Business Intelligence

Vol. 7, No. 2

CUTTER CONSORTIUM

Open Source Business Intelligence

by Brian J. Dooley

Open source business intelligence has moved into the limelight with the recent release of software with improved functionality. Key features like standardization, simplicity, and cost are particularly attractive to smaller organizations and also helping to bring BI to areas where it might not have been considered previously. There is great promise in open source BI, but it needs to be considered carefully, as it is neither as free nor as simple as it might initially appear.

Executive Report

Cutter Business Technology Council

Rob Austin | Tom DeMarco | Christine Davis | Lynne Ellyn | Tim Lister | Lou Mazzucchelli | Ken Orr | Sheleen Quish | Ed Yourdon

About Cutter Consortium

Cutter Consortium is a unique IT advisory firm, comprising a group of more than 150 internationally recognized experts who have come together to offer content, consulting and training to our clients. These experts are committed to delivering top-level, critical, and objective advice. They have done, and are doing, ground-breaking work in organizations worldwide, helping companies deal with issues in the core areas of software development and agile project management, enterprise architecture, business technology trends and strategies, enterprise risk management, business intelligence, metrics, and sourcing.

Cutter delivers what no other IT research firm can: We give you Access to the Experts. You get practitioners' points of view, derived from hands-on experience with the same critical issues you are facing, not the perspective of a desk-bound analyst who can only make predictions and observations on what's happening in the marketplace. With Cutter Consortium, you get the best practices and lessons learned from the world's leading experts, experts who are implementing these techniques at companies like yours right now.

Cutter's clients are able to tap into its expertise in a variety of formats including print and online advisory services and journals, mentoring, workshops, training, and consulting. And by customizing our information products and training/consulting services, you get the solutions you need, while staying within your budget.

Cutter Consortium's philosophy is that there is no single right solution for all enterprises, or all departments within one enterprise, or even all projects within a department. Cutter believes that the complexity of the business technology issues confronting corporations today demands multiple detailed perspectives from which a company can view its opportunities and risks in order to make the right strategic and tactical decisions. The simplistic pronouncements other analyst firms make do not take into account the unique situation of each organization. This is another reason to present the several sides to each issue: to enable clients to determine the course of action that best fits their unique situation.

For more information, contact Cutter Consortium at +1 781 648 8700 or sales@cutter.com.

Open Source Business Intelligence

Business Intelligence Advisory Service

Executive Report, Vol. 7, No. 2

by Brian J. Dooley

Open source business intelligence (BI) is attractive for a number of reasons, especially a lower initial cost, perceived lower overall cost, simplicity of operation, and adherence to standards. The power and capability of business intelligence is growing, and the capacity to perform sophisticated data analysis with easy-access reporting — hallmarks of BI — is increasingly attractive to smaller businesses that do not have the budget to undertake full-scale commercial implementations. Another key emerging market is the incorporation of BI capability directly within applications, either fully embedded or accessed. Open source products are attractive to developers because they can be customized under open source licenses and because many of them are written in Java and conform to standards.

Complete integrated BI suites are a relatively recent development. The prominence of open source BI solutions has risen lately as several integrated suites have come of age. The principle integrated open source solutions today are Pentaho and JasperSoft. The BIRT (Business Intelligence and Reporting Tools) Project has also gained notice, though it is less well advanced than the other two and is currently mainly a reporting solution.

Integrated BI suites bring together a variety of existing open source tools (such as Mondrian OLAP and JPivot), add to them, integrate them, and provide them with an accessible BI platform. The major open source BI projects are offered with a combination of open source and commercial licenses, with the commercial licenses offering added features or added service. Although this departs from a pure open source philosophy, it provides revenues needed to integrate these diverse platforms, as well as an improved ability to try out the software. The "free" version can generally be upgraded later, and the costs of the commercial version of the open source software (OSS) are much lower than those of high-end commercial alternatives.

Open source BI solutions are primarily targeted at smaller businesses and enterprises and at limited implementation scenarios. However, as BI capability continues to be extended to new

territories, open source provides an avenue for experimentation.

For companies investigating the use of an open source solution, it is important to note that the available components are modular and generally meet important standards. This means that incremental installation is possible, beginning with only the capabilities that are required. It is also generally possible to mix open source and commercial products within the same environment, using, say, an open source database with a commercial analytic tool or a commercial extraction, transformation, and loading (ETL) tool with an open source BI platform.

To understand open source BI, it is important to consider the overall implications of implementing open source products as mission-critical applications and to look at the basic requirements of a BI solution. Therefore, this *Executive Report* begins with a discussion of the open source environment before laying out the components of a BI solution. The report then offers an overview of open source BI solutions before focusing on the three major projects introduced above: Pentaho, JasperSoft, and BIRT. The report also includes a brief look at other tools, including Bizgres, the BEE Project, and Jedox Palo-Server. The report concludes with a discussion of selection and implementation considerations.

THE OPEN SOURCE ENVIRONMENT

Open source software is different in culture and economics from commercial software. Development occurs across a wide range of individuals and institutions, many of whom are also users. For purists, payment is in ego points, but, in a practical sense, most payment arrives through service contracts. Development and distribution, though guided by an orchestrating group, tend to be haphazard, and new ideas come from many directions.

Commercial software development tends to be monolithic. Management is hierarchical, and special procedures must be put into place to ensure maintenance is done, bugs are fixed, to make optimum use of available programmers, and so forth. In OSS, these issues are served by the community of interested users, augmented by the license that requires free distribution.

These issues tend to make open source software more standardized, easier to modify, and closer to the programming environment than commercial solutions. While commercial solutions win in complexity and capability, open source can generally win — or put up a strong fight, at least — in flexibility and customization capabilities.

A number of studies have been done over the years on the relative costs of running both open source and commercial products. The results have generally been ambiguous, heavily dependent upon the particular situation under consideration — and on who paid for the survey. The true cost of commercial software is often buried in bundling deals and licensing details; the true cost of open source software tends to lie in a potentially larger implementation effort. In both cases, for complex solutions, the cost — which must include ramp-up, training, customization, trials, installation, and so forth — is likely to be far greater in implementation than in purchase.

Money in computing has been derived, through much of the industry's history, from service, installation, and support. It comes from customization, from modification of software to unique circumstances, and from direct sales to organizations that require a prebuilt or precustomized solution. This could be a product assembled by specially configuring and tweaking a number of commercial

The ***Business Intelligence Advisory Service Executive Report*** is published by Cutter Consortium, 37 Broadway, Suite 1, Arlington, MA 02474-5552, USA. Client Services: Tel: +1 781 641 9876 or, within North America, +1 800 492 1650; Fax: +1 781 648 1950 or, within North America, +1 800 888 1816; E-mail: service@cutter.com; Web site: www.cutter.com. Group Publisher: Chris Generali, E-mail: cgenerali@cutter.com. Managing Editor: Cindy Swain, E-mail: cswain@cutter.com. ISSN: 1540-7403.

products developed by different organizations. Or, it could just as easily be a preconfigured open source product, such as an integrated BI suite.

The basic software models, characterized by distribution type, are as follows:

- **Proprietary/commercial distribution, in binary form with no available source code.** This has been extended to include cases where source code is included for reference only or where source code is included under the assumption that only the vendor or vendor-licensed operators will make changes. Sale is generally through licensing, which may be indefinite for smaller products or have a set time frame with conditions for renewal.
- **Free, semi-free, and partly free distribution, much of which can be considered either open source or allied with open source.** Within the "free" camp, there are a number of important sub-models, the most important of which claim to be open source. In addition to a distribution model, open source licenses have important consequences that encourage product development by an extended community of engineers. "Free" software variants include the following (not all are "free" in the sense of code sharing):
  - *Shareware software* — typically provided in binary code only, at no cost (and often with some functional limits) for an initial period, but requiring a purchased license after the passage of that period.
  - *Freeware* — provided with no license fee at all and generally released only as binary code. Freeware is often used as a "loss leader" to draw attention to a vendor's commercial products.
  - *Open source software* — provided with source code and permission to modify. Remuneration differs. Most is available via download for free, but there are often configurations available for sale by vendors. The main characteristic is that the users can freely use, modify, and redistribute the software. A variety of licenses are available under this general category, some of which are GNU Public Licenses (GPLs).

