



The ABCs of Extracting from ERPs

Introduction

Many organizations have made enormous investments in Enterprise Resource Planning (ERP) applications like JDE, PeopleSoft, and SAP. Designed to manage the daily operations critical to their businesses, these systems are not optimized for business intelligence (BI). Accordingly, their data must be moved to an optimized BI solution for robust analysis. This brief white paper outlines a strategy for extracting data from ERP systems for use in BI systems. It also highlights how Breadboard BI Solution Slices™ content can help with this strategy.

The Challenge

The challenges of adding ERP data to a business intelligence system are considerable. They include:

- ERP database tables are numerous.
- ERP database tables and columns are inconsistently and/or cryptically named.
- ERP database tables often lack primary and foreign key constraints.
- ERP systems are implemented differently across organizations (sometimes differently across business units within the same organization).
- ERP systems change over time. The location and types of data will vary across different versions of a particular ERP.

A Solution

A solution to successfully add ERP data to a BI system involves the following iterative steps:

1. **Identification.** Identify the subset of ERP tables needed to fulfill the requirements.

2. **Selective Staging.** Selectively stage the data from the ERP system into the BI environment.
3. **Data Exploration.** Explore the staged data to logically map to the BI structures (facts, dimensions, bridges, etc.).
4. **BI Data Load.** Load the staged data into the appropriate BI structures.

Identification

Identifying the ERP tables needed to fulfill requirements is deceptively simple. The large number of tables and the inconsistent and cryptically named objects make table identification difficult. For example, it would seem a simple task to find the master customer table in an ERP system. However, it isn't quite as easy as looking for a table named CUSTOMER. In JDE, one might consider the F03012 table; in PeopleSoft one might examine the PS_BO table, and in SAP one needs to look at KNA1. While these table names may be intuitive to some (e.g., the developers that created them), for many people stumbling across them in a set of hundreds of similarly named tables is like finding a needle in a haystack.

Understanding the relationships between the ERP tables may also be a difficult task. For example, basic customer data by itself has limited value, but when combined with order information, it becomes far more interesting. Identifying the order table and then understanding how to relate it to the customer table is not obvious. In JDE, the F4201 table would be a good starting place; in PeopleSoft one should consider the PS_RO_HEADER table, and in SAP one needs to begin with VBAK. Joining these order tables back to their related customer tables can be tricky. For example, there may be multiple customers associated with an order, e.g., 'bill to', 'ship to'. In JDE the 'bill to' is in SHITAN column, while the 'ship to' should be in the SDSHAN column. In SAP, one would need to go to a third table, KNVP, to determine the multiple customers associated to a single order.

There may be internal resources available to help with the identification step. Documentation associated with the original ERP data migration exercise may be available. Technical resources involved in the support of the ERP

package may also be helpful. Finally, operational report writers may prove to be a great resource.

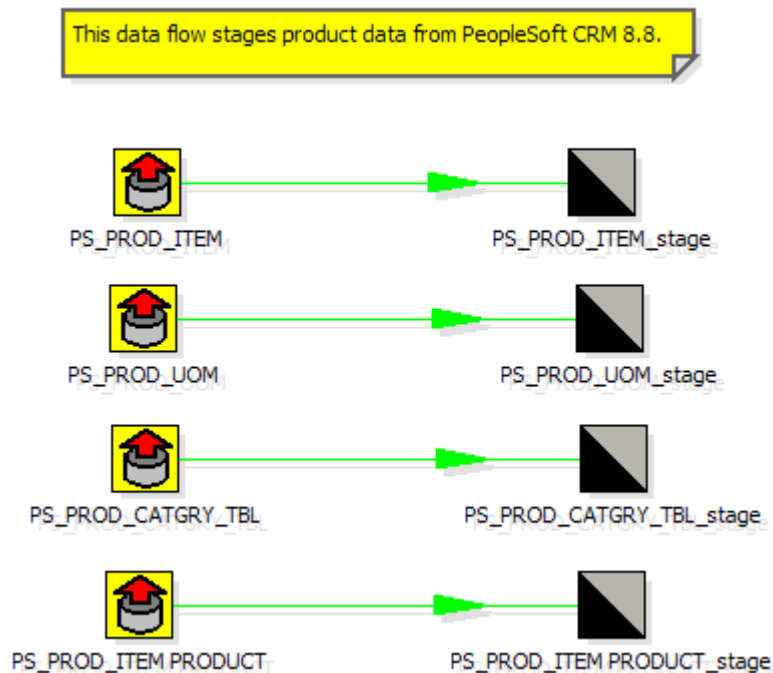
Selective Staging

Once the correct tables have been identified, a process to copy these tables to the BI environment is needed. To minimize the impact to the ERP system, network, and BI system, the data movement should be scheduled during a period of low ERP activity, and include only the tables identified in Step 1. The BI team should work closely with the ERP team to ensure the selective staging process runs smoothly. The use of a data integration (ETL tool) to selectively stage the data will help to speed development.

Identification & Selective Staging with Breadboard BI

Identifying the required tables is a more straightforward task with Breadboard BI Slices. For example, if requirements call for PeopleSoft product information, then our PeopleSoft product transformation set would be helpful. Using the Pentaho Data Integration tool, it lists the product tables most likely needed, and includes the code necessary to stage the data in the BI environment. Figure 1 illustrates this product example using a set of Pentaho Data Integration transformations.

Figure 1: A set of transformations to stage PSFT product information using the Pentaho Data Integration tool.



