



Pentaho Open Source Business Intelligence Platform Technical White Paper

Copyright © 2006 Pentaho Corporation. Redistribution permitted. All trademarks are the property of their respective owners.

For the latest information please visit our web site at www.pentaho.org

Last Modified on March 28, 2006

Contents

- Contents 2
- Abstract 3
- The Problem 3
- Traditional BI Solutions and Their Shortcomings..... 5
- The Pentaho Open BI Suite..... 6
- Open Source Technology 7
- Architecture 7
 - Pentaho Server 7
 - Server Repositories 11
 - Pentaho Pre-Configured Installation 11
 - Pentaho Design Studio 11
 - Embedded Architecture 13
 - The Pentaho Solution 13
- Pentaho’s Role 14
- Summary 15
- Appendix A: Glossary..... 17
- Appendix B: Open Source Software 19

Abstract

Organizations are all looking to increase revenue, lower expenses, and improve profitability by improving efficiency and effectiveness in their business processes and overall performance. Business Intelligence (BI) software vendors claim that they have the technology that can provide this improvement. Vendors concentrate on selling products or tools that can be used to build these solutions but rarely concentrate on the problem the customer is trying to solve. As new requirements are realized, new vendors are brought in, new tools are purchased and new consultants arrive to make it work. Eventually, the corporate BI initiative becomes a collection of disjointed point solutions using a combination of expensive monolithic commercial applications and difficult to maintain custom code.

Using this current approach, each tool is designed to operate on a very specific task. The actual business problem must be broken into pieces and segregated into tasks like Reporting, Analysis, Data Mining, Workflow, etc. There is no application responsible for initiating, managing, verifying or coordinating results. People and procedures are called upon to make up for these deficiencies.

This white paper describes the Pentaho Business Intelligence Platform: a solution-oriented BI platform that integrates open source components and open standards with a process-driven engine. It shows how this BI platform solves BI problems by combining BI with workflow and process management, making significant improvements to both, and by offering this platform as Open Source.

The Problem

Traditional business intelligence (BI) tools are costly, complex and fall significantly short of enabling enterprises to achieve the sought-after benefits in efficiency and effectiveness. Software vendors promise that BI will provide the aggregation, analysis, and reporting capabilities necessary to transform data into the high-value insight that allows management to make more timely and informed decisions. Unfortunately this amounts to little more than reporting and reporting alone is not enough.

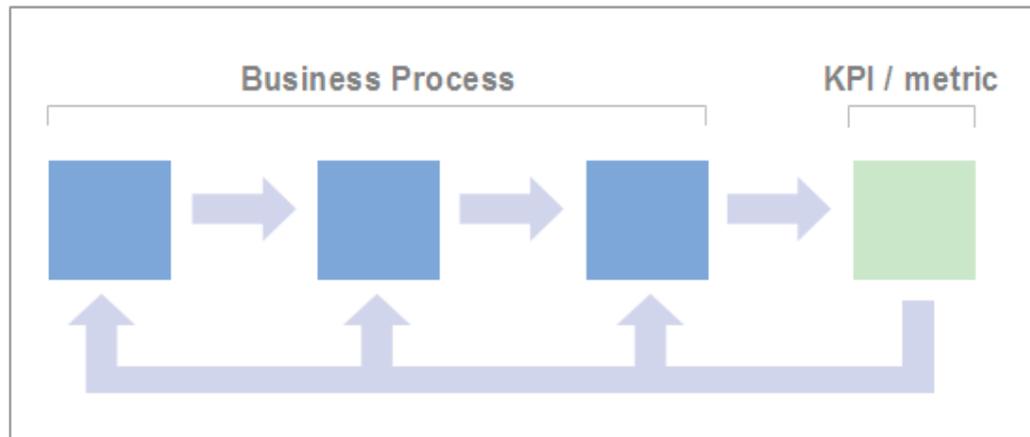
When a report is delivered or a particular situation is encountered in the data something specific needs to happen: a decision must be made, causes must be discovered, or a process must be started. In these cases the information presentation, analysis, and delivery (business intelligence) is a part of a larger process. This process exists to solve the business problem.

To clarify:

- Often the solution to a business problem is a process that includes business intelligence (BI).
- BI, by itself, is rarely the complete solution to the problem.
- If BI is part of the process, then BI tools are, inevitably, also part of the process.
- A BI tool that does not understand processes, or how to be part of one, will be hard to integrate into a larger solution.

A company can maximize business value by moving information to the workers who can use the information to affect the way the company operates. Key benefits to the company will be realized when the company

optimizes the many processes that affect the Key Performance Indicator (KPI) metrics that management is reporting and analyzing.



