Deutschsprachige
SAP® Anwendergruppe

Signifikante Datenreduktion mit SAND Nearline Lösung in einem IS-U BW

Von PSA und Cube-Archivierung in BW3 zu EHP1 NLS in BI7 mit SAND/DNA

Jens Schwarz, entega Service GmbH
Dr. Michael Hahne, SAND Technology

Agenda

1. Challenge – Running Growing SAP BI systems
2. Solution – ILM and SAP BI Nearline Storage
3. Best Practice: Nearline Storage in a SAP BI Enterprise Data Warehousing (EDW) Architecture
4. Best Practice: Nearline Storage and Reporting
5. Case Study entega Service
6. Summary, Q&A

Agenda

1. Challenge – Running Growing SAP BI systems
2. Solution – ILM and SAP BI Nearline Storage
3. Best Practice: Nearline Storage in a SAP BI Enterprise Data Warehousing (EDW) Architecture
4. Best Practice: Nearline Storage and Reporting
5. Case Study entega Service
6. Summary, Q&A

The Challenge

- “With projected compounded annual growth rates for databases exceeding 125%, organizations face two basic options:
  - 1) Continue to grow the infrastructure (e.g., server size, storage capacity)
  - OR
  - 2) Develop processes [and architectures] to separate dormant [archive-ready] data from active data.”

Meta Group Report
Databases on a Diet

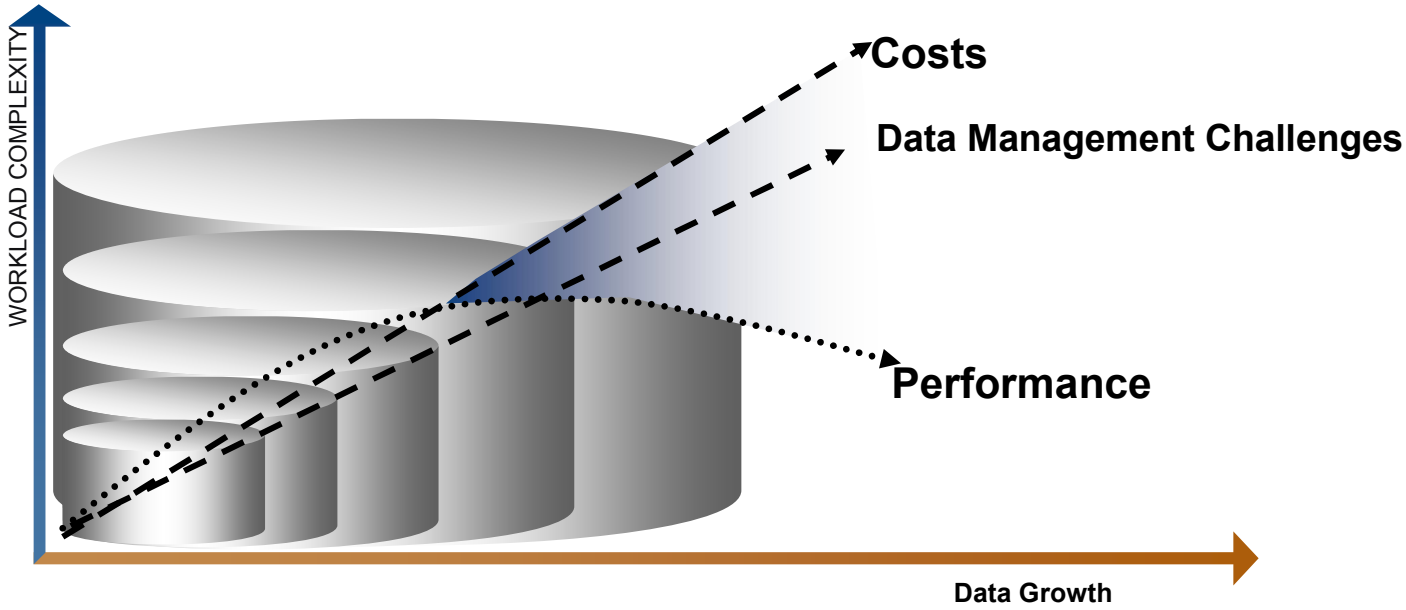
What Companies are Facing Today…

- Explosive data growth & increased performance requirements
  - Corporate expansion and increased sales – more transactions, more customers, etc.
  - New data types, e.g. RFID, IM, logs (transaction logs, web logs, system logs)
  - Increased user expectations, e.g. for more detailed analyses for longer periods
  - Data Remodeling
  - More ad hoc reporting
  - New legal regulations such as SOX, Basel II
  - “Controlled" redundancy within the EDWs

- ➔ Data Warehouse Management challenges
  - Decreased performance
  - increased TCO
  - Increased complexity
  - Failure to provide required levels of service

Result: Missed Service Levels

- Performance Can't Keep Pace
- "Batch Windows" for Data Preparation Unmanageable



WHAT ARE THE OPTIONS????

Traditional Solutions

- Increase the Hardware Landscape
  - Adding processing power
  - Adding Memory
  - Adding Storage capacity
- Data Model Optimization
  - Implement Aggregates
  - Table Partitioning
  - In Database Compression

Moore's Law (and Kryder's Law) and a huge exception

Moore: transistors on IC double every 2 years

Kryder: density of information on discs double every 18 months

Growth factors:

***Transistors/chip*:**

>100,000 since 1971

- ***Disk density:***

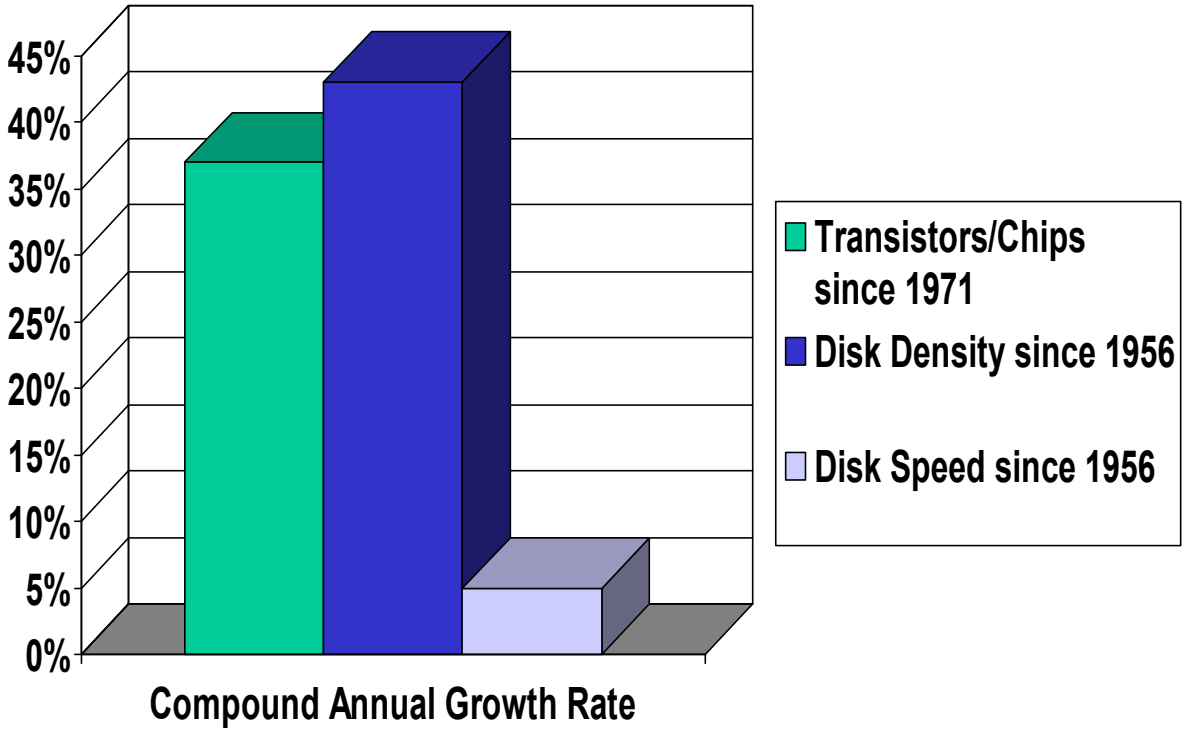
  >100,000,000 since 1956

- ***Disk speed:***

  12.5 since 1956

The disk speed barrier dominates everything!

Source: Monash Research



Why not Just Add More Storage ?

- Data volumes are in growing faster than the price/performance ratios of disk storage technology.
- Fast disks are still expensive
- Data stored in production environments requires failover and backup technology
- For every dollar a company spends on data storage devices, an estimated additional $5 to $10 is required to manage those devices over the lifetime of the equipment
- ➔ Total costs > $ 150.000 per TB per year

- More importantly, large volumes of data have adverse effects on system responsiveness, in areas such as:
  - Data loading performance
  - Performance of change runs, rollups, and so on
  - Backup and recovery times
  - Migration and upgrade times.

