



---

June 2009

## Open Sesame:

Why Open Source BI, Data Integration, and  
Data Warehousing Solutions are Gaining in  
Acceptance

Claudia Imhoff, Ph.D.

---

# Executive Summary

In the beginning, open source solutions were the shiny play things of the techno crowd. After all, they were fun, new, and most importantly, free. No one would have considered using them in any real world corporate IT shops...

My, how times have changed. Today, open source solutions are not only being considered, they are being implemented by large and small enterprises -- at rates that are mind-boggling. This paper examines the challenges faced by organizations today regarding their BI, data integration, and data warehousing environments, why traditional solutions fall short, and the rise of upstart open source companies.

## Introduction

Open source software (OSS) has gained significant traction in both acceptance and popularity in recent years. It seems you can't open a technology magazine, attend a technology conference, or perform a search on BI, data integration, or data warehousing without reading or hearing about open source. Just to give you some idea of its popularity, here are a few facts:

- According to Aberdeen, 25% of survey respondents will adopt open source BI in next 12 to 24 months<sup>1</sup>
- Open Source CEOs agree – open source is a worldwide growth story in 2008<sup>2</sup>
- First nine months of 2007, open source deal flows doubled each quarter<sup>3</sup>
- Sun has made a significant commitment to open source with its \$1 billion purchase of MySQL<sup>4</sup> and now Oracle with its purchase of Sun<sup>5</sup>

---

<sup>1</sup> "The TCO of Business Intelligence – Open Source Takes on Traditional BI", [www.aberdeen.com](http://www.aberdeen.com).

<sup>2</sup> [www.OpenSolutionsAlliance.org](http://www.OpenSolutionsAlliance.org).

<sup>3</sup> [www.the451group.com](http://www.the451group.com)

<sup>4</sup> <http://www.sun.com/aboutsun/pr/2008-02/sunflash.20080226.1.xml>

<sup>5</sup> <http://www.oracle.com/us/corporate/press/018363>

Not familiar with OSS? Wikipedia<sup>6</sup> defines OSS as:

“Computer software for which the human-readable source code is made available under a copyright license (or arrangement such as the public domain) that meets the Open Source Definition (used by the Open Source Initiative<sup>7</sup> to determine whether or not a software license can be considered open source). This permits users to use, change, and improve the software, and to redistribute it in modified or unmodified form. It is very often developed in a public, collaborative manner”.

Certainly, a major component in the definition is that the source code is viewable to all, but it means much more than just this access. The distribution terms of open-source software must comply with strict criteria, some of which are:

- Free Redistribution of the software
- Inclusion of source code in any program including any derived works
- There can be no license specified for a particular product and restrictions cannot be placed on other software distributed with the open source
- Licenses must be technology neutral.

The bottom line is that OSS licenses are *free* – without exception. While many companies using OSS for production purposes elect to pay for enhanced features, technical support, or other services, the total cost of ownership for OSS solutions are typically a fraction of the cost of proprietary solutions. For these reasons and others, open source solutions are not only being considered, they are being implemented by large and small enterprises at increasing rates.

These enterprises are facing difficult challenges today – especially mid-sized ones or those focused on web analytics – traditional solutions may fall short. This paper starts with a description of these challenges and why the traditional approaches don’t always measure up.

Next, we examine the open source BI, data integration and data warehousing options including both the barriers to its acceptance and its potential benefits. We end with a section on the considerations for adopting open source components.

---

<sup>6</sup> [http://en.wikipedia.org/wiki/Open-source\\_software](http://en.wikipedia.org/wiki/Open-source_software)

<sup>7</sup> <http://www.opensource.org/>

# Challenges Facing Companies Today

Companies today face difficult circumstances, which are particularly exaggerated in mid-sized companies who are really feeling the squeeze of the present economic downturn:

