 *Breadboard BI*

Unlocking ERP Data Using Open Source Tools

By Christopher Lavigne

Introduction

Organizations have made enormous investments in ERP applications like JD Edwards, PeopleSoft and SAP. These applications are highly optimized to manage the daily operations critical to their business. However, they are not optimized to share data with other systems, nor do they support sophisticated exploration of their data. At the same time, we have witnessed rapid advances in the capability of open source business intelligence (BI) tools. This poses the question; can open source BI be applied to the ERP space?

This white paper details an effort to unlock ERP data using open source tools. The demand for such methods is made clear in a recent study by Ventana Research. This study stated that 86% of organizations reported value from pre-built report templates for ERP and CRM applications. In this paper, two methods will be briefly explored: 1) report templates running over ERP application databases, and 2) the use of ETL to extract ERP data and load dimensional table structures.

The Methods

There are pros and cons to using either the "Report Templates over ERP application databases" or the "ETL extract and load of dimensional tables" method. Table 1 lists a few examples of these pros and cons; a more detailed discussion is outside the scope of this white paper.

Table 1 – Two Methods to Unlock ERP Data

Method	Pros	Cons
Report Templates over ERP application databases	• Direct access to data(only the correct SQL is needed). • Data is real-time. • No ETL layer required.	• Business questions may be too complex to encapsulate in a SQL statement. • Queries may not run in a

	• All reporting is done against a single database. • Utilizes existing database software and hardware.	reasonable amount of time. • Queries may significantly slow or crash the transaction system. • Queries can not include data from other systems (internal nor external).
ETL Extract and Load of Dimensional Tables	• Loads data into dimensional table structures designed for fast response. These may serve as a foundation for subsequent data mining, cubes, and dashboards. • Cleanses and combines data from multiple, heterogeneous systems into one database.	• ETL processes must be developed. • ETL processes must be maintained. • Dimensional database objects must be designed. • A separate database is required. • Data may not be real-time. • May require 1 – 2 additional servers.

For most organizations with significant amounts of data and reasonably complex reporting requirements, the "Report Templates over ERP application databases method" is not the favored option. However, it is included in this white paper as a possible short-term, tactical solution.

The Prototypes

To properly explore these methods, prototypes were built to unlock a subset of PeopleSoft CRM 8.9 data stored in Microsoft SQL server using the Pentaho BI stack. (The choice of SQL server as source DBMS was somewhat arbitrary

as PeopleSoft runs over SQL Server, Oracle, or DB2.) The Pentaho components leveraged include the Report Wizard, Pentaho ETL (Kettle), and Pentaho OLAP (Mondrian). The Breadboard BI Solution Slices that were used include ERP Connects, ETL Maps, Data Models, Analysis Cubes, and Dashboards. All Slices were from the Customer 360° Solution Area.

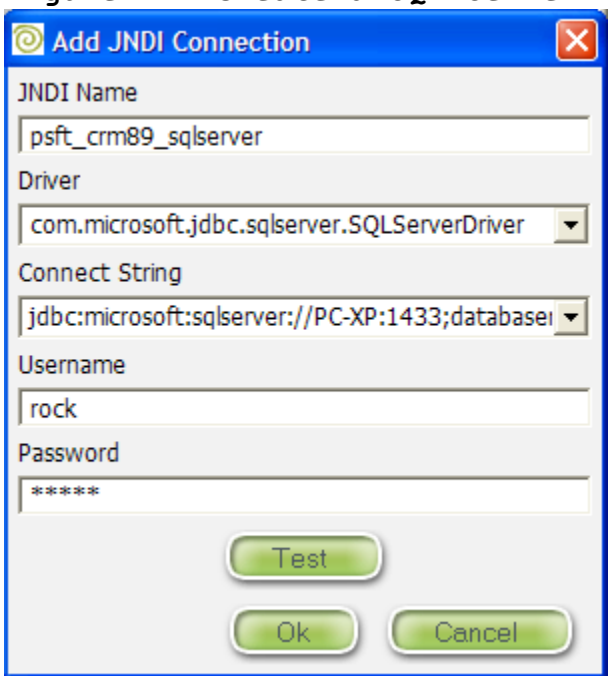
Method 1 – Report Templates over ERP Application Databases

The Pentaho Report Design Wizard 1.1.5 was used to create a sales order report template. It was a fast and easy process to establish a stand-alone report against the PeopleSoft CRM application using this tool. Aside from the wizard, Breadboard BI ERP Connectors were also used. These connectors, built by former employees of the ERP vendors, consist of ANSI Standard SQL statements that can be easily edited to adjust to different implementations.

