



Published in DM Review Online in July 2006.

Printed from DMReview.com

Open BI Forum: Live ... From the World Technology Community ... It's Open Source BI!

by Steve Miller

Steve would like to thank Dave Reinke, co-founder of OpenBI, LLC, for authoring this month's column.

Back in the mid 70s, Saturday Night Live (SNL) arrived on late night TV with a radically different approach to live entertainment. The first few seasons represented a true experiment in network television, departing from the cookie-cutter prime time variety shows of the era and introducing a new brand of raw, energetic - and irreverent - comedy. Political and social satire with unapologetic impersonations created a new genre of stars. The venue was New York, not Hollywood. The "Not Ready for Prime-Time Players" cast was mostly unknown comedians. The band appeared in blue jeans, not tuxedos. Fast forward 30 years and SNL is seen as a permanent fixture on late night TV - a reliable, consistent source of, and a well-worn springboard to, fame and fortune for its cast members. Are there lessons to be learned from this in the BI technology world?

In the past few years, a similar upstart has emerged to confront the mature, staid proprietary software market. Open source (OS) technology has demonstrated the same experimental success as SNL, finding a bold way to create and deliver high-quality software to the contemporary commercial enterprise. The OS model has been a disruptive experiment in the software marketplace, diverging radically from the commercial paradigm. The traditional value propositions of proprietary software vendors are under attack, with volunteer-based, coordinated, global development communities supplanting small, corporate - owned enclaves of highly paid developers. New OS players are emerging, and software powers such as Microsoft, IBM and Oracle are making serious accommodations to the growing threat. Many of the early "unknown" developers of the open source world have utilized

this growing platform to achieve technical celebrity and professional fortune as a result of their early contributions.

Linus Torvalds' Linux operating system was the bellwether of early OS success, establishing a user beachhead and a credibility that could not be dampened, despite both frontal and rear assaults from existing market powers. Quickly following were a torrent of OS alternatives to established proprietary offerings, each building on the next, progressing up the software hierarchy. The Apache Web server, MySQL database and Perl/Python scripting languages have combined with Linux to create the "LAMP" stack for Web applications, which has a significant presence in the marketplace and has been adopted and touted by some of today's most important technology-savvy companies. Now the OS phenomena has moved beyond these core infrastructure building blocks and offers compelling options for critical BI components that include ETL (extract, transform and load), reporting, OLAP and statistical modeling/graphics (see Figure 1). The alternatives in each of these categories merit exploration and consideration by today's value-minded CIO.

	Open Source Examples	Proprietary Examples
Databases	<ul style="list-style-type: none"> • EnterpriseDB • MySQL • PostgreSQL 	<ul style="list-style-type: none"> • DB2 • Oracle • SQL Server
BI Platforms	<ul style="list-style-type: none"> • JasperIntelligence • Pentaho 	<ul style="list-style-type: none"> • BusinessObjects • Microsoft • Oracle
ETL	<ul style="list-style-type: none"> • Kettle • KETL 	<ul style="list-style-type: none"> • DataStage • DTS • Informatica • Oracle Warehouse Builder
Query & Reporting	<ul style="list-style-type: none"> • BIRT • JasperReports • JFreeReport 	<ul style="list-style-type: none"> • BusinessObjects • Cognos • MS Reporting Services
Statistics & Graphing	<ul style="list-style-type: none"> • R 	<ul style="list-style-type: none"> • Insightful/S-Plus • SAS • SPSS
OLAP	<ul style="list-style-type: none"> • Mondrian w/ JPivot 	<ul style="list-style-type: none"> • Hyperion • MS Analytical Services • Oracle Analytical Workspaces
Languages	<ul style="list-style-type: none"> • Perl • Python • Ruby 	<ul style="list-style-type: none"> • C# • Java* • VisualBasic

* Sun has announced that they plan to migrate Java to Open Source

Figure 1: Open Source Alternatives for Proprietary BI Components

MySQL and PostgreSQL are the leading OSBI database technologies. Each has been used for years by commercial enterprises and leverages worldwide development and support communities. Both provide most, if not all, of the basic functionality expected of a relational database management system (RDBMS) for BI. A common fear, uncertainty and doubt (FUD) argument raised against OS databases is scalability; however the OS community has been quick to respond. With its 8.1 release, PostgreSQL now supports relevant BI features that include dynamic bitmap index scans, table partitioning and high-performance bulk loading. Moreover, recent surveys suggest that the majority of data warehouses support well below 1TB, perhaps mitigating scalability concerns for many users.

Kettle (now Pentaho Data Integration after its recent inclusion in the Pentaho project) is an intriguing option for ETL technology. A repository-based, GUI-driven ETL development and deployment technology, Kettle supports all core features expected from an ETL tool. Given its programming paradigm, anyone who has used Informatica or Oracle Warehouse Builder can quickly learn Kettle. More than 40 prebuilt mapping objects that can be combined in modular ways to create complex transformations are provided. Further, JavaScript integration is available for custom-developed mappings. Kettle is able to source from and write to leading RDBMSs and a wide variety of flat file types including Excel, csv, XML and fixed format. Plug-ins are available that enable connectivity to SAP. Kettle also provides the ability to create and execute jobs that sequence transformation execution and catch/respond to processing errors. When compared to commercial competitors, Kettle emerges as a highly functional ETL solution.

