



# **Pentaho Consulting Services BI Pilot Program Overview**

April 2008

## Overview

Getting accurate, useful information into the hands of business people is one of the chronic challenges facing businesses that want to drive decisions using facts and data. Information stored in operational systems is optimized for recording transactions, not for reporting. Most organizations have multiple applications that are focused on the needs of the sales, manufacturing, accounting and customer service functions, making it difficult to answer business questions that require data spread across different databases. Application upgrades and data quality issues make it difficult to create reports that have lasting information value and result in a fluctuating environment that requires constant change to reports and spreadsheets – or risk rendering them obsolete.

## The Pentaho Business Intelligence Pilot

Pentaho offers an onsite professional services engagement that uses the data integration and business intelligence features of the Pentaho BI Suite to solve these problems. Working intensively with the customer's business and IT analysts, Pentaho consultants create a business-focused, "end to end" solution that achieves the following goals:

- Document the customer's information analysis requirements using the language and point of view of the business
- Develop a stable, subject-oriented information model that supports analysis of data from multiple applications
- Integrate and rationalize data from the customer's transactional systems, consolidating it into a regularly-updated reporting repository that is optimized for trend analysis and historical reporting
- Deploy an easy-to-use, web-based analysis tool that allows business analysts to provide early feedback on the usefulness of the information model
- Create a foundation of reports, analysis views and dashboards and that can be used by executives, managers, analysts and operational users over the web and through e-mail
- Provide functional tool training to lead end-users and business analysts
- Provide operational training that allows the customer's IT department to sustain and expand the BI deployment over time

Pentaho's business intelligence pilot is a fixed-price, fixed duration engagement that includes a standard set of deliverables including a working BI application. Subscription-based product support is provided during the pilot and can be extended to support implementation and maintenance.

## The Pentaho Open Source Advantage

Historically, license fees for traditional business intelligence and data integration tools have consumed the majority of customers' project budgets. Pentaho offers enterprise-capable technology using a professional open-source model, and customers may leverage this model to spend a greater percentage of their budget on professional services. Allocating more budget to services allows BI consultants to spend more time with analysts in the critical business discovery and information modeling phases, resulting in a superior design that has a more lasting impact

on the business. Pentaho's professional open source approach also allows customers to expand deployment without concern for per-user or per-seat costs. Pentaho's support subscriptions provide developer assistance, certified software builds, problem resolution and indemnification using an affordable "per-CPU" model.

## Pentaho Business Intelligence Products

The advantage of open source extends to the Pentaho products that are deployed during the pilot. Pentaho manages, facilitates, supports, and takes the lead development role in the Pentaho BI Project. A global community of partners and contributors assist with new feature development and bug fixes.

### Pentaho Data Integration

- Use: extraction, transformation and loading (ETL)
- Data warehouse population with built-in support for slowly changing dimensions
- Export of database(s) to text-file(s) or other databases
- Import of data into databases, ranging from text-files to excel sheets
- Data migration between database applications
- Exploration of data in existing databases. (tables, views, synonyms ... )
- Information enrichment through lookup of data in various information stores (databases, text-files, excel sheets ... )
- Data cleansing by applying complex conditions in data transformations

### Pentaho Analysis Services

- Use: high performance, interactive analysis of large or small volumes of information
- "Dimensional" exploration of data, for example analyzing sales by product line, by region, by time period
- High-speed queries through the use of aggregate tables in the RDBMS
- Parsing of Multi-Dimensional eXpression (MDX) language into Structured Query Language (SQL) to retrieve answers to dimensional queries
- Advanced calculations using the calculation expressions of the MDX language

### Pentaho Reporting

- Use: creation and formatting of standard parameterized operational reports
- GUI-based tool for report creation
- Full on-screen print preview
- XML-based report definitions
- Output to the screen, printer or various export formats:
  - PDF
  - HTML

- RTF
- Excel
- Fast, with minimal footprint and dependencies

#### Pentaho BI Platform

- Use: provides the architecture and infrastructure required to build solutions to business intelligence (BI) problems
- Solution engine integrates reporting, analysis, dashboards and data mining components to form a sophisticated and complete BI application
- Provides core BI infrastructure services including:
  - Authentication
  - Authorization
  - Solution execution
  - Logging
  - Auditing
  - Workflow
  - Web services

More information on the Pentaho product line and open source community can be found online at <http://www.pentaho.com>.

## Methodology

The Pentaho business intelligence pilot uses a standard BI methodology that has been proven to deliver good results when appropriate business and IT participants are actively involved and consultants have access to the relevant data. A key aspect of this methodology is regular communication between the Pentaho consulting team and business stakeholders in the customer organization. Development consists of a series of iterative “design / build / review” exercises that validate the business information model at regular intervals as data is added to the repository.

The primary tasks of the engagement are as follows:

1. Define the Project Scope. Document high-level business objectives and success criteria for the project; identify key performance indicators (KPIs) that drive (or could drive) decision making; determine key business and technical contacts within the company; create an inventory of applications that could contain data which would be useful for analysis.
2. Configure the Pentaho Infrastructure. Establish connectivity to available data sources using Pentaho Data Integration. Install and configure the Pentaho BI application on one of the customer’s servers; enable the Pentaho application for web access on the customer’s intranet.
3. Conduct Interviews with Business Analysts and Scope the Business Questions. Facilitate an exploratory workshop to discover the business processes that need to be measured; identify

primary data fields in the company's transactional applications; follow-up with 1:1 sessions with analysts to review the purpose of existing reports & spreadsheets and their target audiences. Produce and circulate a list of business questions.

