Finance and Risk Data Management in Banking
Best-Practices Approach to Implementing an Integrated Information Management Platform
# Table of Contents

1. Content
2. Purpose of This Document
3. Logical Reference Architecture for Information Management
   - Layers of the Information Management Architecture
   - Data Integration and Data Modeling
4. Physical Deployment of Reference Architecture
   - Source Systems
   - Data Acquisition Layer
   - Propagation Layer
   - Transformation and Access Layer
   - Modeling and Data Architecture
5. SAP's Vision and Its Impact on Information Architecture
   - Consolidation of the System Landscape on SAP HANA
   - Less Materialization, More Virtualization
   - New Options for Integration
   - From ETL to ELT
   - Data Aging
   - Data Federation
6. Find Out More
Content

The contents of this document shall remain the confidential property of SAP and may not be communicated to any other party without the prior written approval of SAP. This document must not be reproduced in whole or in part. It must not be used other than for evaluation purposes only, except with the prior written consent of SAP and then only on condition that SAP’s and any other copyright notices are included in such reproduction. No information as to the contents or subject matter of this proposal or any part shall be given or communicated in any manner whatsoever to any third party without the prior written consent of SAP.

The furnishing of this document is subject to contract and shall not be construed as an offer or as constituting a binding agreement on the part of SAP to enter into any relationship. SAP provides this document as guidance only to estimate costs and time scales of the predicted delivery project. This will be subject to confirmation prior to any contractual or delivery commitment by SAP.

SAP warrants that to the best of its knowledge those who prepared this material have taken all reasonable care in preparing it and believe its contents to be true as of the date of this document. SAP cannot, however, warrant the truth of matters outside of its control and accordingly does not warrant the truth of all statements set out in this document to the extent that such statements derive from facts and matters supplied by other persons to SAP. The statements in this document are qualified accordingly.

SAP uses a layered approach in its reference architecture for finance and risk data management. It is designed to account for the complex, transformational, and deeply analytical processes of the banking industry.
Purpose of This Document

The information management landscape of banks is typically very heterogeneous. Departmental silo solutions and diverse technologies drive up cost and make it difficult to meet enterprise-wide reporting needs. Adding to these challenges are the recent mortgage and European monetary crises, which have led to new regulatory requirements that focus on finance and risk. Innovations such as in-memory technology now provide opportunities for banks to optimize and simplify their solution architecture.

Today’s banking institutions have many reasons to consolidate their information management landscapes, ranging from cost reduction, to greater business insight, to improved market agility. As a result, installing an integrated finance and risk data platform as a single source of truth is now of high strategic importance throughout the banking industry.

This document describes SAP’s reference architecture for retail and commercial banks that want to implement a consolidated finance and risk data platform. It specifically focuses on how to integrate finance and risk data marts and applications into the architecture.

In addition to explaining the logical reference architecture, this document details the physical implementation of the reference architecture based on SAP® software. It also provides an outlook of how technology innovations are expected to change physical implementations in the future.

It is not the goal of this document to provide a blueprint fitting for all customers. Individual customer circumstances may require specific approaches.

We built this architecture in collaboration with some of our biggest customers. Based on industry best practices, its design is derived from our experience in successfully delivering data warehouses to enterprises over the years.
SAP uses a layered approach in its reference architecture for finance and risk data management (see Figure 1). It is designed to account for the complex, transformational, and deeply analytical processes of the banking industry that often need to be delivered through off-the-shelf analytical calculators.

The reference architecture is based on what we consider best practices in the industry. We built it through collaboration with and feedback from some of our biggest customers. Its design is also derived from our experience in successfully delivering data warehouses to enterprises over the years.

With the invention of in-memory databases, not all the layers below need to materialize. This especially holds true for the transformation and access layer and its data marts. Whereas the logical architecture remains the same when you add an in-memory database, the physical implementation is less complex, and you can address new requirements more quickly.

The layered architecture follows a practical and incremental approach to building out a solution. Some of the other benefits offered with the layered approach include:

- **Low total cost of ownership**
  A layered approach helps ensure data flow from source systems to business users. Reporting and analysis happens in a well-thought-out manner. This approach negates the need to constantly revert back to source systems for data, as it is likely found in the propagation layer and ready for business-user consumption.

---

**Figure 1: Logical View of Architecture for Finance and Risk Data Management**

<table>
<thead>
<tr>
<th>Data modeling</th>
<th>Data integration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Enterprise data warehouse</strong></td>
<td></td>
</tr>
<tr>
<td>Data acquisition layer</td>
<td></td>
</tr>
<tr>
<td>• Extraction and staging</td>
<td></td>
</tr>
<tr>
<td>• Cleansing</td>
<td></td>
</tr>
<tr>
<td>Propagation layer</td>
<td></td>
</tr>
<tr>
<td>• Normalization and storage</td>
<td></td>
</tr>
<tr>
<td>Transformation and access layer</td>
<td></td>
</tr>
<tr>
<td>• Transformation and calculation</td>
<td></td>
</tr>
<tr>
<td>• Performance and access</td>
<td></td>
</tr>
<tr>
<td>BI abstraction and reporting layer</td>
<td></td>
</tr>
<tr>
<td>Dashboard and reports</td>
<td></td>
</tr>
<tr>
<td>Ad hoc analysis</td>
<td></td>
</tr>
<tr>
<td>Mobile</td>
<td></td>
</tr>
</tbody>
</table>
• Flexibility in adapting to changing business and regulatory requirements
New and changing business requirements often require employees to cross-reference and report on data from multiple sources. In the layered architecture approach, all source systems converge at the propagation layer, which is the well-defined, well-understood, single-point-of-truth data store for most source systems within an enterprise. New projects can satisfy their data requirements from the propagation layer and avoid time-consuming and expensive integration with the source systems.

• Data consolidation and harmonization in heterogeneous system landscapes
You can integrate data from multiple sources in a heterogeneous system landscape. The architecture provides a clear separation in which the data acquisition layer represents a one-to-one copy of the source system, and the propagation layer provides a harmonized and cleansed view of data across your system landscape.

