Data Acquisition Techniques in SAP Netweaver BW BI

Applies to:
SAP BW 3.5, SAP BI 7.0 etc. For more information, visit the [EDW homepage](http://www.sap.com)

Summary
This paper discusses the various sources available for the data transfer to BI and how the sources are connected to the BI system as source systems. They also describe how the data can be transferred from the sources.

Author: Ramam Velivala
Company: Deloitte Consulting
Created on: 18th June 2011

Author Bio
Ramam Velivala is a Lead consultant with SAP BW/BI, BPC and EPM practice at Deloitte. He is currently leading the SAP-BW track integration with EPM tools. He has extensively worked on PCM modeling to extract data to SAP-BW and BPC systems. He has more than 12 years of SAP experience in providing analytic solution in various SAP functional modules like Finance, Controlling, Procurement etc.
# Table of Contents

Table of Contents .................................................................................................................. 2

Business Requirement .......................................................................................................... 3

Source Systems ....................................................................................................................... 3
  SAP Source Systems: ........................................................................................................... 4
  Non-SAP Source Systems: ................................................................................................... 4

Integration ............................................................................................................................... 4
  Data Extraction from SAP Source Systems ......................................................................... 5
  Available Integration Mechanisms & Experience ................................................................. 8
  Database (DB) Connect .......................................................................................................... 8
  Universal Data (UD) Connect - Architecture ....................................................................... 9
  Legacy Applications .............................................................................................................. 11
  File System ........................................................................................................................ 12
  Web Service (XML DataSources) .......................................................................................... 13

Data Acquisition Capabilities .............................................................................................. 15
  Real Time Data Acquisition ................................................................................................. 17

Related Content ..................................................................................................................... 19

Disclaimer and Liability Notice ............................................................................................ 20
**Business Requirement**

Data retrieval is one of the data warehousing processes in BI. BI provides mechanisms for retrieving data (master data, transaction data and metadata) from various sources.

We will discuss the various sources available for the data transfer to BI and how the sources are connected to the BI system as source systems. They also describe how the data can be transferred from the sources.

The graphic shows the sources and transfer mechanisms.

![Data Acquisition Techniques Diagram](image)

From the above diagram we can see that there are multiple ways of pulling data to BI system. These all systems are named in BI as Source Systems.

**Source Systems**

All systems that provide BI with data are described as source systems. These can be:

- SAP systems
- BI systems
- Flat files for which metadata maintained manually and transferred to BW is using a file interface.
- Database management systems which data is loaded into from a database supported by SAP using DB Connect, without using an external extraction program

- Relational sources that are connected to BI using UD Connect.
- Web Services that transfer data to BI by push.
- Non-SAP systems for which data and metadata is transferred using staging BAPIs.
SAP Source Systems:
- We can pull the data from SAP systems like SAP ECC, SAP CRM, SAP SRM systems using Extractors.
- We can use application-specific extractors or create generic extractors according to our requirement.
- Communication between the SAP source system and BI is only possible if the appropriate plug-in is installed in the source system.

Non-SAP Source Systems:
Business Executives and Analysts have long been merging non-SAP information with SAP information for measuring the pulse of the business. This merging of information has been done for a multitude of reasons including understanding the impact of trade promotions on retail sales for a CPG company, the efficacy of a transportation route planner, or even merging internal financial information that resides in multiple business systems.

Our clients with SAP BW footprint have integrated non-SAP data with SAP data in SAP BW to meet many of the above requirements. Integration of non-SAP data to BW is similar to integrating data from multiple sources in an Enterprise Data Warehouse. Though, there are some nuances specific to the tool and the available integration mechanism, the fundamental approach remains similar. Deloitte has a number of years of experience implementing BW with Non-SAP Sources and External Data Source.

Selecting the right approach for Integration depends on a number of factors including:
- Business Requirement (Purpose of use, Frequency of Load, Latency etc.)
- Integration Methods Available in the Source
- Data Volumes etc.

This is brief information on our implementations, experiences, and issues or challenges that we face and the benefits that we observed of using SAP BW from our experiences.

