Achieving more together!

German-Speaking SAP® User Group e.V.

Information Lifecycle Management

Optimization of a Global Enterprise Data Warehouse Architecture

Dr. Michael Hahne
SAND Technology

5th SAP GLOBALIZATION SYMPOSIUM

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. 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. 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

The Challenge

"In the compliance age, the answer lies in any technology which meets all three of these criteria:

- Large Stored data volume
- Quick Availability
- Fast Query Response Time

and can do so within the seven-figure cost range"

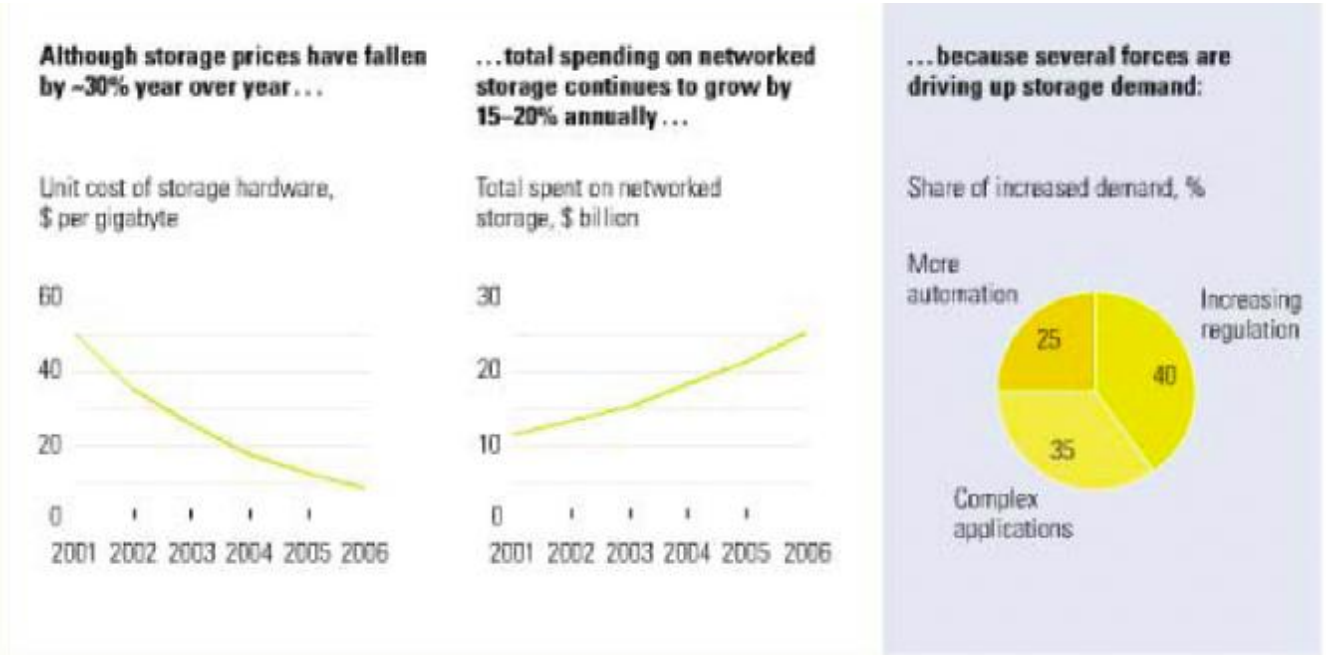
SOX Journal 2005

Challenges....

- "We Can't Meet our Batch Windows"
  - Monthly / Daily Preparation of Revised KPI's & Reporting
  - Backing Up Data
  - Rebuilding Warehouse Data
- "Our Costs are Spiraling"
  - Storage Hardware / Replication
  - Processors to Handle Storage
  - Floor Space / Power / Air Conditioning
  - Data Administration
- "The Targets Keep Changing"
  - New Business Directions
  - Special Project Demands
  - External / Internal Audit Responsiveness

Total Corporate Spending on Storage …

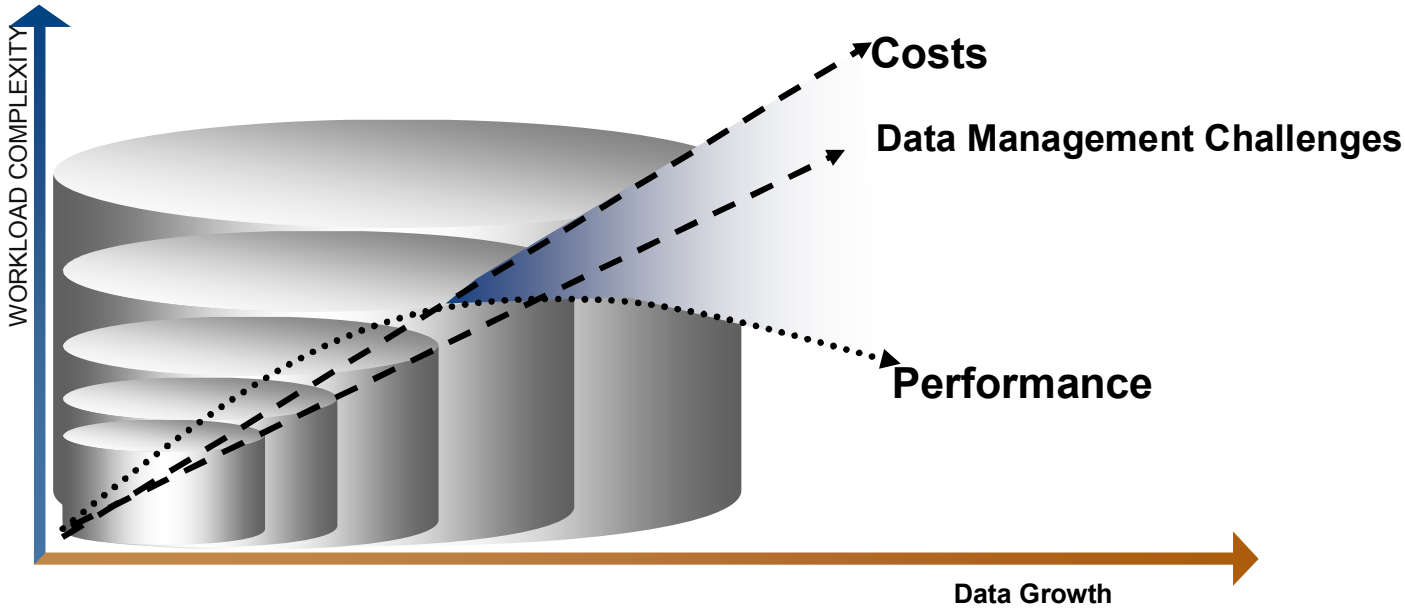
… (disk drives, tape systems, specialized network gear, and the people and software to manage them) grows by 15 to 20 percent every year, even though the unit cost of storage drops by about 30 percent annually



Deutschsprachige SAP® Anwendergruppe

Result: Missed Service Levels

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



WHAT ARE THE OPTIONS????

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.

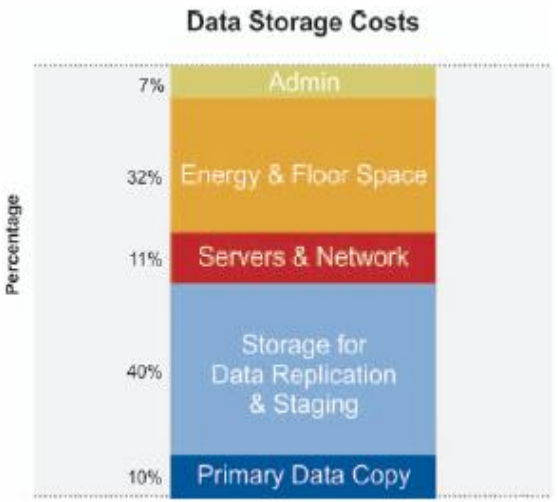
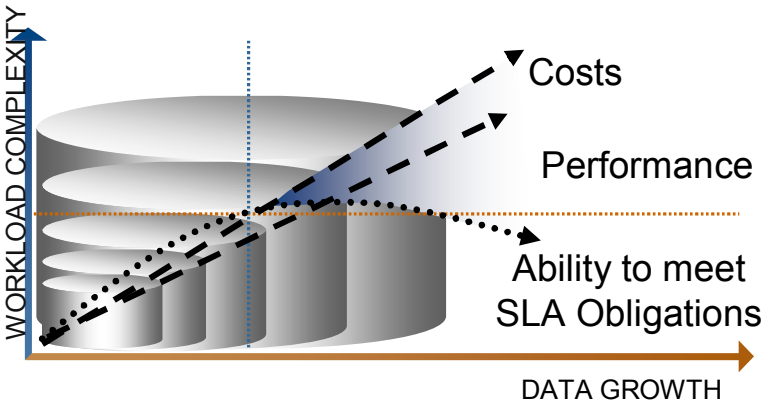


Figure 1: Breakdown of Data Storage Costs

What Companies are Facing Today…

- Unprecedented growth in data –
  - Driven by business growth - more transactions, more customers, more everything
  - Driven by need to keep new types of data – IM files, RFID
  - Driven by user demands – for more in-depth and on-demand analysis/reporting
  - Driven by regulatory mandates - e.g. SOX, Basel II compliance
  - Driven by reluctance to purge data – "just in case"

- ➔ Data Warehouse Management is challenged to meet SLA obligations
  - Traditional solution: Either invest heavily in hardware and consulting, or exclude data from the warehouse
  - Compromising analytical requirements arising from increasingly complex business processes
  - Disturbing the decision-making process
  - Disregarding regulatory obligations



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. Summary, Q&A

What is Data Aging?