BI implementations typically have many stages:

1. Fulfilling reporting and information delivery needs
2. Dashboards and KPI displays
3. Advanced analysis of the probable causes of exceptions
4. Integration of reporting with Process Management/Workflow
5. The analysis and tuning of those processes

Unfortunately, many customers only get as far as implementing reporting and the lack of proper process management and workflow coupled with the absence of methodologies precludes them from going further. They spend massive amounts of money on proprietary BI solutions in the hope that these software products will help. Unfortunately, traditional BI tools do not meet these goals

Traditional Solutions and Their Shortcomings

| | |
|-------------------------------------|---|
| Price | The software license, maintenance, support, and services are too expensive. |
| Usability | Too difficult to use for most users. |
| Skills | Lack of adequate skills transfer from vendor to customer. Lack of implementation methodologies. |
| Customization | Too difficult for customers to develop solutions and integrate business rules. |
| Tool-Set orientation | The 'solutions' are tool sets and not a solution at all. |
| Extensibility | The solutions are proprietary and difficult or impossible for customers and aftermarket suppliers to extend and direct the system. Customers did not buy the software, they paid upfront for the right to use it. This is like getting a lease on a car but making all the payments on day one: it's the worst of both worlds. |
| Reporting and analysis focus | The solutions are focused on the reporting and analysis of KPIs, and ignore the performance of the processes that affect the metric. |
| Process influence | They are unable to ensure driving changes in a business process. They assume that the delivery of a report will have the side effect of influencing a business process. |
| Tracking and Auditing | They are unable to provide complete tracking and auditing. Who got the report? What action did they take? How long did it take? Was a business process initiated as a result? How far along is that process? What is the performance of the process? |
| Prototyping | The software pricing models do not support the prototyping phases necessary to ensure the success of Business Intelligence projects. Significant financial outlay and contractual agreements must be signed before full evaluation and prototyping can be done. |
| License flexibility and fit | Traditional BI license models are built with the assumption that a named user spends much of his/her day using the BI system as a standalone application. The best example of where this model breaks down is a deployment to extranet users – infrequent, casual users, outside the firewall, where a multi-thousand dollar named user license is nonsensical. |

The Pentaho Open BI Suite

The Pentaho Open BI Suite is different from traditional BI offerings. It is a process-centric, solution-oriented platform with BI components that enable companies to develop complete solutions to BI problems.

The Pentaho BI Platform, the core architecture and foundation of the Pentaho Open BI Suite, is process-centric because the central controller is a workflow engine. The workflow engine uses process definitions to define the business intelligence processes that execute within the BI Platform. The processes can be easily customized and new processes can be added. The BI platform includes components and reports for analyzing the performance of these processes.

The BI platform is solution-oriented because the operations of the platform are specified in process definitions and action documents that specify every activity. These processes and operations collectively define the solution to a business intelligence problem. This BI solution can be easily integrated into business processes that are external to the platform. The definition of a Solution can contain any number of processes and operations.

The Pentaho Open BI Suite consists of a BI platform, end-user BI capabilities, and the Pentaho Design Studio:

- The BI platform provides an execution framework and services that include logging, auditing, security, scheduling, ETL, web services, attribute repository and rules engines.
- The end-user BI capabilities include reporting, analysis, workflow, dashboards, and data mining.
- The Pentaho Design Studio is a set of design and administration tools that are integrated into the popular Eclipse environment. These tools allow business analysts or developers to create reports, dashboards, analysis models, business rules, and BI processes.
- The BI platform and end-user BI capabilities form the Pentaho Server. BI solutions are as designed using the Pentaho Design Studio and deployed to the Pentaho Server. The Pentaho Server is the runtime engine, driven by the workflow engine, which coordinates the execution and communication between all the BI Components.

Centralized control by a workflow engine is essential to the architecture of the BI Platform:

- The platform is built on processes and process definitions. The BI platform understands the nature of processes because everything in it executes as one.
- The processes are defined in a standard process definition language that is externally viewable, editable, and customizable. All business logic is exposed, and nothing is hidden.
- Logging, auditing and security are built in at the core and are utilized automatically to ensure that there is always an accurate audit trail available for both governance and performance monitoring.

The architecture is a combination of original source code and mature open source components that have been integrated to form a complete, scalable, sophisticated BI platform.

The Pentaho BI platform is built upon a foundation of servers, engines, and components. These provide the J2EE server, security, portal, workflow, rules engines, charting, collaboration, content management, data integration, analysis, and modeling features of the system. Many of these components are standards-based and can be replaced with other products.

To create a truly integrated, single-source solution, Pentaho adds the following:

- Common metadata in the form of solution definition documents
- Common user interfaces and user interface components
- Security
- Email and desktop notifications
- Installation, integration and validation of all components
- Sample solutions
- Application connectors
- Usage and diagnostic tools
- Design tools
- Customization and configuration

Open Source Technology

The Pentaho Open BI Suite uses a number of third-party Open Source components. The Architecture section shows the architecture and the relationship of those open source components to the Pentaho components. Open Source components allow Pentaho to focus on the value-added functionality of the system: reporting, analysis, workflow integration, business rules, auditing, user notification and interfaces, and analysis / modeling of workflow.