• Historized data
You can historize data to reproduce data snapshots back in history. This is an important requirement in many scenarios based on regulations.

• Governance
The clear semantical definition and modeling rules for the layers provide a governance that helps you manage complex environments, reduce total cost of ownership (TCO), and maintain flexibility to address future requirements.

LAyERS oF The InFoRMATIon MAnAgeMenT ARChITeCTuRe

Our reference architecture for integrated finance and risk data management consists of the following layers.

Source Layer
This layer demarcates the full range of heterogeneous data sources. The scope of data is as broad as possible, including sources such as:
• Master data
• Transactional data from core banking systems
• Enterprise resource planning (ERP) system
• Customer relationship management (CRM) system
• Local data such as Excel spreadsheets
• Structured and unstructured data sources
• Social media sources

Data Acquisition Layer
The data acquisition layer is the first layer in the finance and risk warehouse. This is a crucial layer responsible for providing a single landing point, as well as verifying and cleansing data for all source systems in scope. Without this layer, an enterprise tends to have multiple silos of information that are impossible to reconcile or cross-reference. Having this layer in place avoids the necessity to revert back to the source systems for every new requirement.

The exact nature of this layer can vary from one enterprise to another. It is often dependent on the source systems, processes, and mechanisms in place for the change-data-capture process. If the source layer does not provide a consistent change-data-capture delivery mechanism, then this layer must also provide that functionality within the information management architecture.

The staging sector of this layer receives raw data from the source systems, with a complete set of data elements. The data remains raw and untransformed. The data is typically retained for only a brief period of time, long enough to allow the data to be processed through to subsequent layers and facilitate resolution of loading issues.

Propagation Layer
The propagation layer acts as the single source of truth for all enterprise data that is deemed relevant for strategic decision making and multipurpose by nature. Some of the characteristics of the propagation layer include the following:
• Data is held at the lowest level of fact data. This helps ensure that future project teams are able to find the same level of detail in the propagation layer that they would find in the source systems.
• Data is considered to be immutable and indisputable. No interpretation should have been placed on the data, nor data changed.
• **Data is stored in a normalized data model.** As a result, data is not constrained by the application or source system data model. A normalized data model is usually the best way to store data in the propagation layer. Not only does it reduce data redundancy, but it also helps capture business relationships between different entities within an organization.

**Transformation and Access Layer**
The transformation and access layer takes into account application-specific data formats and performance optimizations. Application-specific logic is applied to the data to enrich it with calculation results. Applications that use the propagation layer as a data hub belong logically to the transformation and access layer. Multipurpose calculation results are added to the propagation layer. The transformation and access layer usually consists of, but is not limited to, the following:

- Data marts to support specific business processes
- Analytical calculators for complex business logic
- Valuations in accordance with applicable generally accepted accounting principles (GAAP) for financial accounting purposes
- Denormalized data expressed as views for easier interpretation and consumption by the business community
- Support for extract, transform, and load (ETL) processes to refresh data marts and data in analytical calculators

With the advent of in-memory database technologies, the access and performance component of this layer is often not materialized, but rather is a virtual view of data from the propagation layer. This delivers huge benefits by dramatically reducing the number and complexity of ETL processes to maintain this layer.

**Business Intelligence Abstraction and Reporting Layer**
The business intelligence (BI) abstraction and reporting layer should provide the broadest range of tools for data access. It should not be limited to just Web reporting and dashboards but should also include tools for data exploration and prediction. In addition, it should include well-defined interfaces for consumption of data from the finance and risk warehouse.

Another objective of this layer is to shield business users and applications from change to the physical structure of the finance and risk warehouse. To meet this objective, this layer should contain a layer of abstraction in the form of a logical data model that is separated and segregated from the physical data model of the finance and risk warehouse.

The BI abstraction and reporting layer also provides federation functionality for information from disparate sources. The true federation should be delivered from the finance and risk warehouse. However, you can use federations in this layer to deliver quick business value as the finance and risk warehouse is incrementally built out.

**DATA INTEGRATION AND DATA MODELING**

Our reference architecture includes tools for enterprise information management. These tools support data integration and quality management by consolidating data from multiple sources into the finance and risk warehouse and across its diverse layers.

In heterogeneous system landscapes, modeling tools add incredible value. Effective oversight of data representations, semantics, and models is key for successful implementations. Our reference architecture includes tools for data modeling and metadata management. These tools help establish governance processes and translate business requirements into IT requirements.
Physical Deployment of Reference Architecture

This section describes the physical implementation of SAP’s reference architecture for retail and commercial banks that want to deploy a central finance and risk warehouse using SAP software. It places specific emphasis on integrating finance and risk data marts and applications. The architecture allows an incremental approach to build out the information architecture across different lines of business. Figure 2 depicts SAP’s proposed solution.

SOURCE SYSTEMS

To build out a data warehouse, one of the first steps is to identify the source systems that need to supply the data to meet business and regulatory requirements.

DATA ACQUISITION LAYER

The data acquisition layer provides three functions:

- Extraction and staging
- Normalization and cleansing
- Changed data delivery (delta interfaces), where possible

Details for each key function are provided below.

Extraction and Staging

Once you identify the source systems, you must bring the data from these sources into the acquisition layer. How the source systems deliver the data is primarily dependent on the following aspects:

- Nature of the source application and the platform that hosts it
- Data latency requirements of the data warehouse

The following governing principles hold true with regard to the extraction solution:

- As noninvasive as possible with regard to the source systems
- As simple as possible to reduce complexity
- As cost-effective as possible

Methods for Data Extraction from Source Systems

The following four SAP products represent the most important data integration technologies in our reference architecture. The SAP Landscape Transformation replication server and SAP Sybase® Replication Server® software are predominantly used for data integration on a database level. SAP Data Services software can also be used for data integration via the application layer. The SAP Sybase Event Stream Processor (SAP Sybase ESP) is the right choice for scenarios that require high-speed data streaming and event handling. Other integration interfaces (such as the BAPI® programming interface and Web services) exist and might be considered for specific use cases as well.