- Data warehousing is a very powerful concept for creating a unified and consistent view of the business
- In a data warehousing environment, it is typical that:
  - Data is amassed and analyzed at an increasing rate
  - As time progresses, companies face the dilemma of storing more and more historical data
  - Over time, data tends to lose its “day-to-day” relevance and is therefore accessed less frequently
  - The costs associated with maintaining historical data are high
- Data aging is a strategy for managing data over time, balancing data access requirements with TCO
- Each data aging strategy is uniquely determined by the customer’s data and the business value of accessing the data
- Need: solution that provides alternatives for the typical “cost vs. business data availability” conundrum

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

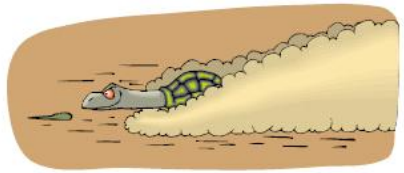
	Online Database Storage	Near Line Storage	Data Archiving
Frequently read/updated data	✓		
Infrequently read data	✓	✓	
Very rarely read data	✓	✓	✓

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

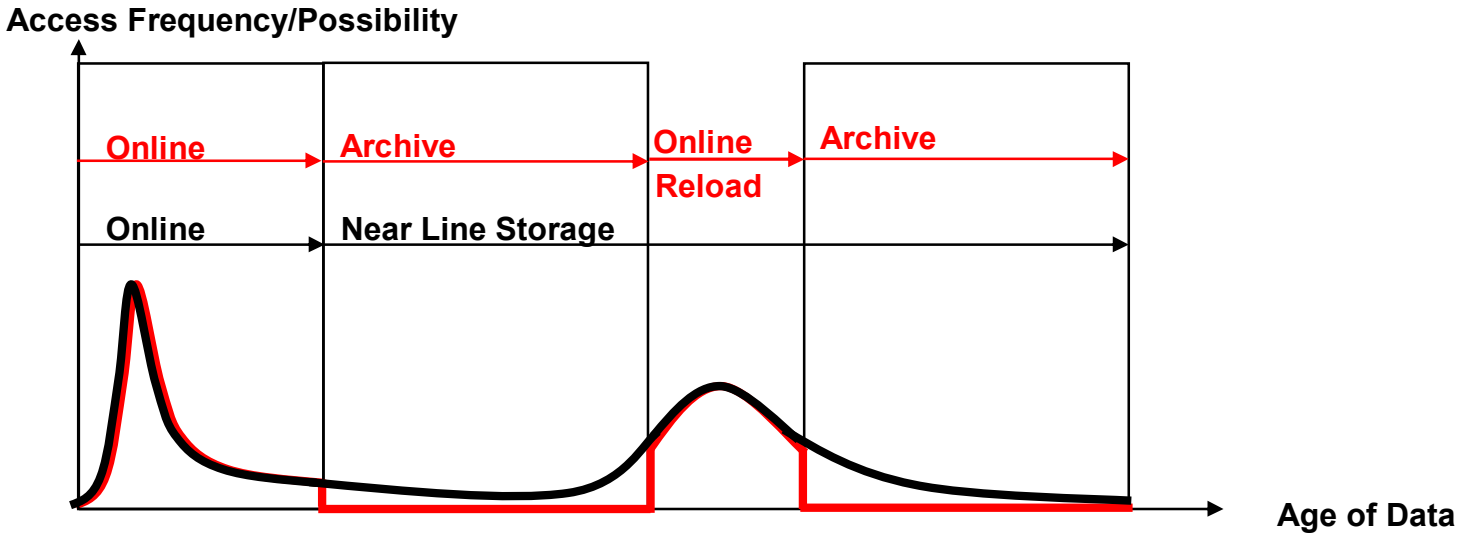
Motivation for a Data Aging Strategy: Benefits



- 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.

Classic Archive vs. Nearline



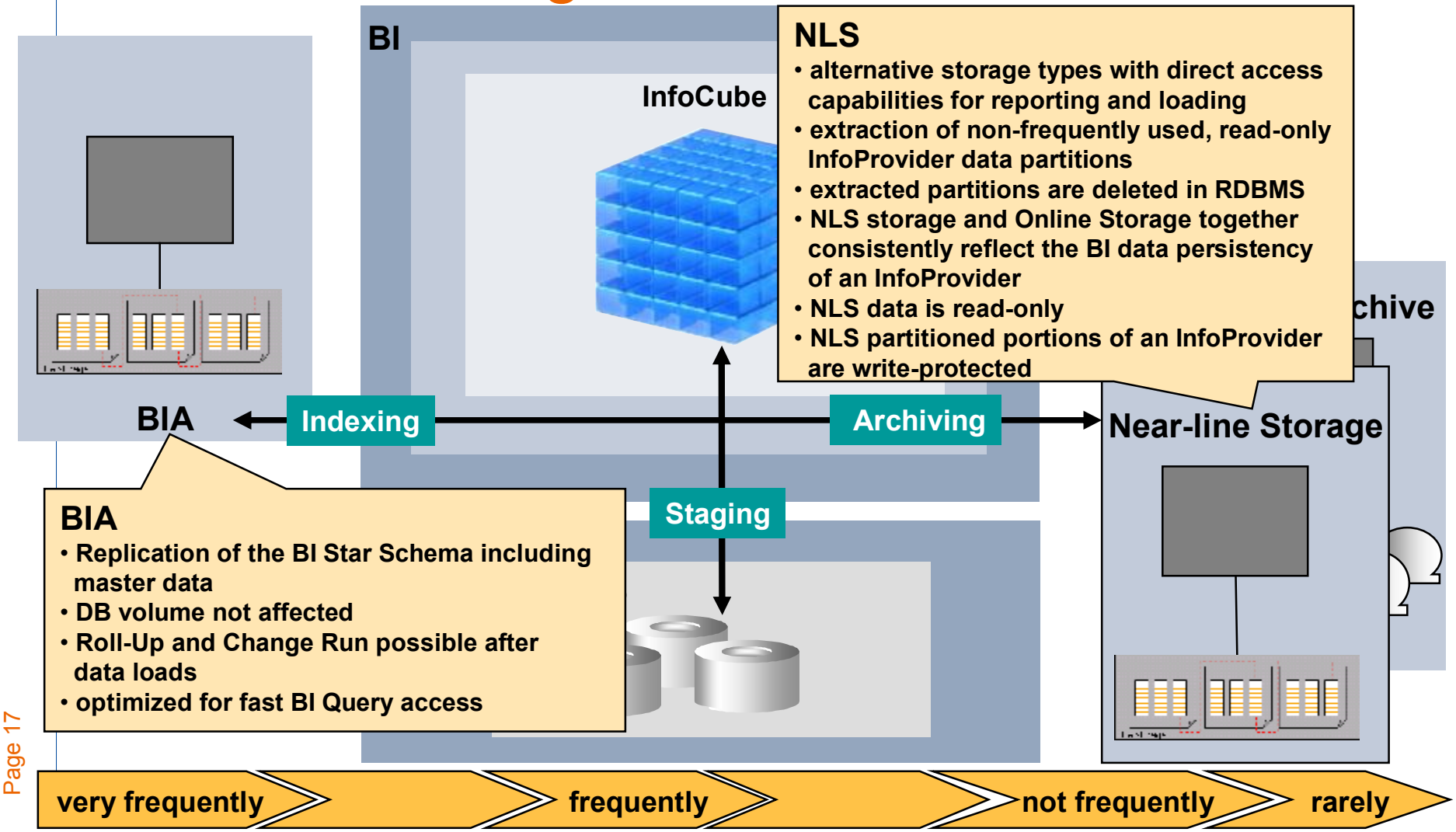
Archiving (SAP BW 3.X)

- ADK-based (Archive Development Kit) archiving solution for InfoCubes and ODS objects
- Cost-reduction due to storing data on alternative storage media
- Archived data must be reloaded into the SAP NetWeaver BI database for analysis purposes

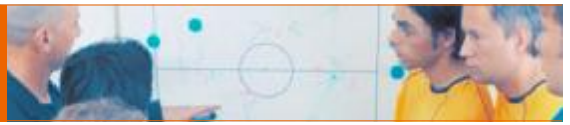
NLS (SAP NW 2004s BI)