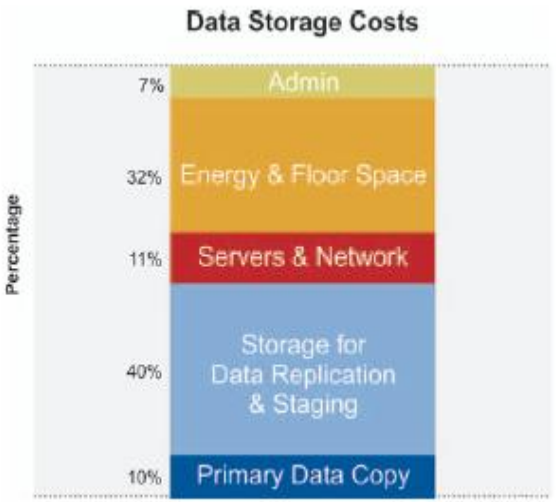
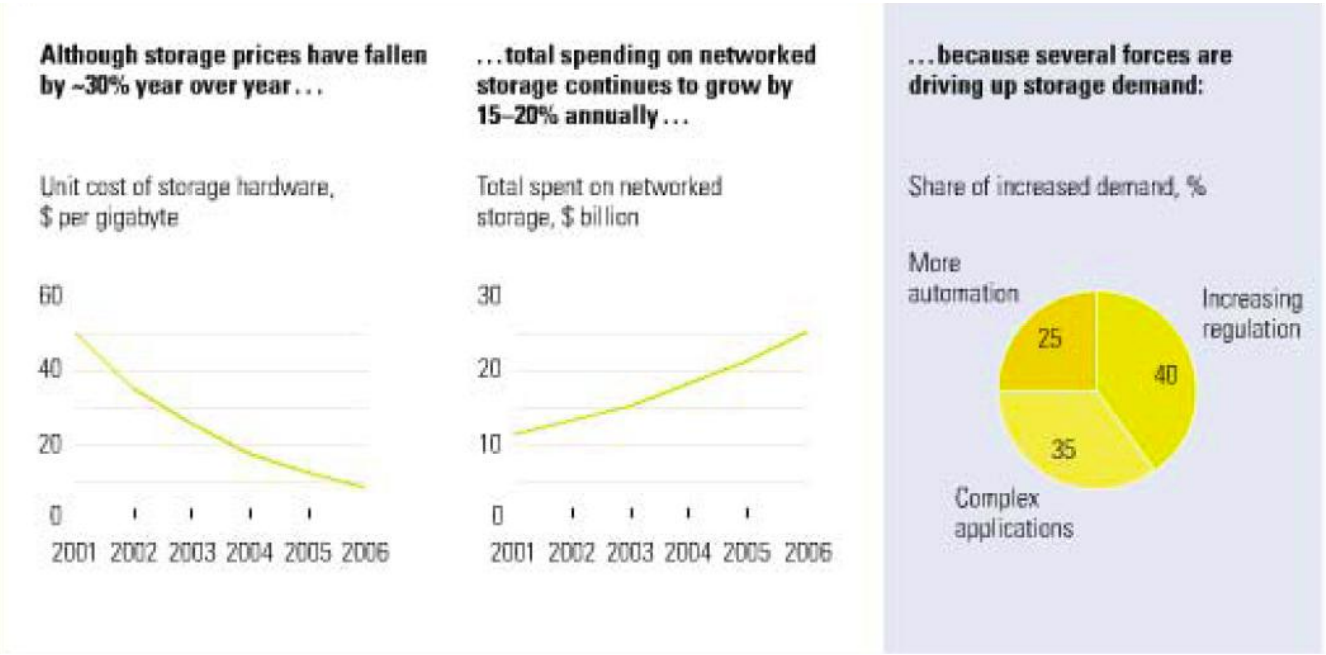


Figure 1: Breakdown of Data Storage Costs

Total Corporate Spending on Storage …

… (disk drives, tape systems, specialized network gear, and the people and software to manage them) grows by 15 to 20 percent every year, even though the unit cost of storage drops by about 30 percent annually



Bill Inmon's Opinion about Performance Issues and NLS

"Indeed, leaving infrequently accessed data on disk storage greatly HURTS performance. … Data warehouse performance is hurt because mixing infrequently used data with actively used data is like adding lots of cholesterol into the blood stream."

Information Lifecycle Management
for Data Warehousing:
Matching Technology to Reality
An Introduction to
SAND Searchable Archive
By W.H. Inmon

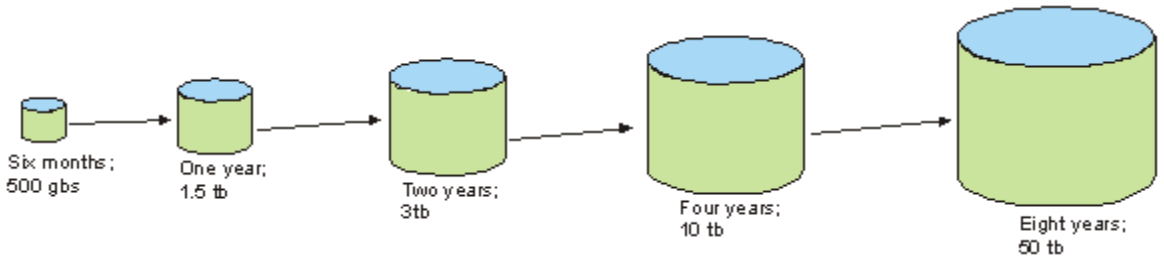
Figure 8: Cholesterol's Effect on Circulation

Agenda

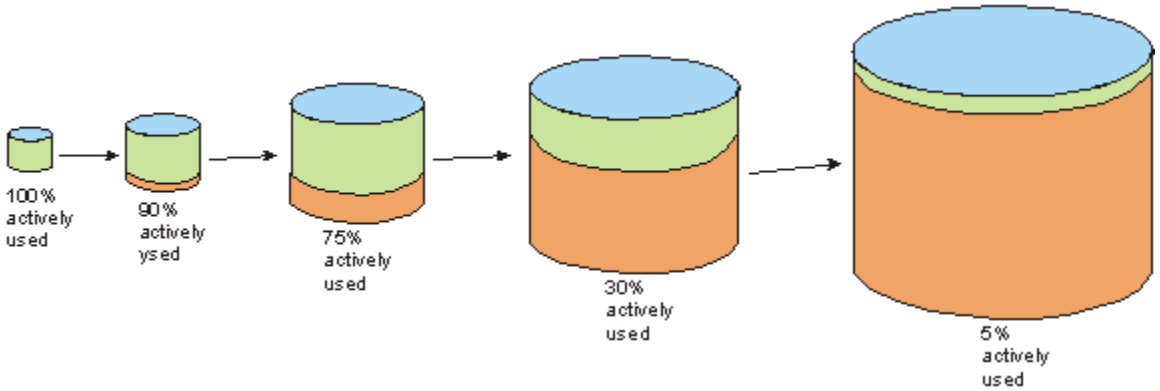
1. Challenge – Running Growing SAP BI systems
2. Solution – ILM and SAP BI Nearline Storage
3. Best Practice: Nearline Storage in a SAP BI Enterprise Data Warehousing (EDW) Architecture
4. Best Practice: Nearline Storage and Reporting
5. Case Study entega Service
6. Summary, Q&A

Data Access vs. Data Growth

- Typical Data Growth



- Typical Data Access vs Data Growth

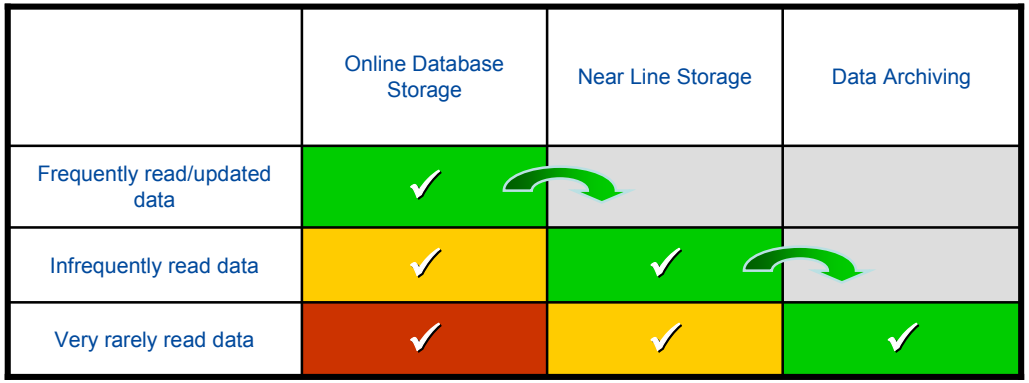


- As data grows in volume, the probability of access of data changes dramatically

The solution: SAP recommended ILM / Data Aging Strategy

ILM for SAP BI:

Split the data according to age or frequency of access into the following areas, moving data to the next level after a specified retention period

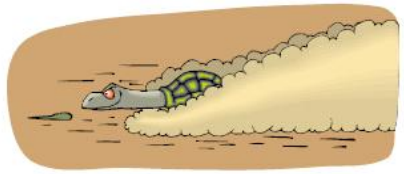


Source: SAP 2006

SAP has introduced an Information Life Cycle (ILM) architecture that enables SAP BI Data Warehouse Managers to:

- Keep a “skinny”, responsive relational database within SAP BI
- Keep *all* their data accessible and usable over time
- Satisfy analytic and legal requirements
- Control their budget
- Ensure system availability according SLA obligations