Figure 2: Physical View of Architecture for Finance and Risk Data Management
SAP Data Services
For an ETL-based integration with source systems, SAP Data Services provides data integration and data quality processing functionalities. The solution allows you to integrate, transform, improve, and deliver trusted data for both SAP and non-SAP software systems. It supports many connectivity options, ranging from SAP applications, to databases, to other vendor applications (such as Microsoft Excel). SAP Data Services can integrate with the data acquisition layer via the application layer (for instance, the SAP NetWeaver® Business Warehouse application) or the database layer (for instance, SAP Sybase IQ database software). It is also the primary data integration technology for the SAP HANA® database. SAP Data Services consists of two products: SAP Data Integrator software and SAP Data Quality Management software. Both are leading products in their market segments.

SAP Landscape Transformation Replication Server
This replication server is for all SAP HANA database customers who need real-time or scheduled data replication sourcing from SAP and non-SAP sources. It is possible to migrate data into SAP HANA while replicating data in real time.

SAP Sybase Replication Server
This software enables real-time data extraction and delivery to all major databases. It uses a log-mining concept for change data capture and extraction. Because log-mining works asynchronously on the redo logs of databases, it is noninvasive to the source system and proves to be a very efficient way to get hold of changed data.

Both SAP Landscape Transformation and SAP Sybase Replication Server enable real-time data extraction into the acquisition layer. Which solution is optimal for your organization depends on your unique requirements (that is, the need to use the solution for data storage in the acquisition layer, the number of tables to be replicated, and your existing product knowledge).

SAP Sybase ESP
This solution supports scenarios that require analyzing and reacting to streams of event data. It lets you rapidly stream data into the finance and risk warehouse and analyze events.

For more and more business scenarios, a real-time replication mechanism is mandatory to properly address the analytical requirements. SAP solutions support replication without any impact on the performance of the source systems. A real-time approach helps ensure that data is trickle-fed into the acquisition layer as it is changed or updated in the source layer.

As a principle, you should extract all relevant data (that is, underlying tables) from the source and bring that data into the acquisition layer. This is similar to the well-known and well-understood principle of “prefetching” in computer science. For example: While a project may require only three attributes from your source systems on day one, it will likely need additional attributes on day two. By extracting (prefetching) as much of the data as possible up front, you can meet the project’s data requirements entirely from the data warehouse tier – without the need to resort to source systems.

Data Cleansing and Quality Management
One of the most challenging requirements of a data warehouse solution is to embed data validation and cleansing activities into data movement to ensure that only reliable data is processed into the data warehouse. Obviously, the best location for these data cleansing and quality improvements to take place is the operational source system itself by setting up a data quality firewall. However, in some cases the direct integration into all source systems is not possible. In such cases, the quality gates are set up at the first layer of the data warehouse – the data acquisition layer. The objective is to avoid sending erroneous data to the propagation layer.

The data cleansing stage is the time to perform some basic checks and improvements. One common category of checks is “column property enforcement.” These checks simply look at the validity of data within columns and the conformance of data to simple rules. For example:

• Does a customer record have a value in the First Name and Last Name column?
• Is the Key or Date column not null?
• Does the Gender field contain values other than M and F?
• Is the country code in the Address Line valid?

Data that does not pass these simple checks should not be accepted by the propagation layer; this data should either be filtered out and rejected or corrected if possible.
Data Quality
With SAP Information Steward software, your data analysts can perform interactive data profiling directly on the staging tables of the data acquisition layer to get a first insight into the content of these tables (see Figure 3). This insight is based on statistical information (such as minimum or maximum values, string length, distribution frequency of content, and words or patterns within columns), as well as advanced profiling of the uniqueness, dependency, and redundancy of the data.

Based on these initial findings and your organization’s existing data quality standards and requirements, you can define and approve validation rules that are embedded in scorecards (see Figure 4). These rules can then be bound to the tables in the staging area of the data acquisition layer for ongoing assessment and monitoring of the data quality level. The rules can also be bound to the operational source systems, so you can assess data quality scores there as well and determine whether the source data is fit for use in your data warehouse. As a result of executing the validation rules against the tables, you get several perspectives on the assessed data quality scores — gaining transparency into how good the data quality level really is.

Figure 3: Profiling Results from SAP® Information Steward Software
The data acquisition layer is the first layer in the finance and risk warehouse. This is a crucial layer responsible for providing a single landing point, as well as verifying and cleansing data for all source systems in scope.
In addition to its data quality assessment functionality, SAP Information Steward also integrates metadata from any source systems, data warehouse models, and reporting solutions into one single repository. Based on this consolidated and linked information, SAP Information Steward can create data lineage and impact analysis graphs. You can view this impact analysis information via data quality scorecards to see the impact of identified data quality issues on downstream systems and reports (see Figure 5).

Using the SAP BusinessObjects™ BI platform, business users can see where their report data is coming from and how good the data is at the source or in the data acquisition layer. That’s because the lineage diagram is natively integrated within the platform’s launch pad.

Due to the tight integration of SAP Information Steward and SAP Data Services, you can directly export all the defined and centrally maintained validation rules into the repository of SAP Data Services. The ETL job designer within SAP Data Services can embed the same validation function within the data movement job to confirm that the data moved from a source to a target is fully compliant with your organization’s data quality requirements.

A key advantage of performing these elementary data quality checks using SAP Data Services is that you can follow the extract-load-transform (ELT) paradigm rather than the extract-transform-load (ETL) paradigm. In a traditional architecture, data is first extracted from the staging system into an ETL server. Then the data is sorted, cleansed, transformed, and finally delivered to the target database. Although the ETL architecture is widely used in the industry, its shortcomings – especially those of scalability with large volumes of data and its inability to keep up with real-time processing requirements – are obvious.

The propagation layer itself is a better place to do structural checks, such as foreign key and primary key checks. These checks are an inherent property of the normalized data model.