The GPL is the most widely used open source license. Products under this license include the GNU Project and Linux. Key points of the GPL are:

- Software licensed under the GPL can only be copied and distributed under this license.
- Products licensed under the GPL may be sold.
- Users can alter the source code, but if the result is distributed or published, it must be made available under the GPL.
- Ancillary technology can be developed, and as long as such products do not include code licensed under GPL, they need not themselves be licensed or made available under the GPL.

The GPL represents the Free Software Foundation's philosophy. Several less restrictive variants exist, including the GNU Lesser General Public License (LGPL) and vendor licenses based on the Mozilla Public License (MPL). While these licenses impose some GPL-like restrictions on the use of software, licensed software can be incorporated into products that can then be licensed without "importing" similar restrictions. Some of the common licenses found in open source BI software are described in Table 1.

OSS and its variants are often also referred to as "free software." Free software refers to the user's freedom to run, copy, distribute, study, change, and improve the software, rather than to lack of cost.

OSS has largely entered the enterprise through the back door. Much like the early personal computers, it has been confined in the past to technical areas, scientific workstations, and special applications such as intranet Web servers. Recently, with support from major vendors increasing and with a significant presence developing in key areas — combined with significant publicity — open source software has reached a level of acceptance in the enterprise.

Table 1 — Open Source Licenses

Reference	Specific License	Description
Apache	Apache License Version 2.0	Free software[1] license, incompatible with the GPL due to a specific requirement regarding certain patent issues.
	Apache License Version 1.1	Free software license, non-copyleft,[2] with some requirements making it incompatible with the GNU GPL.
	Apache License Version 1.0	Free software license, simple, permissive, but non-copyleft. Not fully compatibility with the GNU GPL.
Artistic	Artistic License 2.0	Free software license compatible with the GNU GPL. This license is being considered for use in Perl 6.
	Original Artistic License	Ambiguously worded license, later modified in version 2.0.
Eclipse	Eclipse Public License Version 1.0	Free software license, incompatible with the GPL due to a specific requirement regarding certain patent issues.
LGPL	Lesser GNU Public License	Free software license, with limited copyleft rights permitting linking with non-free modules. It is compatible with the GNU GPL.
BSD	Original BSD License	Free software license, non-copyleft, with an advertising clause that makes it incompatible with the GNU GPL.
	Modified BSD License	Free software license, non-copyleft. Modified by removal of the advertising clause and compatible with the GNU GPL.
MPL or Mozilla	Mozilla Public License	Free software license, weak copyleft. Some restrictions that make it incompatible with the GNU GPL. MPL 1.1 possibly removes these restrictions, but the issue can be complex.
Sun	Sun Public License	Free software license, weak copyleft. Some restrictions that make it incompatible with the GNU GPL. Same as the Mozilla Public License (note that the Sun Community Source Licensing Program is not a free software license).
CPL	Common Public License Version 1.0	Free software license, incompatible with the GPL due to a specific requirement regarding certain patent issues.
EnterpriseDB		Commercial license permitting code modification but no release to the public.
GPL	GNU Public License	This is a free software license and a copyleft license.

[1]Free software means the user is free to run, copy, distribute, study, change and improve the software.
[2]Copyleft means that anyone who redistributes the software, with or without changes, must pass along the freedom to further copy and change it.

Open source and current commercial software models can coexist. Open source is likely to create pressure for lower costs, as well as to continue to develop an alternate pool for innovation. This can be demonstrated in the growing pool of open source BI products.

Open source is likely to continue its movement into the enterprise, ushered in partly by Linux and partly by the fact that, after all, it is not so different from the other

ways that software has been developed and distributed in the relatively brief history of computing.

COMPONENTS OF A BI SOLUTION

The basic definition of a BI tool is as follows: an application used for transforming data into information to derive knowledge through reporting and analysis of structured information. This definition has been broadened somewhat in practice and under pressure of competition.

Most BI tools are designed to operate within a data warehouse environment, with structured information brought together through the use of metadata providing data mappings, transformation rules, and control logic. Standard query and reporting tools generate reports based upon the collected data. This area has evolved from single-use products to suites of applications designed for different types of analysis and reporting tasks. It has also spawned lower-cost solutions that do not have the architectural requirements of the full-blown suites.

No matter how complex the means used to achieve it, a successful BI system is really the automation of a set of simple actions that are designed to process business data. The basic elements of a BI system are to gather data, analyze it, and report meaningful results using information locked in a company's many data stores. The five processes are as follows:

1. Storing, organizing, and grouping information multidimensionally through metadata
2. Aggregating, or gathering information and assembling it in a form that permits analysis
3. Analyzing using tools that provide detailed what-if analyses as well as developing more complex statistical relationships
4. Presenting using "visualization" tools that make data meaningful to users, including charts, graphs, and other appropriate output formats
5. Reporting and providing usable information for planning and decision making

A conventional BI solution consists of two parts: (1) a data warehouse and supporting infrastructure and (2) the tools used to extract data and present it in a meaningful way. The data warehouse is a read-only database that holds consolidated and integrated data that has been extracted from systems throughout the enterprise and entered in a standardized and analyzable form through the use of ETL tools. Data residing in the data warehouse is the most correct, transformed, and validated data available; it is designed for ad hoc queries, analysis, and reporting, and it resides within a database system. For smaller businesses, a true data warehouse represents an ideal that may not always be achievable. However, standardization upon a single database system backed by rules for data use can provide the main functionality: a "single version of the truth" (that is, one set of figures used by all) and standardization of data sources through the use of the database system's tools.

Business intelligence has now reached a point where it is almost universally accepted as an important part of overall IT strategy, particularly among larger corporations. Over the past several years, there has been consolidation within the industry and a commoditization of ETL tools. The major enterprise resource planning (ERP) firms have attempted to integrate BI into their massive framework systems and have succeeded to a great extent, but the idea that a single vendor will provide a single, integrated, corporate-wide BI suite is beginning to fade, as companies recognize that different areas require different approaches. This has led to a decrease in ad hoc solutions, greater integration and improved standardization between systems, and the growing popularity of "executive dashboard" or "enterprise portal" solutions for bringing data analysis from different sources to the users that require it in a manner that suits their understanding, capabilities, and requirements. All of this favors the openness, flexibility, and smaller size of upcoming open source solutions.

One important recent development reflects a growing recognition that "analysts" are not the only users of BI information. As the technology continues to infiltrate the enterprise, it is being used by an increasing array of different operators with different capabilities and requirements. This has led to improvements in user interfaces, simplification of choices, and the provision of different levels of presentation. This, too, tends to favor a solution that is smaller, supports standards, and can be easily customized.

BI has now stretched out across the enterprise, with no single product dominating the market. Major uses include management of corporate performance, monitoring of business activity, reporting or regulatory compliance, and customer analysis.

Another important implementation trend is the growing need to establish ROI and cost justification for BI systems. This results from an atmosphere of increased concern over capital expenditure, combined with a realization that today's BI solutions can be expensive to deploy and that experimental systems or systems that do not provide valuable information may be even more expensive. This has become the primary motivator for development and attention given to open source BI solutions. Although they may not be a panacea for cost issues, they can be implemented initially under the radar, so that the details and magnitude of a potential solution are better known before the project needs to be budgeted.

OVERVIEW OF OPEN SOURCE BI SOLUTIONS

There are a number of open source solutions available, most running on the Linux platform. Available tools currently range from database systems with some BI capability to complete BI suites. The main open source databases are MySQL and PostgreSQL, with MySQL being the preferred system for analytics and integration projects. These are joined with a number of Java and XML-based tools to fill out open source BI suites.

ETL tools that bring data into the database exist in open source mainly as Java applets providing transformations based on XML. There is a wide range available, according to the specific problem to be solved.

For analysis, the leading open source tool is Mondrian, which provides OLAP capability. Reporting and query tools are also available as Java applets; one of the most interesting is JPivot, which works with Mondrian as well as with other OLAP servers.

Implementation of an open source solution may begin with a relatively low cost due to the availability of "free" source code, but it can turn into an adventure. Usability almost always turns upon the capabilities of internal data processing departments or of consultant-engineered vertical market solutions that provide a close fit to the business concerned. Additional programming will almost always be required to fit the solution to the business, and some of it can be complex. Maintenance issues are also likely to arise over the life of the solution, and the original developer may not always be available. However, where such resources are available, open source may provide a convenient and low-cost option for creating a customized BI solution.