Integration
DataSources are used to extract and stage data from source systems. The DataSources divide the data provided by a source system into self-contained business areas.

The Source Systems can be mainly categorized into two types.
1) SAP Source Systems
2) Non-SAP Source Systems

The following graphic provides an overview of the data transfer sources supported by BI and shows the interfaces that you can use:
Data Extraction from SAP Source Systems

Sources
SAP applications, for example mySAP ERP, mySAP CRM, SAP NetWeaver BI.

Implementation
1. Available Business Content will heavily reduce implementation effort
   - Installation of Business Content DataSources on the SAP source system using the Plug-in (PI) software component. ECC 6.0, they will be part of SAP ECC.
   - Installation of technical interfaces on the source system using the SAP Basis Plug-in (PIB).

2. Detailed knowledge of business data and Service API is only required if additional data fields or DataSources are needed.

Purpose
Extractors are part of the data retrieval mechanisms in the SAP source system. An extractor can fill the extraction structure of a DataSource with the data from SAP source system datasets.

Replication makes the DataSource and its relevant properties known in BI.

For the data transfer to the input layer of BI, the Persistent Staging Area (PSA), define the load process with an InfoPackage in the scheduler. The data load process is triggered by a request IDoc to the source system when the InfoPackage is executed.
Process Flow

There are two types of extractions from Source System. Below diagram depicts the same.

**TYPES OF EXTRACTIONS**

- **APPLICATION SPECIFIC**
  - BW CONTENT EXTRACTOR
  - FI
  - HR, CO
  - SAP CRM
  - LO COCKPIT
  - DB TABLES
- **CROSS APPLICATION**
  - GENERIC EXTRACTOR
  - DF VIEW
  - INFOSET
  - FUNCTION MODULE
  - DB TABLES
  - ENHANCEMENTS
  - DELTA METHOD
  - BASICS

There are application-specific extractors, each of which is hard-coded for the DataSource that was delivered with BI Content, and which fill the extraction structure of this DataSource.

In addition, there are generic extractors with which we can extract more data from the SAP source system and transfer it into BI. Only when we call up the generic extractor by naming the DataSource does it know which data is to be extracted, and from which tables it should read it from and in which structure. This is how it fills different extraction structures and DataSources.

You can run generic data extraction in the SAP source system application areas such as LIS, CO-PA, FI-SL and HR. This is how LIS, for example, uses a generic extractor to read info structures. DataSources are generated on the basis of these (individually) defined info structures.

Regardless of application, we can generically extract master data attributes or texts, or transaction data from all transparent tables, database views or SAP query functional areas or using the function module. We can generate user-specific DataSources here.

The OLTP extraction tables form the basis of a DataSource on an R/3 OLTP system. The structure is written in the OLTP using the data elements that describe the available data, usually from a table view. For an R/3 OLTP source system, the ‘DataSource Replication’ step is provided to duplicate the DataSource which is replicated with its relevant properties in BW. The user can request the Metadata for a DataSource, the Metadata for an application component, or all the Metadata of a source system.
Replication of DataSources

Replication of the Entire Metadata
a. Choose Application Component Hierarchy and DataSources of a Source System
b. Choose the option Replicate DataSources in the Data Warehousing Workbench in the source system tree through source system context menu

Replication of the Application Component Hierarchy of a Source System
a. Choose Replicate Tree Metadata in the Data Warehousing Workbench in the DataSource tree through the root node context menu.

Replication of the Metadata
a. (DataSources and Possibly Application Components) of an Application Component Choose the option Replicate Metadata in the Data Warehousing Workbench in the DataSource tree through an application component context menu.

Replication of a DataSource of a Source System
a. Choose Replicate Metadata in the Data Warehousing Workbench in the DataSource tree through a DataSource context menu. or
b. In the initial screen of the DataSource repository (transaction RSDS), select the source system and the DataSource and then choose DataSource  Replicate DataSource.

All application data must be described in SAP BW using Meta data. The Infoobjects used for this are not just transaction and master data but also relationship sets such as attributes or hierarchies for master data. The SAP BW extractors carry out a number of functions in the SAP OLTP, in order to guarantee smooth communication between SAP OLTP and SAP BW.

