SAP NetWeaver 7.0: ETL and EII
Overview

Product Management SAP NetWeaver BI
November 2007
1. Overview

2. ETL
   2.1. Extraction
   2.2. Transformation
   2.3. Data Distribution
   2.4. Real-Time Data Acquisition

3. Distributed Query / EII
1. Overview

2. ETL
   2.1. Extraction
   2.2. Transformation
   2.3. Data Distribution
   2.4. Real-Time Data Acquisition

3. Distributed Query / EII
Conceptual Layers of Data Warehousing

- **Operational Data Store**
- **Data Warehouse**
- **(Architected) Data Marts**
- **(Persistent) Staging Area**
- **Any Source**
- **Information Access**
- **DataSources**

Any Source → (Persistent) Staging Area → Operational Data Store → (Architected) Data Marts → Data Warehouse → Information Access
Data Acquisition Layer – DataSource Types

- **RDBMS**
  - e.g. Oracle, Informatix, MSSQL
- **DB Connect**
- **UD Connect**
- **BI Service API**
- **File Interface**
- **Web Service**
- **Staging BAPI**
- **RDBMS DBSL**
  - e.g. Teradata
- **SAP Source**
  - e.g. SAP Business Suite
- **File data**
- **XML data**
- **Business Objects Data Services**
  - any Source, e.g. Oracle Applications
New BI DataSource concept with SAP NetWeaver 7.0

Highlights
- unique look and feel for all of the DataSource Types
- PSA is attached to DataSource
  - InfoPackage writes to PSA
  - Data Transfer Process writes from PSA to data targets
- direct/remote access is optional
- preview feature is standard
- automated conversions (e.g. date format detection)
Source System Tree

Source systems categories:

- SAP vs. non SAP
- File vs. database
- Relational vs. Multidimensional DB
- ABAP vs. Java
- XML vs. Text/Binary
- Pull vs. Push
- Realtime vs. Batch
DataSource Example – One fits all approach

General Information
- Descriptions
- Reconciliation flag (not functional)
- Opening Balance (inventory)
- Error handling (duprecs)
Data Flow Concept in SAP NetWeaver 7.0

SAP NetWeaver BI

Process Chain
Data Transfer Process
InfoPackage

InfoProvider
Transformation
DataSource / PSA

Source System 1
Source
Data Flow Concept in SAP NetWeaver 7.0
Simplified

SAP NetWeaver BI

Source System 1

Source

InfoProvider

Transformation

DataSource / PSA

Restrictions:
Not optimized for mass data transfer
No packaging of data
Full Mode Only
1. Overview

2. ETL
   2.1. Extraction
   2.2. Transformation
   2.3. Data Distribution
   2.4. Real-Time Data Acquisition

3. Distributed Query / EII
Data Flow in SAP NetWeaver 7.0 BI

SAP NetWeaver Business Intelligence

InfoProvider

Transformation

DataSource / PSA

InfoPackage

Non-SAP

SAP NetWeaver BI

Any Source

SAP

SAP NetWeaver PI

Downstream Systems

Open Hub Destination

Transformation

Data Transfer Process

DTP

DTP

© SAP 2007 / Page 13
Universal transformation from source to target objects

Transformation types:
- Move, aggregate, constant, master data look up, …
- Business rules, e.g. unit + currency translation
- Formula builder with rich predefined functions library
- ABAP routines incl. regular expressions

SAP NetWeaver 7.0 Enhancements
- Intuitive UI
- Unit conversion
- Unified transfer + update rules into all-in-one capability
- Integration of Open Hub Service

* optional
Transformation – Definition

Access from the Data Warehousing Workbench

- New transformation
  - Unification of transfer and update rules
  - InfoSource not mandatory anymore
- Former concept of update rules
  - Small square next to the transformation icon
  - Access from context menu via ‘additional functions’
- Links sources and target
  - New source: InfoSet
  - Other sources: DataSource, InfoCube, DataStore object, InfoObject, InfoSource
  - Targets: InfoCube, DataStore object, InfoObject, InfoSource, Open Hub Destination

Transformation Change

Transformation

Update rule
### Transformation Change

<table>
<thead>
<tr>
<th>Transformation</th>
<th>Source</th>
<th>Target</th>
<th>Version</th>
<th>Active</th>
<th>Saved</th>
<th>Executable</th>
<th>Edited Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDS/PM_SD_C03</td>
<td>PM_SD_C03 (PM_SD_C03)</td>
<td>Demo Cube 'Delete Master Data' (PM_SD_C03)</td>
<td>Active</td>
<td>Saved</td>
<td>Executable</td>
<td>Edited Version</td>
<td></td>
</tr>
</tbody>
</table>