There are several capable OS report-development options, most notably BIRT, JasperReport and JFreeReport, that provide similar functionality to proprietary cousins such as Crystal Reports and Actuate. Each connects to leading commercial and OS databases, provides a GUI-based editor with wizards and enables report bursting. One of the special characteristics the OS tools share that proprietary vendors have been slow to adopt is an obsession with openness. These reporting technologies were built to be integrated within Java and J2EE frameworks and thus may be better suited for embedded, operational reporting needs. Indeed, OSBI's current strength directly addresses

the growing demand for flexible BI application development in support of performance and process management initiatives.

An area of OS reporting functionality that is still relatively immature is the semantic-based, ad hoc query tool capabilities similar to BusinessObjects with its Universe construct. The reporting options listed above enable rapid development of highly parameterized and modular managed query applications but lack a strong semantic layer with the dynamic SQL generation required to support completely ad hoc end-user query functionality. There are solutions for this omission, however. Today, OS-based, user-friendly, ROLAP solutions are readily available. Mondrian, now also the analysis component of the Pentaho platform, is the leading OS ROLAP server while JPivot is a popular user interface that together support access to all leading proprietary and OS databases for source data, in addition to consuming both MDX and JOLAP queries and enabling easy-to-use drilling and slice-and-dice functionality familiar to OLAP users.

Sophisticated statistical analysis and charting has long been the domain of expensive proprietary languages from vendors such as SAS, SPSS and Insightful. However, for almost a decade there has been a free OS alternative, R, that has legions of almost fanatical worldwide devotees. The R project grew as a fork from the original S language developed at Bell Labs and commercialized as S-Plus by Insightful. Today, R and its worldwide community of contributors support a variety of pre-packaged statistical models, applications and graphics. Indeed, R's core supports many of the mundane capabilities of data management, manipulation, and presentation pervasive across business intelligence. In addition, R has a flexible API that enables both Java integration for decision management application development and support for leading agile languages such as Python and Perl. The latter can be a huge productivity boost for the data preparation tasks that often plague analytical modeling efforts.

Together, OSBI technologies are increasingly relevant and competitive to their commercial cousins. A frequent inhibitor to OSBI adoption, however, has been concern over consistency of support, component integration and assurance of continued maturation. Fortunately, in the past year several well-funded commercial open source vendors have launched to fill this need. The OSBI marketplace is now inhabited by

commercial open source database vendors such as EnterpriseDB and Pervasive that provide extended functionality and commercial support for PostgreSQL, as well as BI platform vendors such as Pentaho and JasperSoft who offer preconfigured, complete OSBI stacks - often composed of the technologies enumerated above. These new players add value by unifying OS project brands, simplifying the integration and installation of various OSBI components, providing commercial-grade support, and stewarding the core OS development communities to ensure that feature evolution is prioritized by market needs.

Although we have not exhausted the list of available OSBI offerings , the BI user market should be intrigued by the possibilities. The emergence of legitimate OSBI technology options supported by well-funded commercial open source vendors provides a new option for IT departments looking for more affordable BI infrastructures. And while the technologies may not yet have all of the bells and whistles offered by their proprietary brethren, sufficient base functionality built on open platforms does exist. With the rate of OSBI innovation and investment accelerating, it leaves one to wonder... Will those who ignore OSBI today be left to the same, inevitable fate that continually befell Mr. Bill? "Ohhhh nooooooooooooooooo...."

Dave Reinke is co-founder of a Chicago-based business intelligence (BI) services firm [OpenBI, LLC](#), that specializes in delivering analytic solutions with both open source and commercial technologies. Reinke has more than 20 years of experience creating intelligence and Internet-based solutions, starting with his object-relational focused graduate work at University of Illinois in the late 80's, to database consulting with Oracle Corporation, to executive and thought leadership positions at Braun Consulting and Fair Isaac Corporation. You can reach him at david.reinke@openbi.biz. Reinke would like to thank colleagues Bryan Senseman and Steve Miller for their contributions to this article.

Steve Miller is co-founder of a Chicago-based business intelligence (BI) services firm [OpenBI, LLC](#), that specializes in delivering analytic solutions with both open source and commercial technologies. Miller has more than 30 years of experience in intelligence and analytics, having migrated from health care program evaluation, to database

consulting with Oracle Corporation, to running a fast-growing BI services business at Braun Consulting. Advances in technology over that time have fundamentally enabled the use of quantitative methods for business differentiation. OpenBI, LLC, is all about helping customers attain that differentiation. You can reach him at steve.miller@openbi.biz.

Copyright 2006, SourceMedia and DM Review.