The following summarized steps were followed to install the wizard and create the first report:

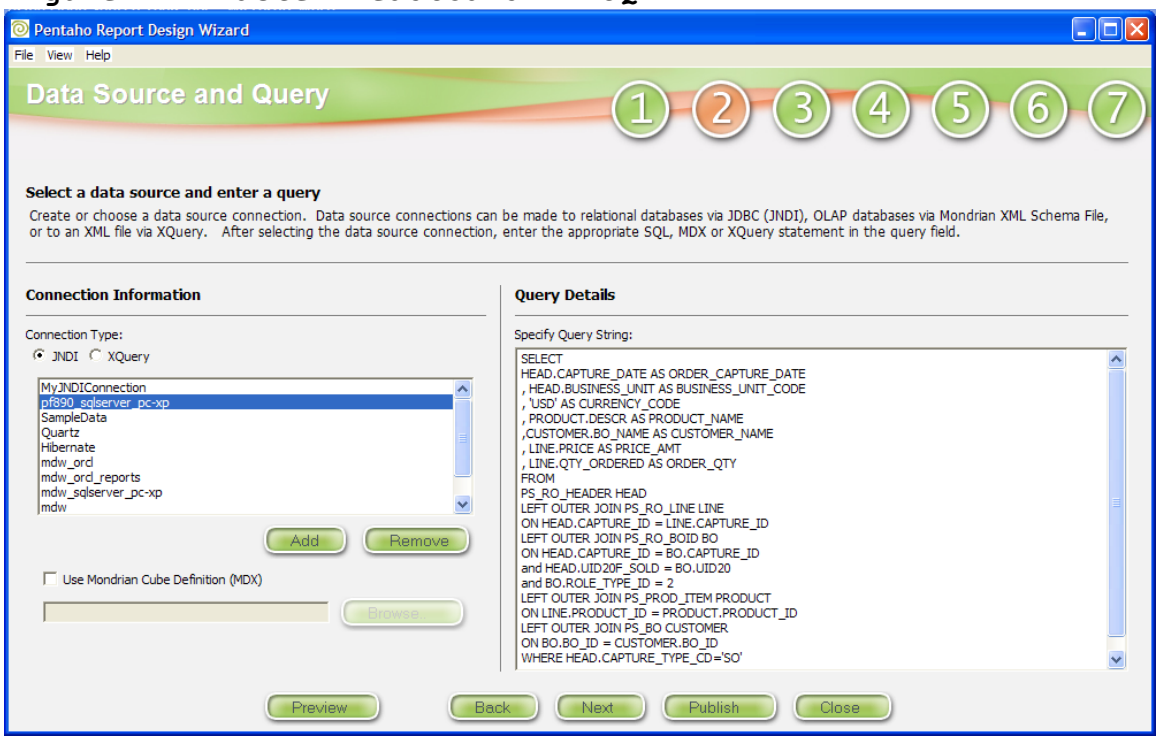
1. Downloaded and unzipped the Report Wizard .zip file from Pentaho's web site - http://www.pentaho.org/download/latest.html
2. Added the SQLServer JDBC drivers (file names msbase.jar, mssqlserver.jar, and msutil.jar) to the "lib\jdbc\" path where the Report Wizard was unzipped.
3. Launched the wizard and created a connection to the SQL Server source database, see figure 1 for details.