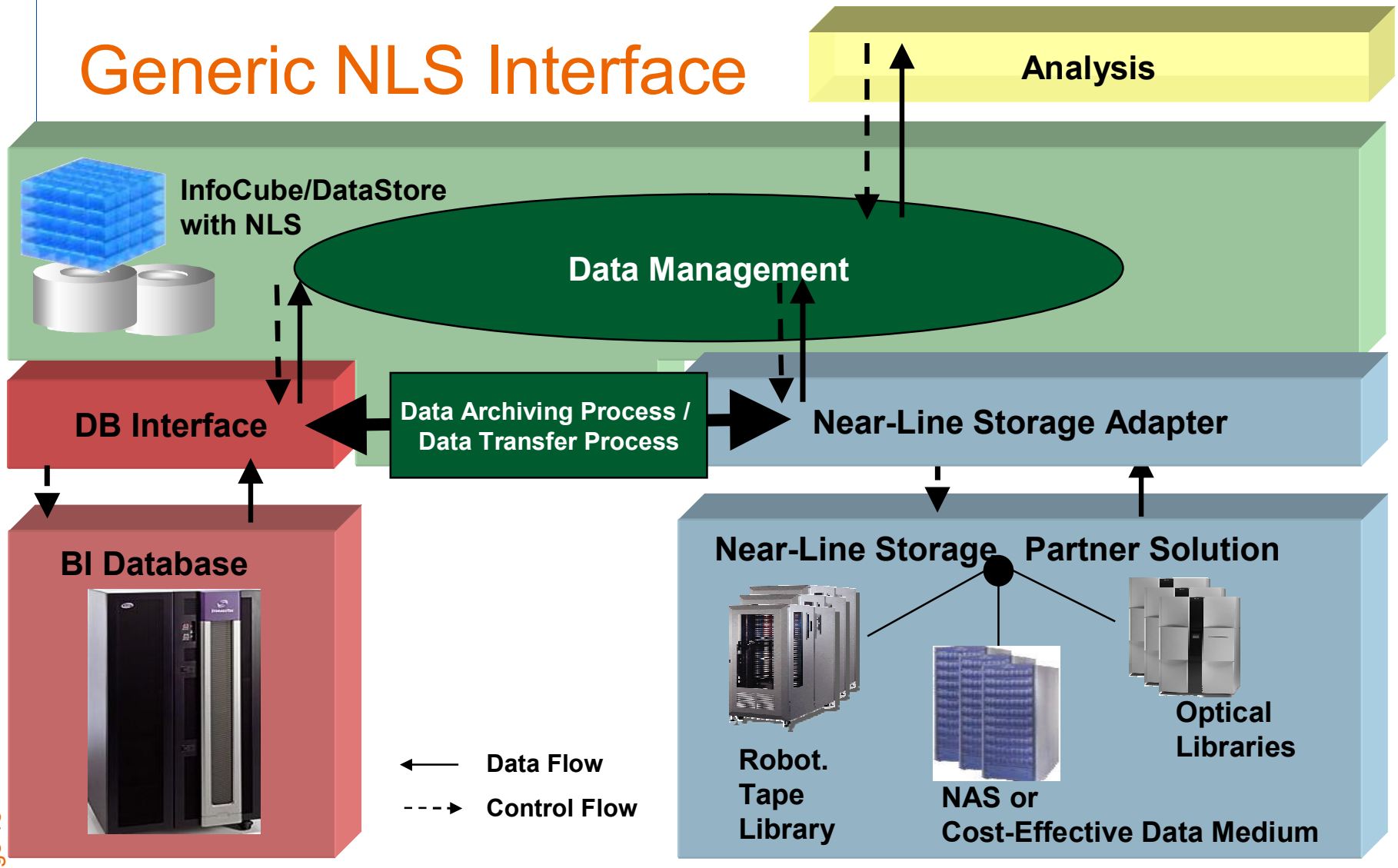
- SAP NetWeaver BI analyses have direct access to NLS data
- Availability of historic data while reducing costs
- Reloading of data into the InfoCube or DataStore Object only necessary in exceptional cases

Near-line Storage and BI Accelerator



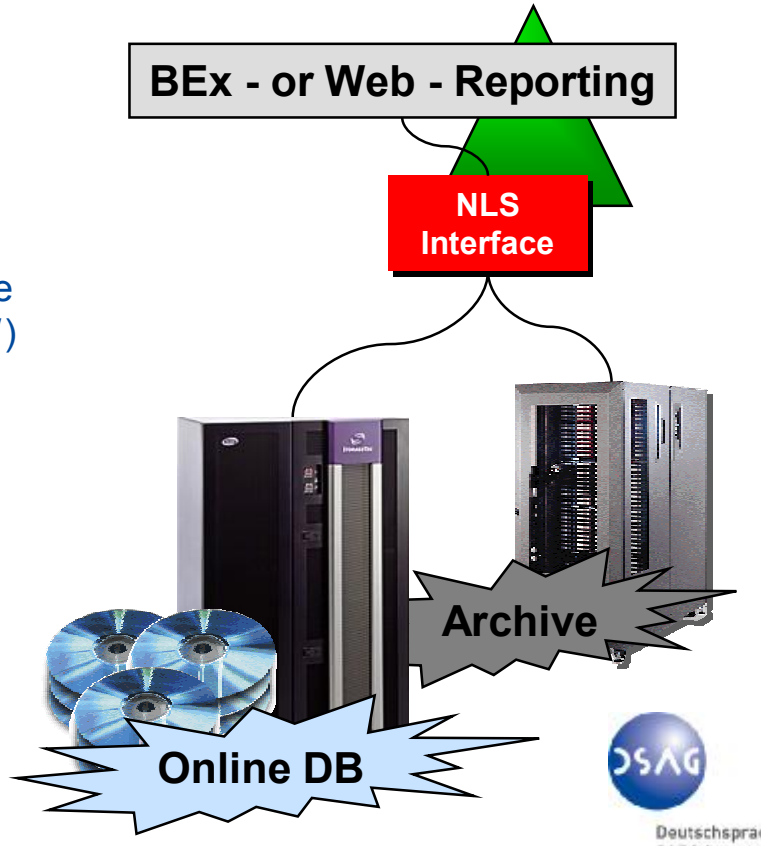
Generic NLS Interface





Consistency between nearline and online

- Analysis and Reporting operate on a combination of online- and near-line datasets. The consistency of the data is an absolute prerequisite.
- Archiving processes into different near-line storage levels have to fulfill transactional requirements with regard to maintaining consistency
  - Archiving and deletion of data in the online database form a logical unit of work (LUW)
  - Rollback mechanisms available for individual archiving steps.
  - The „archive“ gets the character of a database.
  - The archive data are usually ‚read only‘