For a list of web sites that discuss the nature, and advantages, of open source software see Appendix B. In return for the advantage that Pentaho gains by using these open source components in its platform Pentaho contributes resources to these projects and releases components of the BI platform under an open source license.

Architecture

To deliver the solution described above, the Pentaho BI Platform is comprised of the Pentaho Server, and an Eclipse-based Design Studio.

Pentaho Server

The Pentaho Server is made up of a BI platform and libraries that deliver end-user BI capabilities. The Server runs inside a J2EE compliant web server such as Apache, JBOSS AS, WebSphere, WebLogic and Oracle AS. These can also be embedded within other servers or applications. The diagram which follows in a couple of pages shows the relationships of the major components within the server.

The Pentaho Server enables the many functions of the BI platform to be presented to users in a consistent, familiar look and behavior. For example, one component generates the list of reports that a user has access to, a second lists the task-related deadlines in a calendar, and a third shows the current tasks that the user needs to complete. The content generated by each component is relevant for each user's roles. Component content can be retrieved as XML, HTML, or displayed by the included JSR-168 portlets. The Pentaho portlets can be embedded into any portal that supports the JSR 168 standard such as IBM WebSphere, OracleAS Portal, and BEA WebLogic Portal. The XSL and CSS stylesheets used by the components to generate online and report content are accessible and can be fully customized.

The Pentaho Server contains the engines and components for reporting, analysis, business rules, email and desktop notifications, and workflow. These components are integrated together so that they can be used to solve a Business Intelligence problem. Within a solution the behavior, interoperation, and user interaction of each sub-system is defined by a collection of Solution Definition documents.

The Solution Definition documents are XML documents that contain

- Definitions of business processes (XPDL)
- Definitions of activities that execute as part of processes, on demand, or called by web services. These activities include definitions for: Data sources, queries, report templates, delivery and notification rules, business rules, dashboards, and analytic views
- The relationships between all of the above items

The components in the server rely upon a Solution Engine for information about the available solution documents, for security, for information about reports and workflow items, for data, and for auditing. More than one solution can execute in the server. The Solution Definition documents can be copied from one server to another and may be freely distributed.

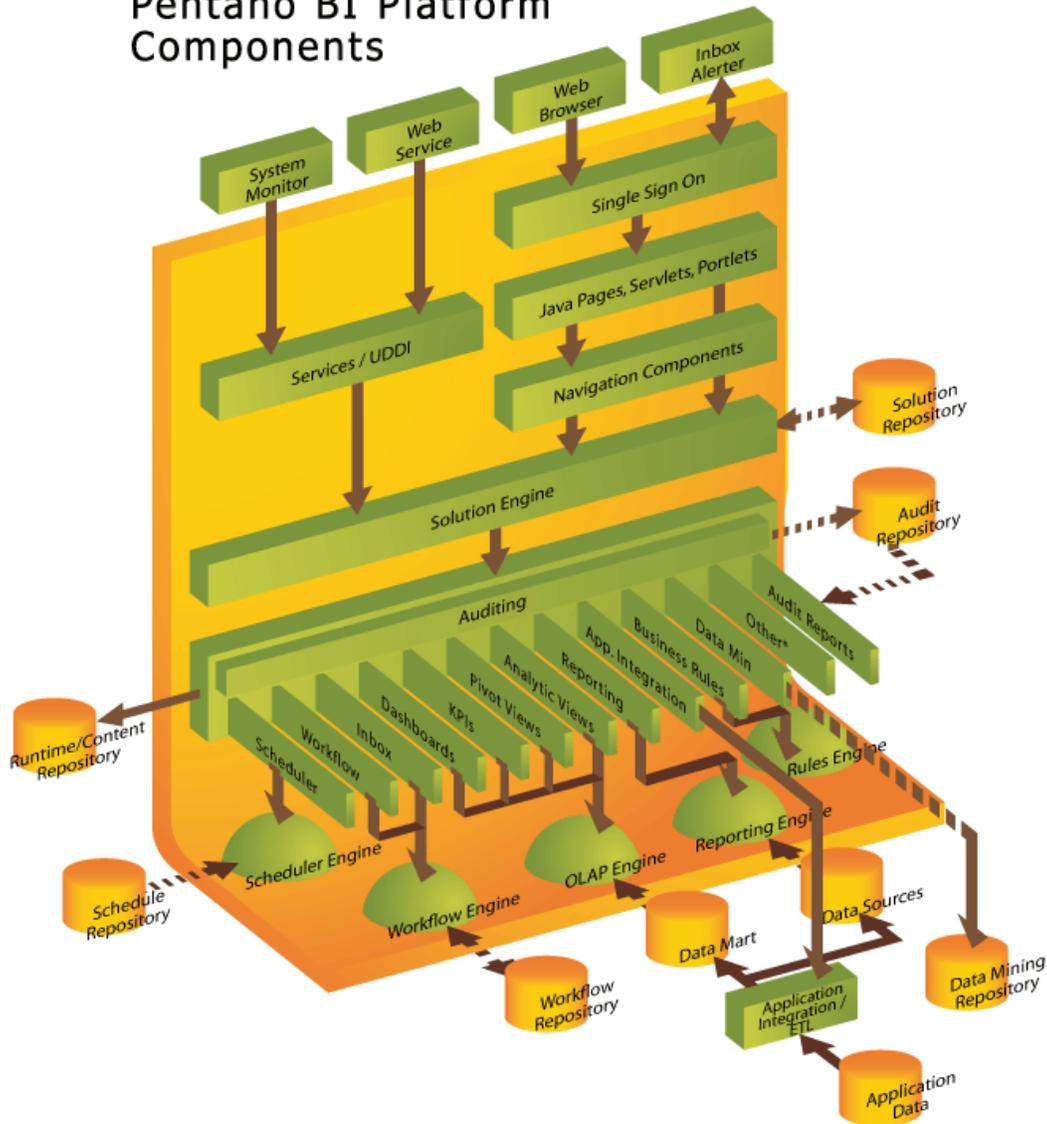
The Pentaho Server contains infrastructure that provides advanced system administration. This includes system monitoring (SNMP) services, usage reports, Web Service support, configuration validation tools, and diagnostic tools.