Figure 1 – Create a SQL Server Database Connection



4. Pasted the Breadboard BI SQL into the Query Details Box. See figure 2 for details.

Figure 2 – Paste Breadboard BI SQL



5. Formatted and ran a simple report. See Figure 3 for a screenshot.

Figure 3 – Breadboard BI Simple Report

Order Date	Business Unit	Customer	Product	Currency	Price	Quantity
05/21/2006	NE	SI Metalowiec	.. Wireless Router G	USD	2,012.22	5
05/21/2006	NE	SI Metalowiec	.. Scanner	USD	1,977.55	4
05/21/2006	NE	SI Metalowiec	.. Acme Laserjet Printer	USD	2,134.54	1
05/20/2006	NE	SI Metalowiec	.. Wireless Router G	USD	2,012.22	8
05/20/2006	NE	SI Metalowiec	.. Acme Laserjet Printer	USD	2,134.54	10
05/19/2006	SE	Breadboard BI	.. 30" Flat LCD Monitor	USD	321.65	2
05/20/2006	SE	Breadboard BI	.. Scanner	USD	1,977.55	6
05/20/2006	SE	Breadboard BI	.. Wireless Router G	USD	2,012.22	3
05/20/2006	SE	Breadboard BI	.. Digital Camera	USD	987.52	4
05/21/2006	WE	Breadboard BI	.. Wireless Router G	USD	2,012.22	3
Grand Total					17,582.23	46

Method 1 Conclusions

Independent of the serious limitations inherent in sourcing directly against a source database, Method 1 is an approach that allowed us to successfully expose ERP data in just a few hours. (Subsequent reports could be created in minutes.) Although outside the scope of this white paper, very sophisticated and graphically pleasing reports could have been produced using Pentaho's new Report Designer tool.

Method 2 - ETL Extract and Load of Dimensional Tables

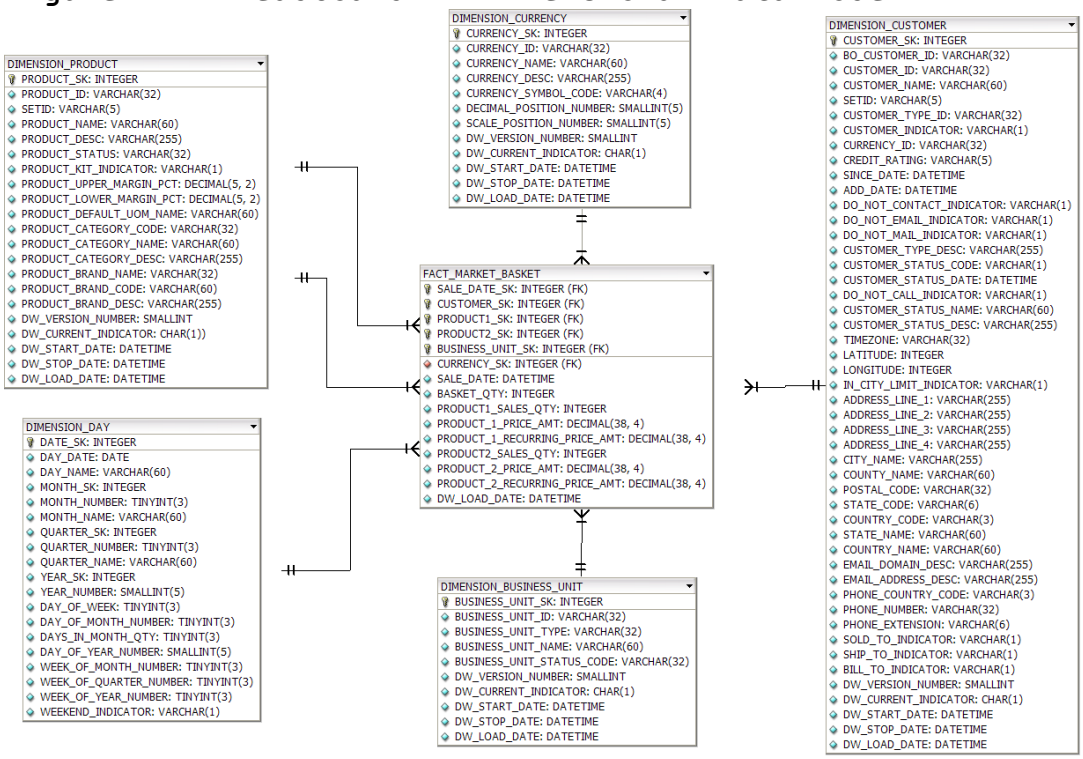
The Pentaho Kettle 2.3.0 ETL tool was used to extract market basket/product affinity data from the PeopleSoft CRM 8.9 application database running over SQL Server. The target database was Oracle 10g R2 Enterprise Edition with the Partitioning option. Kettle reads and writes from a variety of open source (e.g. MySQL) and proprietary databases, so the choice of writing to an Oracle target was a matter of convenience (it was running on the development server that was available for the prototype). The Kettle ETL server was running on the same server as the Oracle database.