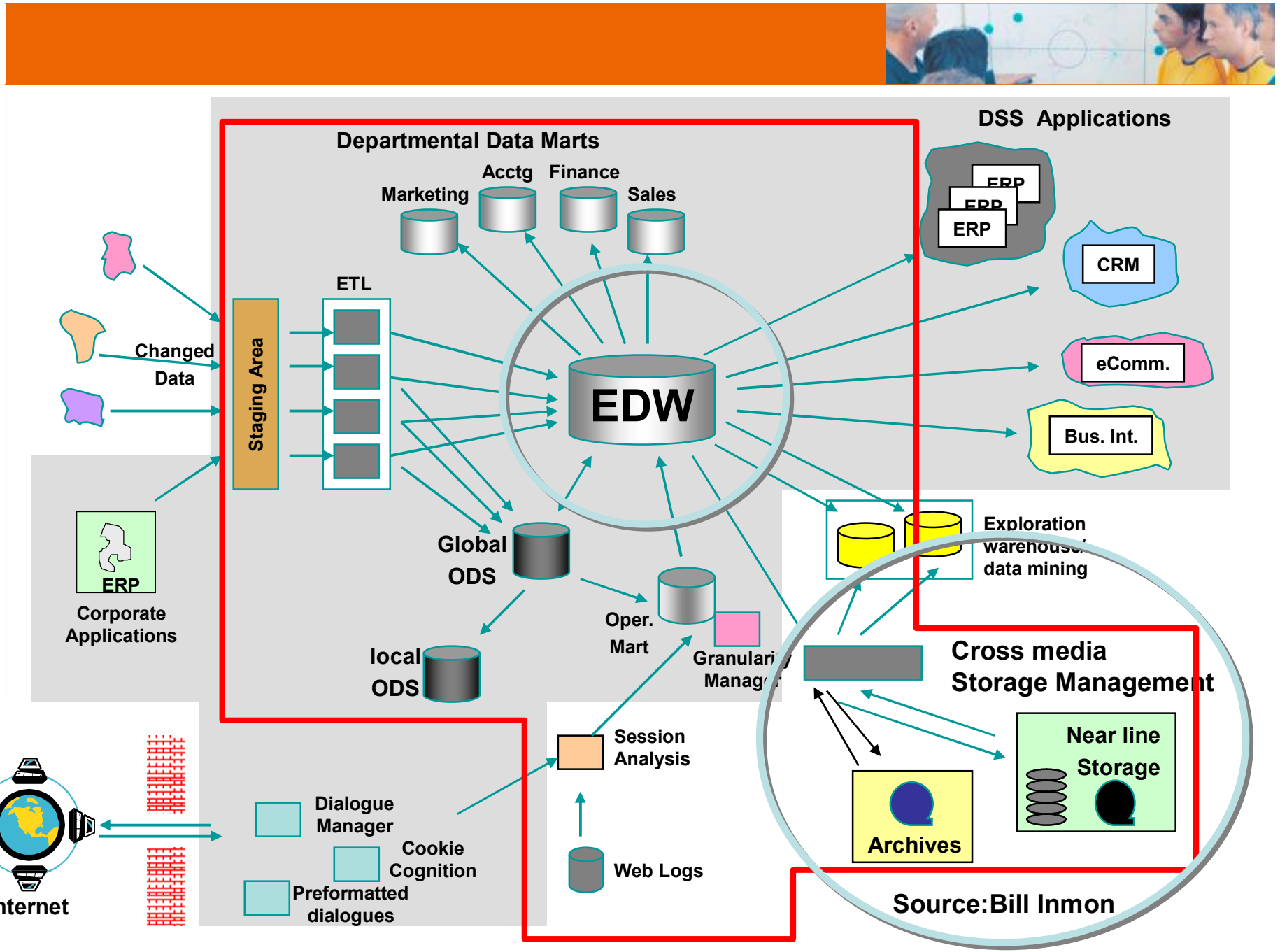


Fundamental ILM Strategy for BI - Benefits

- 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

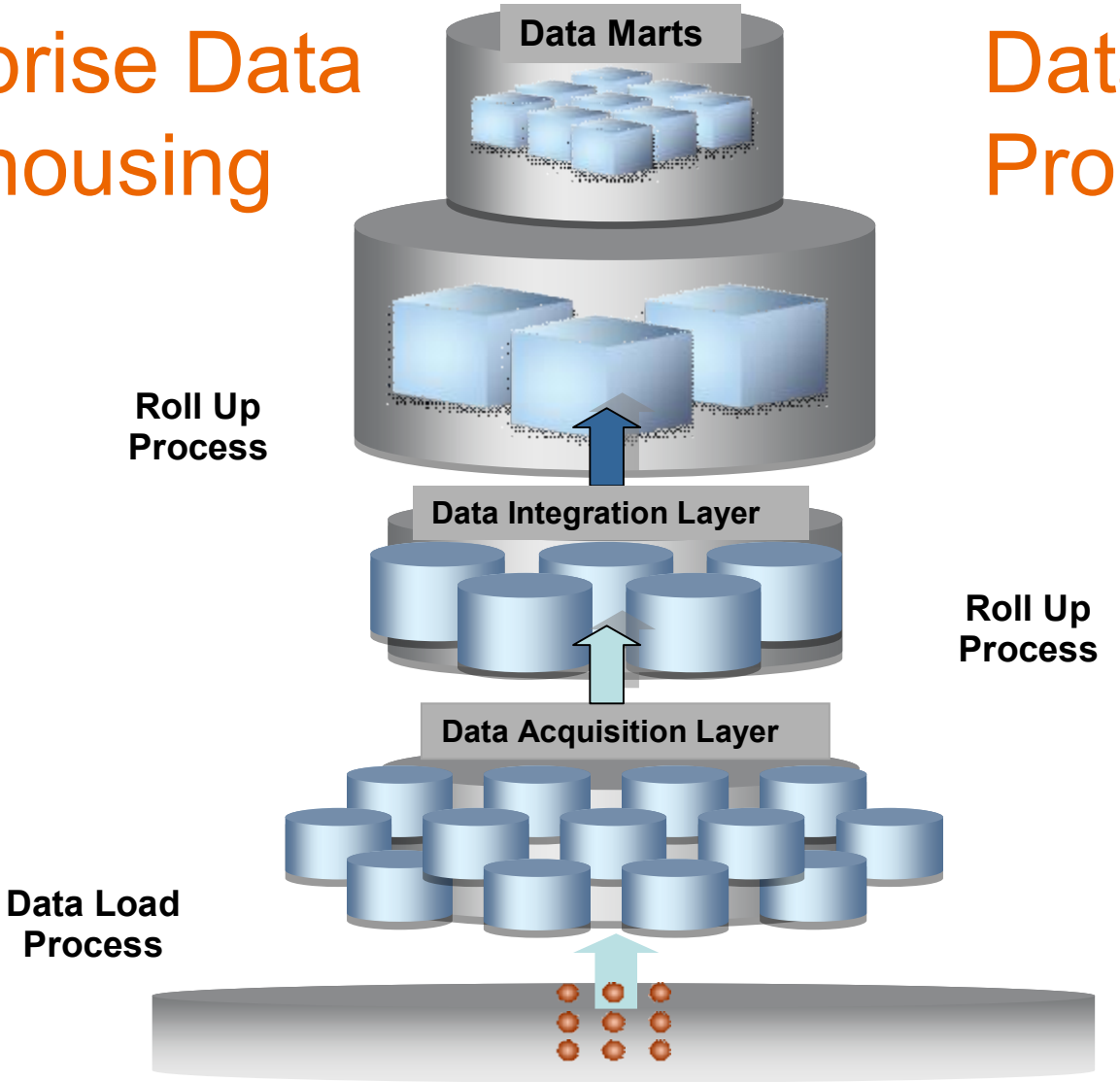
Agenda

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. Summary, Q&A


DSS Applications
Departmental Data Marts
Marketing
Acctg
Finance
Sales
ETL
Staging Area
Changed Data
EDW
ERP
CRM
eComm.
Bus. Int.
ERP
Corporate Applications
Global ODS
local ODS
Oper. Mart
Granularity Manager
Exploration warehouse/ data mining
Cross media Storage Management
Near line Storage
Archives
Session Analysis
Web Logs
Dialogue Manager
Cookie Cognition
Preformatted dialogues
Internet
Source:Bill Inmon

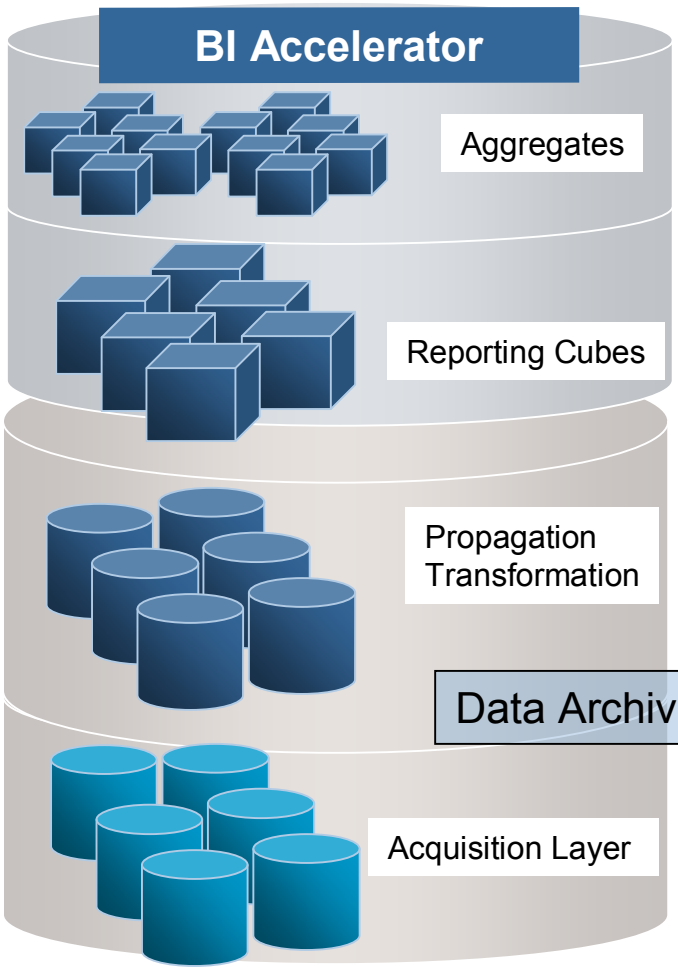
Enterprise Data Warehousing

Data Processing



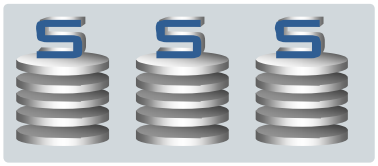
Deutschsprachige SAP® Anwendergruppe

Lesson learned : Nearline on Detailed Data



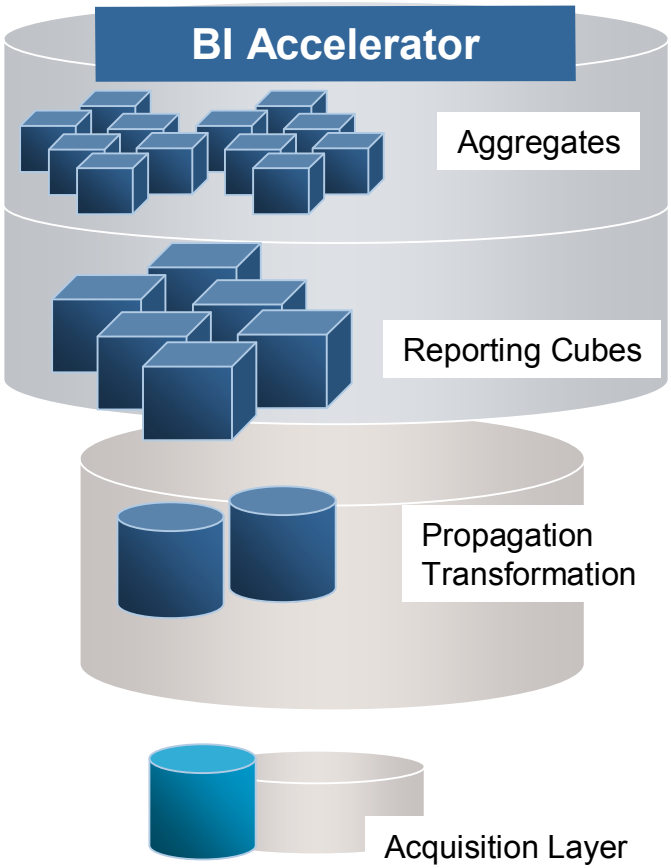
- Relieving SAP BI from detailed data
- Compressed by more than 85%
- Used as a „Corporate Memory“
  - Details in its “pure” form
  - Infrequently used detailed data
  - “Just-in-Case” data
  - Aged and historical data
  - Legacy data