**Figure 5: Impact Analysis from Data Quality Scorecard**
In addition to core data validation, SAP Data Services provides a set of data quality transforms. You can embed these transforms in data flows or ETL jobs to parse, standardize, cleanse, enhance, make error tolerant, and similarity match the data before or while it is moved to the next layer.

Typical data cleansing or data improvement steps are based on external reference data (like address cleansing and standardization) or use the data cleanse transform to standardize or map different variations of the same meaning on one standard form. In many cases, the data profiling of different sources indicates that the same attributes have different data domains in different systems (for example, one source system stores color codes in numeric form, while another stores them in words). You can directly embed all these data standardization and data quality improvement steps into the data flows of SAP Data Services.

**Changed Data Delivery**

Once the data passes through some basic quality control, you need to pass it on to the propagation layer and store it in a time-variant (history-preserving) manner.

The most efficient way to send data to the propagation layer is by only sending data that is different from the last time data was sent. This is usually referred to as change data delivery. As a rule of thumb, you should minimize data movement. Following this rule, the best place to provide change data capture is at the source. Failing that, you should handle change data capture and change data delivery in the data acquisition layer, before you forward data to the propagation layer.

You can utilize all methods of data integration that have been introduced before to not only capture changes at the source but to also enrich the data with a time stamp in case the source does not do so.

**Data Storage in the Acquisition Layer**

For data storage, you can build out the acquisition layer with three different technologies:

- SAP Sybase IQ
- SAP HANA
- SAP NetWeaver Business Warehouse (SAP NetWeaver BW) when running on SAP HANA

The following section describes the pros and cons of using these technologies to build out the acquisition layer.

**SAP Sybase IQ**

The usage of SAP Sybase IQ has the following advantages:

- SAP Sybase IQ can scale to 256 PB of data. If the acquisition layer must hold the history of the source data in source data format for a long time, it is essential that the data store is able to scale without running into performance problems.
- SAP Sybase IQ is a columnar database that provides up to 10X data compression, helping to keep storage costs at a minimum.
- SAP Sybase IQ is designed specifically to perform analytical processing tasks (as opposed to transactional processing) and to optimize the performance of these tasks.

Things to consider when using SAP Sybase IQ:

- Using one technology stack throughout the finance and risk warehouse has several advantages. The following sections suggest using SAP NetWeaver BW running on SAP HANA for both the propagation layer and the transformation and access layer. To employ this single-stack approach, you can also use SAP NetWeaver BW running on SAP HANA for the data acquisition layer.

**SAP HANA**

The usage of SAP HANA has the following advantages:

- SAP HANA is a columnar database that provides high data compression rates of 10X and beyond, helping to keep storage costs at a minimum.
- SAP HANA provides the best query performance of all three alternatives.
- SAP HANA lets you integrate applications at the database level – for example, via Structured Query Language (SQL). This provides the highest level of flexibility and best performance of all three alternatives. SAP HANA Live offerings complement the SAP HANA database tables in the propagation layer and provide models (based on views in SAP HANA) that simplify the consumption of tables.
- If SAP NetWeaver BW is using the same database instance of SAP HANA that is used for the acquisition layer, users can directly access data in the acquisition layer from SAP NetWeaver BW (via the transient and virtual provider concept).
- Using SAP HANA for the acquisition layer is very much in line with SAP’s future vision. SAP plans to allow transactional and analytical systems to be ported to SAP HANA and share one database instance. This is expected to promote simplification, fewer redundancies, and better overall performance.
Things to consider when using SAP HANA:

- Persisting data in the acquisition layer for a long time blows up the data volume. Keeping all data as “hot data” in SAP HANA results in comparably high costs.
- SAP NetWeaver BW on SAP HANA has built-in features for historization and delta handling. When using SAP HANA for the data acquisition layer, you must build these features from scratch.

SAP NetWeaver BW on SAP HANA
The usage of SAP NetWeaver BW on SAP HANA has the following advantages:

- Using SAP NetWeaver BW on SAP HANA for the acquisition, propagation, and transformation and access layers provides a homogeneous technology stack across the complete data warehouse (that is, one toolkit for data movement, one metadata repository, and one set of interfaces to the BI abstraction and reporting layer).
- Using SAP NetWeaver BW on SAP HANA for the acquisition layer is very much in line with SAP’s future vision, in which all transactional data is expected to be present in its original data formats in SAP HANA.
- Using SAP NetWeaver BW on SAP HANA for the acquisition layer, you automatically get all the advantages of SAP HANA (for example, 10X compression rates, faster load times, high query performance, and more).

Things to consider when using SAP NetWeaver BW on SAP HANA:

- Persisting data in the acquisition layer for a long time blows up the data volume. Keeping all data as “hot data” in SAP HANA results in comparably high costs. You can mitigate this by either using SAP Sybase IQ as near-line storage or using the not-active data concept.

SAP NetWeaver BW running on SAP HANA provides the following advantages compared to running it on a conventional database:

- Supercharged business warehouse with dramatically improved performance
- Query acceleration via in-memory column storage
- Optimized information cubes (“infocubes”) and data store objects (DSOs) to significantly reduce load time and data latency
- Simplified administration and a streamlined landscape thanks to less complexity (fewer persistencies, no indexes, no aggregates)
- Reduction in data layers and data volume through data compression
- Self-service access to all enterprise information at the most granular level
- Ability to combine data from SAP NetWeaver BW with data from non-SAP software to increase the breadth of data for analysis

The propagation layer acts as the single source of truth for all enterprise data that is deemed relevant for strategic decision making and multipurpose by nature.
The propagation layer is the key component of the finance and risk warehouse. The foundation of this layer is a business data model that provides a consistent and stable view of business information relationships. The data model strives to remove redundancies that are usually expected when dealing with many source systems. Entities, attributes, and relationships appear only once in this model. This principle forces all data from source systems to be conformed and normalized. Conformed and normalized data is the cornerstone for creating a single source of truth within the enterprise.