These tasks must be carried out by external software in non-SAP Systems that is set up in the BAPI interfaces of the SAP BW. Various businesses have already been certified as official third party providers in this field. A complete list can be found on the SAP net BW homepage.

There are two methods for transfer of data to BW system from SAP Systems.

- With the IDoc method, IDoc interface technology is used to pack the data into IDoc containers.
- With the PSA transfer method, IDoc containers are not used to send the data. Instead, the data is transferred directly in the form of a transfer structure.

Information is sent from the source system (no data) through the IDoc interface (info IDoc‟s). This information can be, for example, the number of data records extracted or information on the monitor.

The differences and advantages between the two methods are listed below.

**PSA**
1. Data record length Max. 1962 bytes.
2. Number of fields per data record: Restricted to 255
3. Uses TRFC as transfer log.
4. Advantage: Improved performance since larger data packages can be transported. Error handling is possible.
5. More common technology since it brings with it a better load performance and gives the user the option of using the PSA as an inbound data store (For Master and Transaction data).

**ALE (data IDoc)**
1. Data record length Max. 1000 bytes.
2. Uses TRFC as transfer log
3. Advantage: More detailed log through control record and status record for data IDoc.
4. Use with hierarchies.

**Operation**

Systems are physically connected using ALE / tRFC technology allowing for mass data transfer

Data transfer to the BI system is done in parallel (degree of parallelization can be configured).
Available Integration Mechanisms & Experience

Integrating data between various systems where small volumes of data need to flow on a regular basis so as to support real-time data requirements

Approach Used: PI interface to connect various systems to SAP BW.

Client(s): FDOR, Caterpillar and Checkpoint.

Database (DB) Connect

Sources

Native SQL connection to all DB platforms supported by the SAP NetWeaver AS

Implementation

System connection is enabled via the multi-connect feature of the SAP NetWeaver Web Application Server DB Interface.

- SAP DBSL (Database Shared Library) has to be installed on the DB to be connected to SAP NetWeaver BI. Additionally, the relevant DB Client is required.
- Central maintenance of DB connection parameters in SAP NetWeaver BI (server address, etc.)

Meta data of source tables and views as automatic proposal (since SAP NetWeaver 2004s BI) allowing for easy implementation of DataSources.

DB specific skills are necessary for implementation of DB views (recommendation) for extraction of required information.

Purpose

The DB Connect enhancements to database interface allow the user to transfer data straight into BI from the database tables or views of external applications. The user can use tables and views in database management systems that are supported by SAP to transfer data. The user uses Data Sources to make the data known to BI. The data is processed in BI in the same way as data from all other sources.

By default, when the BI application server starts, SAP kernel opens a connection to the database on which the SAP system is running. In the remainder of this section, this connection is referred to as the (SAP) default connection. All SQL commands that are submitted by the SAP kernel or ABAP programs (irrespective of whether they are open or native SQL commands), automatically refer to this default connection; they run in the context of the database transaction that is active in this connection. Connection data, such as database user name, user password, or database name are taken either from the profile parameters or from the corresponding environment variables (this is database specific).
We use DB Connect to open other database connections in addition to the default connection and use these connections to transfer data into a BI system from tables or views.

There are 2 types of classification. One is the BI DBMS & the other is source DBMS. The main thing is, both these DBMS are supported on their respective operating system versions, only if SAP has released a DBSL. If not, they don’t meet the requirements and hence can’t perform DB Connect. In this process we use a Data source, to make the data available to BI & transfer the data to the respective Info providers defined in BI system.

Further, using the usual data accusation process we transfer data from DBs to BI system.

Using this SAP provides options for extracting data from external systems, in addition to extracting data using standard connection; the user can extract data from tables/views in database management systems (DBMS).

**Operation**

Systems are physically connected on DB level allowing for powerful data transfer.

Persistent and transient (since SAP NetWeaver 2004s) data handling possible. Data does not need to be stored in BI but can be accessed remotely.