The Pentaho Server contains systems and components that provide advanced process performance reporting and analysis. This includes slice-and-dice, what-if, and data-mining capabilities on the attributes of workflow items, individual tasks, employees and services involved in workflow tasks.

The Pentaho Server also supports Enterprise Application Integration (EAI) for live integration with operational applications as well as Extract, Transform, and Load (ETL) capabilities for creating data warehouses and data marts.

A server architecture diagram is provided on the next page.

Pentaho BI Platform Components



- The Solution Engine is central to the architecture and manages access to the BI components.
- The services of the BI platform:
 - Provide web services to external applications
 - Have access to the same Solution Engine as the user interface components
 - Are called by the workflow engine and scheduler to execute system actions
- Auditing is built into the platform components. The platform can provide process performance reports by extracting historical and real-time data from the workflow and auditing repositories.
- Components are modules that can be added to the system.
- Each engine has corresponding component(s) that integrate the engine into the platform. Engines can be switched out for other engines or added to the platform if the necessary components are created.
- The Server includes the components and technologies required to build a business intelligence solution: reporting, workflow, business rules, dashboards/analysis, web services, scheduling, a mix of convenient web and desktop user interfaces, and auditing.

- The Pentaho BI platform integrates with external systems that provide data to drive the reporting engine and that receive events from the workflow engine.
- The Pentaho BI platform provides system monitoring via Simple Network Management Protocol (SNMP).
- The repositories are stored inside an RDBMS that is outside of the Pentaho platform. The embedded repositories in the preconfigured installation are stored inside an open source database, either FireBird or MySQL. These repositories can be replaced with other relational databases such as Oracle, SQLServer or DB/2 if required.
- Multiple rules engines are delivered with the Pentaho BI Platform so that business logic is exposed and can be customized easily. Additional rules engines can be added to the system. The business rules engines are external to the components, and any component can utilize any rules engine.
- Not all components are shown on the diagram, other components include email, printing, message formatting, workflow instance attribute management, and process performance reporting and 'what-if' analysis.
- Some parts of the architecture use a combination of technologies or technologies that can be readily swapped for equivalent ones:
 - The J2EE Server provided is JBoss AS, but any Java JDK 1.4 compliant application server can be used.
 - The BI platform provides user interfaces built with Java Server Pages (JSPs), servlets and portlets. Third party or customized JSPs, servlets or portlets can be added
 - The BI Suite includes an open source OLAP engine (mondrian), but any MDX-compliant OLAP server could be used (we plan to test with Microsoft OLAP Services and Hyperion Essbase)
 - The Platform provides a javascript-based and SQL-based rules engines and plan to provide support for ILOG JRules, Drools and Jess based on market demand
- The Pentaho BI platform integrates and enhances existing popular third-party Open Source components such as:
 - Mondrian OLAP Server and jPivot Analysis Front-End
 - Firebird RDBMS
 - Shark and JaWE Workflow
 - Kettle EII and ETL
 - JBoss Application server, Hibernate and Portal
 - Weka Data Mining
 - Eclipse Workbench and BIRT reporting components
 - JOSSO single sign-on and LDAP integration
 - Mozilla Rhino Javascript Processor
- The BI Platform utilizes open standards and protocols including:
 - XML – W3C's Extensible Markup Language
 - JSR-94 – JCP's Rules Engine API
 - JSR-168 – JCP's Portlet Spec
 - SVG – W3C's Scalable Vector Graphics
 - XPDL – W3C's XML Process Definition Language
 - XForms W3C's Web Forms
 - MDX – Microsoft's OLAP Query Language
 - WSBPEL - Oasis's Web Services Business Process Execution Language
 - WSDL - W3C's Web Services Description Language