The propagation layer is based on SAP NetWeaver BW running on SAP HANA, which is an enterprise data warehouse (EDW) application. SAP NetWeaver BW on SAP HANA features many data warehousing features and functions that are not offered by database-only offerings such as SAP Sybase IQ, Oracle 11g, or IBM DB2. Some of the prebuilt functionality delivered by SAP NetWeaver BW on SAP HANA includes (but is not limited to):

- Delta handling
- Fine-grained access control
- Load request metadata
- Information lifecycle management through near-line storage
- Data modeling in the propagation and data access layers
- Data integration with the propagation layer

### Delta Handling

SAP NetWeaver BW on SAP HANA provides comprehensive functionality for delta handling, which is often one of the most complex aspects of a propagation layer. Delta handling is closely tied to the business. The deciding factor on the delta approach comes down to how the business wants to reflect deltas, as opposed to how the database physically represents them.

Consider the following example. A business receives a contract for 100 units on 8/28/13, only to have it increased to 1,000 units on 8/29/13. Delta handling needs to deal with this situation. The business must determine which of the following methods it should use in the propagation layer to represent the change in the contract value:

- A complete overwrite (28/08/2013: 100, 29/08/2013: 1000)
- A new transaction with a negating transaction for the previous one (28/08/2013: -100 & 29/08/2013: +1000)

Both representations are correct. It’s a business decision as to which representation is preferred. SAP NetWeaver BW on SAP HANA provides a configuration-based approach for delta handling. For each DSO in SAP NetWeaver BW, the business can define delta handling as it sees fit. This functionality would usually have to be built in to any other ETL tool.

### Fine-Grained Access Control

SAP NetWeaver BW on SAP HANA features a robust access and authorization model for accessing business data. It provides a configuration-based (rather, coding-based) approach to security management and lets you restrict data access at the following levels:

- Infocube level
- Characteristic level
- Key figures level
- Nodes in hierarchy level

Achieving the same sophistication of access and authorization control in a non-SAP NetWeaver BW environment would require implementing complex database procedures and SQL code that utilize row-level and column-level security concepts.

### Load Request Metadata

Every time a load request is made for data from the acquisition layer, a “load log” table needs to be maintained. The purpose of this table is to identify when and which rows in the propagation layer were added or modified. In the event of a failed load process, the table is extremely helpful in identifying affected rows. Without such a mechanism, the propagation layer could end up in an inconsistent state with partial loads of data. The load log table in SAP NetWeaver BW on SAP HANA is part of the standard metadata for each DSO. If you use a “naked” database such as SAP Sybase IQ or DB2 as the propagation store, you need to build the load log table in the data model for each of the entities.
The propagation data store is time variant (history preserving). A consequence of this characteristic is the vast amount of historical data that the propagation layer needs to keep. Therefore, an information lifecycle management strategy is of utmost importance not only to keep costs in check (for example, by “aging” or moving old data from expensive to inexpensive storage), but also to help ensure optimal performance. In an ideal world, you should place the current and most frequently accessed data that is queried for business reports on a daily basis in the quickest type of storage possible. Meanwhile, infrequently accessed data can be placed in slower, and cheaper, storage mediums.

SAP NetWeaver BW integrates fully with near-line storage solutions such as SAP Sybase IQ (see Figure 6). What’s more, SAP NetWeaver BW can manage the automatic aging of data from the quickest tier (in memory on the SAP HANA database) to slower, less expensive tiers that can support up to 256 PB of data (SAP Sybase IQ).

Data Modeling in the Propagation and Data Access Layers
Figure 7 shows the logical data model of a loan contract that is modeled in a relational and normalized data format.
Figure 8 shows the propagation layer and the data access layer.

Each entity (for example, a loan or repayment option) is physically split up in semantically separated data clusters that are related to each other. These clusters can have a flat or tabular structure.

Some of the data clusters that describe the loan are:
- Basic data (flat)
- Financial conditions (tabular)
- Credit exposure data (flat)
- Cash flows (tabular)

The transformation and access layer takes into account application-specific data formats and performance optimizations. Application-specific logic is applied to the data to enrich it with calculation results.
The information needs of business users working on the propagation layer usually differ very much. Some might need data for a complete business entity (such as a loan contract), while others might only need the cash flows of a loan contract. To address these different information needs, it makes sense to model the data access layer in multiple layers of views, ranging from the most granular level (for example, data clusters such as basic data or credit exposure data) to complete business entities (see Figure 9). Depending on the complexity of the business entities, it might also make sense to work with several levels that refer to each other. This allows business users to select the view layer that addresses their information requirements in the most appropriate way.

SAP offers several banking-specific logical data models that you can use as a reference when physically implementing the data model in the propagation layer. One of these data models is the finance and risk data model of SAP Bank Analyzer. Parts of the physical implementation in SAP NetWeaver BW on SAP HANA are already available as business content in SAP NetWeaver BW (http://help.sap.com/saphelp_nw73ehp1/helpdata/en/41/42/440c6f203014e10000000a114e5d/frameset.htm). SAP is planning to build a complete finance and risk data model (including data access layer) for SAP HANA.

Figure 9: Modeling Layers Using Multiple Layers of Views

<table>
<thead>
<tr>
<th>Customer views:</th>
<th>Individual views</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Join data across SAP® BusinessObjects® software in finance and risk model</td>
<td></td>
</tr>
<tr>
<td>• Join data with data from nonfinance and risk model</td>
<td></td>
</tr>
</tbody>
</table>

| Level 2 views: | |
|----------------|-----------------
| • Loan | S40CASHFLW – Basic data |
| • Credit default swap | S40CASHFLW – CF_HEADER |
| • Business transaction: flow transaction | S40CASHFLW – CF_ITEMS |

<table>
<thead>
<tr>
<th>Level 1 views:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Template for repayment option</td>
<td>S40CASHFLW – Basic data 0..n S40CASHFLW – CF_HEADER 1..n S40CASHFLW – CF_ITEMS</td>
</tr>
<tr>
<td>• Cash-flow data</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level 0 views:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Loan: basic data</td>
<td>S40CASHFLW – Basic data</td>
</tr>
<tr>
<td>• Loan: cash-flow header data</td>
<td>S40CASHFLW – CF_HEADER</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Physical database table</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Loan: basic data</td>
<td>S40CASHFLW – Basic data</td>
</tr>
<tr>
<td>• Loan: cash-flow header data</td>
<td>S40CASHFLW – CF_HEADER</td>
</tr>
<tr>
<td>• Loan: cash-flow item</td>
<td>S40CASHFLW – CF_ITEMS</td>
</tr>
</tbody>
</table>
Data Integration with the Propagation Layer
Transferring data from the acquisition layer to the propagation layer requires transforming that data into the normalized data model of the propagation layer. Key entities in the propagation layer’s data model should have a “time” aspect added to ensure that data is stored in a time-variant (history-preserving) manner. During the transformation and normalization phase, data that does not satisfy structural conformance has to be rejected and written in the reject log.