**Universal Data (UD) Connect - Architecture**
Sources

- Java connectors for different drivers, providers, and protocols available (JDBC, OLE DB for OLAP) providing connectivity to a wide range of relational (MS SQL Server, Oracle) and multi-dimensional sources (Hyperion).
- Based on the SAP NetWeaver J2EE connection framework.

Implementation

- Basis skills in configuring connectors on the SAP NetWeaver J2EE engine are required.
- Meta data of source tables and views as automatic proposal (since SAP NetWeaver 2004s BI) allowing for easy implementation of DataSources.
- Source system specific skills might be necessary if for extraction purposes additional objects have to be implemented or created in the source system (drivers, views).

Purpose

UD Connect (Universal Data Connect) uses Application Server J2EE connectivity to enable reporting and analysis of relational SAP and non-SAP data. To connect to data sources, UD Connect can use the JCA-compatible (J2EE Connector Architecture) BI Java Connector.

Firstly, the user can extract the data, load it into BI and store it there physically.

Secondly, provided that the conditions for this are met, the user can read the data directly in the source using a Virtual Provider.

Process Flow

1. Create the connection to the data source with the user relational or multi-dimensional source objects (relational database management system with tables and views) on the J2EE Engine.
2. Create RFC destinations on the J2EE Engine and in BI to enable communication between the J2EE Engine and BI. In the Implementation Guide for SAP NetWeaver →Business Intelligence →UDI Settings by Purpose →UD Connect Settings.
3. Model the Infoobjects required in accordance with the source object elements in BI.
4. Define a DataSource in BI.

Prerequisites
We have to install the J2EE Engine with BI Java components.

Operation
Persistent and transient data handling Possible. Data does not need to be stored in BI but can be accessed remotely.

Available Integration Mechanisms & Experience
Integrating data from another non-SAP BW data warehouse and/or database.
Approach Used: UD Connect & DB connect to connect to a Oracle DBMS
Client(s): Suncor.

Legacy Applications

Sources
• Potential sources depend on the capabilities of the Certified Software Partner tool. Usually, specialized tools offer connectivity to a wide range of sources (example: various ERP applications such as Siebel or PeopleSoft).

Implementation
• In the area of data migration, Certified Software Partners and ETL (Extraction, Transformation and Loading) vendors connect their solutions to SAP NetWeaver BI using an open interface, the Staging BAPI (Business Application Programming Interface).
• Tools can connect to DataSources in SAP NetWeaver BI including Meta data transfer.
• Additional connectivity and features (if applicable) require additional license fees, skills and hardware. Alternatively, using UD Connect or DB Connect should be investigated.
• The new DataSource concept in SAP NetWeaver 2004s BI is not available in a BAPI. Alternatively, you can continue using the SAP BW 3.x DataSource by still using new features such as the new transformation

Purpose
In order to enable extraction of data and metadata from non-SAP sources on the application level, SAP BW provides open interfaces - staging BAPIs. BAPIs (Business Application Programming Interface) is standardized programming interfaces that offer external access to the business processes and data of a SAP system. These interfaces enable connection between various third-party tools (such as Extraction, Transformation and Loading) and SAP BW. In this way, for example, data from an Oracle application can be transferred to SAP BW and can be evaluated there.

Operation
Administration of a possibly high number of heterogeneous sources on ETL tool level possible. Integration into BI monitors is provided, too.

File System

Sources
CSV (1234;AB;20,00;EUR) or ASCII (1234AB2000EUR) files

Implementation
• Meta data of source files as automatic proposal (since SAP NetWeaver 2004s BI) allowing for easy implementation of DataSources
• Automated conversion of external data types provided
• Rapid implementation (example: prototyping)
• No further skills or system connections are required, but conversion to CSV or ASCII format might be necessary

Purpose
SAP BW supports the transfer of data from flat files, files in ASCII format (American Standard Code for Information Interchange) or CSV format (Comma Separated Value). For example, if budget planning for a company’s branch offices is done in Microsoft Excel, this planning data can be loaded into SAP BW so that a
plan-actual comparison can be performed. The data for the flat file can be transferred to SAP BW from a workstation or from an application server.