It should also be noted that an increasing number of mainstream BI applications and database systems have migrated to Linux, largely from the proprietary Unix world. However, these are not open source solutions in themselves.

Some of the more popular open source BI solutions are listed in Table 2.

The next section takes a look at the three most advanced open source BI suites: Pentaho, JasperSoft, and the Eclipse BIRT Project. (Eclipse is an open source development initiative led by BI vendor Actuate.) Of the three, Pentaho and JasperSoft are the most complete initiatives.

MAJOR OPEN SOURCE BI PROJECTS

Today, the major open source BI projects are generally considered to be Pentaho, JasperSoft, and

Table 2 — Open Source Solutions for Business Intelligence

Product	License	Web Address	Description
Reporting			
JasperReports	LGPL	www.jaspersoft.com	JasperReports is a feature-rich, high-performance report development and execution product for developers and end users. It consists of a powerful graphical report design tool and a comprehensive XML-based report definition and execution library. It is a high-performance and massively scalable standards-based system supported by a large and active community of report designers and developers.
JFreeChart	LGPL	www.jfree.org	JFreeChart is a free Java chart library. JFreeChart supports pie charts (2D and 3D), bar charts (horizontal and vertical, regular and stacked), line charts, scatter plots, time series charts, high-low-open-close charts, candlestick plots, Gantt charts, combined plots, thermometers, dials, and more. JFreeChart can be used in applications, applets, servlets, and JavaServer Pages. This project is maintained by David Gilbert.
JFreeReport	LGPL	www.jfree.org	JFreeReport is a free Java report library. JFreeReport reads data from a TableModel and generates formatted output, supporting features that include headers, footers, page numbering, grouping, totals, averages, embedded images, and more. Reports can be previewed on screen or saved in Acrobat PDF, Excel, HTML, XML, or text format. This project is maintained by Thomas Morgner.
BIRT	EPL	www.eclipse.org/birt	The BIRT project is intended to provide a next-generation reporting technology with a Web-centric design metaphor that is open source and extensible and provides an XML report design format. It is designed to act as a foundation for commercial products. The project's chief sponsors are Actuate, IBM, and InetSoft.
Analytics			
Mondrian	CPL	http://mondrian.pentaho.org	Mondrian is an OLAP server written in Java. It enables you to interactively analyze very large data sets stored in SQL databases without writing SQL. It serves as the Analytic tool of the Pentaho BI suite. The latest release of Mondrian, version 2.3.2, is now available.
JPivot	CPL	http://jpivot.sourceforge.net	JPivot is a JavaServer Pages custom tag library that renders an OLAP table and lets users perform typical OLAP navigations like slice and dice, drill down, and rollup. It uses Mondrian as its OLAP Server. JPivot also supports XMLA data source access.
OpenI	MPL	http://openi.sourceforge.net	The first release of OpenI is a simple Web application that does out-of-box OLAP reporting. It can be deployed on any J2EE server, and interactive OLAP reports can be created from existing cubes immediately. Future versions will include data sources other than OLAP cubes (relational databases, data mining models, and so on).
ETL			
Clover	LGPL	http://cloveretl.berlios.de	Clover.ETL is an open source, Java-based ETL framework that can be used to transform structured data. While using Java technology it allows for platform independence. It can be used standalone — as an application run from command line — or can be embedded in applications. Clover.ETL is accompanied by CloverGUI visual transformations designer, which is integrated with the Eclipse platform.
Octopus	LGPL	tk	Enhydra Octopus is a Java-based extraction, transformation, and loading (ETL) tool. It may connect to any Java Database Connectivity (JDBC) data sources and perform transformations defined in an XML file.

(Table continues on next page.)

Table 2 — Open Source Solutions for Business Intelligence (continued)

Product	License	Web Address	Description
ETL (continued)			
Kettle	LGPL	http://kettle.pentaho.org	Kettle is the ETL component of the Pentaho BI suite. It provides: • Data warehouse population with built-in support for slowly changing dimensions, junk dimensions, and much, much more • 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 by looking up data in various information stores (databases, text files, Excel sheets) • Data cleaning by applying complex conditions in data transformations • Application integration
Database			
EnterpriseDB	Enterprise DB	www.enterprisedb.com	EnterpriseDB Advanced Server is an enterprise-class relational database management system (RDBMS) that is compatible with applications written for Oracle. EnterpriseDB Advanced Server is based on PostgreSQL, the world's most advanced open source database, ensuring the world-class data integrity, security, and performance necessary for enterprise environments. In addition, the total cost of ownership (TCO) of an enterprise database solution powered by EnterpriseDB Advanced Server is only a small fraction of the TCO of a comparable Oracle-powered solution.
Ingres	GPL, commercial	www.ingres.com	Ingres RDBMS is a full-featured, enterprise-class database. Among its feature highlights are: • 64-bit architecture support for large-scale enterprise deployments • A small footprint requiring fewer system resources • Industry-standard data access via JDBC, ODBC, and .NET as well as support for the latest open source development environments including Eclipse, PHP, Perl, Python, and Ruby • Transaction journaling providing point-in-time recovery • Parallel query processing • Key range table partitioning • High availability cluster support • Advanced query optimization techniques
MySQL	GPL, commercial	www.mysql.com	The MySQL database provides consistent fast performance, high reliability, and ease of use. It is used in more than 10 million installations ranging from large corporations to specialized embedded applications. It is also the database of choice for a new generation of applications built on the LAMP stack (Linux, Apache, MySQL, PHP/Perl/Python). MySQL runs on more than 20 platforms, including Linux, Windows, OS/X, HP-UX, AIX, NetWare, giving you the kind of flexibility that puts you in control.
PostgreSQL	BSD	www.postgresql.org	PostgreSQL is a powerful, open source relational database system. An enterprise class database, PostgreSQL boasts sophisticated features such as multi-version concurrency control (MVCC), point-in-time recovery, tablespaces, asynchronous replication, nested transactions (savepoints), online/hot backups, a sophisticated query planner/optimizer, and write ahead logging for fault tolerance.

BIRT, though the latter is currently primarily a reporting solution. Each of these projects has major industry support, has a commercialization strategy, and offers a suite of integrated open source products with an established roadmap for growth.

The Pentaho approach is based on Eclipse and provides enhancements to Mondrian OLAP — a Dashboard component, data mining, reporting, workflow, and business framework (including a business rules engine and document repository). JasperSoft's open source BI suite variant is offered under a curious "commercial open source" license. In fact, all three of these product suites offer either a commercial license or commercial product version that provides extended capabilities.

Let's start by examining Pentaho.

Pentaho

The Pentaho BI Project provides a comprehensive BI suite. This includes reporting, analysis, dashboards, data integration, data mining, and a BI platform for production deployment. The suite is built around the Pentaho BI Platform, which provides the architecture and infrastructure required to build solutions to BI problems. The platform includes an embedded workflow engine and can be easily integrated into business processes. Core platform services include authentication, logging, auditing, workflow, Web services, and rules engines.

The included solution engine integrates reporting, analysis, dashboards, and data mining components to form a sophisticated and complete BI infrastructure. Components are all open source components, written in Java. The solution engine executes BI Process Flows. BI Process Flows are also available as Web service calls that may be used by orchestration technologies such as the Business Process Execution Language (BPEL).

The server architecture has been built for J2EE. The client design environment is built around the Eclipse workbench, with end-user access provided through HTML and other thin-client technologies.

The Pentaho BI Platform is normally deployed as a standalone server with a standalone Design Studio. In order to solve business problems, a solution will need to be deployed as well. New solutions can be created with the Pentaho Design Studio, or preexisting solutions can be obtained from Pentaho or other sources. Customers can start with one component, such as Reporting, and add other components, such as Analysis and Dashboards, as required.

The Pentaho Design Studio requires the Eclipse framework and can only be deployed to platforms supported by Eclipse. Currently all versions of Windows from ME to XP are supported as are many versions of Linux, Mac OS X, Solaris, and other Unix platforms.