- **Shrinking IT departments** – While large companies are seeing their IT resources get downsized, mid-sized companies have very little IT “fat” to remove. Even so, these companies are trying to reduce their IT staff wherever possible. While these IT departments were never very large to begin with, they are now being asked to do much more with many fewer resources. Building a BI environment is still critical to these companies, but the ability to build these critical applications is increasingly difficult. They need all the help and technology they can get.
- **Reduced budgets** – Funding for IT is under severe scrutiny today. While BI still remains the number one initiative for most enterprises, CIOs are struggling to find the funding for these projects. Outsourcing, budget cutbacks, spending freezes are all in play. The ability to purchase necessary BI technology is becoming increasingly difficult. Again, the availability of free BI software is quite attractive.
- **Minimal tolerance for IT failures or trial and error projects** – Gone are the days when the business would give IT some leeway to see if something “might work”. Testing BI technologies or potential applications was a useful way for IT to learn what worked and what didn’t, but it was also fairly expensive and led to disillusionment in the business users. The ability to test whether a technology might work was quite limited; implementers either had to buy the technology outright or get a limited time license which may have been too short for a proper trial.

## Traditional BI Solutions Often Fall Short

Given the challenges stated in the previous section, you can see why traditional BI solutions are having a difficult time. Certainly, any price tag may be the stumbling point. By the time you buy licenses for the various technologies – data integration, data quality, DBMS, hardware, BI application, etc. – the total cost of ownership can place these technologies completely beyond the budget horizon.

In addition, the complexity of the overall BI technology stack can be overwhelming. In July, Forrester issued their Wave report for Enterprise BI Platforms<sup>8</sup> in which they show the many different components (more than 40!) that go into delivering large enterprise-class BI environments. The complexity of these environments is enough to frighten away a great number of potential customers of BI. Fortunately, we now have large vendors who offer end-to-end suites that cover all these moving parts. The downside is that they are quite expensive and are still not completely integrated.

## What About Open Source BI Solutions?

Enter the open source movement. What started out as a group of developers who were fed up with proprietary software practices has grown into a full-fledged sophisticated community with standards, codes of conduct, rules, and formal procedures for what constitutes open source products. The road to this end is still not finished, and there are still a number of doubts, misconceptions, and even myths about whether OSS is a truly viable form of software development. Let's look at a few of these barriers as well as the benefits of OSS.

### Barriers to the Adoption of OSS

There are barriers to the adoption of open source software by enterprises. These barriers include the perception that i) open source licenses are somewhat viral in nature, ii) they lack formal support and training, iii) the velocity of change is high and unstable, and iv) OSS lacks a long term roadmap. Most of these relate to risk and pose a serious hurdle for many companies considering the adoption of OSS.

It should be noted that not all OSS licenses are viral especially when the technology is used in a standard BI application deployment mode, and most OSS companies make money from paid support, consulting, enhanced functionality and documentation. It must also be stated that many traditional, proprietary vendors do not disclose their exact future plans either, leaving their customers guessing about the future of their software.

Opponents to BI OSS state that it lacks the features or functionality needed to succeed right now. Further, they question how much the BI OSS community is actually contributing to the overall

---

<sup>8</sup> [www.forrester.com](http://www.forrester.com)

offerings. It behooves a company considering the adoption of BI OSS to determine whether the available functionality being offered is suitable for their BI problem. Granted, features and new functionality are constantly being contributed by the OSS community but those considering BI OSS should not assume that a particular feature of function will be available soon.

Others question whether the BI OSS vendors have the same quality of professional support that is found in the proprietary vendor's environment. After all, proprietary vendors have armies of professional services consultants either on the payroll or via tight relationships with consulting partners. It is true that many open source vendors are small and may lack internal professional services resources. Therefore, another criterion for adoption consideration should be the demonstration of support by the BI OSS vendor.

Another concern relates to the fears of patent infringement, led by large proprietary software vendors, and copyright infringement, so ungracefully demonstrated by SCO when it filed a lawsuit against IBM. While justification of these fears is often anecdotal, they have led to demands for more audits of open source code and legal indemnification from open source software vendors.

## Benefits of Open Source Software

Many business models exist around OSS to provide a complete offering or product to help reduce these barriers. A complete offering should include support, commercial licenses, professional services, training, certification, partner programs, references and use cases.

These business models range from 'services only' organizations that do not participate in the development of the software to models where the majority of the software is created by full-time implementers that are employed by a central organization.<sup>9</sup> BI and data warehouse open source vendors have such business models, though some have only recently implemented them.