Flat files are data files that contain records with no structured relationships. Additional knowledge is required to interpret these files such as the file format properties. Modern database management systems use a more structured approach to file management (such as one defined by the Structured Query Language) and therefore have more complex storage arrangements.

Many database management systems offer the option to export data to comma delimited file. This type of file contains no inherent information about the data and interpretation requires additional knowledge. For this reason, this type of file can be referred to as a flat file.

For example, .csv is comma separated flat file, .txt, .lis, .lst etc...

**Operation**

- Data quality and availability of files (on a central server) are main factors for efficient periodic data loads
- Persistent and transient (since SAP NetWeaver 2004s) data handling possible. Data does not need to be stored in BI but can be accessed remotely.

**Available Integration Mechanisms & Experience**

Historical data extraction to BW from non-SAP systems for a “one-time” load during the implementation.

Approach Used: “Flat File load”

Client(s): ACH Foods (Historical Data Extraction for Integrated Planning) and AIG systems.

**Web Service (XML DataSources)**
Implementation

- Easy Meta Data Maintenance of the DataSource in SAP NetWeaver BI including automatic generation of a Web Service and SOAP enabled function module.
- BI Data Load Web Services allow for rapid implementation of easy XML based data staging scenarios.
- SAP NetWeaver XI (exchange infrastructure) technology is essential for optimal integration of business applications with SAP NetWeaver BI using the Web Service source system type.
- Depending on the scenario, additional skills on Web Service implementation or SAP NetWeaver XI are necessary.

Purpose

Data is generally transferred into SAP BW by means of a data request, which is sent from SAP BW to the source system (pull from the scheduler). You can also send the data to SAP BW from outside the system. This is a data push into SAP BW.

A data push is possible for various scenarios:

- Transferring Data Using the SOAP Service SAP Web AS
- Transferring Data Using Web Services
- Transferring Data Using SAP XI

In all three scenarios, data transfer takes place using transfer mechanisms that are sufficient for Simple Object Access Protocol (SOAP); the data transfer is also XML-based.

Operation

- Periodic loads are scheduled in SAP NetWeaver BI loading the data from the PSA to InfoProviders.
- SAP NetWeaver XI is a central platform for integrating internal and / or external applications and offers additional central monitoring.
- SAP NetWeaver XI is scalable to handle high data volumes

Available Integration Mechanisms & Experience

Integrating with an Enterprise Web service
Approach Used: Web Services within SAP BW

Client(s): Carlsberg, Heineken
Web Service integration with SAP BW via BoDI
Client(s): Sysco.
SAP BW Integration with Teradata, Informatica and 3rd party Tools:

- Integration with Teradata by using virtual cubes. Deloitte consultants actually worked with the Product Development team in developing this integration mechanism (TVAS).
  Client(s): cCCE, Minsheng Bank and Bayer

- Integration with Teradata and Informatica via Open Hub method.
  Client(s): Warner Brothers.

- Essbase, Redwood and Cognos Reporting tool through the interfaces
  Client(s): Wal-Mart, Peabody Energy and The Irwin Company.

Data Acquisition Capabilities
“Pulling” vs. “Pushing” data into SAP NetWeaver BI

There are two mechanisms in which data can be retrieved into BI. They are

1) Data Pull
2) Data Push

**Data Pull:**
This is a standard mechanism in which data can be pulled into BI by executing an InfoPackage, as shown in the above diagram.

**Data Push:**
In data push mechanism, the data is transferred into BW via a generated RFC enabled Function module. Based on that there are the following implementations:

- SOAP-DataSource (XML-DataSource)
- Web Service
- BW-XI-Integration
**Real Time Data Acquisition**

Real-time data acquisition supports tactical decision-making. It also supports operational reporting by allowing you to send data to the delta queue or PSA table in real-time. You then use a daemon to transfer the data to Data Store objects to the operational Data Store layer at frequent regular intervals. The data is stored persistently in BI.

Real Time Data Transfer is recommended if transfer of data to BI at shorter intervals (every minute) than scheduled data transfers and you need up-to-date data to be regularly available in reporting (several times a day, at least).