### Source fields

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BUCHUNGSKREIS (BUCHUNGSKREIS)</td>
</tr>
<tr>
<td>2</td>
<td>VERRECHNUNG (VERRECHNUNG)</td>
</tr>
<tr>
<td>3</td>
<td>SPARTEN (SPARTEN)</td>
</tr>
<tr>
<td>4</td>
<td>Material</td>
</tr>
<tr>
<td>5</td>
<td>PLANT (PLANT)</td>
</tr>
<tr>
<td>6</td>
<td>VERKUPLERGRUPPE (VERKUPLERGRUPPE)</td>
</tr>
<tr>
<td>7</td>
<td>VERKAUFSPROGANBA (VERKAUFSPROGANBA)</td>
</tr>
<tr>
<td>8</td>
<td>AUFTRAGGEBER (AUFTRAGGEBER)</td>
</tr>
<tr>
<td>9</td>
<td>WARENMANG (WARENMANG)</td>
</tr>
<tr>
<td>10</td>
<td>REGELER (REGELER)</td>
</tr>
<tr>
<td>11</td>
<td>VERSION (VERSION)</td>
</tr>
<tr>
<td>12</td>
<td>WERKART (WERKART)</td>
</tr>
<tr>
<td>13</td>
<td>DELEKCLASSO (DELEKCLASSO)</td>
</tr>
<tr>
<td>14</td>
<td>CRED_DEBUCHUNG (CRED_DEBUCHUNG)</td>
</tr>
<tr>
<td>15</td>
<td>CALMONTH (CALMONTH)</td>
</tr>
<tr>
<td>16</td>
<td>GESCHWSTJHARESVA (GESCHWSTJHARESVA)</td>
</tr>
<tr>
<td>17</td>
<td>GAUFFRUMN (GAUFFRUMN)</td>
</tr>
</tbody>
</table>

### Target fields

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GCAL_MONTH (GCAL_MONTH) Calendar Year/Month</td>
</tr>
<tr>
<td>2</td>
<td>GCAL_YEAR (GCAL_YEAR) Calendar Year</td>
</tr>
<tr>
<td>3</td>
<td>GFSVAR (GFSVAR) Fiscal year variant</td>
</tr>
<tr>
<td>4</td>
<td>0D_D_CODE (0D_D_CODE) Company Code (SAP Demo)</td>
</tr>
<tr>
<td>5</td>
<td>0D_DEB_CRED (0D_DEB_CRED) Credit/Debit posting (CD) (SAP Demo)</td>
</tr>
<tr>
<td>6</td>
<td>0D_DCH_Channel (0D_DCH_Channel) Distribution Channel (SAP Demo)</td>
</tr>
<tr>
<td>7</td>
<td>0D_DIV (0D_DIV) Division (SAP Demo)</td>
</tr>
<tr>
<td>8</td>
<td>0D_DOCCLASS (0D_DOCCLASS) Docu class order/invoice (SAP Demo)</td>
</tr>
<tr>
<td>9</td>
<td>0D_PAYER (0D_PAYER) Payer (SAP Demo)</td>
</tr>
<tr>
<td>10</td>
<td>0D_PLANT (0D_PLANT) Plant (SAP Demo)</td>
</tr>
<tr>
<td>11</td>
<td>0D_SALE_GRP (0D_SALE_GRP) Saba Group (SAP Demo)</td>
</tr>
<tr>
<td>12</td>
<td>0D_SALE_Org (0D_SALE_Org) Sales Organization (SAP Demo)</td>
</tr>
<tr>
<td>13</td>
<td>0D_SHIPTO (0D_SHIPTO) Ship-to Party (SAP Demo)</td>
</tr>
<tr>
<td>14</td>
<td>0D_SOLD_TO (0D_SOLD_TO) Sold-to-Party (SAP Demo)</td>
</tr>
<tr>
<td>15</td>
<td>0D_VERSION (0D_VERSION) Version (SAP Demo)</td>
</tr>
<tr>
<td>16</td>
<td>0D_VTYPE (0D_VTYPE) Value Type for Reporting (SAP Demo)</td>
</tr>
<tr>
<td>17</td>
<td>0D_COUNTRY (0D_COUNTRY) Country (SAP Demo)</td>
</tr>
<tr>
<td>18</td>
<td>0D_REGION (0D_REGION) GIS-Region (SAP Demo)</td>
</tr>
</tbody>
</table>