So what draws people to open source offerings? There are a number of solid benefits established for OSS:

- Price! In a recent survey of BI implementers<sup>10</sup>, 84% said software license costs were a major or moderate inhibitor to

---

<sup>9</sup> [http://en.wikipedia.org/wiki/Open-source\\_software](http://en.wikipedia.org/wiki/Open-source_software)

<sup>10</sup> Pentaho webcast, <http://www.b-eye-network.com/events/details/1965>

expanding their use of BI. Given the cutbacks in IT budgets occurring today, you can see why the price tag of software licenses inhibits potential BI implementations. And just when BI is needed more than ever!

The fact that OSS is free is perhaps the most compelling benefit – you don't need to justify significant up-front fees before you even get started with a BI project. Instead, you download the software, begin building a prototype of a BI application, and demonstrate it to the business community, long before you have to fork over a single dime to a software company. It is a nice way to try before you buy.

Robin Schumacher, MySQL's Director of Product Management, states in a recent blog posting, "Just as open source has revolutionized many areas of software these days ..., it's now doing the same thing in the area of data warehousing and BI. Witness the fact that data warehousing (as indicated on MySQL's last major community and customer poll) is now the fifth most common use case for MySQL".<sup>11</sup>

Of course, most OSS vendors have commercial open source offerings as well and while these do cost money they are typically priced below products from proprietary vendors. These offerings usually include more features, functions, or extended professional services support beyond what is available for free.

- The expanding open source BI ecosystem. Over the past several years we have seen the emergence of open source products up and down the BI stack. Companies such as Pentaho, Jaspersoft, and Talend have validated the market for open source BI reporting and ETL tools. Most recently, Infobright has introduced Infobright Community Edition (ICE), an enterprise class, open source data warehouse. This expanding open source ecosystem signals the start of the mainstreaming of open source in the BI market. This is good news for users as open source vendors work together to ensure interoperability and offer joint solutions that simplify deployment and operations.
- Open Source software's reliability and scalability. As I wrote in a previous white paper<sup>12</sup>, the advantage of column-store

---

<sup>11</sup> [http://dev.mysql.com/tech-resources/articles/datawarehousing\\_mysql\\_infobright.html](http://dev.mysql.com/tech-resources/articles/datawarehousing_mysql_infobright.html)

<sup>12</sup> [http://www.infobright.org/Learn-More/better\\_way/](http://www.infobright.org/Learn-More/better_way/)

databases for data warehousing is the simplicity and ease of deployment; the database design requires minimal design effort. With column-based databases, you don't need to design a complicated star or snowflake schema just to get performance. In fact, you can simply replicate the original transactional schema to yield a usable prototype. Also, there is no performance hit simply because you want to load the lowest level of granularity. These characteristics are particularly important for smaller companies with limited IT resources to get started.

- Commercially available open source data integration. Anyone who has built a BI environment knows that the integration of the data constitutes a significant effort and expense (some analysts estimate this at over 65% of the entire project). The emergence of new open source ETL technologies has certainly reduced the cost of this effort substantially. And these new technologies have incorporated the innovations and productivity improvements garnered over the years to make the effort less onerous as well.
- Open source community has grown to a significant size – and it is really active. One of the primary factors driving customer adoption has been a rebellion of sorts against vendor lock-in or the payment of a software license. The combination of a lower cost of production with the freedom and flexibility prevalent in open source deployments has been driving both adoption and production. The yin of BI software commoditization and the yang of customer demands for lowered costs of ownership have created a perfect opportunity for open source software and the open source community. If you adopt OSS, then it behooves you to become active in the OSS community. The vendors constantly receive input from their communities on product direction and roadmap.

## Data Integration – Going Beyond BI

Data integration is certainly at the heart of a reliable and stable BI environment. With enterprise-grade open source technologies like Talend Integration Suite, IT organizations are now able to perform ETL processing on very high volumes of data in no time. Data warehouses and analytic applications can be updated faster with low latency, granular data so necessary for operational BI solutions. Through these improvements in ETL processing, BI environments now support the



enterprise's ability to make more informed and precise business decisions. But data integration technology has also taken on a much wider role than just ETL processing for BI and analytics.

Many enterprises now look to innovative companies to help with data integration across operational applications in addition to traditional BI ETL processing. Data integration for operational applications is used to create standard sets of data to be used by multiple operational applications. Since data integration not only standardizes the data, it can also be used to ensure the quality of the data as it is integrated between operational applications.