The Pentaho platform was built to be open and flexible and to allow developers to integrate third-party components. The Pentaho BI framework uses Acegi Security for a pluggable framework to implement and combine multiple authentication schemes and credential stores with central authentication storage for platform-neutral single sign-on.

The major investment themes of 2007 for Pentaho include ease of use, maintenance and administration, and platform extensions.

BI Platform

The Pentaho BI Platform provides the infrastructure and core services that integrate business intelligence components to complete the BI suite. This includes the infrastructure necessary to build, deploy, execute, and support applications. It is a comprehensive development and runtime environment for building complete solutions to BI problems.

The Pentaho BI Platform centers its solutions around a workflow core and a service-oriented architecture (SOA). Pentaho BI solutions are made up of collections of XML documents. The Pentaho Design Studio is built with plug-ins to make editing and managing these documents easier; however, the solution documents can also be edited with any simple text editor, if necessary.

The Pentaho BI Platform uses a sophisticated combination of business rules, services, assured messaging, workflow, clustering, and auditing to ensure efficient operation.

The main open source components that have been integrated into the Pentaho BI infrastructure are listed in Table 3.

Reporting

Pentaho Reporting is designed to meet a wide range of business reporting needs, from simple reports on a Web site to highly formatted reports for applications like financial reporting. It also provides both scheduled and on-demand report publishing in formats including Adobe/PDF, Microsoft Excel, HTML, RTF, and text. Reports can be filtered and delivered to targeted users.

Pentaho Reporting can access data from relational and OLAP data sources as well as XML or Web services sources. It is a highly extensible platform with flexible customization points and a sophisticated architecture based on the JFreeReport project, which Pentaho owns and sponsors.

Technical capabilities include:

- Generating and delivering thousands of 1-to-20-page documents in a single BI process
- Generating reports in Microsoft Excel, HTML, PDF, RTF, and text formats
- Generating report content from relational, OLAP, XML, and Web services data sources
- Providing intuitive interface components for business users

Table 3 — Pentaho Open Source Components

Project/Component	Description and Usage	License
Apache Commons Logging	Logging	Apache
Apache HttpClient	Server-to-server HTTP communications	Apache
OpenSymphony Quartz	Scheduler	Apache compatible
Apache log4j	Logging	Apache
Chiba	Server-side xForm to HTML conversion	Artistic
Eclipse Platform	Desktop workbench	Eclipse
Eclipse Modeling Framework	Workbench Modeling Framework	Eclipse
Eclipse Graphical Editing Framework	Workbench Graphical Editor	Eclipse
JBoss Application Server	Application server used for sample deployments	LGPL
JBoss Hibernate	Object persistence layer	LGPL
JBoss Portal	JSR 168 compliant portal server used for sample deployments	LGPL
JFreeChart	Chart engine (www.jfree.org/jfreechart)	LGPL
Acegi	Single sign-on and LDAP integration (pro version)	BSD
MetaStuff dom4j	XML parser	BSD
Mozilla Rhino	JavaScript processor	Mozilla
Sun JavaMail	E-mail delivery	Sun
Sun Java Database Connectivity	Database access	Sun
Sun JIMI	Image management	Sun

- Providing a drag-and-drop graphical report design environment using the Pentaho Report Designer
- Integrating with Pentaho Analysis components, such as Dashboards and pivot views
- Supporting deployment of reporting as an embedded Java library, as a packaged Web reporting application, or as part of an entire BI suite
- Conforming to all standards and interfaces provided by the Pentaho BI Platform

The main open source components that have been integrated into Pentaho Reporting are shown in Table 4.

In addition to these components, Pentaho Reporting takes advantage of Pentaho BI Platform components such as JBoss, JFreeChart, and Eclipse.

Analysis

Pentaho Analysis provides extensive analysis capabilities that include a pivot table viewer (JPivot), advanced graphical displays using SVG or Flash, integrated dashboard widgets, data mining integration, portal integration, and workflow integration. These capabilities are supported by scheduling, Web services, content navigation and management, security, application integration, and auditing. Analysis can be performed on relational data sources using Pentaho Analysis Services (based on Mondrian OLAP).

Pentaho Analysis provides tools that make it possible to freely explore business information in a Web-based environment; to drill down into data; and to cross-tabulate data. The incorporated Pentaho Spreadsheet Services allow users to browse, drill, pivot, and chart against Pentaho Analysis Services from within Microsoft Excel.

Technical objectives include:

- Providing pivot table views of OLAP data
- Providing advanced analytic graphical views of OLAP data
- Providing intuitive interface components for business users
- Providing drill-through to and from reporting content
- Integrating analysis with business processes
- Providing user-by-user customization and preferences
- Using the best open source analysis components available
- Providing technology encompassing an out-of-the-box reporting server, embeddable analysis components, and anything in between

The main open source components that have been integrated into Pentaho Analysis are shown in Table 5.

In addition to these components, Pentaho Analysis takes advantage of the open source software included in the Pentaho BI Platform.

Dashboards

Pentaho Dashboards display, arrange, and control BI content. All Pentaho components including Reporting and Analysis can contribute content to Pentaho Dashboards. Dashboard widgets can be created to display dials and gauges. External content such as Web pages, third-party applications, and RSS feeds can also be integrated. Filter controls can be added to provide subject-based content filtering.

Table 4 — Open Source Components in Pentaho Reporting

Project/Component	Description and Usage	License
Apache FOP (Formatting Objects Processor)	PDF document production	Apache
Eclipse BIRT	HTML and PDF report designer and engine	Eclipse
Sun JIMI	Image management	Sun

Table 5 — Open Source Components in Pentaho Analysis

Project/Component	Description and Usage	License
JPivot	Web pivot tables/charts	Common Public License
Mondrian	ROLAP modeler	Common Public License

Role-based security and filtering is easily built in.

Pentaho Dashboards can be embedded into applications, JavaServer Pages (JSPs), or within JSR 168–compliant portals using the provided portlets.

Technical objectives include:

- Providing browser- and portal-independent dashboard displays
- Providing reusable display widgets (gauges, dials, charts, etc.)
- Integrating reporting, analysis, and dashboard content
- Providing configurable, common filtering controls
- Providing role-based security and filtering
- Providing user-by-user customization and preferences
- Using the best open source components available
- Providing technology encompassing an out-of-the-box reporting server, embeddable dashboard components, and anything in between
- Adhering to all the design goals of the Pentaho BI Platform

Pentaho Dashboards is a Pentaho project. Third-party open source components are limited to utilities like XML parsing and the common libraries. The Eclipse framework is used for the dashboard creation tool.

Data Mining

Pentaho Data Mining incorporates Weka, a collection of machine-learning algorithms applied to data mining tasks. These algorithms are combined with OLAP technologies to provide machine-intelligent data analysis to end users. Data mining tools can analyze historical data to create predictive models. Pentaho Reporting and Analysis components can then be used to distribute this information.

Technical objectives include:

- Integrating open source data mining with reporting and OLAP data sources
- Integrating data mining with business processes
- Enhancing compliance and corporate governance by applying data mining tools to business process data
- Using the best open source data mining components available
- Adhering to all the technical objectives of the Pentaho BI Platform

The main open source component integrated into Pentaho Data Mining is Weka, which has a GPL license.

JasperSoft BI Suite

JasperSoft provides reporting, analysis, and data integration in a system that is meant to be easily operated by anyone, from casual business users to analysts to executives, in a form that can be incorporated in businesses of all sizes. It is designed to provide a comprehensive, flexible, seamlessly embeddable, and affordable solution. Key features include:

- Interactive data analysis/OLAP
- Interactive and managed reporting
- High-performance data integration
- Reports for screen or print

It incorporates a production report designer and provides

advanced on-demand reporting for Salesforce.com.

JasperSoft is available in two versions: the "free" open source system and a professional edition, which provides added value. The professional edition provides several assurances and conveniences that may be important to many customers, including:

- Advanced features
- Broader platform support
- Managed release cycles
- Premium support subscriptions
- Commercial license with indemnities
- Rights to bundle with other commercial products

JasperSoft BI Suite is designed principally to be a standalone BI solution for workgroups and small organizations. It includes a fully functional, production-ready reporting and BI server that deploys quickly. Faster deployment yields a faster ROI. It is built on a modular architecture and designed to leverage existing enterprise software and inhouse skills.