- SOAP - W3C's Simple Object Access Protocol

Server Repositories

The Pentaho Server includes embedded repositories that store the data necessary to define, execute and audit a solution:

- Solution Repository: The metadata that defines solutions
- Runtime Repository: Items of work that the workflow engine is managing
- Audit Repository: Tracking and auditing information

Pre-Configured Installation

A preconfigured sample deployment is provided so that the platform can be tested quickly and easily.

The deployment includes

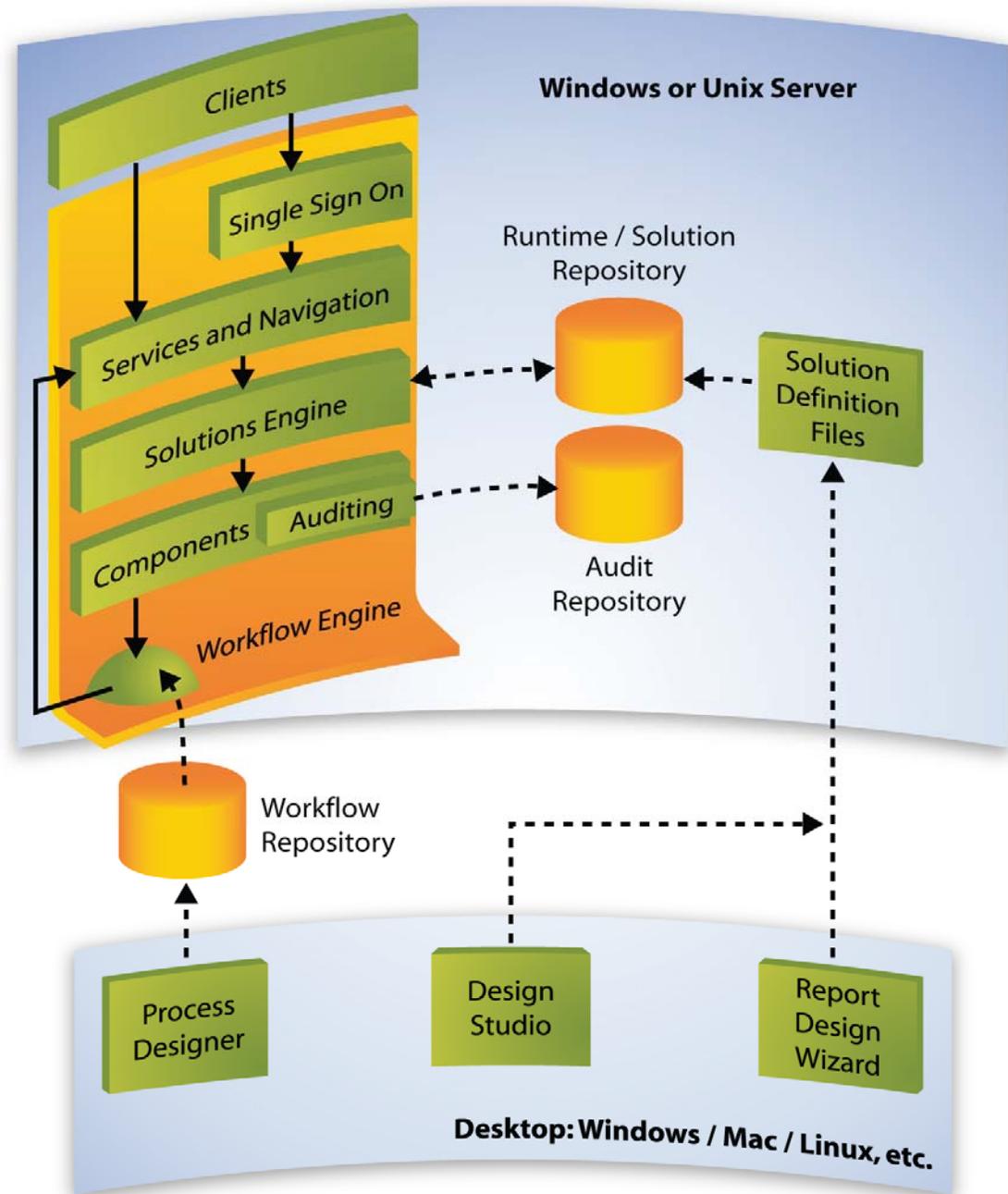
- JBoss Application Server
- JBoss Portal V2.0, a JSR-168 certified portal server
- Example JSPs that demonstrate platform component usage
- Sample data
- Sample reports and BI processes
- Users and roles used in the examples