It should be noted that operational data integration is significantly different from standard BI ETL by its very nature. Therefore, an enterprise needing to integrate data across its operational applications must pick a data integration technology that recognizes and handles these differences which include:

- A limited number of sources for the data. Unlike data warehouse data integration, operational data integration needs data from only a few other systems. However, it is critical that these systems reliably receive the integrated data, that the transfer of data can be assured.
- The volume of data to be integrated is low but the speed of the data integration is much faster. Batch processing of data is usually not fast enough. Many traditional ETL technologies struggle with this low latency requirement and is where the new open source data integration vendors have an advantage.
- Operational data integration is also generally performed on fairly routine sets of data usually involving transactions like POS, ATM, or CDR transactions. In addition, these sets of data usually do not require massive transformations but rather select fields only may need to be transformed.
- Finally, operational data integration may be used for intermittent work only rather than the constant updating occurring in BI environments.

Operational data integration is involved in projects such as upgrades of operational databases or migrations from one database to another. The projects may be for database consolidation, synchronization or B2B data exchange.

# Considerations in Adopting BI Open Source Software

One consideration in adopting BI OSS should be whether the BI OSS can coexist with a company's current (proprietary) BI environment. Many companies considering the implementation of an open source data warehouse, for example, must determine whether the new environment will leverage the existing [SQL Server / DB2 / Oracle / fill in the blank vendor name] database currently housing your data warehouse. Can the new product sit on top of your traditional environment and use its data? Will you have to rip and replace parts of your BI environment to accommodate the open source software? What about your existing reports? Will the new software be able to support existing [Business Objects / Cognos / SAS/ fill in the blank vendor name] reports and queries?

What we have seen are four models of deployment for companies adopting OSS offerings:

- Coexistence with traditional proprietary BI technologies. There do not appear to be any problems with this model. It can minimize a company's risk or exposure with consolidated vendors. However, IT resources should do their due diligence and prove the fit of the technology and services with their internal environment.
- Co-deployment with traditional proprietary BI technologies. BI OSS vendors have recognized the need for their technologies to use or leverage an enterprise's existing BI investments. They have open APIs and interfaces that allow the enterprise to use all that they have while reducing the total cost of ownership with the BI OSS additions.
- Replacement of traditional proprietary BI technologies. Where it makes sense, enterprises are ripping and replacing their existing proprietary technologies with BI OSS. Many justify this move due to the need to upgrade their BI capabilities, reduce their overall costs, expansion into new areas of the organization, etc. They capitalize on the fact that they are facing some "disruptive" event (software or hardware upgrade, licensing change, movement from departmental to enterprise architecture, etc.) anyway and the time is propitious to change out technologies.
- New installations. One of the advantages of open source previously cited was the ability to quickly and easily download software and

get started – with no upfront cost. For small and mid size companies who are implementing BI and data warehousing for the first time, open source solutions that have vendors standing behind them provide an excellent option.

An obvious consideration must be your IT department's stance on open source. If there is push back on its adoption, you can always point out that open source is everywhere. If you use TiVO, Google, YouTube, or even a cell phone, you have experience with open source. Many enterprises use Linux, Apache, or Firefox – all of which are open source solutions. Finally, even Business Objects, Cognos, and several other traditional BI solutions have components that are open source in their offerings.

A final consideration concerns the question of support from your existing IT team for the open source BI capability. Of concern is whether the existing skills of the IT implementation team will transfer to an open source offering. For example, if a programmer / analyst is already versant in similar software, then the transition from one environment to another is relatively straight-forward. Of course, the BI OSS vendors all offer training but it is important to understand that many of the skills used in BI environments today are easily transferable to the open source environment.

## Wrap-Up

Given the stresses of today's economy and the increasing need for better decision making environments, IT departments are looking for alternatives to expensive, traditional BI technologies. Often, traditional vendor offerings are cost prohibitive for many smaller enterprises. Even for large enterprises looking to expand their BI capabilities throughout their organizations, an open source BI solution is an attractive solution.

I encourage you to take a look at such innovative and cost-effective solutions. Determine your company's needs, its acceptance of open source, the specific features and functions of such a comprehensive offering, and come to your own conclusions about its fit for your BI environment.