4. Create a Dimensional Model (Measures, Dimensions & Hierarchies). Using the information gained in the workshop and interviews, define a logical data model to contain the model's numeric (fact) and character (descriptive) data. Identify and mark common dimensions that exist across multiple business units (e.g. Customers, Products, Time ...). Define commonly used hierarchies (groupings) based on the company's organizational structure and standard reporting intervals.
5. Perform Initial Data Integration & Load the Repository. Translate the logical data model into a physical star schema and implement in the database of the customer's choice. Connect to the transactional systems using Pentaho Data Integration; extract data and load into the dimensional model. Identify high-level data quality issues (if any). Design and test multidimensional schemas for Pentaho Analysis.
6. Explore & Refine the Business Data Model (Cubes). Reconvene with business analysts, provide training on Pentaho Analysis; conduct a detailed walkthrough of the information model and test the ability to answer business questions. Identify additional data elements and revise the dimensional model.
7. Perform Secondary Data Integration & Reload the Repository. Based on feedback from business analysts, acquire new fact and descriptive data and reload the repository. Reconvene with business analysts and ensure that dimensional models are sufficient and produce valid answers. Identify analyses for which data does not exist. Determine high-level information access rules based on roles.
8. Create Reports, Analysis Views & Dashboards. Create operational reports, analysis views and dashboards that are appropriate for a larger audience. Determine the default folder structure for web-based content. Mentor IT and business analysts on the development process and authoring tools.
9. Publish BI Content to the Web. Connect to the company's security provider (LDAP or MS Active Directory) or use Pentaho's default capability; implement access control rules based on roles. Organize reports, analysis views and dashboards within the Pentaho repository.
10. Conduct End-User Training Sessions. Provide functional training to end-user communities. (Focus on mechanics of access, types of analysis and output formats.)
11. Perform IT Administrative Hand-off. Provide training and mentoring to system / database administration staff. (Focus on scheduling and maintenance of Pentaho Data Integration jobs.)
12. Assess Effectiveness & Outline Next Steps. Provide a written summary and presentation to executive sponsor(s). Identify areas for improvement (e.g. data quality) and future investments (additional subject areas, BI workflow).

## Participants & Roles

Business intelligence is an inherently cross-functional development exercise. It is critical to the success of the pilot that the right individuals are available to participate in the discovery and development phases of the project. Employees should be identified for the following roles during creation of the “Definition of Project Scope” phase:

Business Analyst(s). The person(s) who are responsible for day-to-day reporting within the organization -- typically reports directly to a line of business executive. Recognized internally as the expert in finding data to answer business questions. Often has significant quantitative skills and performs “desktop data integration” using spreadsheet and PC database tools.

IT Analyst(s). Technical individuals who have access to the company’s transactional systems. Often possesses application expertise with the company’s accounting, CRM and ERP systems. Has intermediate to advanced SQL skills; performs custom extracts and data feeds for the business analysts. Has typically worked on projects for multiple business units within the company and maintains cross-functional communication.

DBA / System Administrator. Responsible for controlling access to the company’s systems. Implements systems using company standards. Identifies hardware & software resources and makes them available to the pilot project.

Pentaho consultants act as project managers, architects and BI developers. These roles may be filled by one person or several operating in a small team.

## Milestones

Milestones serve as natural points of communication and feedback between the Pentaho consulting team and the executive sponsor.

Although the BI pilot does not need to be performed in a continuous six-week engagement, doing so helps the team to maintain momentum and stay focused on the design. If needed, a one- or two-week break can occur after the dimensional model has been finished.

### Pre-Engagement

- Approve of Statement of Work

### Week 1

- Business Sign-off on Business Questions
- Business Sign-off on Dimensional Model

### Week 2

- Business Data Model Available

### Week 3

- Dimensional Model Complete

### Week 5

- BI Content Development Complete
- BI Content Published to Web

### Week 6

- End-Users Trained

- IT Administrative Checklist Delivered
- Project Summary Report Delivered

## Pricing

The Pentaho Business Intelligence Pilot is based on a fixed-fee, fixed-deliverables engagement model. The only variable component in the program is for consultant travel expenses. Customized services quotes are available for customers who desire shorter- or longer-term engagements or a different set of deliverables. More information is available through [sales@pentaho.com](mailto:sales@pentaho.com).

## Glossary of Terms

Measure – a quantitative aspect of the business. In a sales analysis application, measures might include “Quantity”, “Unit Price” and “Order Status”.

Dimension – a collection of attributes which describes a business subject or entity. Dimensions are typically categories such as “Products”, “Customers”, “Markets” and “Territories”. *Time* is a special type of dimension.

Hierarchy – provides a structure through which aggregated information can be consistently viewed. For example, the Time dimension might contain several levels including “Year”, “Quarter”, “Month” and “Week”. A Markets dimension might have levels such as “Country”, “State” and “City”.

Cube - in the Pentaho environment a cube is a logical model that provides a point of view to a business user. It consists of subset of measures, dimensions and hierarchies which may be optionally constrained by a role. Pentaho Analysis can host multiple cubes simultaneously.

Star Schema – a method of data modeling that organizes descriptive data (in dimensions) around one or more tables containing numeric data (measures). Star schemas are ideal for business analysis and for ensuring good query performance. Pentaho Data Integration contains special capabilities which allow it to efficiently populate and maintain star schema databases.

Key Performance Indicator (KPI) – a measure which is “rolled up” to a level which is appropriate for executive decision-making. Examples might include “Year to Date Sales”, “Budget Variance”, “Order Backlog”, etc.

Dashboard – a web-based application which organizes a collection of key-performance indicators into a customized view for a specific individual. Dashboards include “point of view” controls which allow the user to quickly change the perspective (dimension filter selection) for multiple KPIs at the same time.