Pentaho Design Studio

The design and administration workbench is an Eclipse-based desktop workbench that provides:

- Easy to use design tools for reports, dashboards, analytic views
- Workflow process designer
- Business rules editors
- Data mining console for data preparation
- OLAP modeling tools

An architecture diagram for the Pentaho BI Workbench is provided on the next page.



Notes on the Pentaho Design Studio architecture:

- The Design Studio is a desktop Eclipse-based design and administration environment.
- The Design Studio generates workflow definitions and Solution Definition files that are used by the server to execute BI solutions.
- The Design Studio runs on multiple platforms.
- The Design Studio audits creation and editing of Solution Definition documents.
- A version control system can be used to maintain the Solution Definition documents and provide synchronization and versioning capabilities.
- The Design Studio allows solutions, reports, queries, business rules, dashboards, and workflows to be viewed and edited graphically. The Design Studio is a Java application that is installed on system administrator's and designer's desktop computers.

Embedded Architecture

- Pentaho technology can be embedded into standalone or server-based Java applications.
- The Solution Engine and components package must be installed.
- Only those components, engines, and repositories that are required need to be configured.
- Specifically these components are optional:
 - Workflow engine, workflow repository and runtime repository
 - Auditing and audit repository
 - Application Integration / ETL for data extract, transformation and loading
 - User interface components
 - Solution repository and solution definition files

The Pentaho Solution

The Pentaho BI platform integrates workflow, business rules, information delivery and notification, scheduling, auditing, application integration, content navigation, user interfaces, design and administration tools with reporting, analysis, dashboards, and data mining components and engines.

The architecture of the Pentaho BI Platform has many advantages. Specifically:

- By building, integrating, and enhancing open source components into a single integrated platform the cost of BI implementations is drastically reduced.
- Lower cost of ownership means resources can be invested elsewhere, such as value added implementation or customization services, or increasing the scope of the BI project and deploying more advanced content and capabilities to end users. In other words, a significantly higher percentage of the project budget can be spent on requirements gathering, implementation, and services increasing the success of the project.
- Delivering the software with no cost for prototyping enables prototyping to be performed for any duration required.
- Delivering the software with no cost for prototyping enables project requirements iterations to be performed for any duration required.
- By building a workflow-based platform a true service-oriented architecture is delivered.
- By building a workflow-based platform it is easy to integrate BI into any business process.
- By building a workflow-based platform the system is easily to cluster and scale. By delivering the software with no cost the system can be deployed on any number of servers without additional software costs.
- By building a workflow-based platform with process performance reports BI projects can be continually tuning and improved
- By building information delivery and notification into the platform reports, analysis, tasks, and decisions points can be routed to anyone involved in a business process.
- By building multiple rules engines into the platform all business logic is customizable.
- By building reporting, analysis, and dashboards into the platform a seamless increase in sophistication of the BI solution is enabled. This can be performed at a pace that is right for the

organization and is not constrained by additional software costs or driven heavily by purchasing cycles or vendor quarter-closing and fiscal commitments.

- By building data mining into the platform the most advanced data analysis possible can be added in a timely basis.
- By delivering an out-of-the-box server and a set of components and engines it is easy to deploy and integrated any required configuration.
- By integrating auditing and audit reports, system monitoring, and administration features into the platform the system is easy to maintain.
- By providing intuitive user interfaces that are readily customizable, the system is to use and the cost of training users is reduced.

To reproduce the Pentaho architecture it would be necessary to:

- Define requirements for entire architecture
- Determine whether to design and build each component or use existing third party ones
- Identify suppliers for each of the many components/projects
- Perform due diligence research on each component/project
- Install and configure each component
- Design and implement an integration layer for each component
- Design and implement consistent user interface components
- Design and implement consistent administration tools
- Design and implement analysis and modeling tools
- Design and implement the common services and infrastructure
- Design and create repositories
- Design and implement new components or enhance existing components with new functionality
- Integrate security
- Integrate auditing
- Design and implement process performance reports
- Create a common definition language
- Have significant experience with this kind of product development

Due to its architecture and licensing, using this BI platform as part of a business intelligence solution:

- Increases the budget and time available for requirements gathering
- Increases the budget and time available for prototyping
- Increases the budget and time available for implementation and project rollout
- Increases the budget and time available for services and training
- Accelerates the start of the implementation phase
- Reduces the range of skills needed to implement the solution

These factors combine to decrease the cost and risk of the project and increase the likelihood of a successful result.

