5:53 Min.	38:45 Min.
Fortschreibung 1.1.05 - 28.2.05:	5:52 Min.	39:23 Min.

Die wesentlichen Einbussen sind beim Zugriff auf die Storage-Objekte des SAND/DNA und deren Bearbeitung zurückzuführen. Durch eine parallele Bearbeitung lassen sich hier Optimierungen erzielen.

Durch den Hersteller wurde eine im Verhältnis 1:2,5 höhere Laufzeit avisiert.

Da diese Notwendigkeit ausschließlich für Neuaufbauten genutzt wird und diese dann im Rahmen eines Projektes erfolgen, wurde diese höhere Laufzeit akzeptiert.

# Laufzeiten

- Zugriff über eine Query auf die Dateninhalte aus dem ODS im Vergleich zu den Storage-Objekten des SAND/DNA. In den Testfällen wurde der Cache mit genutzt.

Im ersten Test-Fall wurde eine Query aus dem laufenden System auf das ODS und anschließend nach einer Archivierung (mit einem Tagesversatz) auf die Storage-Objekte des SAND/DNA ausgeführt.

Bei beiden Zugriffen war die Datenmenge identisch abgegrenzt und nur in genau einem Datenhaltungsobjekt verfügbar.

- Erster Aufruf

Anz. Datensätze / Ergebnismenge	ODS	SAND/DNA
536.868 / 6	1.199.000 ms	1.405.047 ms

- Zweiter Aufruf

Anz. Datensätze / Ergebnismenge	ODS	SAND/DNA
536.868 / 6	1.078 ms	1.156 ms

Die Messungen fanden jeweils im laufenden Betrieb statt. Damit wurden diese nicht unter Laborbedingungen erhoben und waren weiteren Einflüssen unterlegen (parallel laufende Bearbeitungen. Die jeweiligen Datenhaltungsobjekte wurden nicht parallel bearbeitet.

# Laufzeiten

- Zugriff über eine Query auf die Dateninhalte aus dem ODS im Vergleich zu den Storage-Objekten des SAND/DNA. In den Testfällen wurde der Cache explizit ausgeschaltet.

Im ersten Test-Fall wurde eine Query aus dem laufenden System auf das ODS und anschließend nach einer Archivierung (mit einem Tagesversatz) auf die Storage-Objekte des SAND/DNA ausgeführt.

Bei beiden Zugriffen war die Datenmenge identisch abgegrenzt und nur in genau einem Datenhaltungsobjekt verfügbar.

- Erster Aufruf

Anz. Datensätze / Ergebnismenge	ODS	SAND/DNA
536.868 / 6	1.114.750 ms	1.790.624 ms

Die Messungen fanden jeweils im laufenden Betrieb statt. Damit wurden diese nicht unter Laborbedingungen erhoben und waren weiteren Einflüssen unterlegen (parallel laufende Bearbeitungen. Die jeweiligen Datenhaltungsobjekte wurden nicht parallel bearbeitet.

# Signifikante Datenreduktion im SAP-BW

Fazit

# Fazit

- Durch Nutzung der Lösung SAND/DNA konnten nachweislich
  - Plattenbedarfe reduziert
  - Backup-Zeiten – bedingt durch die Datenreduktion – verkürzt
  - Datenhaltungszeiten bei identischen Kosten verlängert werden
- Durch die Nutzung der identischen HW-Komponenten fielen keine zusätzlichen Installations- und Wartungskosten an.
- Die Installation der SW-Komponenten von SAND waren innerhalb von 4 Tagen abgeschlossen (inkl. einspielen der benötigten Hinweise).



## André Paul

EDS Itellium

Theodor-Althoff-Strasse 1

45133 Essen, Germany

+49 (0) 201 8709 2379

[andre.paul@eds.com](mailto:andre.paul@eds.com)

EDS and the EDS logo are registered trademarks of Electronic Data Systems Corporation. EDS is an equal opportunity employer and values the diversity of its people. © 2007 Electronic Data Systems Corporation. All rights reserved.