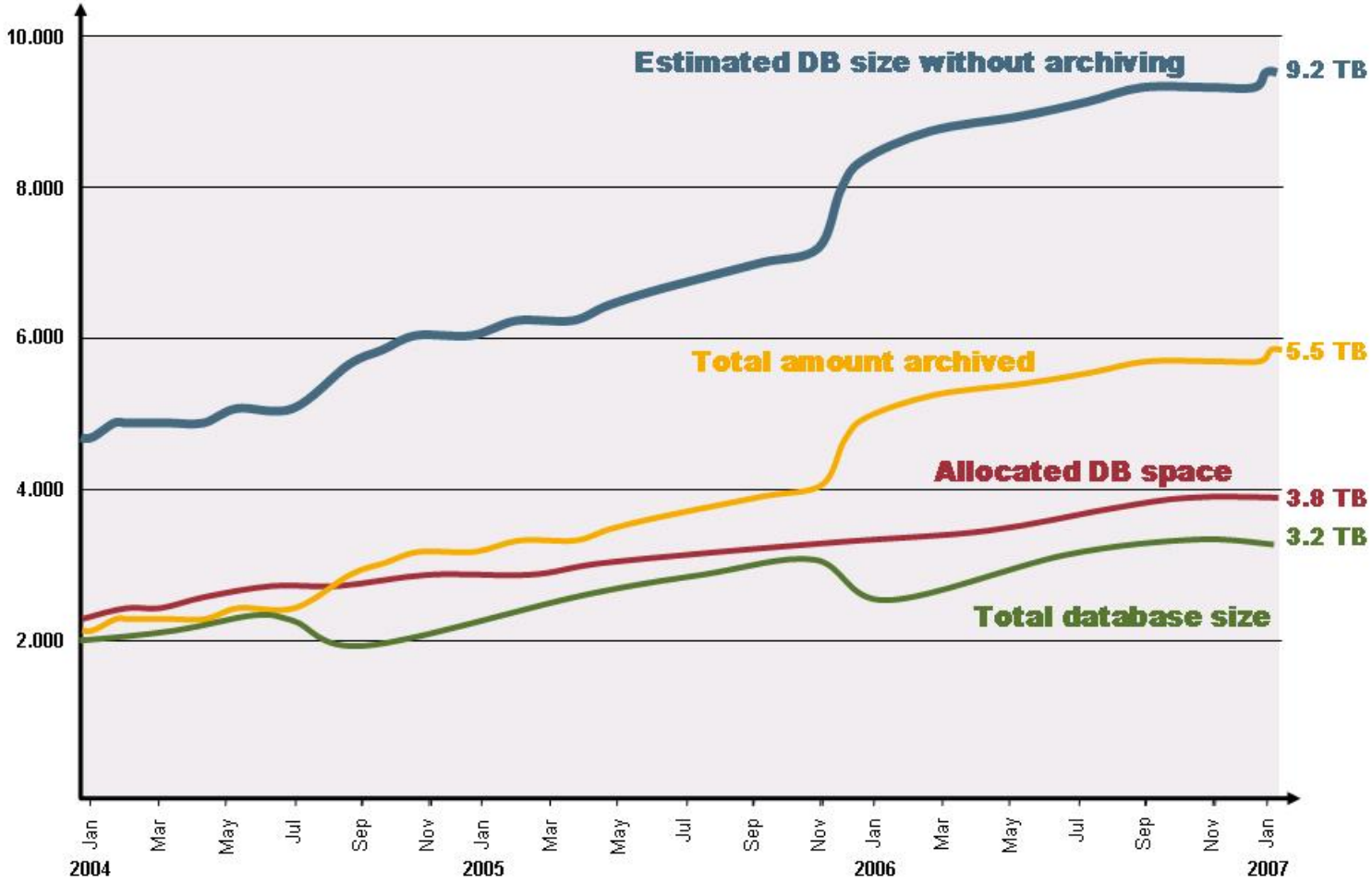
Motivation for a Data Aging Strategy: Benefits



- Performance
  - Faster data load times
  - Faster query execution times
- Cost
  - Storage costs: High availability, high IO disks, etc.
  - Resource and Administration overhead
    - System: CPU, Memory, etc.
    - Headcount: Number of full-time employees, etc.
  - Control of system growth

- Availability
  - Data availability – faster rollups, change runs, etc.
  - System availability – less downtime for backups, upgrades, etc.

Database On an ILM Diet



Deutschsprachige
SAP® Anwendergruppe

Improved SLAs

- Full Backup to tape takes 10 hours
  - 80 % Inactive – reduces to 2 hours!
- Full Recovery takes 15 hours
  - 80 % Inactive – reduces to 3 hours!

Reduced Storage Acquisition Costs

- 5 TB on high-end storage @ $50 per GB
  - Cost = $250,000
- If 80% of data is Inactive
  - Migrate 4TB data to low-end disk $5 per GB
  - Saving = $180,000

Reduced tape Cost

- 5 TB Full Backup to tape @ 1$/GB = $5,000
  - Full Backup every week with 6 months retention
  - Total Annual tape cost: $130,000
- Migrate 80% of inactive data
  - Reduces total annual tape cost to $26,000
- Saving of $104,000

RDBMS SLA Improvement

- Overall Query Performance
- Index Maintenance
- Database Reorganization
- Reduced Batch Windows
- Quicker Disaster/Recovery process
- Data Model Flexibility (more index, more summary)

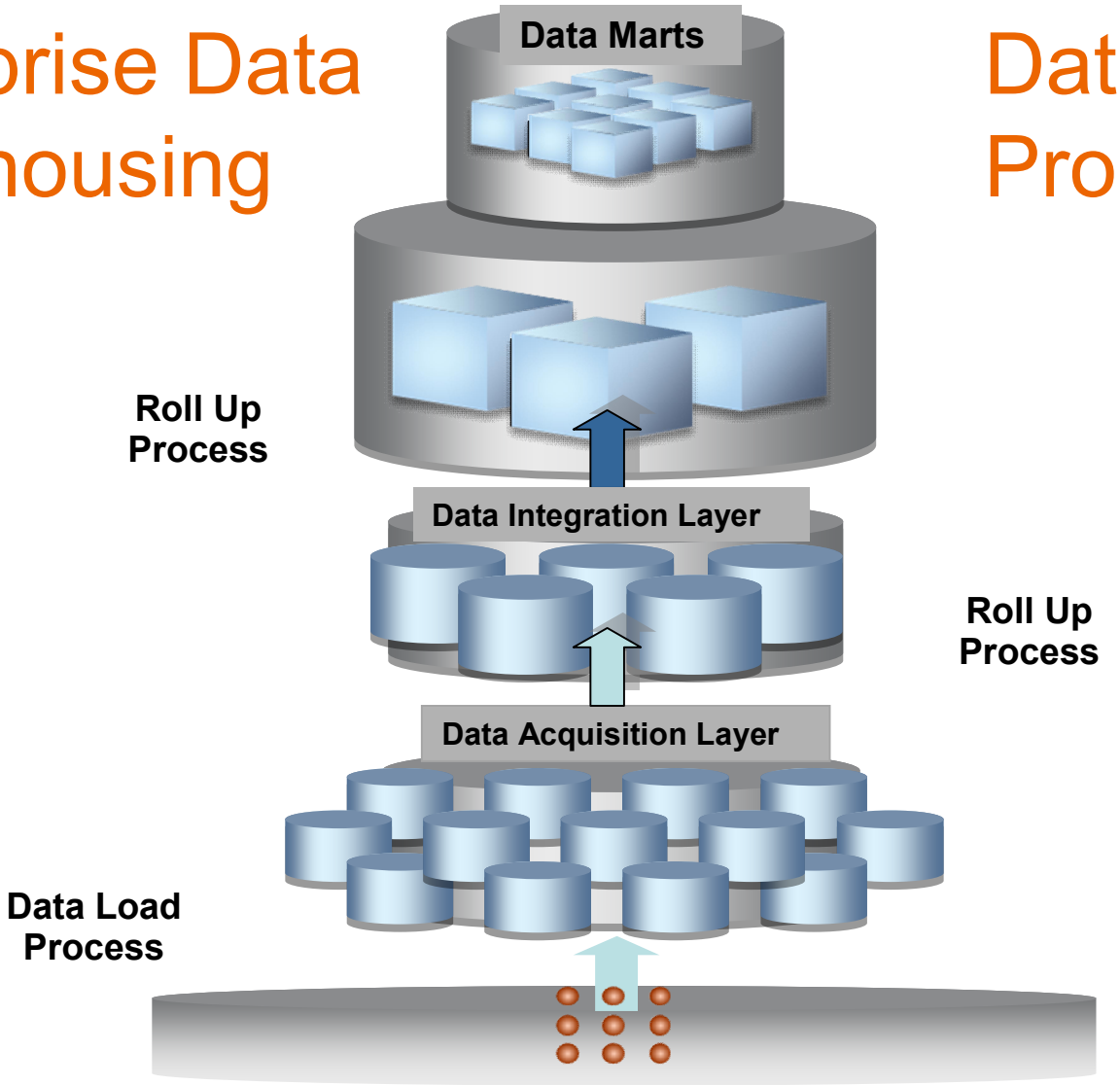
Fundamental ILM Strategy for BI - Benefits

- Increase Volume
  - Manage and use even larger amounts of information more effectively
  - Information available for any time frame for ad-hoc analyses and rebuilds
- Reduce Resource Consumption
  - Reduction of hardware costs for hard drive hardware on the BW side
  - Main memory and CPU as well as costs for system administration
- Increase Availability
  - Quicker, simpler software- and release management in BW
  - Reduced backup- and recovery times
  - Intelligent data access
- Optimize Performance
  - Speed up loading processes in SAP NetWeaver BI
  - SAP NetWeaver BI query response times in the dialog