### Note

Key figures, characteristics and date fields are shown on the same level (transformation group).
Transformation rule details

- Information on
  - Rule type
  - Currency/Unit Conversion
- Source fields
- Target fields
Enhanced Data Flow in SAP NetWeaver 7.0 BI

SAP NetWeaver Business Intelligence

InfoProvider

Transformation (optional)

InfoSource (optional)

Transformation

Data Source / PSA

Transformation

InfoProvider

Transformation (optional)

InfoSource (optional)

Data Transfer Process
InfoSource

- Transformation directly links from a source InfoProvider (or DataSource) to a target InfoProvider
- An InfoSource is usually not needed
- New InfoSource architecture is used (flat InfoObject-based structure)
- Scenarios for (flexible) InfoSource
  - A flexible InfoSource is necessary in order to use currency or unit conversion from the source DataSource → Define InfoSource as an intermediate structure
  - You can use a flexible InfoSource as a uniform source for several targets; the InfoSource can the be target from different sources (see next slide)
- Note: for ‘direct’ InfoSources (for master data updates), there is no difference between ‘old’ and ‘new’ InfoSource, i.e. you can define a transformation as well as transfer rules
  - Pre-requisite: InfoObject is defined as InfoProvider
InfoSource

- Scenario: InfoSource as a uniform source for several targets and as target from different sources
Transformation Groups

- Summarize key figures with the same characteristics assignments
  - All key figures of one transformation are updated based on the same characteristic values
  - If other characteristic updates are necessary for particular key figures, a new transformation is created

Transformation Change

### Rule Group: Standard Group

<table>
<thead>
<tr>
<th>Rule Name</th>
<th>Post Key</th>
<th>Info Object</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSALESORG</td>
<td>3</td>
<td>0SALESORG</td>
<td>Sales Organization</td>
</tr>
<tr>
<td>OCUSTOMER</td>
<td>2</td>
<td>0CUSTOMER</td>
<td>Customer number</td>
</tr>
<tr>
<td>OMATERIAL</td>
<td>1</td>
<td>0MATERIAL</td>
<td>Material</td>
</tr>
<tr>
<td>0MATERIAL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0CUSTOMER</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sales Order Items (RHORDITM)

<table>
<thead>
<tr>
<th>Post Key</th>
<th>Info Object</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0CC_NUMBER</td>
<td>Document number</td>
<td></td>
</tr>
<tr>
<td>0_ORD_ITEM</td>
<td>Sales Order Item</td>
<td></td>
</tr>
<tr>
<td>OMATERIAL</td>
<td>Material</td>
<td></td>
</tr>
<tr>
<td>OCUSTOMER</td>
<td>Customer number</td>
<td></td>
</tr>
</tbody>
</table>
Transformation Groups

Use / Example

- Scenario: overview on bonus-relevant sales of all employees
- An employee generates a certain sales volume, which is the basis for his/her bonus
- The manager of the employee will be assigned 10% of the employee’s bonus as manager’s bonus relevant
- → two transformation groups are generated (e.g. ‘employee’ and ‘manager’)

<table>
<thead>
<tr>
<th>Source</th>
<th>Sales Volume</th>
<th>Manager</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johnson</td>
<td>1000</td>
<td>Giles</td>
</tr>
</tbody>
</table>

Transformation Group 1

Employee → Employee Sales Volume → Bonus-relevant Sales

<table>
<thead>
<tr>
<th>Target</th>
<th>Bonus-relevant Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johnson</td>
<td>1000</td>
</tr>
<tr>
<td>Giles</td>
<td>100</td>
</tr>
</tbody>
</table>

Transformation Group 1

Manager → Employee Sales Volume*0,1 → Bonus-relevant Sales
1. Overview

2. ETL
   2.1. Extraction
   2.2. Transformation
   2.3. Data Distribution
   2.4. Real-Time Data Acquisition

3. Distributed Query / EII
Benefits of New Data Transfer Process

**Data Transfer Process (DTP): data ‘distribution’ within SAP NetWeaver BI**

- Loading data from one layer to others except InfoSources
- Separation of delta mechanism for different data targets
- Enhanced filtering in dataflow
- Improved transparency of staging processes across data warehouse layers (PSA, DWH layer, ODS layer, Architected Data Marts)
- Improved performance: optimized parallelization
- Enhanced error handling for DataStore object (error stack)
- Enables real-time data acquisition
Error Handling Overview

There is no error handling available for an InfoPackage. In case of invalid records, data needs to be reloaded from the source system.