The JasperSoft BI Suite can also be embedded into other business applications, such as CRM, finance, ERP, and human capital management systems. JasperSoft is pure Java and offers a clean business API. Non-Java applications, such as those written in PHP or Perl, easily access JasperSoft BI capabilities through prebuilt Web services connectors.

Technical details are provided in Table 6.

JasperSoft BI Platform

The BI platform provides the framework for products in the modular JasperSoft BI Suite, making it possible to manage, secure, and deliver full business intelligence capabilities to different user communities.

The modular components of the BI suite are:

- **JasperServer** — interactive and managed reporting for JasperReports
- **JasperAnalysis** — interactive data analysis/OLAP server
- **JasperETL** — high-performance data integration system
- **JasperReports** — report generator for screen or print

JasperServer, JasperAnalysis, and JasperETL run on a shared BI platform that provides a common framework to allow reporting, analysis, and data integration. The system is designed to meet the needs of very small to very large organizations. The BI platform is an integral part of the JasperSoft BI Suite. It is a production-ready environment, with security features that include external authentication, role-based authorization, and single sign-on. User interfaces are configurable and replaceable. It provides easy integrations with existing IT infrastructure servers and services and can be embedded in applications using public Java APIs and Web services.

JasperServer

JasperServer is a high-performance, interactive, standalone report server designed to provide managed reporting for workgroups, small businesses, and enterprises. It includes a secure report management repository, standards-based report definitions, integration interfaces, JasperReports, and the iReport graphical report designer. It supports drag-and-drop, Web-based ad hoc reporting; self-service; and interactive parameterized reports. Report scheduling and distribution capabilities are provided, and historical report versioning and auditing is available for regulatory compliance. JasperServer can be seamlessly embedded into other software applications.

JasperServer's enterprise-level RDBMS-based repository is optimized for speed and stores report definitions, images, fonts, data source definitions, OLAP views, and so on for fast and secure access.

Using a standard Web browser, users can create, run, save, and interact with reports. Users only see reports they are authorized to use and can select from multiple prompted parameters to retrieve the data they need.

Table 6 — JasperSoft Technical Overview (Source: JasperSoft)

	JasperReports	JasperServer	JasperAnalysis	JasperETL
Overview				
Product classification	Report design and execution	Interactive and managed reporting	Interactive analysis/ OLAP server	Data integration/ ETL
Architecture	Java API (Library): Used in Server, Java Application, Applet JasperReports Architecture	Servlet Engine, Web services; Future: Portal integration JasperServer Architecture		
Language	Java	Java	Java	Java (generates Perl)
APIs	Java	Java, JSP, SOAP, HTTP, Web services	XML/A, Java	Command line
Embeddable	Yes	Yes	Yes	Yes
Standalone	Yes	Yes	Yes	Yes
User Features				
Output	HTML, PDF, RTF/ Word, XLS, CSV, XML	HTML, PDF, RTF/ Word, XLS, CSV, XML	HTML, PDF, XLS	Perl code
Parameterized reports	Programmable	Built-in and extensible		
Ad hoc reporting	Yes, using iReport	Yes, using iReport or Web interface		
Interactive data analysis			Yes	
Dashboard		Programmable		
Server Features				
BI objects repository	Yes, can access using iReport plug-in for JasperServer	Yes	Yes	Pro. only
Scheduler		Yes		Yes
Administration	Yes, can administer using iReport plug-in for JasperServer	Yes	Yes	Yes
Security		Yes	Yes	Yes
Clustering		Yes	Yes	Planned
Data sources	JDBC, EJB, POJO XML	JDBC, Future Extensions	JDBC, JNDI, Bean	30+ connectors
Custom data sources	Yes	Programmable	Yes	
Data caching and reuse			Yes	Yes

Table 6 — JasperSoft Technical Overview (continued) (Source: JasperSoft)

	JasperReports	JasperServer	JasperAnalysis	JasperETL
Report Design Features				
Graphical designer	iReport	iReport and ad hoc Web interface		Yes
Report definition language	JRXML	JRXML		
Scripting language	Java, Groovy	Java, Groovy		Perl
Query language	SQL, Hibernate (HQL), XPath (XML), EJBQL, MDX	SQL, Hibernate (HQL), XPath (XML), EJBQL, MDX	MDX	SQL
Query builder	SQL, MDX	SQL, MDX		SQL
Business Features				
License	LGPL and commercial	GPL and commercial	GPL and commercial	GPL and commercial
OEM-able	Yes	Yes	Yes	Yes
Technical support	Yes	Yes	Yes	Yes
Consulting services	Yes	Yes	Yes	Yes
Training	Yes	Yes	Yes	Yes

JasperServer includes the following scheduling and distribution features:

- Time-zone selection
- Scheduled run begin and end dates
- Maximum number of recurrences
- Run report every *x* minutes, hours, days, weeks
- Calendar-based recurrence: every day/weekday/day of month
- Output to PDF, HTML, Excel, RTF
- E-mail notifications with report links or optional report attachments

JasperServer Professional features extensions including ad hoc reporting.

JasperAnalysis

JasperAnalysis provides capabilities to explore causes, trends, patterns, anomalies, and correlations using a standard Web browser. It is designed to complement and extend the power of JasperReports by providing drill-down, "slice and dice," pivot, filter, and charting capabilities. It enables dynamic drill-down to detail data and is accessible through an easy-to-use Web interface designed for business users. JasperAnalysis is backed by a high-performance relational OLAP (ROLAP) server engine, and is MDX- and XML/A-compliant. It is pre-integrated with JasperServer.

Features include:

- Analytic operations
- Sort across/within hierarchy
- Show empty rows/columns
- Swap axes (pivot)
- Show chart
- Show source data/drill-through to detail

- Color-coded exception alerts
- Change data cube
- Filter
- Show MDX query
- Configurable output to Excel or PDF
- OLAP server administration
- Manage/tune OLAP server configuration
- Flush OLAP cache
- Secure user and role-based permissions

JasperAnalysis Professional feature extensions include drillable charts, an enhanced user interface, and an OLAP server management utility.

JasperETL

JasperETL is based on the open source Talend engine and provides a complete and ready-to-run data integration platform with high-performance data ETL capabilities. JasperETL works with the JasperSoft BI Suite but can also be used in standalone mode to provide comprehensive ETL capabilities for other applications and systems.

JasperETL simplifies and automates data integration through easily created, managed, and maintained data integration processes. It provides an intuitive GUI and is usable by small, medium-sized, and large organizations. It simplifies and standardizes data warehouse/data mart updates, thus simplifying and standardizing the end-user experience. JasperETL can routinely extract, transform, and load data from operational systems into a "star schema"–style database, where it can be safely and quickly accessed for interactive end-user reporting and analysis.

JasperETL meets operational data integration needs that include data consolidation, duplication, synchronization, quality, migration, and change data capture. Components include:

- **Job Designer —** provides a graphical editor and functional view of the ETL process
- **Transformation Mapper —** provides a graphical editor and view of complex mappings and transformations
- **Business Modeler —** provides a nontechnical graphical view of the business information workflow

Numerous included connectors (30-plus) permit output and input from and to many different data sources including flat files, XML files, all databases, POP and FTP servers, and more. Included metadata configuration wizards help configure heterogeneous data sources and complex file formats including positional, delimited, CSV, RegExp, XML, and LDIF data.

JasperETL Professional feature extensions include a multiuser metadata repository that enables team development of data integration projects.

JasperReports

JasperReports is a feature-rich, high-performance report development and execution product for developers and end users. It consists of a powerful graphical report design tool and a comprehensive XML-based report definition and execution library. It is a high-performance and massively scalable standards-based system supported by a large and active community of report designers and developers. It is capable of creating dashboards, tables, crosstabs, and charts with complex layouts for screen or print. It supports flexible and extensible data sources and a wide range of output formats. Built-in virtualization enables output of arbitrarily large reports, limited only by disk storage resources. JasperReports also delivers detailed printer-ready output and is often used for detailed forms, invoices, and other complex operational reports.

The iReport graphical report designer component for JasperReports provides easy access to all JasperReports capabilities. iReport is a powerful Java client application that includes optional integration with the popular Eclipse integrated development environment (IDE). It provides capability to retrieve, store, and modify reports when

used with JasperServer and Jasper4Salesforce.