Agenda

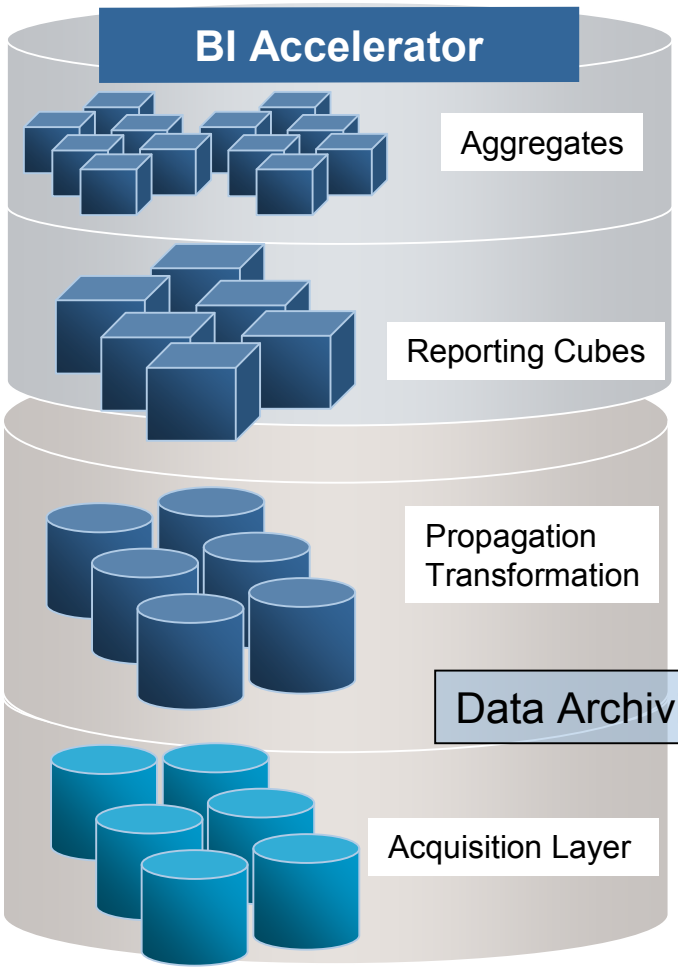
1. Challenge – Running Growing SAP BI systems
2. Solution – ILM and SAP BI Nearline Storage
3. Best Practice: Nearline Storage in a SAP BI Enterprise Data Warehousing (EDW) Architecture
4. Best Practice: Nearline Storage and Reporting
5. Case Study entega Service
6. Summary, Q&A

Enterprise Data Warehousing



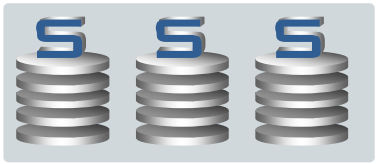
Data Processing

Lesson learned : Nearline on Detailed Data



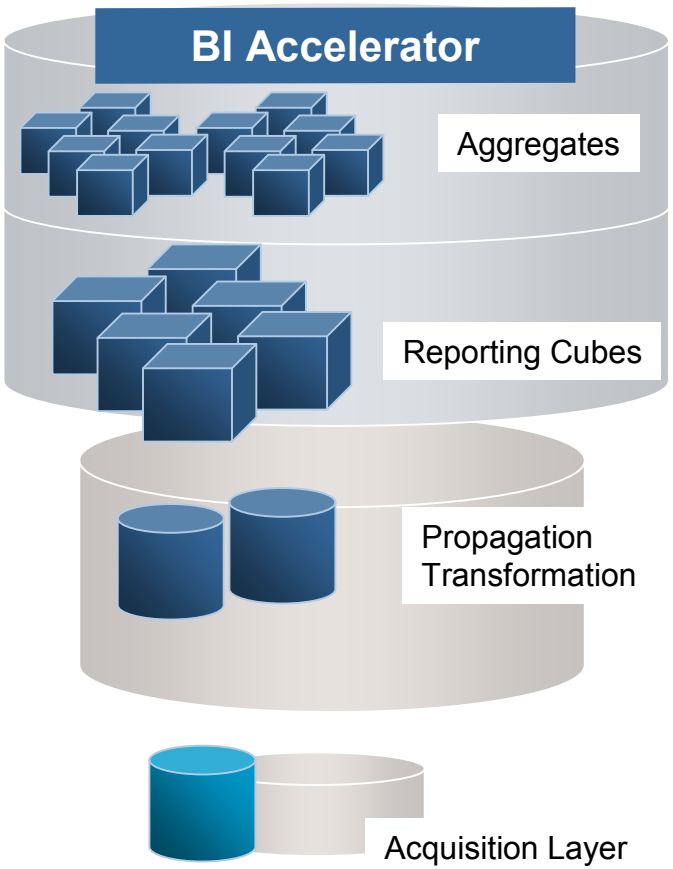
- Relieving SAP BI from detailed data
- Compressed by more than 85%
- Used as a „Corporate Memory“
  - Details in its “pure” form
  - Infrequently used detailed data
  - “Just-in-Case” data
  - Aged and historical data
  - Legacy data

Data Archiving Process (DAP)



Efficient Corporate Memory

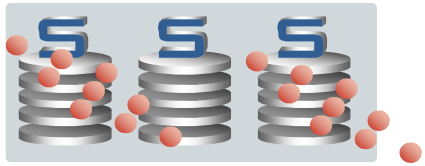
Usage of the „corporate memory“



Greater Flexibility in Responding to New Analytical Requirements

- deriving new InfoCubes or DSO‘s
- building new KPI‘s based on historical data

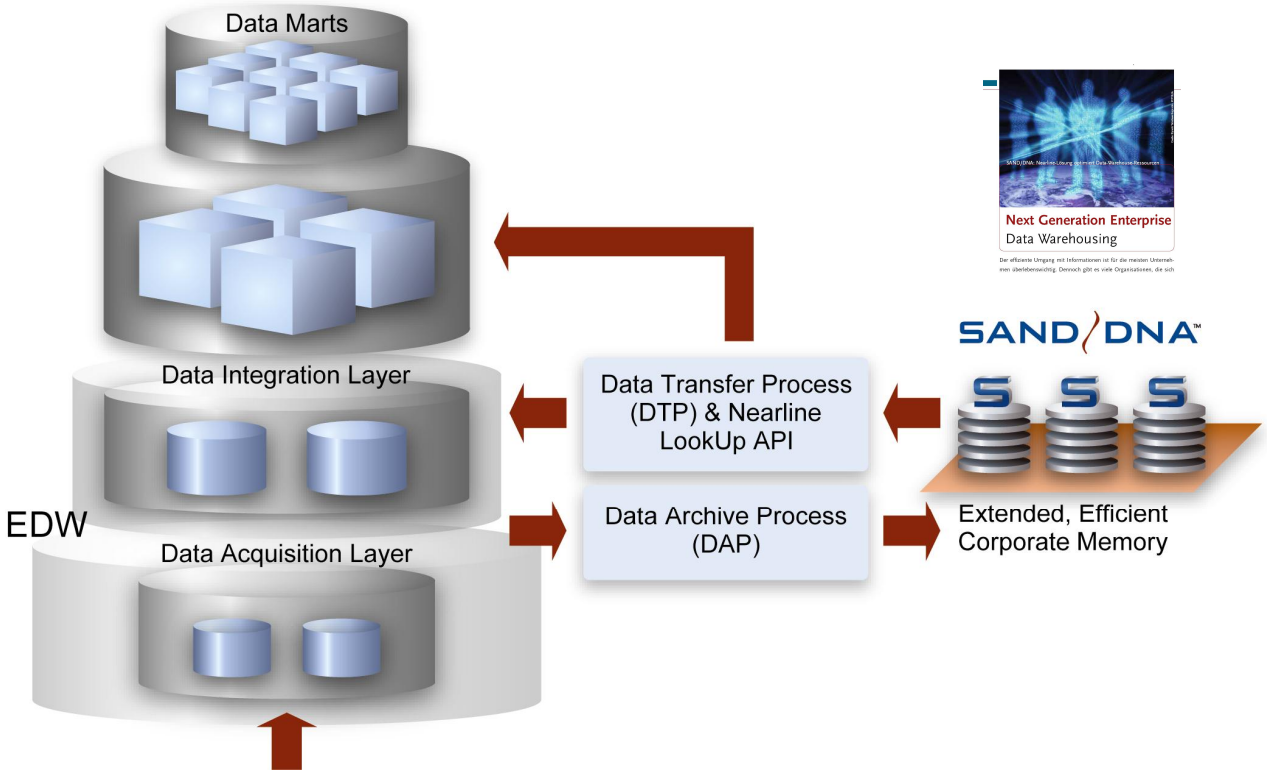
Data Transfer Process (DTP)
& Look Up API



Efficient Corporate Memory

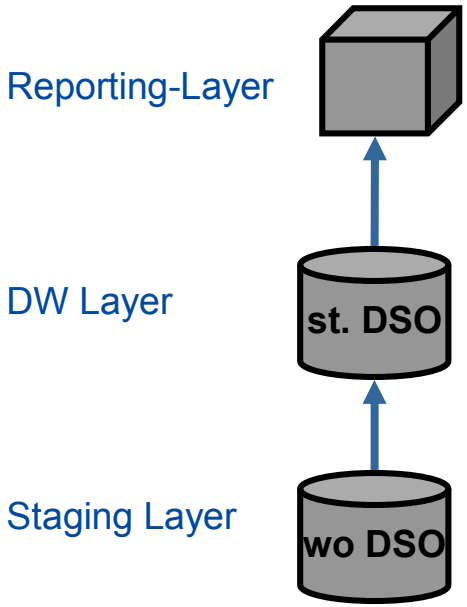
Next generation EDW -Layer

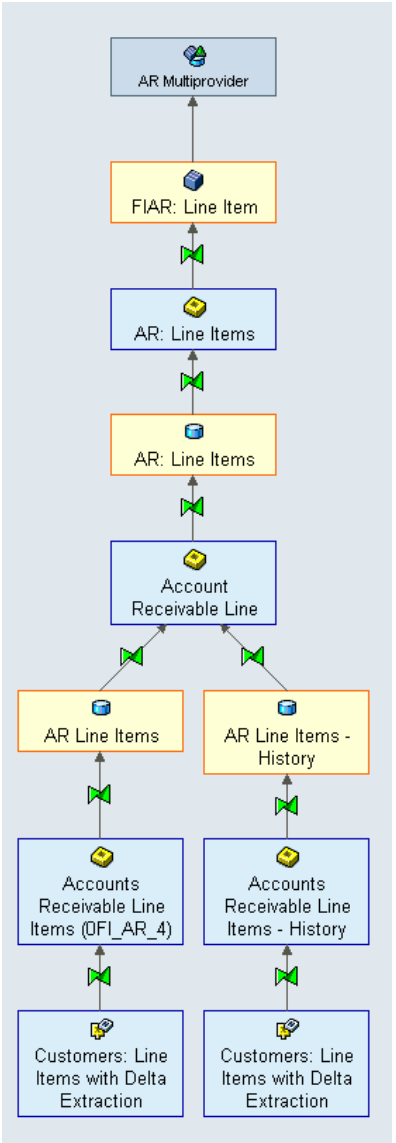
- storing detailed data according business and legal requirements
  ... and not according data management or costs constraints ...