Pentaho's Role

Pentaho facilitates and manages development of the Pentaho BI Platform. Pentaho provides the seed source code, provides full-time engineering and services personnel, and manages the project roadmap.

Specifically, Pentaho:

- Builds components for the open source community
- Contributes to embedded open source projects
- Enhances components developed by others
- Integrates components into cohesive and flexible building blocks that Java developers can use to rapidly assemble custom solutions
- Uses these building blocks to create complete, out-of-the-box products and a comprehensive BI platform for end-users
- Provides comprehensive training, technical support, release management, quality assurance, and enterprise services
- Fosters open source BI collaboration via its website and forums
- Uses professional engineering approaches
- Makes the entire solution available – No hidden upgrades
- Addresses BI shortcomings and Open Source concerns

Summary

| | |
|-----------------------------|---|
| Price | By leveraging open source, Pentaho is able to dramatically lower the cost of ownership of BI solutions. Open source components help deliver a robust platform that can be scaled. The cost of ownership is primarily and appropriately focused on customization of solutions and integration/automation of business processes. |
| Usability | The BI platform has intuitive task-based inboxes, tight integration with external systems, and provides complete customization of the user interfaces. This results in solutions that are much easier to use. |
| Skills | The BI platform uses recognized standards wherever possible so that third-party tools and portable skills can be used. |
| Customization | The BI platform has external rules engines, editable processes, source code availability, re-usable BI components, and fully customizable user interfaces. The platform executes editable processes and can be integrated with external business processes. It also leverages open standards for customization and integration wherever possible. This results in a highly customizable solution. |
| Tool Set orientation | The BI Platform is process-centric and solution-oriented. It is architected not to be a tool-set but to be a fully-embeddable, easily-integrated system. |
| Extensibility | Source code is provided. The infrastructure, components, engines, and interfaces can be modified at will. Money saved on license revenue can be directed towards extending the platform in any required area. The roadmap of the platform's features is public and feature development can be sponsored. |
| Reporting and | The BI platform is process-centric, workflow-based, and directed by |

| | |
|------------------------------------|---|
| analysis focus | business-rules. It can be easily integrated into business processes and has process performance reporting built in. Business rules are used to identify issues and exceptions and audited processes are executed to resolve them. |
| Process influence | The BI platform tracks information delivery and allows customized business processes to be started by the recipients. The delivery of a report is not a 'dead-end'. Process performance reports are generated and process modeling can be done. Actions taken can be plotted against fundamental metrics to gauge the effectiveness of the processes. |
| Tracking and Auditing | The BI platform includes robust process-centric tracking and auditing. The history of a situation can be assessed including the actions taken, discussions/ commentary, and supporting information. |
| Prototyping | The BI platform licensing allows prototyping to be done for as long as necessary without significant financial expenditure. |
| License flexibility and fit | The BI platform licensing is flexible, and reflects real-world customer value, where BI is a part of larger business processes, adding incremental value. Beyond that, CPU-based licensing at a reasonable cost opens up possibilities for extranet applications that would not be realistic with traditional role-based, named-user-only licensing. |

Appendix A: Glossary

| | |
|----------------|---|
| BI | Business Intelligence. A sector of the IT market that includes reporting and analysis applications. |
| BPEL | Business Process Execution Language. A standard system used to orchestrate workflows across multiple services. The web services are called BPEL4WS or WSBPEL. |
| CMS | Content Management System. |
| HTML | Hyper Text Markup Language. The language used to create web pages. |
| IBM WebSphere™ | IBM Corporation's proprietary portal that supports the JSR 168 standard |
| J2EE | Java 2 Enterprise Edition. A platform independent standard for modular multi-tier enterprise software |
| Java™ | A platform independent language for writing software. |
| JBoss™ | A professional open source company that provides a number of open source technologies along with support and services. |
| JBoss™ | JBoss Inc's open source portal that supports the JSR 168 standard. |
| Portal | |
| Jetspeed | Apache Foundation's open source portal that supports the JSR 168 standard |
| JSR 94 | Java Rule Engine API. A standard specification for java-based rules engines |
| JSR 168 | Portlet Specification. A standard specification for creating user interfaced for embedding into a portal. |
| JSR 170 | Content Repository for Java Technology API. A standard for implementing content management systems. |
| Kerberos | A standard network authentication protocol. |
| KPI | Key Performance Indicator. A performance metric tracked by a business. |
| LDAP | Lightweight Directory Access Protocol. A standard protocol for accessing properties about resources, e.g. employees or web services |
| Metadata | Information used to describe the structure and content of data. |
| MS Excel™ | A Windows™ program that executes inside of MS Excel™ |
| Add-In | |
| Notification | Commonly referred to as the 'System Tray'. The area of the Windows™ |
| Tray | Taskbar that shows alerts and services |
| OLAP | On-Line Analytical Processing. An application that allows multidimensional data to be navigated, reported, and analyzed. |
| OracleAS | Oracle Corporation's proprietary portal that supports the JSR 168 |
| Portal™ | standard. |
| Portal | A web based application for aggregating content from different systems to display in one or more multi-functional pages. |
| Portlet | A component of a portal that provides content for portal pages from another system |
| Third Party | A portal server that is not the portal server embedded into the Pentaho |
| Portal | BI Platform |