Features include:

- Dashboards, tables, crosstabs, charts, and gauges
- Report output in PDF, XML, HTML, CSV, XLS, RTF, and text
- Page-oriented or continuous output for screen or print
- Integrated bar code support
- Visual text rotation
- Styles library
- Drill-through/hypertext links, including support for PDF bookmarks
- No limit to report size

The iReport graphical report and chart designer component provides:

- Comprehensive library of chart types, including meter, thermometer, and multi-axis charts
- Built-in expression builder with syntax checker, object methods list, and wizards
- Graphical query builders for SQL and MDX
- Ability to be used in Eclipse or as a pure Java (Swing) application
- Ability to build, test, and run JasperReports from the desktop environment
- Preintegrated with the JasperServer repository

Complex reports and report features are supported, along with internationalization.

JasperReports Professional includes the JasperReports reporting library, iReport graphical report designer, and updates for the iReport User Manual and JasperReports Ultimate Guide.

Jasper4Salesforce

Jasper4Salesforce delivers advanced on-demand reporting for Salesforce.com. This includes exception reports, reports on any combination of standard and custom objects and fields, drag-and-drop ad hoc, and highly complex and ready-to-print reports. It provides on-demand, subscription-based usage and advanced reporting; it is Salesforce.com-certified at the professional, enterprise, and unlimited levels.

Features include:

- Exception reports
- Flexible queries
- Integration of graphs and charts directly into other areas of Salesforce.com
- Drag-and-drop ad hoc reporting
- Ability to incorporate graphical report designer
- High performance
- Advanced report management

Jasper4Salesforce also includes a library of predefined reports.

JasperSoft BI Applications

JasperSoft and its partners offer prebuilt BI applications that integrate with or are embedded in popular applications. With these applications, any CRM, ERP, or other application can have complex reporting, interactive reporting, integrated analysis, visual dashboards, and data integration to and from other data sources.

BIRT

BIRT is an open source Eclipse-based reporting system that integrates with Java/J2EE applications to produce sophisticated reports. It provides core reporting features such as report layout, data access, and scripting. The BIRT Project is a top-level project of Eclipse (www.eclipse.org). The current released version is 2.1.2.

The BIRT Project addresses a wide range of reporting needs within a typical Java application. BIRT aims to address the problem of ad hoc reporting capabilities by providing Eclipse-based open source and extensible tools and frameworks that allow developers to easily incorporate reporting functionality.

The BIRT Project's initial releases have focused upon reporting. Other projects will expand BIRT into other areas of BI such as data modeling, ETL, APIs, and frameworks to build business-user query tools for flexible ad hoc access to data, as well as analysis tools.

BIRT is built on the open source Eclipse platform. Eclipse is an open source software development project dedicated to providing a robust, full-featured, commercial-quality, industry platform for the development of highly integrated tools. The Eclipse platform is designed for building IDEs that can be used to create diverse C++- and Java-based applications.

Actuate Corporation also provides Actuate BIRT, a commercial version of the BIRT Project. It is now MySQL Network Certified and a recommended reporting and analytics component of a MySQL data warehouse scale-out solution set. Actuate BIRT delivers functionality that is identical to the Eclipse BIRT project — its open source Eclipse Public License (EPL) equivalent. Actuate BIRT includes reporting and charting offerings that are immediately available on an annual subscription pricing basis.

BIRT features include:

- A palette of report components — text (character large object, or CLOB; HTML), data, images (binary large object, or BLOB), tables, grids, lists, labels
- Sorting, grouping, filtering, conditional highlighting, mapping
- Scripting in JavaScript/Java
- Cascading and dynamic report parameters
- Hyperlinking, bookmarks
- TOC, paging
- A direct XML source editor for report design
- Multipass processing (top N/bottom N)
- Data source access including POJOs, JDBC, CSV, XML
- A WYSIWYG editor
- An integrated chart wizard
- Report component libraries
- Report templates
- Styles, CSS import, themes
- Wizards for guided development
- Report outline capability
- Preview within report designer for iterative development
- Context pass-through to data source
- Capability to call stored procedures

The BIRT Project is intended to provide a next-generation reporting technology with a Web-centric design metaphor that is open source and extensible and provides an XML report design format. It is designed to act as a foundation for commercial products. The project's chief sponsors are Actuate, IBM, and InetSoft.

In its current version, BIRT is designed to provide a personal desktop report development tool or reporting technology for integration into corporate Web applications, corporate desktop applications, ISV Web applications, or ISV Eclipse-based applications.

Extensibility is a key principle for the BIRT Project. It provides:

- Data source extensibility
- Application-specific, design-time query builders
- Custom design-time and runtime data access
- Custom business logic extensibility
- Capability to incorporate complex business logic scripting
- Capability to access existing and new Java code
- Visualization extensibility
- Capability to build new visual data presentation "widgets"
- Capability to extend charting with new chart types; new output formats
- Capability to target report output for specific devices and formats

The next major release, 2.2, is targeted for June 2007.

Eclipse Report Designer (ERD)

The ERD is an Eclipse-based desktop authoring environment for report development. Eclipse Report Designer enables application and report developers to create simple and complex reports for use within their organization.

The tool caters to the broad range of report development skills, from the nonprogrammer report developer focused on report layout to the application developer looking for sophisticated control over report creation.

Eclipse Report Engine (ERE)

The ERE allows Java application developers to quickly integrate powerful report generation and viewing capabilities into their applications without having to build the infrastructure from lower-level Java components.

The Eclipse Report Engine project enables reports to be generated using the XML report designs created by the Eclipse Report Designer, Web Based Report Designer, or any other tool. To support this, the ERE provides two core services: generation and presentation.

The generation service within the ERE is responsible for connecting to the specified data source(s), retrieving and processing the data (sorting, grouping, aggregations, and so on), creating the report layout, and generating the report document.

The presentation service within the ERE provides a rich set of viewing capabilities for report content. This includes the infrastructure for viewing a document online, for printing a document, and for generating alternate output documents such as PDFs.

Eclipse Chart Engine (ECE)

The ECE project provides a rich business-chart generation capability that can be used as a standalone charting component and also provides the chart generation service within the ERE project. Visual presentation of business data in the form of charts is a common and key aspect of many reports and other forms of business intelligence. As such, a robust charting capability is essential within the overall BIRT Project.

Web Based Report Designer (WRD)

The WRD project delivers a fully customizable and extensible 100% HTML-based tool for creating reports with basic layouts and data manipulation.

The tool will leverage components such as style sheets and templates created using the ERD. In addition, the tool will provide an Eclipse-based customization facility for the user interface for full branding and embedding within Java applications.

The goal of this project is to provide an accessible and easy-to-use report design environment to meet the needs for ad hoc report creation by business users within any Java application.

Comparisons

Comparisons between open source BI suites can be difficult because the suites frequently either incorporate or interoperate with the same open source products, and both the supporting communities and companies involved are in active communication with each other. The result is to create a different spin on competition. The major focus is on collaboration, and products need to be selected according to how well they fit company needs. Implementation and integration issues are often of greater consequence than the software platform selected.

The Pentaho BI Project is currently the most comprehensive BI suite. It includes reporting, analysis, dashboards, data integration, data mining, and a BI platform. JasperSoft is the second most comprehensive, but it lacks the data mining capabilities of Pentaho and focuses upon reporting. Movement between open source components and commercial versions or upgrades is also important and differs among open source BI solutions. The upgrade path from BIRT is to the proprietary Actuate iServer. For JasperSoft, the upgrade path is also to a proprietary system, though it is closer to the open source version and there is more versatility in upgrade paths. Similarly, Pentaho's Professional Edition builds on and extends Pentaho's open source capabilities with additional features added for mission-critical or large-scale deployment.

JasperSoft's approach differs from that of Pentaho and BIRT in that

it focuses more on embeddable software. Embedding eases integration, extends functionality, and stimulates a strong partnership community. The JasperSoft BI Suite is already integrated with products from innovative companies like Novell, Salesforce.com, SpikeSource, and SugarCRM. Additional partners include BEA, IBM, Oracle, Sun, Eclipse Foundation, EnterpriseDB, JBoss Open Source Federation, and MySQL.