Write-Optimized DSO Support

- Available with Enhancement Package 7.01

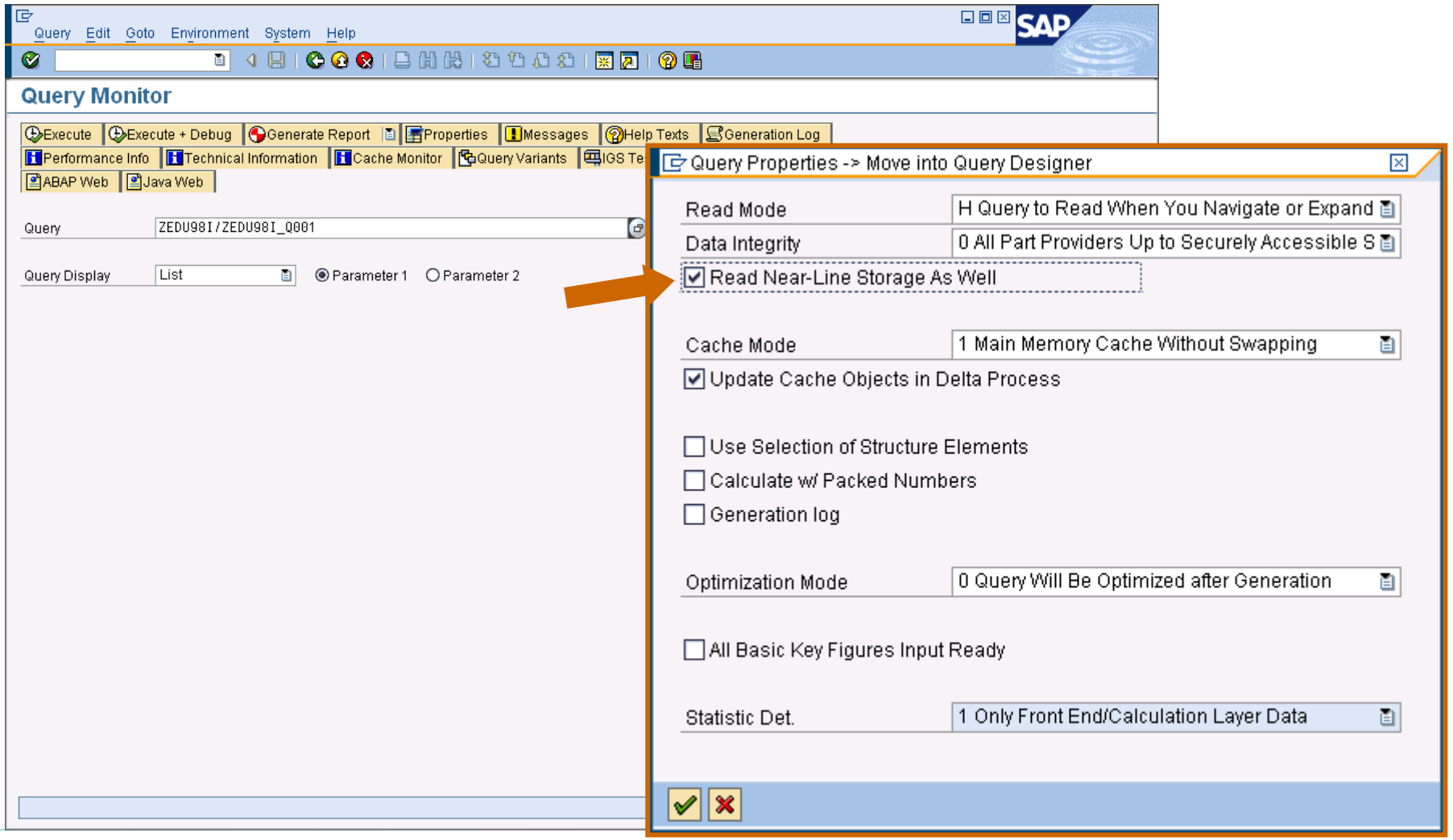




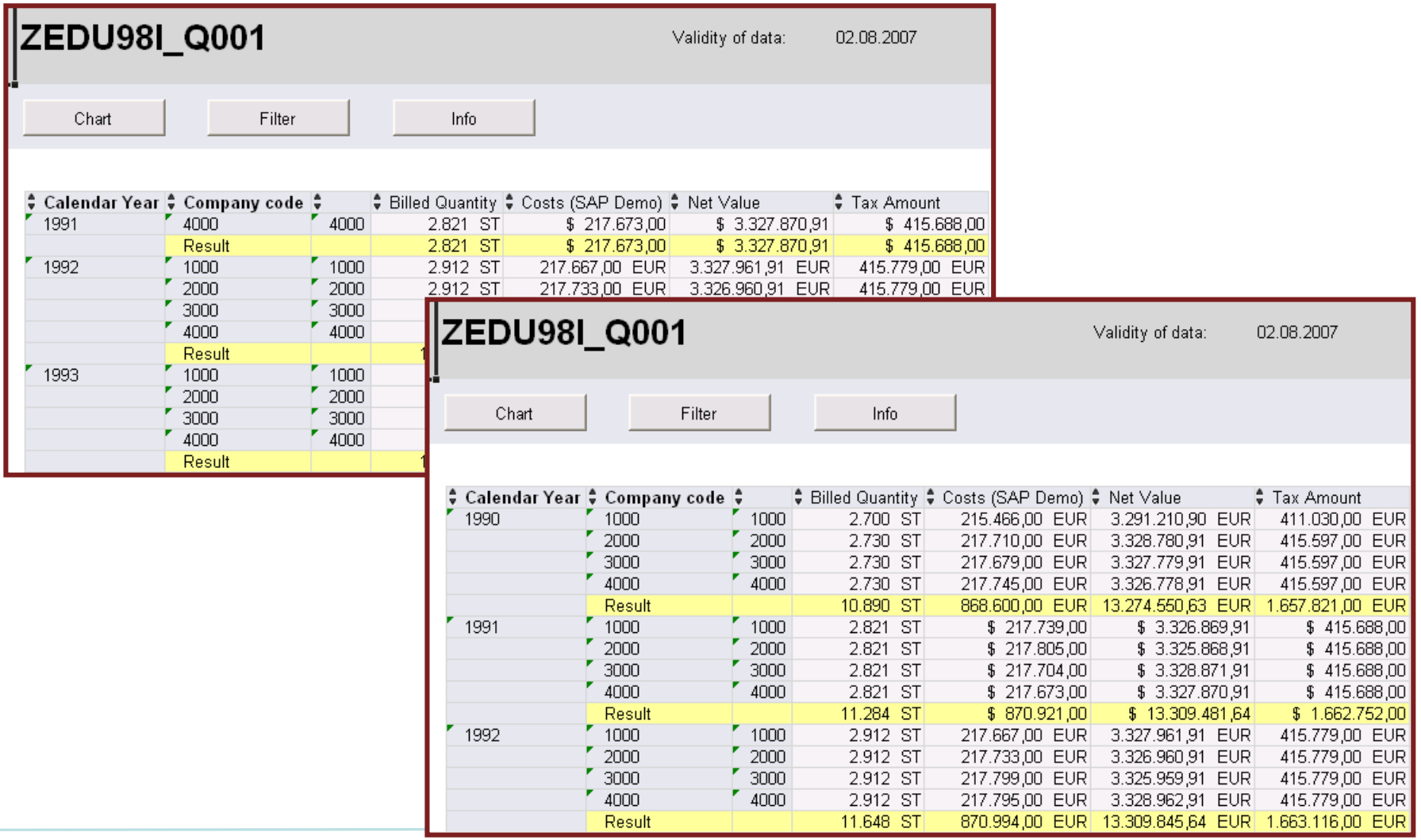
Agenda

1. Challenge – Running Growing SAP BI systems
2. Solution – ILM and SAP BI Nearline Storage
3. Best Practice: Nearline Storage in a SAP BI Enterprise Data Warehousing (EDW) Architecture
4. Best Practice: Nearline Storage and Reporting
5. Case Study entega Service
6. Summary, Q&A

Transparent Query Access



Query Result with and without NLS Flag



ZEDU98I_Q001 Validity of data: 02.08.2007

Chart | Filter | Info

Calendar Year	Company code		Billed Quantity	Costs (SAP Demo)	Net Value	Tax Amount
1991	4000	4000	2.821 ST	$ 217.673,00	$ 3.327.870,91	$ 415.688,00
	Result		2.821 ST	$ 217.673,00	$ 3.327.870,91	$ 415.688,00
1992	1000	1000	2.912 ST	217.667,00 EUR	3.327.961,91 EUR	415.779,00 EUR
	2000	2000	2.912 ST	217.733,00 EUR	3.326.960,91 EUR	415.779,00 EUR
	3000	3000				
	4000	4000				
	Result		1			
1993	1000	1000				
	2000	2000				
	3000	3000				
	4000	4000				
	Result		1			