Depending on whether you use SAP Sybase IQ or SAP NetWeaver BW on SAP HANA for data storage in the acquisition layer, you can use alternative technologies to transfer data from the acquisition layer into the propagation layer.

Transferring Data from SAP Sybase IQ
SAP recommends using SAP Data Services to transfer data from the acquisition layer (SAP Sybase IQ) to the propagation layer (SAP NetWeaver BW on SAP HANA). Figure 10 gives an overview of the architecture.

**Figure 10: Data Transfer from SAP® Sybase® IQ to the Propagation Layer**
The data is initially transferred from the acquisition layer to the persistent staging area of SAP NetWeaver BW on SAP HANA. This step is necessary to help ensure that, in the event of the load process failing, you can rerun the process with data available in SAP NetWeaver BW on SAP HANA without having to fetch data again from SAP Sybase IQ.

Transferring Data from SAP NetWeaver BW on SAP HANA
Figure 11 describes how data in SAP NetWeaver BW transfers from one layer to the other.

Data transfer processes (DTPs) are used to transfer data from one persistent object to another object, in accordance with certain transformations and filters. You can use the same technology to integrate data with downstream systems via Open Hub technology (mass data replication).

Figure 11: Data Transfer from SAP NetWeaver® Business Warehouse to Propagation Layer

<table>
<thead>
<tr>
<th>SAP NetWeaver® Business Warehouse</th>
<th>Downstream systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>InfoProvider*</td>
<td>Open Hub destination</td>
</tr>
<tr>
<td>Transformation</td>
<td>DTP</td>
</tr>
<tr>
<td>InfoProvider</td>
<td>Transformation</td>
</tr>
<tr>
<td>Transformation (optional)</td>
<td>DTP</td>
</tr>
<tr>
<td>InfoSource** (optional)</td>
<td>Data transfer process (DTP)</td>
</tr>
<tr>
<td>Transformation</td>
<td></td>
</tr>
<tr>
<td>Data source/persistent staging area</td>
<td></td>
</tr>
<tr>
<td>InfoPackage***</td>
<td>InfoPackage</td>
</tr>
<tr>
<td>SAP NetWeaver® BW</td>
<td>SAP NetWeaver® Process Integration</td>
</tr>
<tr>
<td>Non-SAP</td>
<td></td>
</tr>
</tbody>
</table>

*Object for physical data storage or view that is relevant for reporting
**Structure in which key figures and characteristics (“InfoObjects”) are logically grouped together
***Object that specifies when and how to load data from a given source system
A DTP makes the transfer processes in the data warehousing layer transparent. With a DTP, you can separate delta processes for different targets and use filter options for the persistent objects on different levels. You can also define error handling for DSOs with a DTP. To simplify error handling, the DTP sorts out incorrect records in an error stack and writes the data to a buffer. It uses transformation rules to transform, clean, and consolidate data while transferring it from one layer to the other.

**TRANSFORMATION AND ACCESS LAYER**

The transformation and access layer has the following tasks:

- Apply complex business logic
- Create data marts for specific purposes
- Simplify data consumption for the business community

**Apply Complex Business Logic**

The following SAP applications address complex business issues as part of the transformation and access layer:

- SAP Bank Analyzer set of applications
- SAP Liquidity Risk Management application powered by SAP HANA

**SAP Bank Analyzer**

SAP Bank Analyzer is a set of finance and risk applications, including solutions for:

- Accounting for financial instruments
- Profitability management
- Credit risk management
- Limit management

SAP Bank Analyzer comes with a preconfigured data model that is based on industry best practices and supports a comprehensive set of finance and risk applications. The software distinguishes between semantically integrated source data (residing in the source data layer, or SDL) and results data (residing in the results data layer, or RDL). Once the data is in the SDL, the various analyzers are able to get to work and write out the results of the calculations to the RDL. You can then use the data from the RDL to enrich the propagation layer with calculation results where relevant (especially if the results are multipurpose). The results are reported through data marts designed for specific purposes that present the data in a multidimensional form fit for reporting and consumption by business users.

SAP Bank Analyzer also integrates with third-party solutions.

Data from the propagation layer is loaded into SAP Bank Analyzer via the data load layer (see Figure 12).

The data load layer feeds data from the propagation layer via standard interfaces (such as BAPI) into the source data layer and the results data layer. The data load layer automatically identifies changed objects in the propagation layer and provides data to SAP Bank Analyzer, taking into consideration the specific delta handling logic of the software.

SAP also offers a set of enterprise services that you can use to upload data into SAP Bank Analyzer. These services have been designed for an event-based, point-to-point integration from source systems to SAP Bank Analyzer. Point-to-point integration contradicts the basic idea of the data warehouse (that is, multipurpose, single source of truth). Therefore, it is not recommended to use these services in the data warehouse architecture.