Data Archiving Process (DAP)



Efficient Corporate Memory

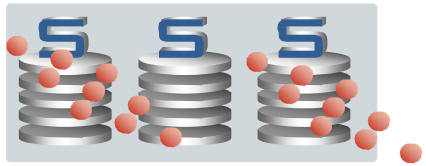
Usage of the „corporate memory“



Greater Flexibility in Responding to New Analytical Requirements

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

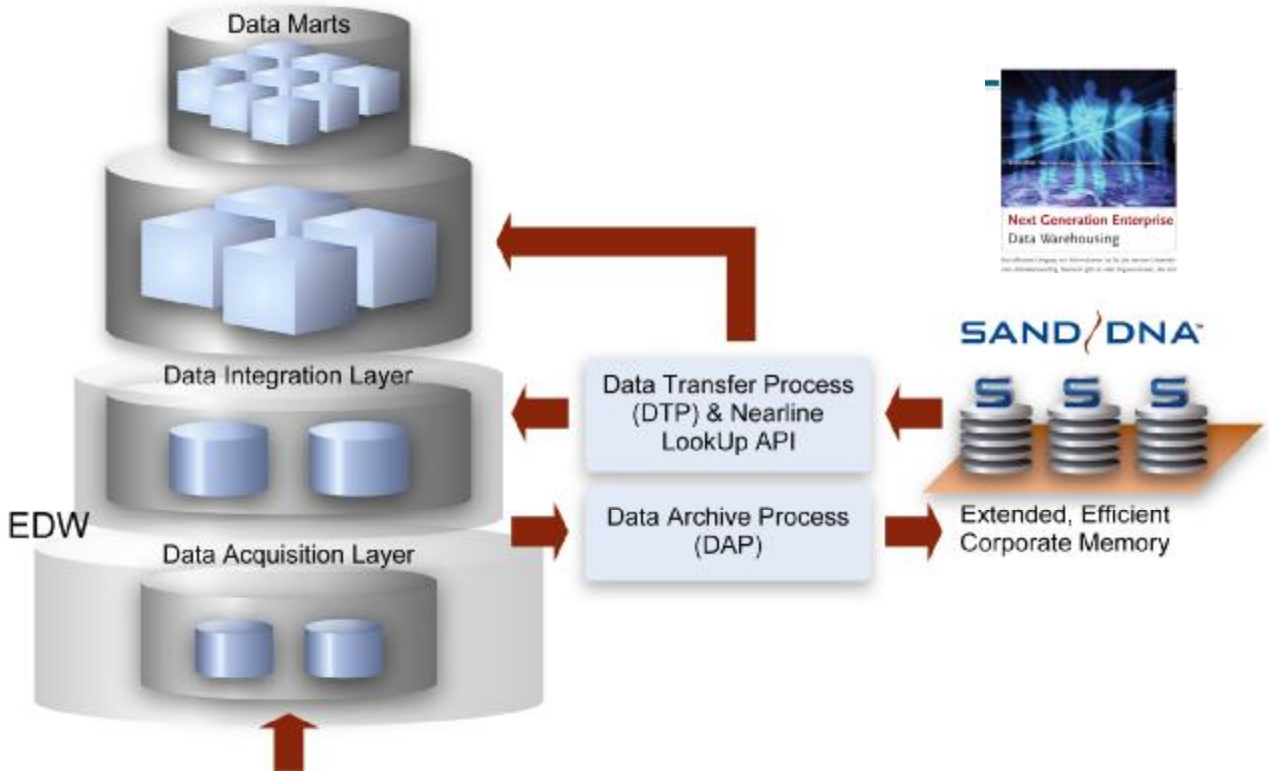
Data Transfer Process (DTP)
& Look Up API