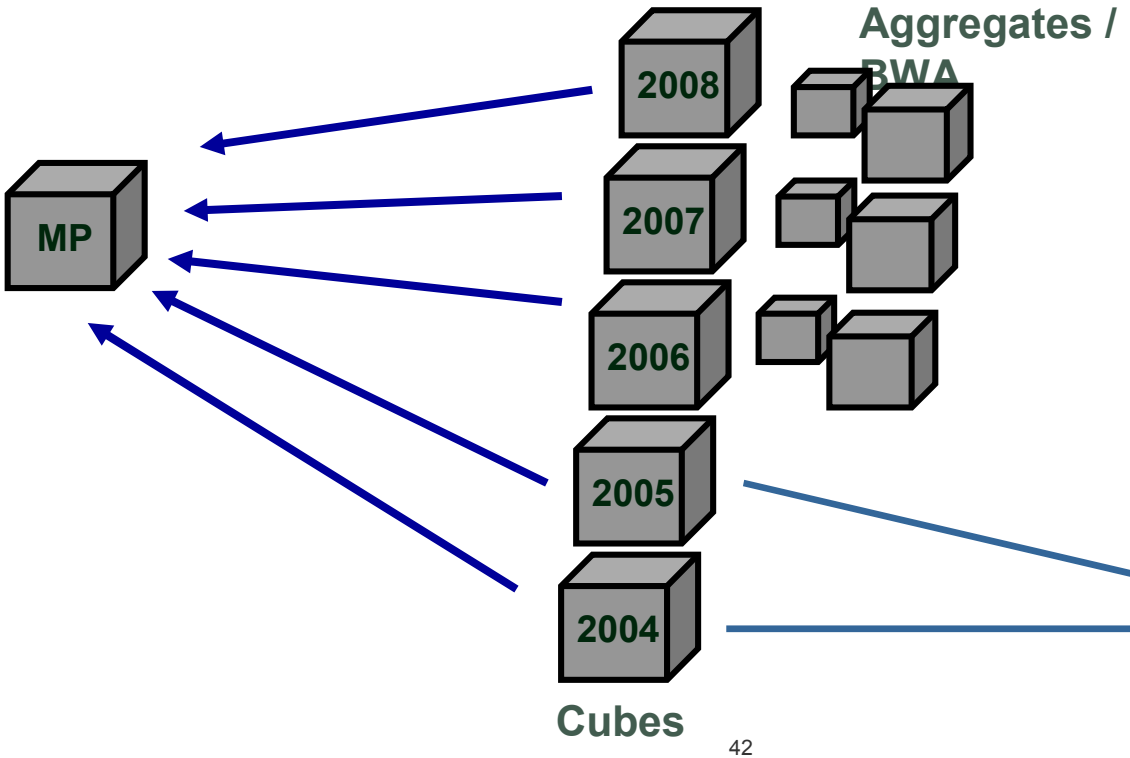
ZEDU98I_Q001 Validity of data: 02.08.2007

Chart | Filter | Info

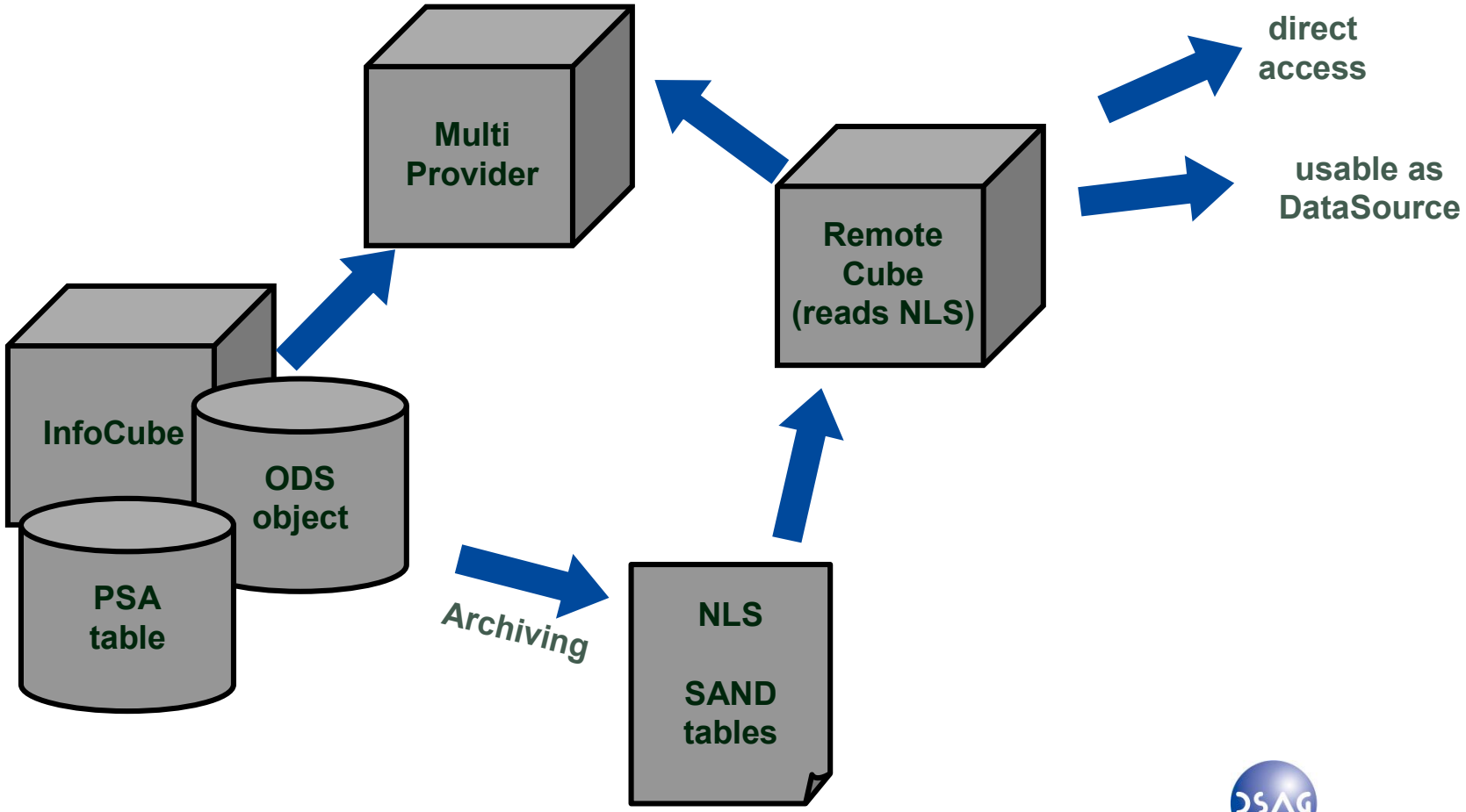
Calendar Year	Company code		Billed Quantity	Costs (SAP Demo)	Net Value	Tax Amount
1990	1000	1000	2.700 ST	215.466,00 EUR	3.291.210,90 EUR	411.030,00 EUR
	2000	2000	2.730 ST	217.710,00 EUR	3.328.780,91 EUR	415.597,00 EUR
	3000	3000	2.730 ST	217.679,00 EUR	3.327.779,91 EUR	415.597,00 EUR
	4000	4000	2.730 ST	217.745,00 EUR	3.326.778,91 EUR	415.597,00 EUR
	Result		10.890 ST	868.600,00 EUR	13.274.550,63 EUR	1.657.821,00 EUR
1991	1000	1000	2.821 ST	$ 217.739,00	$ 3.326.869,91	$ 415.688,00
	2000	2000	2.821 ST	$ 217.805,00	$ 3.325.868,91	$ 415.688,00
	3000	3000	2.821 ST	$ 217.704,00	$ 3.328.871,91	$ 415.688,00
	4000	4000	2.821 ST	$ 217.673,00	$ 3.327.870,91	$ 415.688,00
	Result		11.284 ST	$ 870.921,00	$ 13.309.481,64	$ 1.662.752,00
1992	1000	1000	2.912 ST	217.667,00 EUR	3.327.961,91 EUR	415.779,00 EUR
	2000	2000	2.912 ST	217.733,00 EUR	3.326.960,91 EUR	415.779,00 EUR
	3000	3000	2.912 ST	217.799,00 EUR	3.325.959,91 EUR	415.779,00 EUR
	4000	4000	2.912 ST	217.795,00 EUR	3.328.962,91 EUR	415.779,00 EUR
	Result		11.648 ST	870.994,00 EUR	13.309.845,64 EUR	1.663.116,00 EUR

Multi-Provider Support

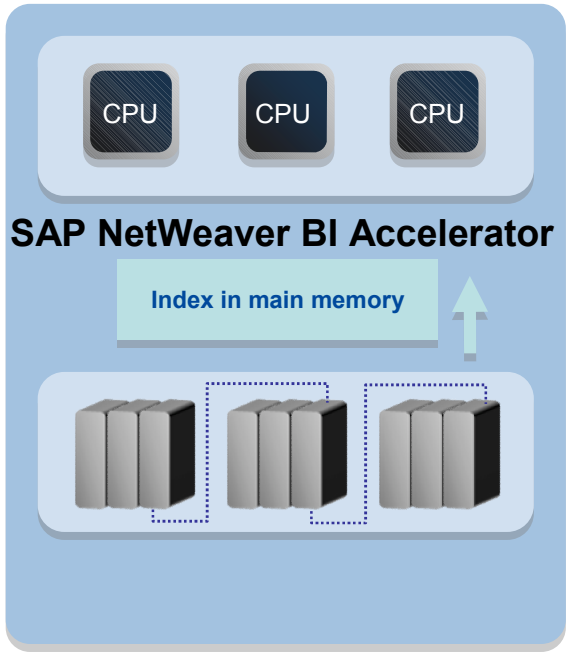
- Complete Multi-Provider support with NW 7.20
- Especially a problem if logical partitioning is used
- Best Practice Solution: Using a Virtual Provider