This set of transformations includes four tables that commonly store product data in PeopleSoft. Once this transformation has been executed, the data from these ERP tables is made available in the BI data stage environment for exploration.

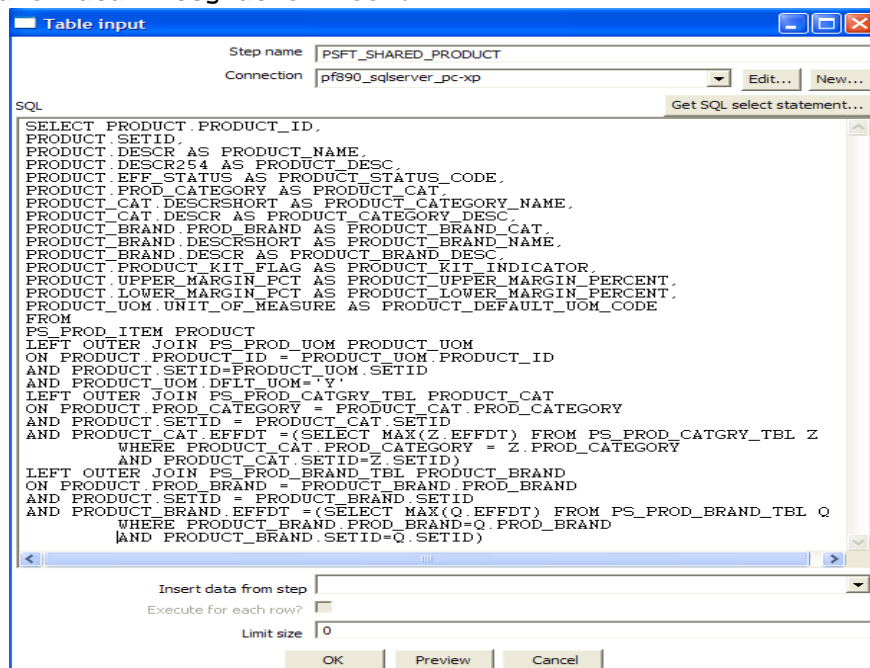
Data Exploration

Once the staging of the data is complete, the active exploration of the data using SQL can begin. This exploration process helps to understand what data is truly stored in the tables, and confirms the relationships amongst the tables. This is important since ERP systems are inconsistently implemented across organizations (sometimes across business units within the same organization). Further, the changes across various ERP versions also make this step important. The output of this process is a logical mapping that links ERP structures (tables and columns) to BI structures (tables and columns).

Data Exploration with Breadboard BI

Our content helps to identify important ERP tables and columns in the BI stage environment. They contain the complex, ANSI standard SQL that define the table, column, and table relations. Figure 2 illustrates a PeopleSoft product example. For more details, refer to our 'Unlock ERP Data with Open Source' white paper.

Figure 2: An Input Step that Exposes PSFT Product Information using the Pentaho Data Integration Tool.



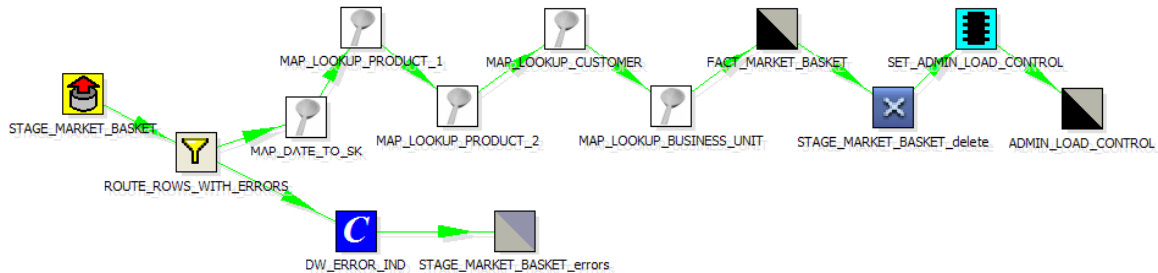
BI Data Load

Once the data exploration step is complete, then the loading of data into the heart of the BI system - the dimension and fact tables - can begin. Like the selective staging step, the BI data load can be implemented using a data integration tool.

BI Data Load with Breadboard BI

Breadboard BI content loads staged ERP data into structures optimized for BI using the Pentaho Data Integration tool. It integrates advanced processes like incremental load strategies, error processing, slowly changing dimension (SCD) type 1 & type 2 updates, surrogate keys, and reusable logic. Figure 3 illustrates a transformation that loads a BI fact table from a stage table. For more details, refer to our "Advanced ETL with Pentaho" white paper.

Figure 3: An ETL Transformation that Unlocks Market Basket Metrics using Pentaho Data Integration (Kettle).



Conclusion

The iterative processes of identification, selective staging, data exploration, and BI system data load should be utilized to fulfill the requirements. Identifying, selectively staging, exploring, and loading ERP data can be difficult and time consuming. However, it is made easier by utilizing Breadboard BI Slices.

About Breadboard BI

Breadboard BI LLC is the leading provider of open source enterprise performance management (EPM) solutions. Their seasoned experts rapidly match client needs to one or more Breadboard BI Solution Areas. A solution is then iteratively customized using license-free tools and their Solution Slices™ content. The result is a modular, scalable, and extendable EPM solution built rapidly and inexpensively.

PeopleSoft and JDE are registered trademarks of Oracle Corporation. SAP is a registered trademark of SAP AG. Pentaho and Pentaho Data Integration are registered trademarks of Pentaho Corporation. Breadboard BI LLC is independent and not associated with Oracle Corporation nor SAP AG.