The following five summarized steps were followed to install the ETL tool, build the target dimensional data structures, and build the ETL transformations (maps).

1. Downloaded and unzipped the Kettle .zip file from Pentaho's web site - http://kettle.javaforge.com
2. Created the target dimensional database tables using the Breadboard BI Market Basket Subject Area Data Model. (The database DDL was generated from this

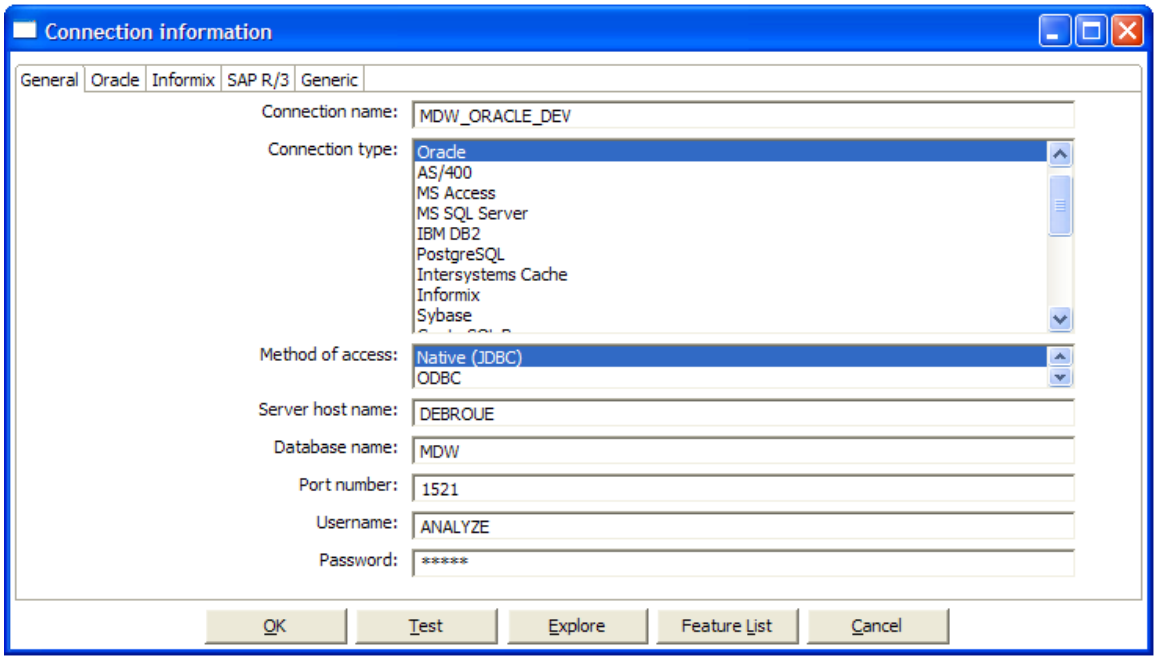
data model.) See figure 4 for the market basket data model.