**Figure 12: Data Integration Architecture of Propagation Layer and SAP® Bank Analyzer**
SAP Liquidity Risk Management Powered by SAP HANA

Powered by SAP HANA, the SAP Liquidity Risk Management application is a high-performance solution that calculates key liquidity risk figures for regulatory reporting (Basel III), as well as for internal management purposes. It helps calculate, analyze, and simulate the short-, mid-, and long-term liquidity situation of banks. The calculations are based on large volumes of granular cash flows. These cash flows have to be stored in dedicated SAP HANA database tables. If multiple applications use the cash flows, it makes sense to integrate these applications with SAP Liquidity Risk Management via the propagation layer. If only SAP Liquidity Risk Management uses the cash flows, a point-to-point solution between cash-flow source systems and SAP Liquidity Risk Management makes more sense. SAP Bank Analyzer can serve as the cash-flow generation engine for SAP Liquidity Risk Management.

Data Integration with SAP Bank Analyzer
You can directly integrate SAP Bank Analyzer with SAP Liquidity Risk Management. This lets you use SAP Bank Analyzer as a cash-flow generation engine for SAP Liquidity Risk Management.

The calculation and valuation process manager (CVPM) is an architecture component in SAP Bank Analyzer that can generate and support multiple-step cash-flow processes between SAP Bank Analyzer and SAP Liquidity Risk Management. SAP Bank Analyzer also provides tools to schedule the processing and monitoring of cash-flow generation and transfer.

Data Integration with the Propagation Layer
If the propagation layer holds cash-flow data that must be integrated with SAP Liquidity Risk Management, you can use SAP Data Services to transfer that data into the corresponding SAP HANA database tables of SAP Liquidity Risk Management.

Create Data Marts for Specific Purposes
The data mart is the classical way to get data out to business users. It represents a subset of the data that is usually oriented to a specific business line or team. You can build data marts based on the propagation layer to address specific business requirements and optimize performance in standardized reporting. Data marts are created in SAP NetWeaver BW on SAP HANA; the data transfer from the propagation layer to the data mart is based on DTP technology in SAP NetWeaver BW (described in more detail in an earlier section).

Simplify Data Consumption for the Business Community
One of the key advantages of SAP NetWeaver BW on SAP HANA is that you can simplify consumption and optimize performance based on the definition of views in SAP HANA. While traditional data warehouse solutions use data marts and indexes to meet performance requirements, the in-memory computing technology in SAP HANA makes it no longer necessary to materialize these structures in the transformation and access layer. This reduces redundancies, lessens administration effort, and optimizes performance. Furthermore, ad hoc reporting requirements can be addressed better and faster.

The BI abstraction and reporting layer should provide the broadest range of tools for data access, exploration, and prediction. It should also include well-defined interfaces for consumption of data from the finance and risk warehouse.

Finance and Risk Data Management in Banking
© 2013 SAP AG or an SAP affiliate company. All rights reserved.
Using views on top of the propagation layer represents a paradigm shift in how data models are built out in the data warehouse. In a traditional architecture, the data model includes materialized data marts in a dimensional model for business consumption. Obviously, this is another layer that needs to be managed and comes with its own complexity.

Although using views instead of physical materialization does not eliminate all work needed to reflect changes in business requirements, it certainly makes the task a lot easier.

**MODELING AND DATA ARCHITECTURE**

The absence of data governance is a critical point of failure in information management architectures based on heterogeneous system landscapes. Most of the time, the diverse systems have been built to satisfy specific business requirements without aligning metadata and structures with other application silos. If enterprises adopt “logical data warehouse” concepts in which data warehouses are split between several physical instances, data governance and data modeling become even more important.

SAP Sybase PowerDesigner® software helps define, implement, document, and maintain the information architecture described in this document. This includes all logical architectural layers ranging from source systems to the BI and reporting layer and corresponding data flows. SAP Sybase PowerDesigner can be used for SAP and non-SAP software systems.

**Modeling**

SAP Sybase PowerDesigner supports different kinds of models:

- Conceptual data model
- Logical data model
- Physical data model
- Data warehouse model
- XML modeling

**Conceptual Data Modeling**

Conceptual data models represent the overall logical structure of a database independent of any software or data storage structure. These models give a formal representation of the data needed to run an enterprise or business activity.

**Logical Data Modeling**

Logical data models describe database-independent relational structures that developers and designers use for optimization and understanding. Logical data models can be developed independently or generated from physical data models. Logical data models can generate to one or more physical data models.

**Physical Data Modeling**

The physical data models supported by SAP Sybase PowerDesigner generate and reverse-engineer structures for over 80 relational database management systems (RDMS). This support includes all database artifacts and advanced techniques, such as XML and Web services in the database, security modeling, specialized handling of views, and more.

SAP’s architecture includes tools for data integration and quality management that consolidate data from multiple sources. Its tools for data modeling and metadata management support governance and translate business needs into IT requirements.
Data Warehouse Modeling
Multidimensional diagrams document the online analytical processing (OLAP) environment by representing cubes, facts, dimensions, dimensional hierarchies, and queries independent of the physical table structures used to store the warehouse or data mart information. The data mapping editor links the source definition, warehouse, mart, and reporting environments together. This provides clear impact analysis and design-time change management of the BI environment.

XML Modeling
XML-specific modeling techniques document, generate, and reverse-engineer XML Schema and XML Document Type Definition (DTD) structures.

Enterprise Glossary
Centrally managed by the enterprise repository, the enterprise glossary helps ensure that business terms and definitions are established and consistently managed throughout all models. This enforces consistent naming standards and business language alignment while supporting data stewardship and governance efforts. The enterprise glossary streamlines the communication process by enabling everyone to speak the same language.

Integration Aspects of SAP Sybase PowerDesigner
As mentioned earlier, SAP Sybase PowerDesigner can centrally manage data models across the whole system landscape. One important feature of SAP Sybase PowerDesigner is its smooth integration with databases (such as SAP HANA, SAP Sybase IQ, and Teradata). You can import data models from these and approximately 80 additional databases (reverse engineering). What’s more, data models created or changed in SAP Sybase PowerDesigner can be automatically generated in the diverse databases.

You can also use SAP Sybase PowerDesigner to document dimensional models (data warehouse models) and the data model for SAP Bank Analyzer.

Finally, SAP Sybase PowerDesigner lets you migrate existing data models from one database to another (for example, from Teradata to SAP HANA).