BIRT's strongest feature is its association with Eclipse. Eclipse is an active, growing open source development organization with support from IBM, HP, SAP AG, and other major players in the IT industry. BIRT is developed by the Eclipse community and features tight integration with Eclipse. The BIRT committee has an active dialog with these other projects to understand where BIRT can leverage existing or proposed functionality.

Other areas of comparison are in overall infrastructure, or how the BI platform brings the open source components together. Infrastructure factors vary in areas such as security, administration, auditing, failover, scalability features, portal, and other functions.

OTHER TOOLS

While Pentaho, JasperSoft, and BIRT have shared the limelight recently, there are other open source BI efforts in development that are worthy of note. Three of these are the Bizgres Project, the BEE Project, and the Jedox Palo-Server.

The Bizgres Project

The Bizgres Project is a commercially sponsored and community supported open source project (www.bizgres.org). The main supporters currently are Greenplum, JasperSoft, Kinetic Networks, and Loyalty Matrix. The goal of Bizgres is to build a complete database system for business intelligence exclusively from free software.

Bizgres is targeted at propelling PostgreSQL into practical, real-world use within mainstream businesses needing high-quality RDBMSs for business intelligence. Bizgres is designed to make PostgreSQL an alternative to Oracle, Sybase, Informix, and Microsoft proprietary databases.

The stated goals of Bizgres are to:

- Create a complete database system for business intelligence with capabilities exceeding competing database systems, built for and by the community
- Build the database system completely from open source software
- Provide a robust development platform for building software, particularly open source software
- Emphasize usability and a "just works" philosophy in selecting default configuration and designing features
- Include a range of popular packages, beyond those included in Greenplum's commercially supported products
- Create an environment where commercial vendors can easily get involved to make PostgreSQL a supported platform for their offering
- Promote a global perspective by supporting as many languages and geographic locales as possible
- Provide binary releases and a robust build environment to enable business users to easily install and test Bizgres

The BEE Project

BEE is a suite of tools supporting BI project implementation within midsized companies under an open source GPL license. It is being developed by Insight Strategy (http://bee.insightstrategy.cz). It aims to optimize data storage for analysis and to focus on ETL process and multilayer applications. The architecture is developed on the ROLAP methodology with the aim of covering projects having data volumes up to 50 GB. BEE is released under the GNU General Public License, with a stable version of the product available under a commercial license (similar to MySQL). For the commercial product, technical support is provided.

The BEE Project ETL tool is based on simple daemon services, ensuring distributed data transmission and transformation. The environment is administered by a centralized supervising application. The suite is designed for distributed processing with robust encryption. A GUI for administration of ETL processes and modeling is available within the system.

The ROLAP server provides its functionality through a SOAP application interface for a potentially rich set of client applications. The incorporated Web portal has been designed as a primary user interface for report design, presentation, and data manipulation through a Web browser. It uses XML, HTML, CSS, and JavaScript to aid in communications and ensure easy administration of the whole environment. Integration with tools of the R Project for Statistical Processing is available.

The BEE project also includes the BeeWebAnalyser 1.0, a new open source tool that uses its own ROLAP technology for data analysis and decision support. It is embedded in the BEE Project decision support system and enables a detailed interactive analysis of Web page visits (Apache Web server logs) using a broad range of different customizable views.

Jedox Palo-Server

Palo-Server is a cell-oriented, multidimensional OLAP (MOLAP) data server, specifically developed for spreadsheet data storage and analysis. Palo-Server is sponsored by Jedox GmbH (www.jedox.com), which develops and markets Worksheet-Server, an Excel-to-Web solution with support for Palo-enabled Excel Workbooks. Palo-Server provides a central database for enterprise spreadsheets in Microsoft Excel and offers interfaces to many third-party databases provided by SAP, IBM, Microsoft, and Oracle.

Palo-Server is based on spreadsheet cells, rather than operating as a traditional relational database. It offers a powerful, multidimensional data model based on OLAP principles (data cubes). A hierarchal data storage structure allows for incremental data inputs into central hierarchies with results computed automatically.

It can be installed locally or in a company network. In network mod,e all users work with the same data; changes in one spreadsheet are immediately visible on other workstations in the intranet.

In addition to providing multidimensional data analysis of existing data, Palo-Server allows inputting of new data to the dataset. This helps differentiate Palo-Server from relational OLAP servers.

SELECTION AND IMPLEMENTATION

Selection of an open source BI approach requires careful determination of the areas in which analysis will prove most fruitful, the actual budget available, and factors of existing infrastructure. Open source systems are provided on the basis of low initial cost and simplicity. In some cases, implementation will be relatively straightforward. If extensive data integration is required, however, a simplified approach may be impossible. To work through the possibilities in this case will almost invariably require a consultant.

Implementation Costs

Implementation and deployment costs for BI projects can contain a high proportion of consulting fees. Depending on complexity, there may also be development, customization, and integration costs. One of the most important and expensive areas of deployment is creation and integration of a data warehouse, if necessary. This is likely to involve a considerable effort, and cost depends upon the type of data required, where it is located, and how it is transformed.

Light open source BI solutions implemented without a data warehouse need to be evaluated carefully, as they impose the risk of developing multiple versions of data that will result in derived figures and analysis being out of sync. This can be remedied through a variety of technologies and data usage rules, but the solution itself can create additional cost. Light solutions may also fail to provide the advantages that

would be available from a full BI suite approach.

Added Requirements

In addition to implementation concerns, deployments of open source BI capabilities require careful consideration of the areas around the implementation. Although cost is a key factor, it is important to avoid unnecessary sacrifices in the rush to find a solution. Key additional areas to consider include:

- Security
- Training
- Scaling

Security issues are always a concern in implementation of a database solution, particularly since the BI solution inevitably incorporates critical data. Access needs to be guarded, and if the solution is delivered online, adequate safeguards need to be installed on the network to prevent break-ins. Training is important and must be considered from the beginning. Not only must information be made available in an easily understandable way, but users must be trained in how to make the most of the system.

Scaling is important in keeping with low costs, because cost will grow according to the breadth of the project and to who must have access to the data. Analysis of possible candidates should be rigorous and should include a careful evaluation of appropriateness, compatibility with existing systems, availability of implementation and maintenance help, availability of training, and any need to modify systems or data to meet requirements of the proposed solution.

CONCLUSION

Open source BI software has come into focus recently as a result of both new advances and increasing need. The software itself is starting to provide useful analytic capabilities in preintegrated suites that fill out capabilities and reduce the problems of selection and implementation. At the same time, the power and sophistication of components — including the open source database products — are continuing to develop. These technological developments are coming at a time when the need for low-cost, testable, and more easily used BI solutions is becoming apparent. The areas in which this need is most visible include a wide range of smaller businesses, which require BI solutions to handle an ever-increasing information and analysis workload; the extension of BI possibilities into nontraditional areas; and as a directly embedded component in software.

Open source BI does not necessarily compete with the major commercial solutions because it is generally targeted at a lower level of user. It does, however, put price pressure on BI solutions aimed at the lower end of the market. The combined commercial/open source initiatives represented in each of the major suites demonstrates how they are moving into this sector. Offering a highly standardized system with low initial cost makes it possible for companies to experiment with the technology without large-scale expenditure — and still ensure that anything developed can be integrated with a more comprehensive offering in the future.

The jury is still out with respect to the overall adequacy of open source BI solutions, particularly for larger organizations. The software has only very recently reached a level where it is likely to provide adequate and reasonably complete capabilities. However, the suite projects — and open source BI software in general — are currently in rapid development mode. They have ambitious roadmaps for rolling out advanced features, particularly in analytics. Lesser-known open source "competition" is continuing to develop, and many integrated solutions are likely to fall below the radar entirely as integrators and consultants build their own suites out of available components.

One thing is certain: open source BI is beginning to have a positive impact upon the development and implementation of BI solutions in general. By providing an alternative model for pricing and delivery, it is shaking the complacency out of the market and yielding new ideas. This is an area of great potential that needs to be watched for developments in the near future.