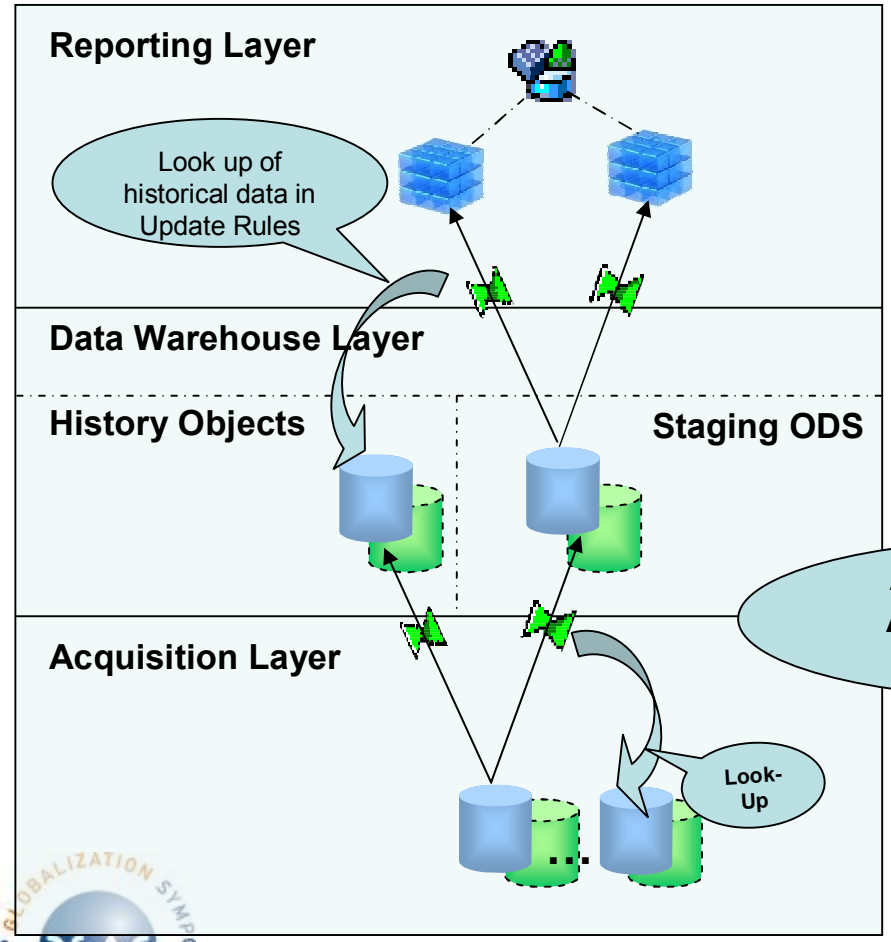
Efficient Corporate Memory

Next generation EDW -Layer

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



Look-Ups in the Data Flow Architecture

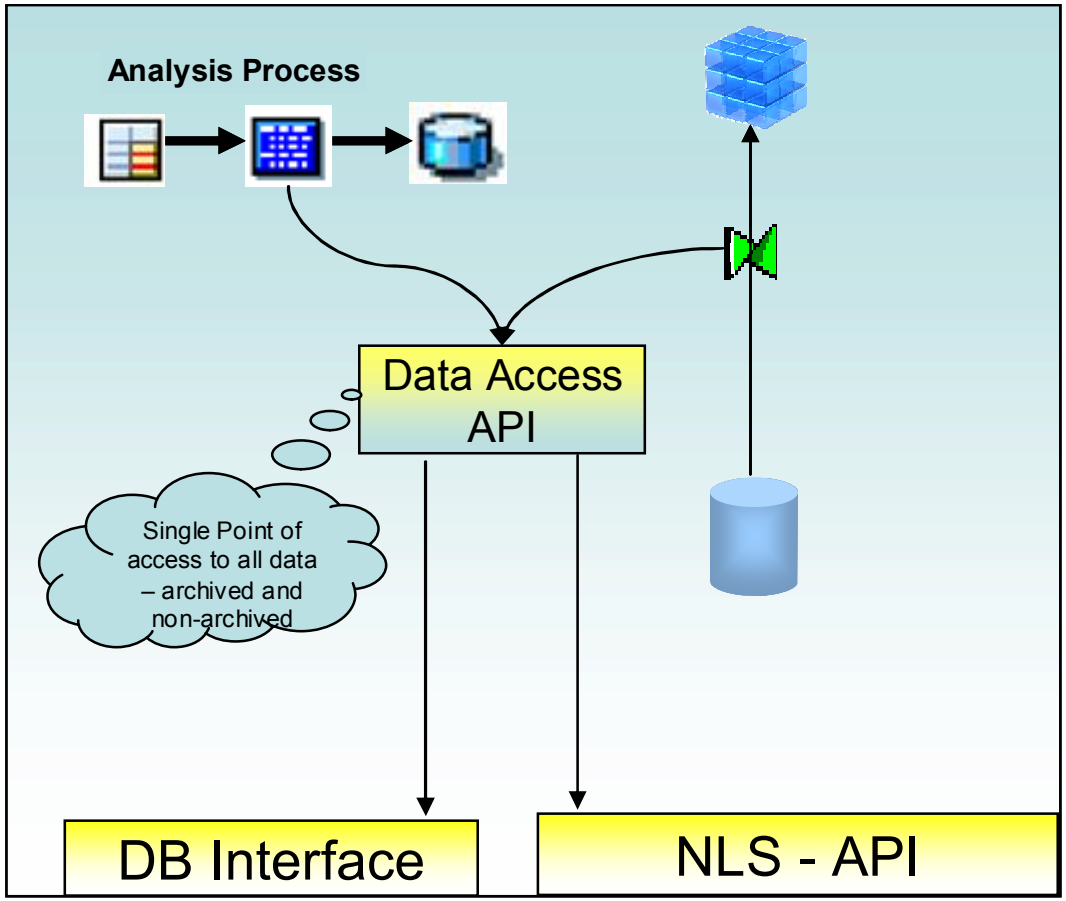


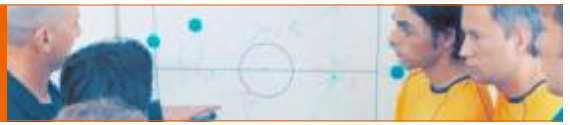
Look-Ups are often used e.g. to extend with derived attributes

Adhoc reporting, Analysis Process Designer

Nearline - Object

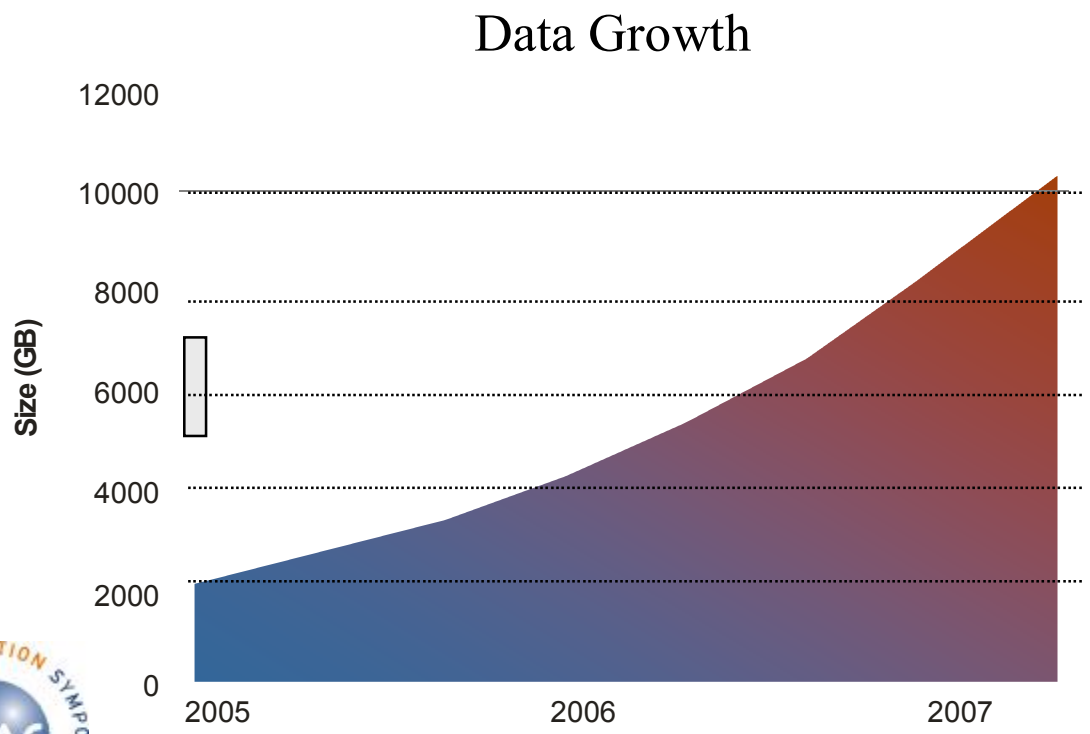
Usage of Look-Up API in Analysis Processes





Volkswagen Bank Case Study

"1 TB of data in our SAP BI production environment generates 5 TB for failover and backup processes." - *Adrian Bourcevet, Volkswagen Bank GmbH Germany*



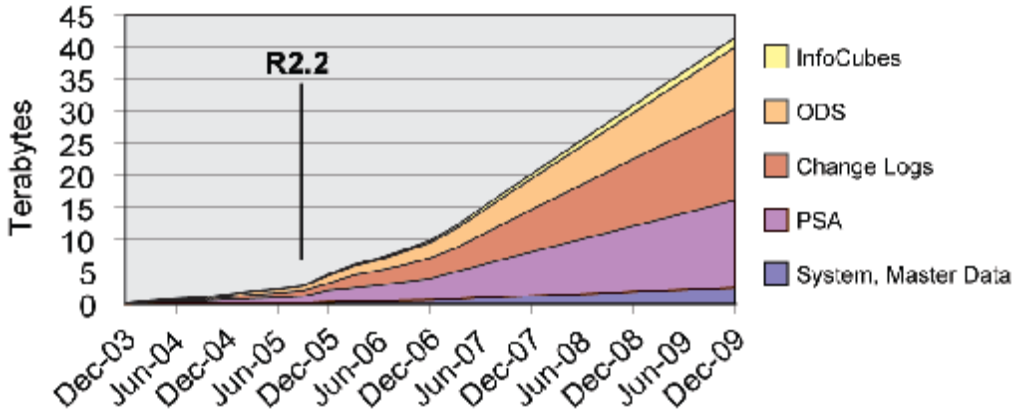
Volkswagen Bank

Deutschsprachige SAP® Anwendergruppe

US Government Case Study

- SAP BW database was growing at an unsustainable rate
- Limited funding for disk resources
- Performance risk

→ Data management strategy urgently required



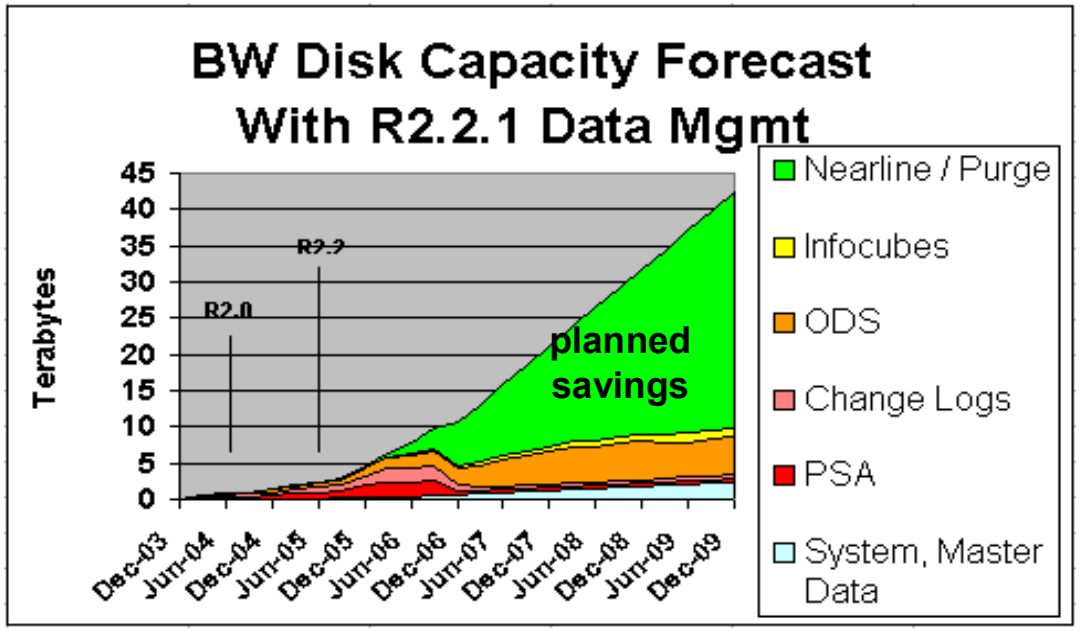
Database Growth:

- SAP BW database currently 5 TB (used space)
- Approximate growth rate at 400 GB/month
- Expected database size 10 TB by Dec 2006

Deutschsprachige SAP® Anwendergruppe

US Government Case Study (cont.)