Best Practice Solution: Virtual Provider

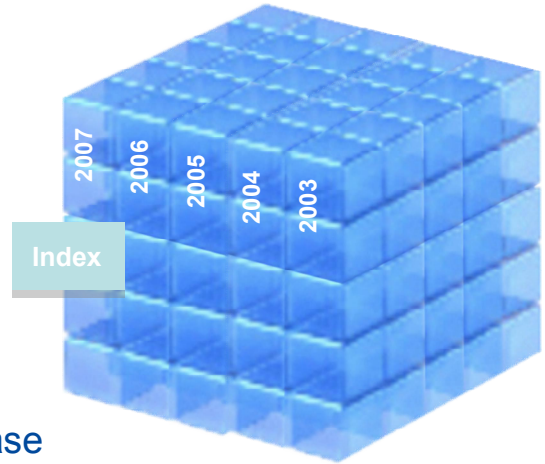


Optimization of BWA by Nearline Storage



- ➔ **Only actual important data is indexed in BWA**
- ➔ **Optimal usage of Resources like CPU**

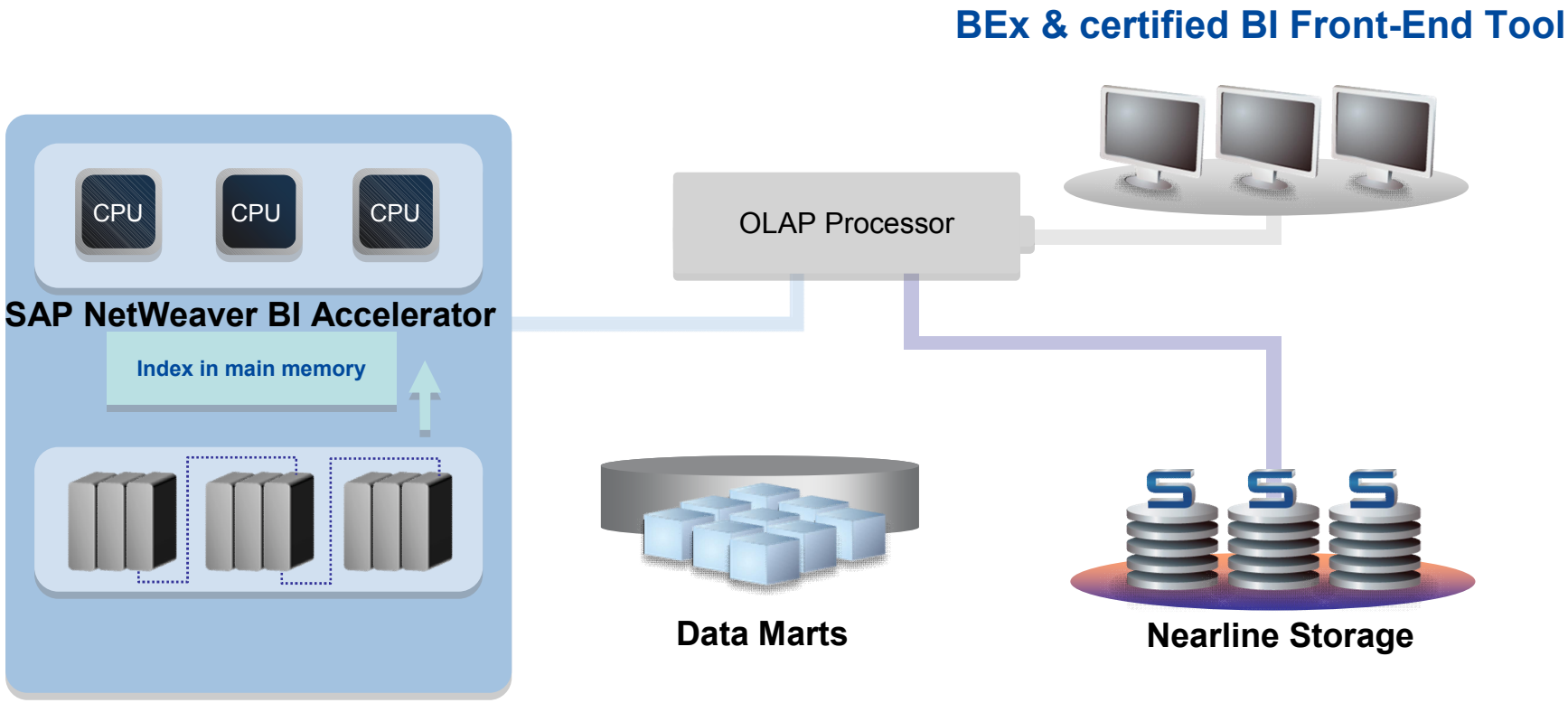
Nearline Storage



- ➔ InfoCubes partially indexed in BWA
- ➔ Data **remains** in the relational Database

- ➔ Archiving a part of the InfoCube via a DAP
- ➔ Deletetion of the corresponding data in the relational database

Transparent Access



Agenda

1. Challenge – Running Growing SAP BI systems
2. Solution – ILM and SAP BI Nearline Storage
3. Best Practice: Nearline Storage in a SAP BI Enterprise Data Warehousing (EDW) Architecture
4. Best Practice: Nearline Storage and Reporting
5. Case Study entega Service
6. Summary, Q&A

ENTEGA Service

- Founded June 2002 as service provider for utility companies with locations in Mainz and Darmstadt
- Affiliate of HEAG Südhessische Energie AG (HSE) and Stadtwerke Mainz AG
- Revenue ~50 Mio. Euro
- 230 employees
- 900.000 customers in service areas for billing, payments and accounts receivable management
- Operating and hosting IT systems and applications with overall ~ 3.250 Users