Figure 4 – Breadboard BI Dimensional Data Model



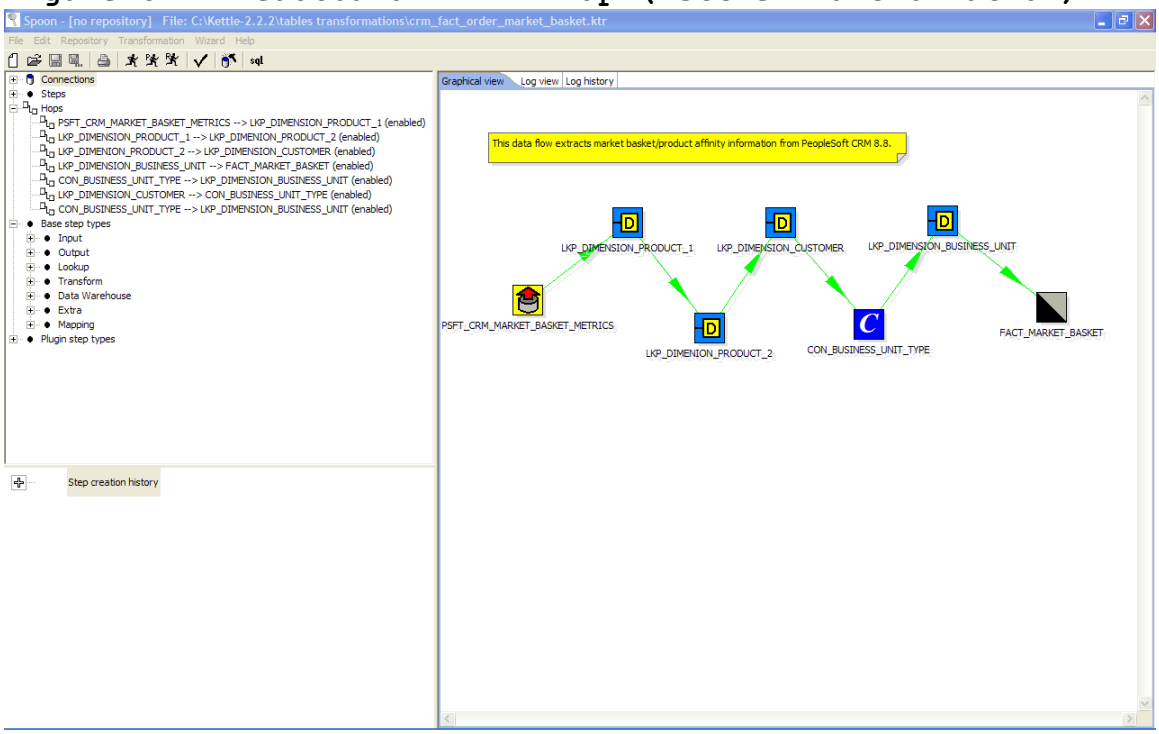
3. Within the Spoon Kettle component, created a new connection for the source database and a new connection for the target database. See Figure 5 for the target database connection details.

Figure 5 – Create Kettle ETL Connections



4. For each target dimension and fact table, one Kettle transformation was created. Breadboard BI ERP Connects were used in each table input step. Used in this manner, the Connects both hide the source system complexity and push join and filter processing to the generally robust ERP Application database server. See figure 6 for a graphical representation of the ETL transformation that loads the FACT_MARKET_BASKET fact table.

Figure 6 – Breadboard BI ETL Map (Kettle Transformation)



Because Method 2 moves data into specialized data structures, the process of adding Analysis Cubes and Dashboards to explore and display the knowledge hidden in data is simplified. Figure 7 displays a Breadboard BI Analysis Cube that utilizes Pentaho OLAP. Figure 8 displays a Breadboard BI dashboard using the Pentaho BI platform. Because screenshots mask the true functionality, readers are invited to visit the Breadboard BI Demonstration Server at http://www.breadboardbi.com/demo.html.