SAP BW Forecast with Data Management Strategy



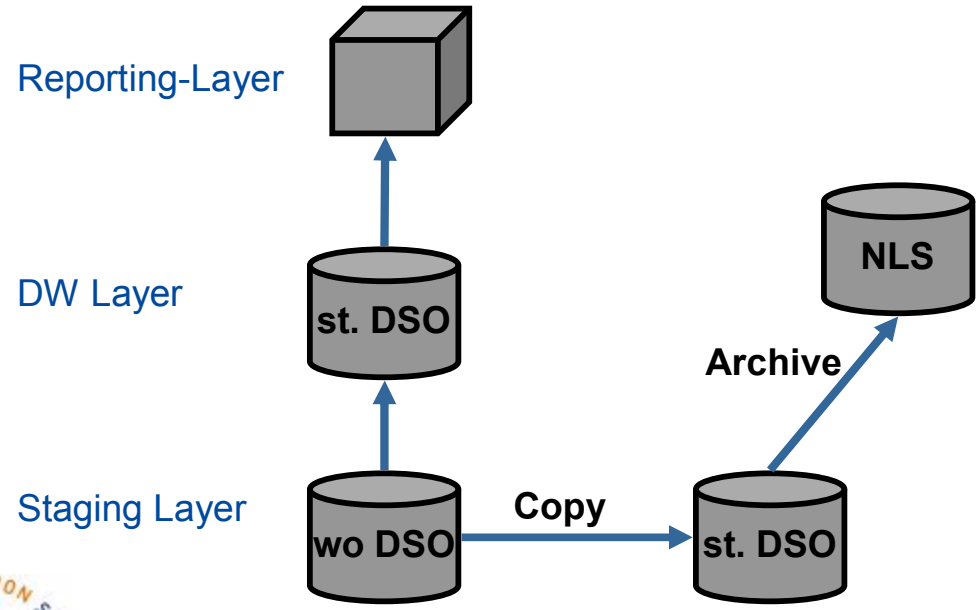
Return on Investment

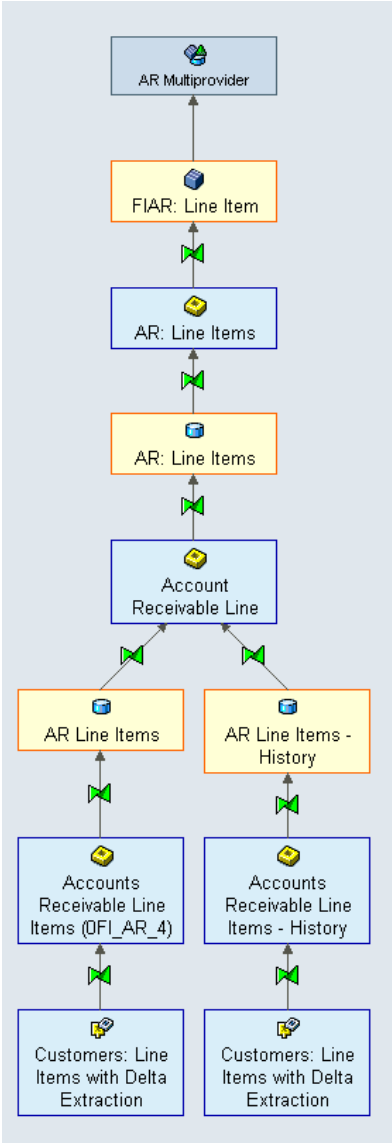
- Volkswagen Bank:
  - 90% data compression (on average), still available for use in reporting or as the basis for new DataStore objects or InfoCubes
  - Low total cost of ownership, due to the need for far less administrative support as compared with standard archiving solutions

- US Government:
  - ROI in less than 6 months (immediately after production go-live)
  - Savings of over 50% on related storage infrastructure thereafter
  - About 95% compression
  - Reduced data footprint eases replication/bandwidth issues

Write-Optimized DSO Support

- Comes with Enhancement Package 7.01
- Best Practice: Workaround via Standard DSO Archiving
  - E.g. at EDS Itellium





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. Summary, Q&A

Transparent Query Access

Query Edit Goto Environment System Help

SAP

Query Monitor

Execute | Execute + Debug | Generate Report | Properties | Messages | Help Texts | Generation Log

Performance Info | Technical Information | Cache Monitor | Query Variants | IGS Te

ABAP Web | Java Web

Query: ZEDU98I/ZEDU98I_Q001

Query Display: List — Parameter 1 — Parameter 2

Query Properties -> Move into Query Designer

Read Mode	H Query to Read When You Navigate or Expand
Data Integrity	0 All Part Providers Up to Securely Accessible S

☑ Read Near-Line Storage As Well

Cache Mode	1 Main Memory Cache Without Swapping

☑ Update Cache Objects in Delta Process

☐ Use Selection of Structure Elements

☐ Calculate w/ Packed Numbers

☐ Generation log

Optimization Mode	0 Query Will Be Optimized after Generation

☐ All Basic Key Figures Input Ready

Statistic Det.	1 Only Front End/Calculation Layer Data

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

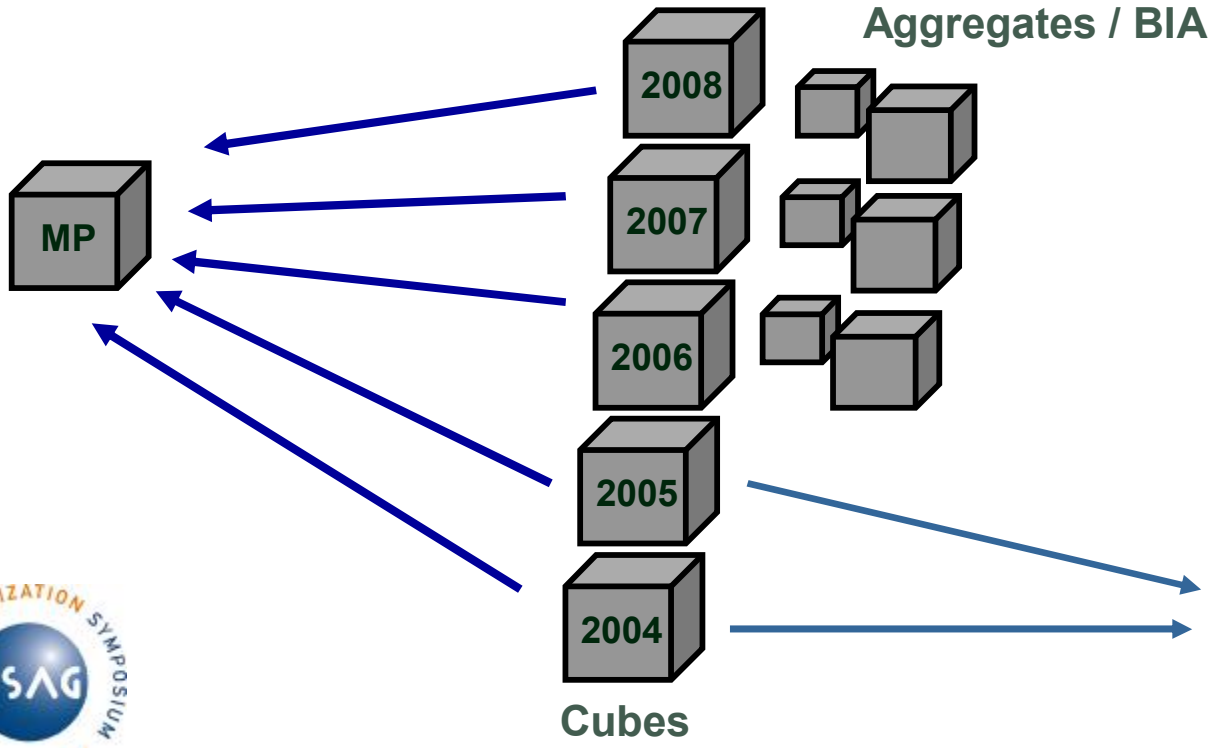
Calendar Year	Company code		Billed Quantity	Costs (SAP Demo)	Net Value	Tax Amount
1990	1000	1000	2.700 ST	215.466,00 EUR	3.291.210,90 EUR	411.030,00 EUR
	2000	2000	2.730 ST	217.710,00 EUR	3.328.780,91 EUR	415.597,00 EUR
	3000	3000	2.730 ST	217.679,00 EUR	3.327.779,91 EUR	415.597,00 EUR
	4000	4000	2.730 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

Query Designer and “Read NLS”