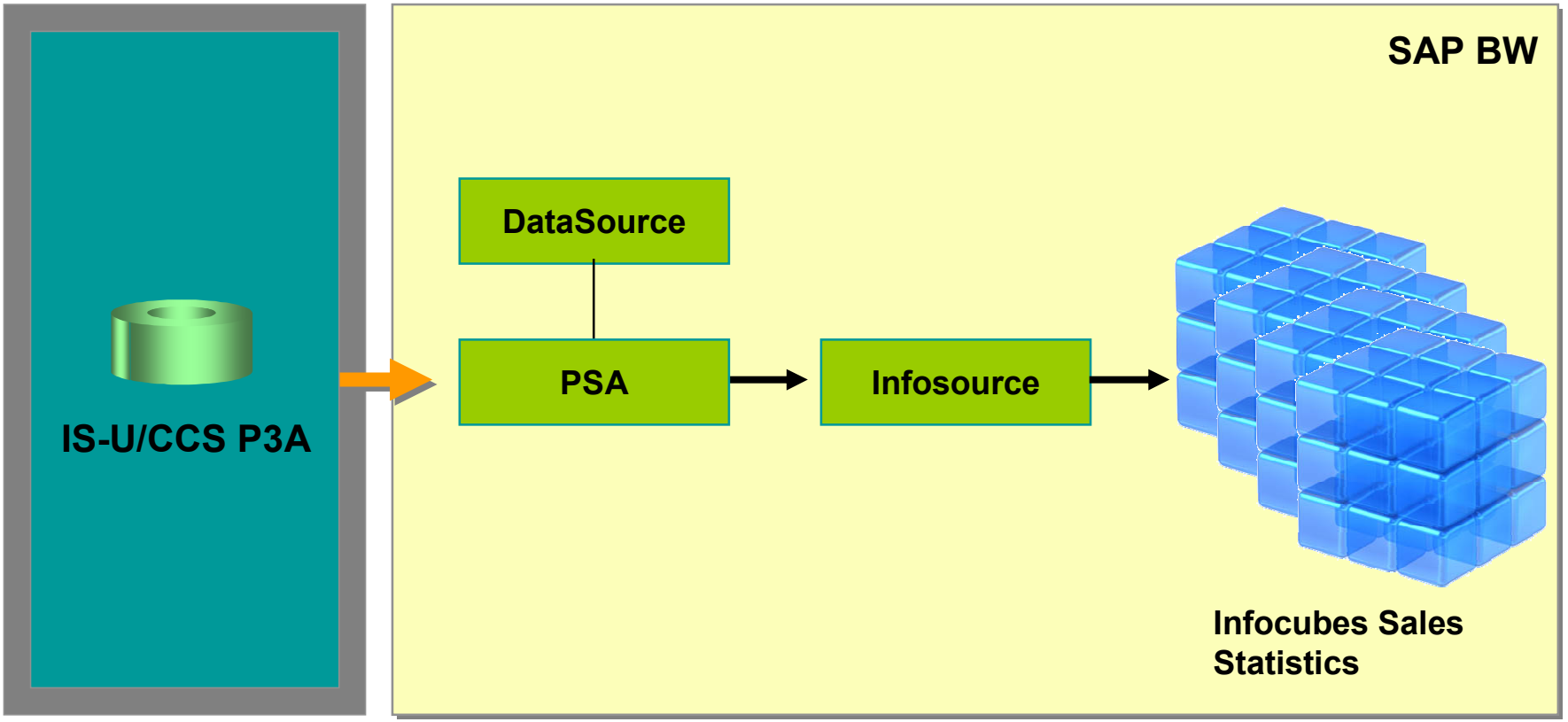
SAND Nearline Solution at ENTEGA

- **Moving less frequently used PSA Data, historical Cubes to SAND/DNA**

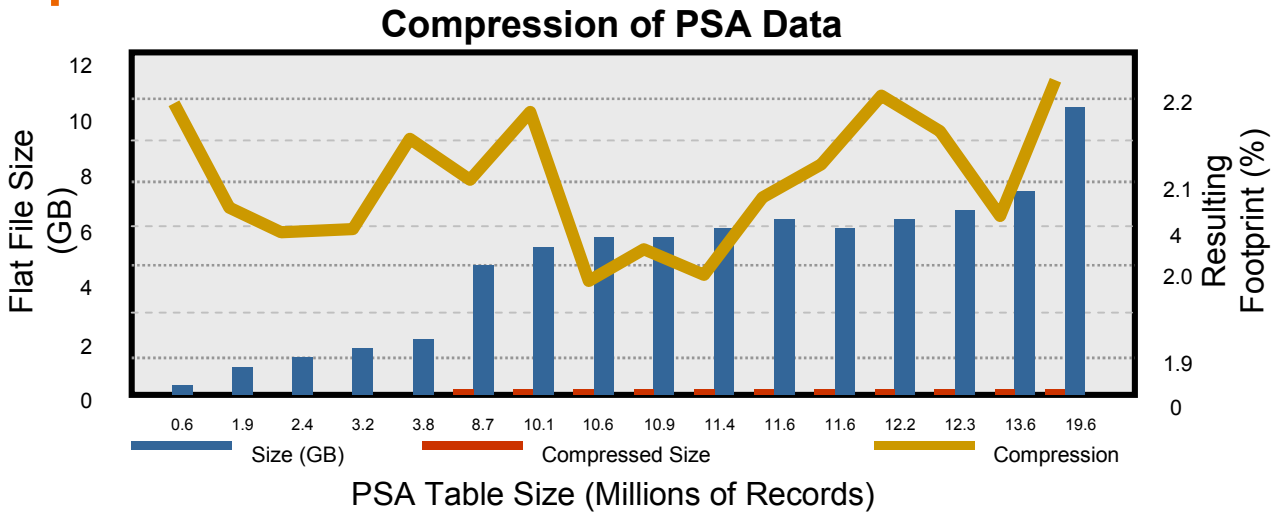
➔ Reducing data volumes and TCO

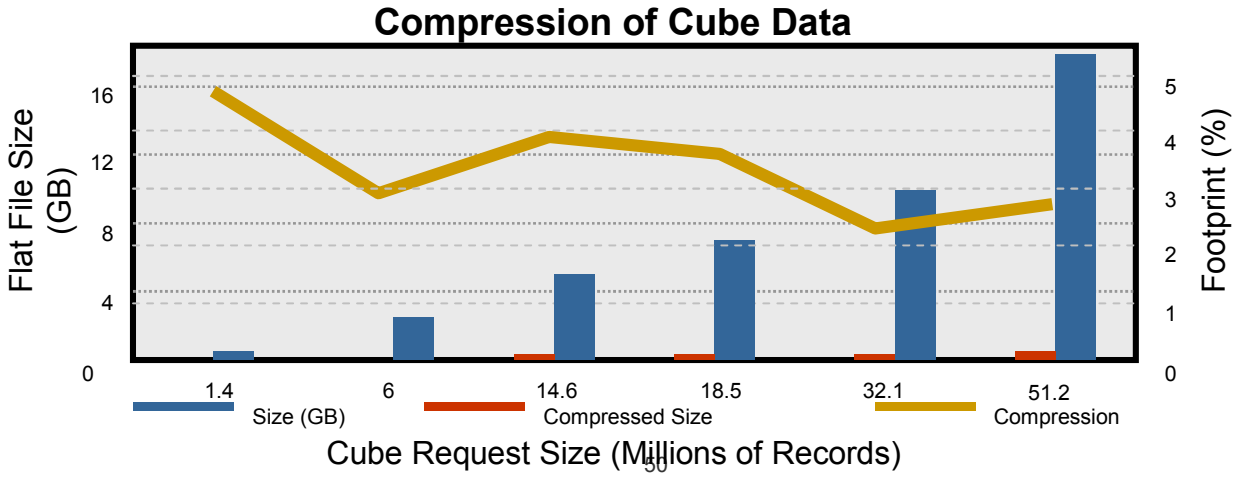
➔ Minimize administrative requirements

Old BW 3.5 Structure Sales Statistics



Compression rates



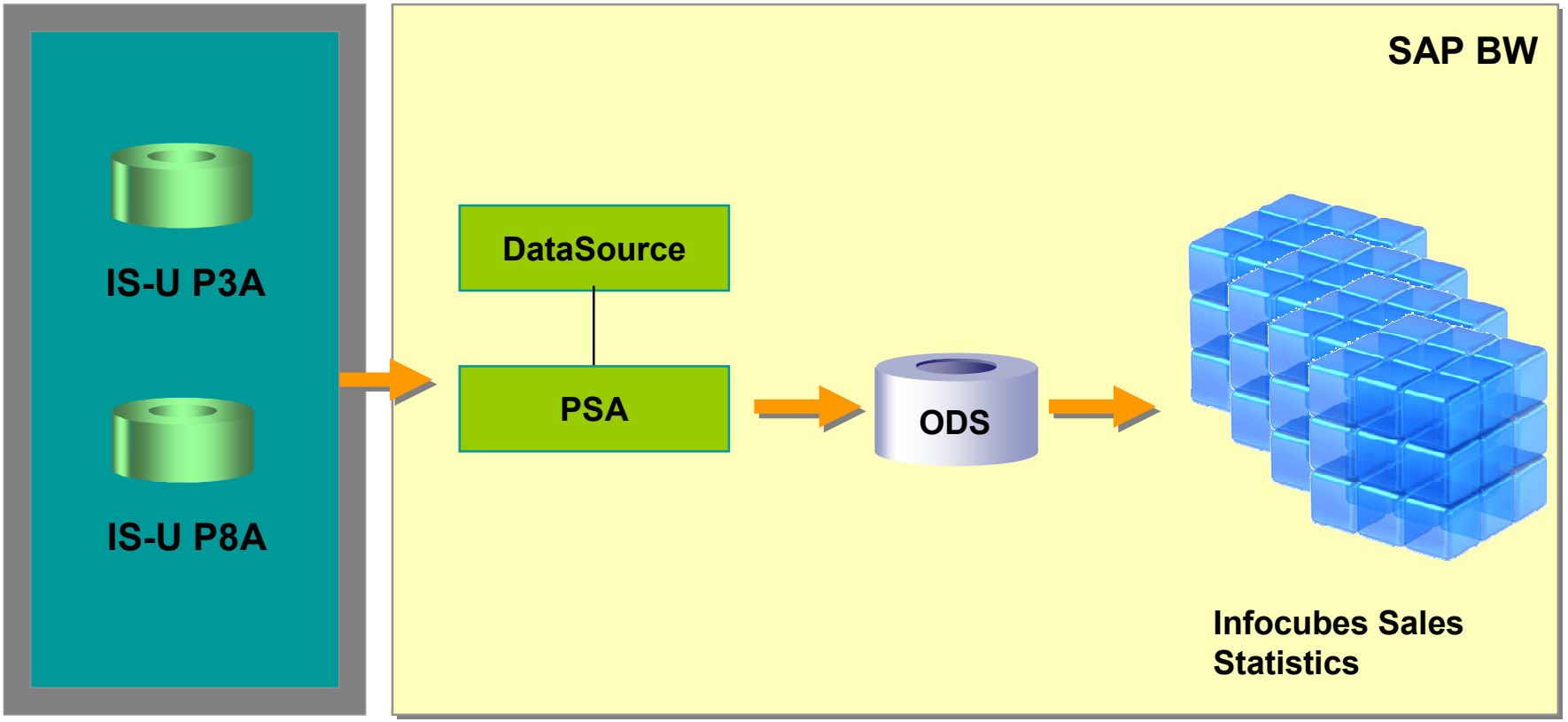


Deutschsprachige SAP® Anwendergruppe

ENTEGA Service – NLS History

- Started BW Archiving to SAND/DNA nearline in 2006
  - Release BW3
  - Archiving of PSA tables and info cubes
- Migration to BI7 in 2008
  - PSA archiving remains the same
  - Cube archive requests migrated to BU7 requests
- Implementing EHP1 in 2009
  - Proper granular staging layer in write-optimized DSOs
  - No more PSA archiving needed

New BI 7 Structure Sales Statistics



Agenda

1. Challenge – Running Growing SAP BI systems
2. Solution – ILM and SAP BI Nearline Storage
3. Best Practice: Nearline Storage in a SAP BI Enterprise Data Warehousing (EDW) Architecture
4. Best Practice: Nearline Storage and Reporting
5. Case Study entega Service
6. Summary, Q&A

Take Away / Conclusion

- You can lower your TCO and improve operational efficiencies with Nearline
- You can keep more data at your fingertips to respond to changing business needs, trend analysis, and regulatory compliance
- You can stop throwing away your data or choosing what data to keep as you upgrade - keep it all!
- Move your infrequently used data to nearline
- Implement a proper Corporate Memory in your Nearline Repository and react appropriately and quickly to unknown needs (anticipate the unknown)
- Have a nearline strategy so you can react quickly to audits or new business directions and avoid penalties, lost revenue and customer dissatisfaction
- Have a SAP NetWeaver ILM Nearline strategy for BI in place before you experience performance or maintenance issues

“Save Yourself Time…”

1 The „healthy“ system

Don’t start thinking about data archiving when your system is about to crash!

2 Timely Planning

Proactive action to prepare sustainable system performance

3 Interdisciplinary Process

Data archiving requires a large amount of coordination between IT- and those responsible for applications.

Additional Resources

- Best Practice Paper
- HowTo Papers
- White Paper
- Case Studies
- Brochures

Available at www.sand.com and at www.sandtechnology.de

Check also the Marketplace and SDN for additional information (ILM and EDW)

Your Turn: Questions?

Jens Schwarz
ENTEGA Service GmbH
Jens.schwarz(at)entega-service.de

Dr. Michael Hahne
Vice President
SAND Technology
michael.hahne(at)sand.com
+49 671 9203662