Figure 7 – Breadboard BI Analysis Cube

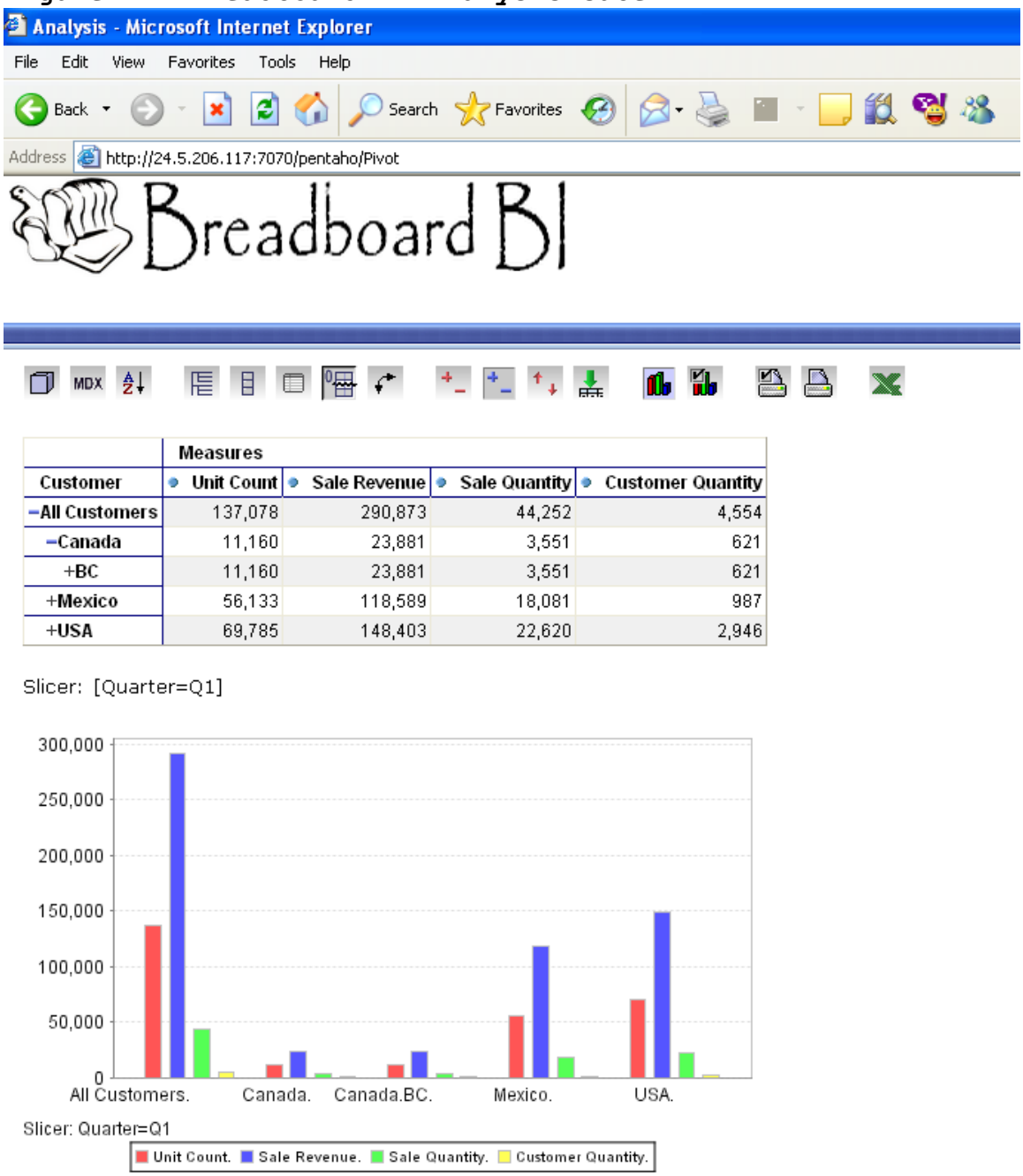
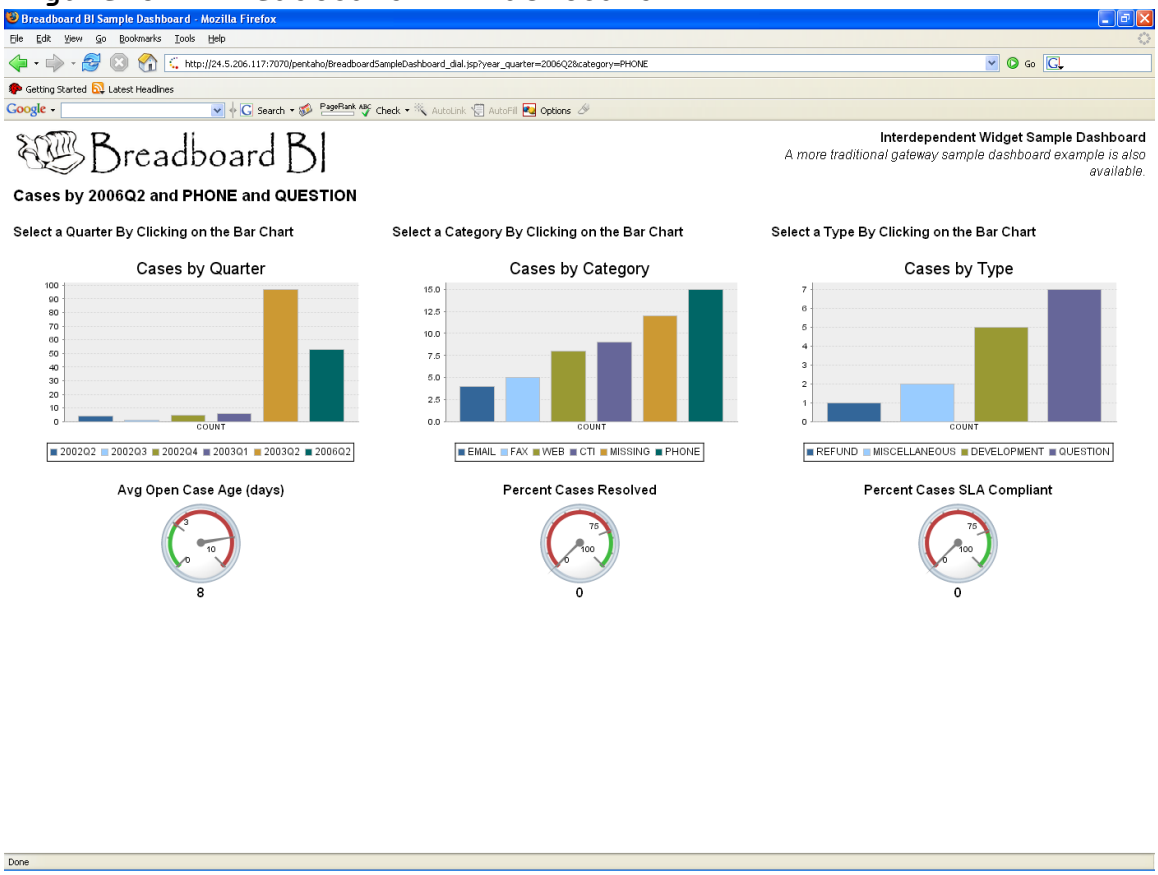