Invalid records can be corrected in the error stack and updated into the data target.
**Error Handling Features**

- Possibility to choose in the scheduler to...
  - abort process when errors occur
  - process the correct records but do not allow reporting on them
  - process the correct records and allow reporting on them

- Number of wrong records which lead to a wrong request

- Invalid records can be written into an error stack

- Keys should be defined for error stack to enable the error handling of DataStore object

- Temporary data storage can be switched on/off for each substep of the loading process

- Invalid records can be updated into data targets after their correction.
Error Stack

- Stores erroneous records
  - Automatic checks: Existence of master data, conversion exit (restricted, e.g. Alpha)
  - Customer-defined checks in transformation routines (see appendix for more information)
- Keeps the right sequence of records → for consistent DataStore handling
- Key of error stack defines which data should be detained from the update after the erroneous data record
- After correction, Error-DTP updates data from error stack to data target
- Note: Once the request in the source object is deleted, the related data records in error stack are automatically deleted
Temporary Data Storage

- Help for tracing the erroneous records and transformations
- Data records from different steps within the data transfer process can be stored temporarily
- Stores complete set of data (erroneous as well as valid records)
- Scenario:
  - If the debugging mode is switched on
  - Trace the erroneous records
  - Trace Transformation
Data Transfer Process Monitor

DTP Monitor

- Integrated in InfoProvider management screen
- Integrated in DTP maintenance
- Additional information: duration of each step
- Temporary storage access – if activated
- Error Stack is displayed in DTP Monitor
1. Overview

2. ETL
   2.1. Extraction
   2.2. Transformation
   2.3. Data Distribution
   2.4. Real-Time Data Acquisition

3. Distributed Query / EI1
Motivation for real-time

- Upload frequency for regular staging not sufficient
  - Number of requests can not be handled by BI system
  - Reporting requirements of data with sub 1 hours actuality

- Operational reporting
  - The clear distinction between tools for analytical applications and tools for operational applications is more and more difficult.
  - See transactional data in reporting as it is created in source system („post & see“)

- Availability of data in source system not known
  - Use „Push“ mechanism (e.g. SAP NW XI)
  - Polling or
  - Trigger data load process via event from source system

- Remote access of data not feasible due to resource consumption in source system (and cross application reporting requirements)

- Splitting of staging processes (EDW) – General business rules are processed during the day whereas scenario specific rules (defined by application, region, time zone, etc.) are processed at customer defined times
Real-time Data Acquisition (RDA)

SAP NetWeaver Business Intelligence

Operational Data Store

DataStore Objects

Architected Data Marts

Data Warehouse Layer

Daemon

Pull ~ 1/min

DataTransfer Process for Real-time Data Acquisition

InfoPackage for Real-Time Data Acquisition

PSA

DataSource

Scheduled Data Transfer Process

Scheduled Data Transfer Process

Real-time Update

Real-time Update

Delta Queue

Service API

Application

SAP Source System

Web Service Push

External System
RDA Scenarios

Real-time data acquisition can be used in two primary scenarios:

- **via the Service API (SAPI)**
  - Incorporates usage of InfoPackage for Real-time Data Acquisition (source to PSA)
  - Then leverages Data Transfer Process for Real-time Data Acquisition (PSA to DataStore Object)

- **via a Web Service**
  - Incorporates usage of Web Services to populate the PSA
  - Then leverages the Real-time DTP to transfer data to the DataStore Object

A system Daemon is used to initiate and control data transfer in both scenarios.
What is a Daemon?

Daemon
- System process fulfils a specific task at regular intervals
- The SAP NetWeaver BI Daemon performs three steps (depending on the scenario chosen):
  - Initiate Service-API (SAPI) data pull via InfoPackage for Real-time Data Acquisition into PSA
  - Track status of data transfer from source system
  - Initiate update of DataStore Object via DTP

- Successful execution of each step is tracked in a control table
  - Allows restarting if necessary
  - Restart can be initiated so it starts at the next step after the last successfully executed step.
Daemon Monitoring

- Maintenance screen for demon control can be called from the Data Warehousing Workbench
- Transaction RSRDA
Latency between source system transaction and availability in DataStore Object is minimal due to “Lean Staging”.

“Lean Staging” means that the DataStore Object handles the data differently internally – less logging and activation activity is allowed when the DTP for Real-time Data Acquisition is connected to the DataStore Object

From an administrative point of view, the DataStore Object is a standard one.