ABOUT THE AUTHOR

Brian J. Dooley is an author, analyst, and journalist with more than 20 years' experience in analyzing and writing about IT trends. He has written six books, numerous user manuals, hundreds of reports, and more than 2,000 magazine features. Mr. Dooley is the founder and past president of the New Zealand chapter of the Society for Technical Communication. He initiated and is on the board of the Graduate Certificate in Technical Communication program at Christchurch Institute of Technology, and he is on the editorial advisory board for Faulkner Technical Reports. Mr. Dooley currently resides in New Zealand, where he maintains a Web site at http://bjdooley.com. He can be reached at bjd@bjdooley.com

ACCESS TO THE EXPERTS

A Collaborative Community of Thought Leaders

Cutter Consortium is a unique IT advisory firm, comprising a group of more than 150 internationally recognized experts who have come together to offer research, consulting, and training to our clients. These experts are committed to delivering top-level, critical, and objective advice. They have done, and are doing, groundbreaking work in organizations worldwide, helping companies deal with issues in the core areas of software development and agile project management, enterprise architecture, business-technology trends and strategies, risk management, metrics, business intelligence, and sourcing.

Cutter's goals are to further the thinking in the business-technology arena and to help organizations leverage technology for competitive advantage and business success. We accomplish our mission by serving as a catalyst for collaboration between business-technology thought leaders worldwide and by giving clients access to this think tank through our research, training, and consulting — all of which is provided exclusively by the world's leading business-IT experts.

Cutter offers a different value proposition than other IT research firms: We give you *Access to the Experts*. You get practitioners' points of view, derived from hands-on experience with the same critical issues you are facing, not the perspective of a desk-bound analyst who can only make predictions and observations on what's happening in the marketplace. With Cutter Consortium, you get the best practices and lessons learned from the world's leading experts — experts who are implementing these techniques at organizations like yours right now.

Cutter's clients are able to tap into this collaborative community of thought leaders in a variety of formats including online and multimedia research, mentoring and inquiries, and training and consulting. And by customizing our research products and training/consulting services, you get the solutions you need while staying within your budget.

The Cutter Difference

Cutter's research, inquiry response, consulting, and training are qualitatively different from that of other analyst firms in many vital ways:

- **Cutter Research gives clients *Access to the Experts* — their latest writings and thinking.** All of Cutter's research — right down to the last *E-Mail Advisor* — is provided exclusively by internationally recognized expert practitioners. Cutter has no desk-bound analysts. Cutter's research allows clients to tap into this brain trust and get the latest thinking on the business-IT issues challenging enterprises worldwide.
- **Cutter Inquiry Privileges give clients *Access to the Experts* — personalized guidance from the world's top practitioners.** Every inquiry is fielded by a Cutter Senior Consultant, Fellow, or Faculty member. Clients can purchase bundles of hours and may allot some of the inquiry



Cutter's focus is on the user community — the people who use technology-related products and services to deliver business results. Vendor independence (particularly in terms of funding) means that Cutter's reports don't read like commercials. The advice given by Cutter is practical and pragmatic — it draws on a wealth of experience from a wide pool of 'hands-on' industry practitioners. Their breadth and depth of coverage is reflected in the frequency and format of reports — the underlying focus is on the supporting principles which enable success, not the latest fads.

— Paul Ramsay,
Service Delivery Manager,
Equinox,
Auckland, New Zealand

time to senior executives, such as the CIO, pairing him/her with a top business-IT strategist so they can brainstorm every month. This allows the two of them to build a rapport and gives the Cutter expert a deep understanding of the issues the enterprise is facing. The Cutter expert quickly becomes a valuable advisor.

- **Cutter Consulting and Training gives clients *Access to the Experts* — hands-on assistance from best-in-class consultants.** Unlike other consulting and training firms that use senior partners to sell the work but junior staff to execute, Cutter uses only its best-in-class experts for every assignment. The Consortium's great strength is that it can draw from its more than 150 internationally recognized experts to assemble the ideal team to help your enterprise tackle any business-technology challenge it faces.
- **Clients call Cutter "the thinking person's research firm"** because of Cutter's dedication to debate of the business-technology issues enterprises face and its success in attracting the thought leaders worldwide to conduct this debate. Clients get cutting-edge thinking plus multiple viewpoints so they can determine what's best for their individual enterprise/department/project. Reliance on current industry trends should not replace the hard work of figuring out what's best for your organization given its business strategy and drivers.
- **Emphasis is on strategies and processes**, not on vendor/product detail. This is based on Cutter's view that it is not the choice of product or vendor that determines success or failure, but sound strategies and business processes, agility, leadership and effective teams, and hard-hitting financial analysis.
- **Written content is likened to "consultancy in print"** since it provides hands-on, actionable solutions from expert practitioners who are successfully implementing these ideas, whether it be IT strategic planning, security strategies, or risk management. This relates back to the fact that the content is written by the people who are at the forefront of their fields and who are guiding companies daily.
- **Cutter Research — from reports to podcasts, *E-Mail Advisors* to Webinars — is prepared with both the IT and the business user in mind.**
- **Information and advice is truly objective.** Cutter is unique in the research space in having no relationships with vendors. It is well known that other analyst firms derive significant portions of their revenues from vendors and that the choice of which vendors to evaluate, and the slant of the research, can be influenced by this factor.
- **Emphasis on agility.** The leaders of the agile project management and software development movement are all part of Cutter. Agile methodologies are evolving in real time, as new projects push the previous limits. Clients take advantage of the latest breakthroughs via Cutter research, training, and consulting.
- **Focus on project management, risk management, and change management** so IT can accomplish the changes business demands. Software engineering's pioneers are key to the Cutter team and they know that delivering results depends on successful project, risk, and change management.
- **Emphasis on IT when it matters — IT as innovation, as the hyper-differentiator** that will give your business the competitive edge it needs to succeed.

Cutter Consortium
Access to the Experts

The peace of mind that comes from knowing you're relying on the world's leading experts.

About the Practice

Business Intelligence Practice

The strategies and technologies of business intelligence and knowledge management are critical issues enterprises must embrace if they are to remain competitive in the e-business economy. It's more important than ever to make the right strategic decisions the first time.

Cutter Consortium's Business Intelligence Practice helps companies take all their enterprise data, augment it if appropriate, and turn it into a powerful strategic weapon that enables them to make better business decisions. The practice is unique in that it provides clients with the full picture: technology discussions, product reviews, insight into organizational and cultural issues, and strategic advice across the full spectrum of business intelligence. Clients get the background they need to manage technical issues like data cleansing as well as management issues such as how to encourage employees to participate in knowledge sharing and knowledge management initiatives. From tactics that will help transform your company to a culture that accepts and embraces the value of information, to surveys of the tools available to implement business intelligence initiatives, the Business Intelligence Practice helps clients leverage data into revenue-generating information.

Through Cutter's subscription-based service and consulting, mentoring, and training, clients are ensured opinionated analyses of the latest data warehousing, data mining, knowledge management, CRM, and business intelligence strategies and products. You'll discover the benefits of implementing these solutions, as well as the pitfalls companies must consider when embracing these technologies.

Products and Services Available from the Business Intelligence Practice

- The Business Intelligence Advisory Service
- Consulting
- Inhouse Workshops
- Mentoring
- Research Reports

Other Cutter Consortium Practices

Cutter Consortium aligns its products and services into the nine practice areas below. Each of these practices includes a subscription-based periodical service, plus consulting and training services.

- Agile Project Management
- Business Intelligence
- Business-IT Strategies
- Business Technology Trends & Impacts
- Enterprise Architecture
- IT Management
- Measurement & Benchmarking Strategies
- Enterprise Risk Management & Governance
- Sourcing & Vendor Relationships

Senior Consultant Team

The Senior Consultants on Cutter's Business Intelligence team are thought leaders in the many disciplines that make up business intelligence. Like all Cutter Consortium Senior Consultants, each has gained a stellar reputation as a trailblazer in his or her field. They have written groundbreaking papers and books, developed methodologies that have been implemented by leading organizations, and continue to study the impact that business intelligence strategies and tactics are having on enterprises worldwide. The team includes:

- Verna Allee
- Stowe Boyd
- Ken Collier
- Clive Finkelstein
- Jonathan Geiger
- David Gleason
- Curt Hall
- David C. Hay
- André LeClerc
- Lisa Loftis
- David Loshin
- Larissa T. Moss
- Ken Orr
- Gabriele Piccoli
- Thomas C. Redman
- Ricardo Rendón
- Michael Schmitz
- Ed Schuster
- Mike Sisco
- Karl M. Wiig