Figure 8 – Breadboard BI Dashboard



Method 2 Conclusions

The Pentaho Kettle ETL tool is an easy to use, capable tool that can be used to unlock ERP data. The ability to source and target most relational database platforms and files, as well as the many built-in steps, greatly assist with the load of dimensional targets. It was clear from our prototype that Kettle supports rapid ETL development and will assist with the maintenance of ETL code. When this ETL functionality is combined with Breadboard BI Solution Slices, ERP data can be quickly added to a new or existing data warehouse or data mart. Finally, the application of Method 2 opens up the world of Analysis Cubes and Dashboards to organizations. **It was clear to us that the combination of Pentaho technology and Breadboard BI content provides not only a low-cost solution, but a technically superior analytics alternative.**

About Breadboard BI
Breadboard BI was founded by seasoned professionals from the packaged analytics software and business intelligence consulting spaces. Our combined expertise includes customer relationship management(CRM), finance, human capital management(HCM)/workforce, and supply chain management (SCM) systems. Breadboard BI's experienced teams work collaboratively with our clients' business and IT professionals to incrementally build business intelligence systems, incorporating open source technology whenever feasible. The result is an extensible, modular, and maintainable solution built for less time and money than packaged analytic or enterprise data warehouse solutions. Our systems open the world of advanced analytics to organizations of all sizes and budgets. Visit us on the web at http://www.breadboardbi.com.

About Pentaho
Pentaho provides a full spectrum of open source Business Intelligence (BI) capabilities including reporting, analysis, dashboards, data mining, data integration, and a BI platform. Formed by a highly experienced team of industry veterans, Pentaho's mission is to bring innovative, high quality technology and professional support to the BI market. Pentaho uses a revolutionary approach to development, distribution and support made possible by an open source business model. Pentaho's technologies support a wide range of business initiatives from sales and profitability analysis, customer analysis, HR reporting, Financial reporting, KPI dashboards, Supply Chain analytics, and operational reporting. For more information: Pentaho's web site is located at http://www.pentaho.org