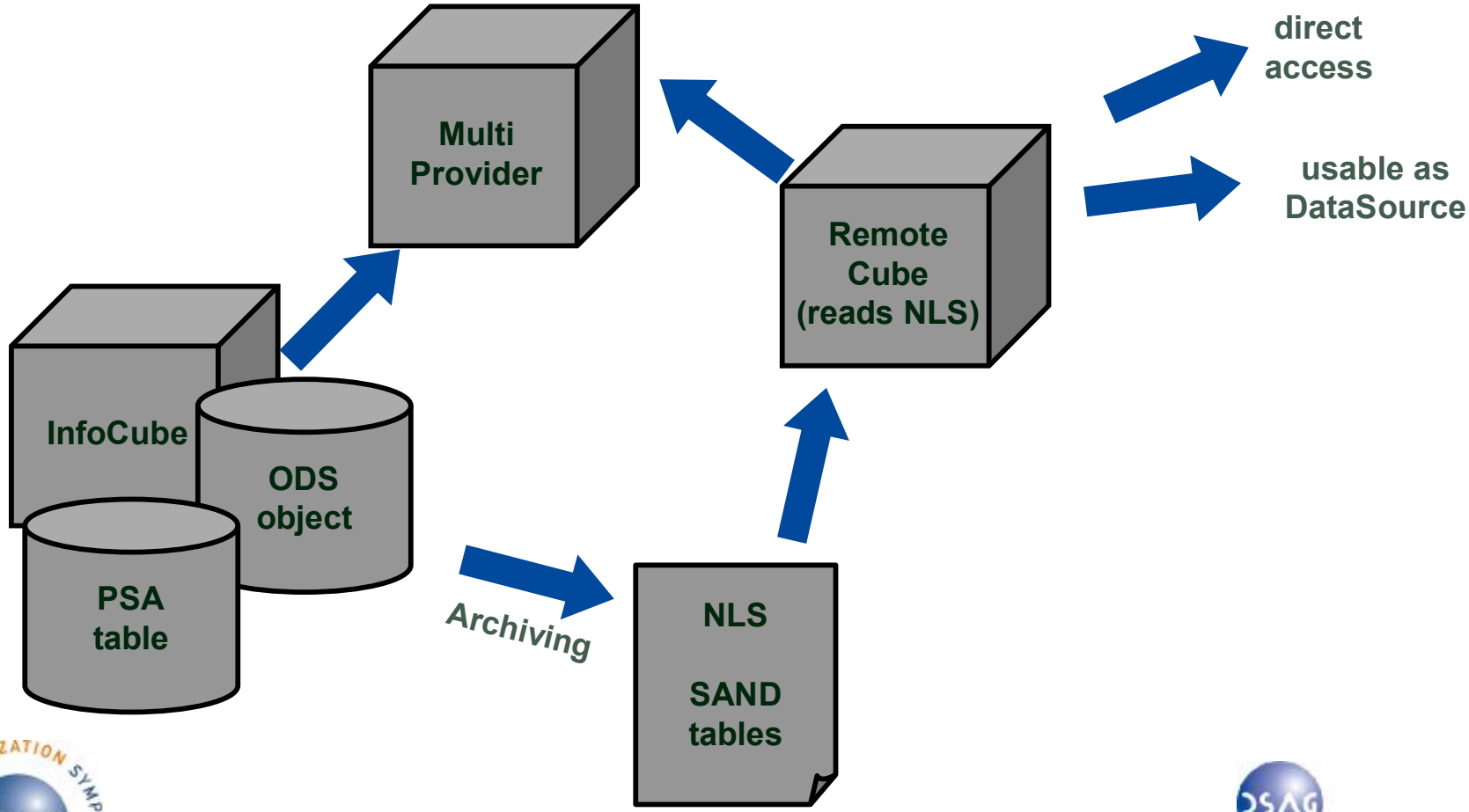
- Issue:
  - Per default, no NLS will be read
  - End-User can’t maintain Query Property, only rsrt is supported
- Impact:
  - Restricts NLS to non-reporting layers in many cases
  - NLS not available for ad-hoc reporting, only for centrally maintained Queries
- Best Practice Solution
  - Usage of Virtual Providers

Multi-Provider Support

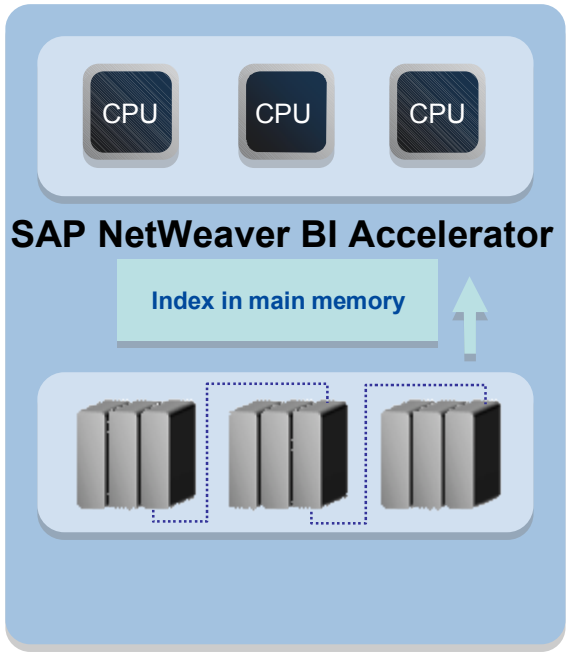
- Complete Multi-Provider support with NW 7.20
- Especially a problem if logical partitioning is used
- Best Practice Solution: Using a Virtual Provider



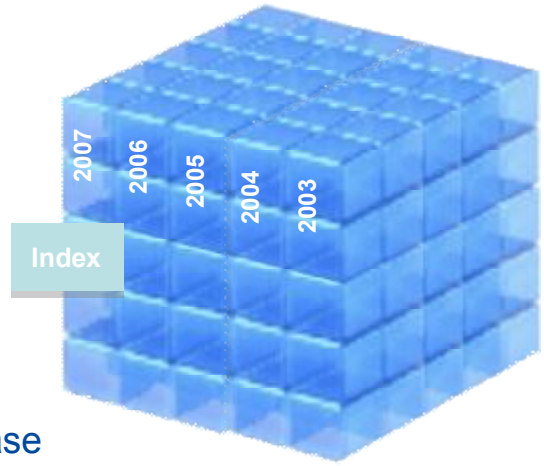
Best Practice Solution: Virtual Provider



Optimization of BIA by Nearline Storage



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

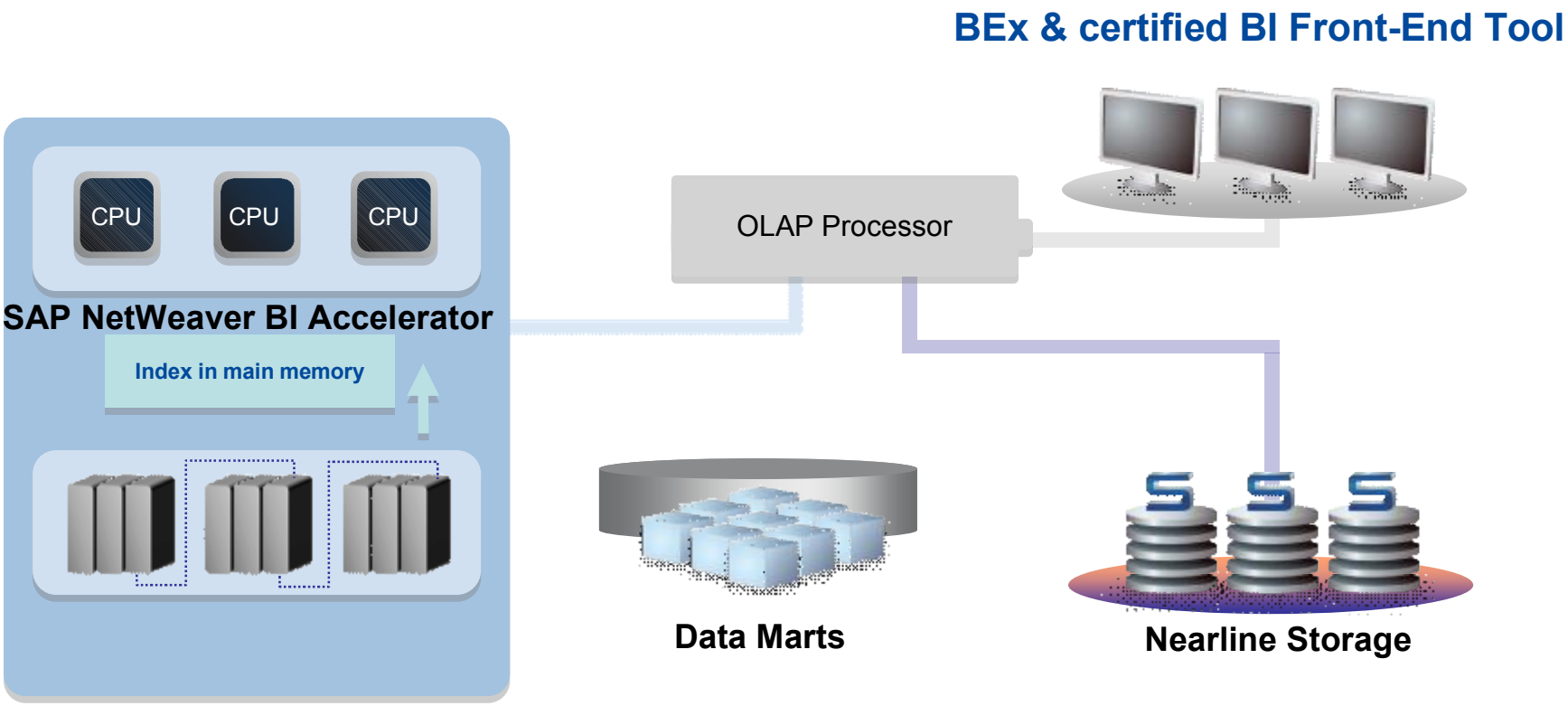


- ➔ InfoCubes partially indexed in BIA
- ➔ Data **remains** in the relational Database

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

Deutschsprachige SAP® Anwendergruppe

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. 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.

Deutschsprachige SAP® Anwendergruppe

Additional Resources

- HowTo Papers
  - How to Access Nearline Data via Multi Providers (planned)
  - How to Archive PSA Data in SAP NetWeaver BI (SDN)
- 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?

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

Deutschsprachige SAP® Anwendergruppe

Deutschsprachige SAP® Anwendergruppe