The following overview displays the differences between standard data acquisition using scheduled data requests and real-time data acquisition:

<table>
<thead>
<tr>
<th></th>
<th>Standard Data Acquisition</th>
<th>Real-Time Data Acquisition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Goal</strong></td>
<td>Strategic Decision-Making (long-term planning)</td>
<td>Tactical Decision-Making (daily decisions)</td>
</tr>
<tr>
<td><strong>Data Acquisition</strong></td>
<td>Request Oriented (nightly background job)</td>
<td>Data Availability Oriented (permanently active background job)</td>
</tr>
<tr>
<td><strong>Load Cycle</strong></td>
<td>1/day ... 1/week</td>
<td>1 minute ... 1/hour</td>
</tr>
<tr>
<td><strong>Resource Consumption</strong></td>
<td>Processing normally nights (load balancing)</td>
<td>Permanent Resource Consumption</td>
</tr>
</tbody>
</table>
But there are some prerequisites to be there for Real Time Data Acquisition to happen. They are:

1) The DataSource has to support real-time data acquisition.

Web service DataSources and DataSources from SAP source systems can support real-time data acquisition. In DataSource maintenance, the Extraction tab page shows whether the Real Time property is set.

DataSources from SAP source systems can be used for real-time data acquisition if the following prerequisites are met:

- BI Content DataSources have to be delivered with the property for supporting real-time data acquisition.
- The Real-Time-Enabl. Indicator has to be set in the generic delta settings for generic DataSources (more information: Delta Transfer to BI).

2) You use a transformation between the PSA and Data Store object and update data into Data Store objects from the PSA using the data transfer process.

Data is loaded into BI at frequent, regular intervals and is then posted to the Data Store objects that are available for operational reporting. In BI, special Info Packages are used for this purpose and data transfer processes for real-time data acquisition are created to further process data from the PSA in the Data Store objects. This is scheduled and executed regularly by a dedicated background process (the daemon). Data is available for reporting as soon as it has been successfully posted to the Data Store object and activated. Refresh the query display to display the recent data. The query shows the time that the query was last updated by a daemon run, even if no new data was posted.

Data can transferred from the source to the entry layer of BI (the PSA) in two ways:

- Using a Web service
  
  We can use the Web service to write the data from the source into the PSA. The transfer of data is controlled externally, without a request from BI. Only an InfoPackage (for full upload) is required to determine specific parameters for real-time data acquisition.

- Using a service API
  
  Data from an SAP source system can be loaded into the PSA using an InfoPackage created specifically for this purpose. This is triggered when the delta queue in the source system requests data. You have to simulate the initialization of the delta process for the DataSource beforehand.

The following two scenarios are possible:

- The source system application writes the data to the delta queue.
  
  In this case, the daemon retrieves the data without calling the extractor.

- The application does not write data to the delta queue automatically; the extractor writes the data to the delta queue at the request of BI.
  
  For extractors that transfer data synchronously from BI to the service API on request (generic extractors, for example), the daemon calls the extractor, and the extractor writes the data to the delta queue. The data is transferred to BI directly from the delta queue.

- Converting an Existing Data Flow to Real-Time Data Acquisition
  
  If we want to integrate the transfer of data with real-time data acquisition into an existing data flow, there are two options:

  - Using two different DataSources
    
    One DataSource executes the standard data transfer. The other DataSource transfers the data with real-time data acquisition. The data is then combined in a MultiProvider.

  - Using a single DataSource
    
    You have to replace the standard data transfer completely with a real-time data acquisition scenario.
Related Content

For more information, visit the EDW homepage
Disclaimer and Liability Notice

This document may discuss sample coding or other information that does not include SAP official interfaces and therefore is not supported by SAP. Changes made based on this information are not supported and can be overwritten during an upgrade.

SAP will not be held liable for any damages caused by using or misusing the information, code or methods suggested in this document, and anyone using these methods does so at his/her own risk.

SAP offers no guarantees and assumes no responsibility or liability of any type with respect to the content of this technical article or code sample, including any liability resulting from incompatibility between the content within this document and the materials and services offered by SAP. You agree that you will not hold, or seek to hold, SAP responsible or liable with respect to the content of this document.