The DataStore Object also handles the availability of this real-time data (e.g. data whose request has not been closed) for reporting.
How to embed RDA in your data flow

Using two DataSources

- BI
  - “RDA” DataStore Object
  - Regular DataStore Object

- OLTP
  - real-time Update
  - Application
  - Delta Queue

- Daemon
  - Pull ~5/min

- PSA
  - DataStore

- Periodic Scheduling
  - 2LIS_02_VAHDR
  - 2LIS_02_VAITM

Using one DataSource

- BI
  - Regular DataStore Object

- OLTP
  - real-time Update
  - Application
  - Delta Queue

- Daemon
  - Pull ~5/min

- PSA
  - DataStore

- Periodic Scheduling
  - 2LIS_02_VAHDR
  - 2LIS_02_VAITM

- Extraction done via RDA and InfoPackage scheduling (“Normal Delta”) can currently not be done in parallel
  - DataSource can only use a single extraction mechanism

- Well established Data Flow can can be used
- Separate (potentially) leaner DataStore Object can be used
- Reporting has to be enabled via MultiProvider
- Data has to be deleted frequently from „RDA“ DataStore Object to avoid redundant data

- Old data flow has to be replaced by new real-time data flow
- No issues with data alignment from different DataSources
- Can facilitate upload of huge data volumes by spreading data load over many smaller data loads

© SAP 2007 / Page 38
1. Overview

2. ETL
   2.1. Extraction
   2.2. Transformation
   2.3. Data Distribution
   2.4. Real-Time Data Acquisition

3. Distributed Query / EII
Direct/Remote Access Scenarios

- Integration of external data into BI Analyses without storing data persistently into BI (Example: Teradata Integration)
- SAP NetWeaver BI is owning the data model
- Persistency of the transient data might be an option for the future (integration can be the first step into an migration scenario)
- Master data of transient transactional data has to be available in the BI environment
VirtualProvider

- Scenarios
  - Data stored outside BI (other SAP or Non-SAP systems)
  - Analysis requires complex algorithms which can be implemented via VirtualProvider
- Support of InfoObjects for direct access
- Types:
  - BAPI
  - Function Module
  - DTP
Direct Access: Simple Example

- BI
- Query
- VirtualProvider
- DataSource (direct access)
- SourceSystem

DataTransfer Process for Direct Access

Transformation
Direct Access: Complex Example II

BI

VirtualProvider

DataSource (direct access)

TR

DataSource (direct access)

Remote InfoObject

DTP for Direct Access

Source System

Transactional Data

Master Data

BI

DTP for Direct Access

DataSource (direct access)

TR

DataSource (direct access)

Remote InfoObject

DTP for Direct Access

Source System

Transactional Data

Master Data
Remote Characteristic

SID table attr. /BIC/XPM_NWCUST

- Medium length text exists
- Long text exists
- Texts language dependent
- Texts are time-dep.

MstDataMaint with Authorization

DS Object for Check

Master Data InfoSource / Data Target / InfoProvider

- InfoSource with Direct Update
  - Application Component ZPM REM_ACCESS Remote Access Scenario
  - Character. is InfoProvider
    - InfoArea RABI70 M02 RABI70 Remote Access
- Characterist. is export data source

Characteristic Type Remote Characteristic

Name der Stammdaten Lese-Klasse CL_RSR_REMOTE_MASTERDATA
No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP AG. The information contained herein may be changed without prior notice.

Some software products marketed by SAP AG and its distributors contain proprietary software components of other software vendors.

SAP, R/3, mySAP, mySAP.com, xApps, xApp, SAP NetWeaver, Duet, Business ByDesign, ByDesign, PartnerEdge and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP AG in Germany and in several other countries all over the world. All other product and service names mentioned and associated logos displayed are trademarks of their respective companies. Data contained in this document serves informational purposes only. National product specifications may vary.

The information in this document is proprietary to SAP. This document is a preliminary version and not subject to your license agreement or any other agreement with SAP. This document contains only intended strategies, developments, and functionalities of the SAP® product and is not intended to be binding upon SAP to any particular course of business, product strategy, and/or development. SAP assumes no responsibility for errors or omissions in this document. SAP does not warrant the accuracy or completeness of the information, text, graphics, links, or other items contained within this material. This document is provided without a warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability, fitness for a particular purpose, or non-infringement.

SAP shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials. This limitation shall not apply in cases of intent or gross negligence.

The statutory liability for personal injury and defective products is not affected. SAP has no control over the information that you may access through the use of hot links contained in these materials and does not endorse your use of third-party Web pages nor provide any warranty whatsoever relating to third-party Web pages.