| | |
|------------------|---|
| WSRP | Web Services for Remote Portlets. A specification allowing one portal to call another portal to get content to display. |
| XML | eXtensible Markup Language. A standard language for creating machine-readable data. Commonly used to transfer and transform data. |
| XPDL | XML Process Definition Language. A standard language for describing a process in XML. |
| XSL Style sheets | A standard for transforming an XML document into another format, e.g. HTML |
| Weblogic Portal™ | BEA Corporation's proprietary portal that supports the JSR 168 standard. |

Appendix B: Open Source Software

The Open Source software movement began in the 1970's and is gaining such significant momentum that:

- Companies such as IBM and Sun are releasing previously proprietary products to the Open Source community for enhancement, performance, and stability improvements. A primary example of this is Eclipse, an Open Source development tool that IBM donated to the Open Source community (formerly IBM Workbench). Here are some characteristics of the Eclipse project, which is illustrative of the kind of resources and support many mature Open Source projects evolve to enjoy:
 - Eclipse is guided by a board of steward companies including IBM, Borland, Rational, Red Hat, Sybase, Fujitsu, Hitachi, Oracle, SAP, OMG, and Intel.
 - Eclipse is the most popular Java development environment in North America (45% market share) and, with an annual 60% increase in Market share, looks to become the worlds leading Java tool.
 - Eclipse has over 20 Universities worldwide contributing to it and eclipse.org has sufficient funding to make research grants.
 - The keynote speakers at the Eclipse 2005 conference were Tim O'Reilly, from O'Reilly & Associates publishing (considered to be the worlds best computer book publisher), and Urs Hoelzle (Google's first V.P. of Engineering)
 - More than 1,200 developers in over 60 countries are working on the project
- A Nov 2004 survey of the world's web servers found that over 70% (13.1 million) are Apache, an Open Source product, and it is the only web server gaining market share.
- Open source products and products including or built on open source: FlexWiki (Microsoft), Windows Installer (Microsoft), Windows Template Library (Microsoft), Cloudscape database (IBM - worth \$85 million), StarOffice (Sun), CVL (Intel), Oracle Internet Application Server, Oracle Enterprise Manager, Ingres (Computer Associates), SADB (SAP), OS X (Apple).
- In an attempt to compete with open source operating systems Microsoft agreed to give the source code for Windows to the government of China.
- Information Week reports that 65% of companies surveyed use open source products and only 17% of companies have no short term plans to use open source.
- IBM and Sun (competitors in most markets) are working on an open source Java implementation together.
- California Governor Arnold Schwarzenegger promised a complete audit of California's procedures and finances. That audit resulted in the California Performance Review. Information technology was considered in the review, and the California Performance Review recommends that the State of California explore Open Source alternatives (<http://www.report.cpr.ca.gov/cprprt/issrec/stops/it/so10.htm>).

Additional Web-based Resources

Open Source.Org

<http://www.opensource.org/>

O'Reilly Publications

<http://opensource.oreilly.com/>

Open Source Advantages and Limitations – Overview

http://www.masternewmedia.org/2002/06/30/open_source_advantages_and_limitations_overview.htm

Open Source Advantages for IT Consultants

<http://www.cyber.com.au/misc/frsbiz/consults.htm>

Perceived Disadvantages of Open Source Models

http://eu.conecta.it/paper/Perceived_disadvantages_ope.html

Advantages of Open Source Software

http://www.cs.virginia.edu/~pev5b/writing/econ_oss/advantages.html

Open Source Funding: \$40 million from IBM alone

<http://news.com.com/2100-1001-275388.html?legacy=cnet>

Ten Myths About Open Source

http://opensource.oreilly.com/news/myths_1199.html

Information Week: Open-Source Software Use Joins The Mix

<http://informationweek.smallbizpipeline.com/trends/51201995>

CIO Magazine: Your Open Source Plan

<http://www.cio.com/archive/031503/opensource.html>

CIO Magazine: Who Are Those Guys?

http://www.cio.com/archive/031503/opensource_sidebar_3.html

Open Source Goes Mainstream

<http://www.minoru-development.com/en/opensource.html>