Advantages of a Layered Architecture for Enterprise Data Warehouse Systems

Thorsten Winsemann, Veit Köppen, Gunter Saake
Otto-von-Guericke-Universität, Magdeburg/Germany
1. Characteristics of Enterprise Data Warehouses

2. Traditional Data Warehouse Architecture
   1. Reference Architecture
   2. Dataflow-Example

3. Architectures for Enterprise Data Warehouses
   1. SAP’s Layered, Scalable Architecture
   2. Layers in Detail
   3. Dataflow-Example
   4. Simple, but Detailed Example

4. Architectural Differences
   1. Overview
   2. Advantages of a Layered Architecture
Characteristics of Enterprise Data Warehouses (EDW)

• Business DW, thus covering all business areas
• Data basis for several applications, such as BI, planning, CRM, …
• Single Version of Truth of company’s data
• Multiple, heterogeneous source systems
• Huge amount of data (granular, detailed, old)
• World-wide scope, different time zones
• 24*7-hours availability
• …
Traditional Data Warehouse Architecture: Reference Architecture

- User
- Data Marts
- Basis Data Base
- Staging Area
- Operational Data Warehouse System
- Data Sources
Traditional Data Warehouse Architecture: Dataflow-Example
Architecture for Enterprise Data Warehouses: SAP's Layered, Scalable Architecture
Architecture for Enterprise Data Warehouses: Dataflow-Example
Architecture for Enterprise Data Warehouses: Layers in Detail (1)

• Data Acquisition Layer
  – „DW Inbox“ (temporary)
  – Data stored immediately without changes

• Corporate Memory
  – „DW Life Insurance“ (long-term, granular, complete)
  – Data for non-predictable demands („master the unknown“)

• Quality & Harmonization Layer
  – Technical and semantical data integration
  – Usually no data storage
Architecture for Enterprise Data Warehouses: Layers in Detail (2)

- **Data Propagation Layer**
  - „Single Version of Truth“
  - Harmonized, integrated data without business logic

- **Business Transformation Layer**
  - Data are transformed according to business‘ needs
  - E.g., combination of sales + finance figures

- **Reporting & Analysis Layer**
  - Data are transformed according to requirements for usage and fast access performance
Architectures for Enterprise Data Warehouses: Simple, but Detailed Example (1)

<table>
<thead>
<tr>
<th>Data Acquisition Layer + Corporate Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sales Order</strong></td>
</tr>
<tr>
<td><strong>DOCNR</strong></td>
</tr>
<tr>
<td>Numc10</td>
</tr>
<tr>
<td>12345</td>
</tr>
<tr>
<td><strong>System A: Ordering</strong></td>
</tr>
<tr>
<td><strong>DOCNR</strong></td>
</tr>
<tr>
<td>Numc10</td>
</tr>
<tr>
<td>12345</td>
</tr>
</tbody>
</table>

| **Sales Invoice**                          |
| **INVNR** | **DATEI** | **CUSTOM** | **ORDNR** | **DATEP** |
| IN02085  | 20100805  | 0000012345 | 20100820  | |
| **System B: Invoicing**                    |
| **DOCNR** | **INVDT** | **PAYER** | **ORDER** | **PDATE** | **SYSID** |
| IN02085  | 100805    | 7410000   | 100820    | SBI |
| **Sales Invoice - Item**                   |
| **DOCNR** | **INVDT** | **PAYER** | **ORDER** | **PDATE** | **SYSID** |
| IN02085  | 100805    | 7410000   | 100820    | SBI |

Data types adapted
Homonyms split
Synonyms merged
System-ID added
Field names changed
Architectures for Enterprise Data Warehouses: Simple, but Detailed Example (2)

Report Execution (no persistence!)

Sales
- CUSTOM
- MATNR
- DATEO
- QUABU
- UNITB
- DATEP
- AMDCI
- CURRD
- AMDC
- PRPDC

0007410000 ABT0471 ART 20100730 4,000 PC 20100820 5,000 EUR

Business Transformation Layer

Sales Orders
- ORDNR
- ITMNO
- DATEO
- CUSTOM
- MATNR
- QUABU
- UNITB
- DATEP
- AMDCI
- CURRD

0000012345 0001 ABT0471 20100730 2,000 BOX 20100820 300,00 EUR

Data Invoices
- INVNR
- ITMNI
- MATNR
- DATEI
- CUSTOM
- ORDNR
- DATEP
- QUABU
- UNITB
- AMDCI
- CURRD

IN02085 0001 ABT0471 20100805 0000012345 20100820 4,000 PC 285,00 EUR

Data Propagation Layer

Sales Order
- ORDNR
- DATEO
- CUSTOM

0000012345 20100730 0007410000

Sales Order - Item
- ORDNR
- ITMNO
- MATNR
- QUABU
- UNITB
- AMDCI
- CURRD

0000012345 0001 ABT0471 2,000 BOX 20100820 5,000 PC 20100820 300,00 EUR

Sale Invoice
- INVNR
- DATEI
- CUSTOM
- ORDNR
- DATEP

IN02085 0001 ABT0471 20100805 20100805 0000012345 20100820 4,000 PC 285,00 EUR

Harmonization & Quality Layer

Further information added
Data combined (according to usage)
Data configured (according to business' needs)
Additional information added
## Architectural Differences: Overview

<table>
<thead>
<tr>
<th>Matter</th>
<th>Reference Architecture</th>
<th>Layered Architecture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complexity</td>
<td>Medium</td>
<td>High (several layers)</td>
</tr>
<tr>
<td>Data volume</td>
<td>High</td>
<td>Very high</td>
</tr>
<tr>
<td>Conceptual work</td>
<td>Medium (requirement-driven)</td>
<td>High (overall concept view)</td>
</tr>
<tr>
<td>Implementation effort</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>Enhancement &amp; scalability</td>
<td>Possible</td>
<td>Supported</td>
</tr>
</tbody>
</table>
# Architectural Differences: Advantages of a Layered Architecture

<table>
<thead>
<tr>
<th>Matter</th>
<th>Reference Architecture</th>
<th>Layered Architecture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change of transformation rules (e.g., changed key-figure calculation)</td>
<td>Reload/-build from source system</td>
<td>Rebuild from propagation layer</td>
</tr>
<tr>
<td>Change of data (e.g., new key-figure calculation)</td>
<td>Reload/-build from source system</td>
<td>Rebuild from propagation layer</td>
</tr>
<tr>
<td>Need for new data</td>
<td>Dataflow enhancement and reload/-build</td>
<td>Load from propagation layer or corporate memory</td>
</tr>
<tr>
<td>„Single Version of Truth“</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Decoupling of data load and availability</td>
<td>No/limited</td>
<td>Yes/supported</td>
</tr>
<tr>
<td>Detailed previous data</td>
<td>Limited</td>
<td>Available (corporate memory)</td>
</tr>
</tbody>
</table>
Appendix

CSDM 2011
Poster Layout

Slides' Arrangement on Panel

Panel: 150x125cm
Slides: A3 + A4