Innovations such as in-memory technology now provide opportunities for banks to optimize and simplify their solution architecture for finance and risk data management.
A LOOK AHEAD

SAP’s Vision and Its Impact on Information Architecture

This section describes parts of SAP’s future vision for data management and its potential impact on the information architecture described in this document. The forward-looking statements are subject to various uncertainties that could cause actual results to differ from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates.

CONSOLIDATION OF THE SYSTEM LANDSCAPE ON SAP HANA

One of the main reasons to set up and maintain BI systems in a separate environment is the technical deficiency of operational systems to handle transactional and analytical processes in one system. With the introduction of SAP HANA based on the latest hardware inventions, SAP believes that operational systems can be empowered to provide complex analytical functionality and still fulfill their operational tasks. Operational reporting can be performed on the operational system. The need for redundant data structures, like cubes, does not exist anymore; views or virtual cubes can replace them. Calculations and transformations formerly addressed in the ETL process can move to query execution time. As a result, operational systems can fulfill more and more operational reporting tasks, rather than the data warehouse. This reduces data redundancy, data latency, and system landscape complexity. Needless to say, this also has tremendous impact on costs.

Another reason to set up and maintain BI systems in a separate environment is the need to integrate data from many different sources. This holds true even after the introduction of SAP HANA. Nevertheless, we expect that more and more applications will be able to share one database instance. This will have a positive impact on costs but will also open up new possibilities. Applications won’t be restricted to accessing data only from the SAP HANA database schema that they manage themselves, but they will also be able to access data from the schemas of other applications. This is already possible in SAP NetWeaver BW on SAP HANA today. Not only does this mean that you don’t have to replicate data from one database to the other to integrate applications, you don’t even have to replicate data from one table to the other on the same database instance to make it available for other applications deployed on the same database. Some SAP for Banking solutions are already available on SAP HANA (such as SAP Liquidity Risk Management); others (like SAP Bank Analyzer) are expected to follow soon. In addition to reducing data redundancy and latency, this opens up the architecture for direct consumption of data from other applications.

LESS MATERIALIZATION, MORE VIRTUALIZATION

Given the processing power of SAP HANA, many aggregations, transformations, and calculations can happen on the fly. We expect this power to lead to less materialization – and more virtualization – of classical data warehouse layers that still have relevance.

Views are the key technology behind this more virtualized architecture. You can stack views in multiple layers. An application can access data from the virtually unified data store using views. Views store no data; they only store transformation descriptions (such as computations), possibly drawing upon data from several sources including other views or actual database tables. Analytical views are similar to views but provide the same interface as real cubes. Over time, this can mean that less data is materialized, different kinds of applications work on the same data sources, and the architecture is more ready for self-service business intelligence.

For more information on virtualization and its impact on the layered architecture in SAP NetWeaver BW, please log on to www.saphana.com/docs/DOC-2815.

NEW OPTIONS FOR INTEGRATION

In the past, SAP to SAP as well as SAP to non-SAP software integration has always been realized through the application layer. SAP provides a large diversity of interfaces for this purpose, ranging from low-level application programming interfaces (APIs) to highly standardized enterprise Web services. This kind of integration has numerous advantages that are still valid. With the introduction of SAP HANA, however, SAP plans to also support integration scenarios that are implemented on a database level.
Figure 13 describes these alternatives based on the application architecture of SAP Bank Analyzer.

SAP HANA is storing the data of the complete application stack that is illustrated in Figure 13. An application can use the interfaces provided by SAP HANA to read and write data from the database level. These technologies include SQL and MultiDimensional eXpression (MDX) interfaces. Since the data model of many SAP applications is quite complex, there is no easy way to consume the data from these data models. SAP plans to simplify the consumption of these data models by providing preconfigured consumption views and proper documentation.

Deciding which layer of the architecture to use when integrating an application is based on your organization’s requirements. One of the key benefits of integrating at the application layer (for example, SAP Bank Analyzer) is the richness of business services (such as reuse of the cash-flow engine) and technical services (including authorizations and delta handling). The key benefits of integrating at the database level are maximal flexibility and performance to operate on the data.

FROM ETL TO ELT

ELT stands for extract, load, and transform. In contrast to ETL, transformation work occurs after the data has been loaded into the target database in its raw format. Transformation either happens at the time data is requested by any client or in a second step to a materialized table in the data warehouse. In both cases, SAP HANA provides significant advantages with regard to performance, since the transformation logic is processed on a database level. This represents a big performance advantage compared to transformations on the application level. Moving from ETL to ELT further reduces data latency.

With SAP HANA based on the latest hardware innovations, SAP believes that operational systems can provide complex analytical functionality and still fulfill their operational tasks – reducing data redundancy, data latency, IT complexity, and cost.
DATA AGING

When banks are dealing with Big Data in their data warehouses, technologies that can compress data volume, reduce storage costs, and still allow easy data access and analysis are key. Near-line storage is such a technology. Today, SAP NetWeaver BW on SAP HANA already supports near-line storage and utilizes SAP Sybase IQ database software for that purpose. SAP plans to use near-line storage concepts in other parts of its information architecture – thereby reducing storage costs without any loss of functionality.

DATA FEDERATION

There is a growing recognition that the notion of a single data warehouse is outdated. This is due to several factors. On the one hand, few companies can reengineer their complete system landscape. Most of these landscapes have grown organically over time, and the costs to consolidate systems in one central place often exceed the benefits. Yet given today’s explosive growth of information, existing systems are often at their physical limits. In these cases, IT can use data federation technologies to aggregate data from disparate sources in a virtual database for business intelligence or other analysis. Traditionally, data federation solutions are often recognized to be slow. SAP is planning to natively support data federation functionality on a database level (across all SAP databases). This is expected to dramatically improve the performance of data federation for applications that are based on an SAP database.

FIND OUT MORE

To learn more about how to achieve excellence in financial management, call your SAP representative or log on to www.sap.com/financialexcellence.

Near-line storage can compress data volume and reduce storage costs. SAP NetWeaver BW on SAP HANA supports near-line storage by utilizing SAP Sybase IQ.