



Open Source Business Intelligence Costs and Benefits

What You Need to Know

Jaspersoft

539 Bryant Street, Suite 100

San Francisco, California 94107

phone: 415.348.2300 fax: 415.281.1987

jaspersoft.com

Introduction

For over a decade, Business Intelligence (BI) has been sold as an investment—typically a six-to-seven figure investment¹ that promised to make organizations smarter and more competitive. Admittedly, the approach is valid: many enterprises have achieved handsome returns on investments in data warehousing and BI. But the “high investment/high return” scenario priced BI out of the reach of most organizations.

A 2006 survey of IT and line of business professionals conducted by Ventana Research² concludes that interest in and adoption of open source BI is widespread and growing, driven primarily by cost considerations. The report is careful to point out, however, an interesting disparity. While 48% of respondents expect open source BI to cost half as much as equivalent proprietary BI software, one-fourth of the respondents expect it to cost the same.

This guide will help you understand the hidden and explicit costs associated with open source BI in comparison with familiar proprietary BI solutions. At the end, you should understand whether open source BI is a viable option for your organization.

Evaluation is Never Cost-Free

The first thing to recognize is that evaluating BI is never cost-free. The nature and extent of those costs can vary widely when evaluating open source and proprietary software.

Consider that the typical enterprise software buying cycle is well over 6 months long.³ Much of the evaluation period is spent building the business case to encourage executives to fund the project. As the evaluation team builds the business case, they inevitably add requirements to the project—before even seeing any tangible results. The idea is to project greater returns on that large investment. This process is repeated again and again, complicating and extending projects. The longer it takes between project inception and rollout, the greater the lost opportunity cost, the more upfront risk, and the less value the business gets from the software.

The technical evaluation can also cost real money if a vendor offers “paid pilots,” requires professional services architects to install the solution, or if the solution is so proprietary that organizations must take training classes in order to get started.

Evaluating open source BI is not cost-free, either, but the software is freely available for download, and the labor costs should be dramatically lower. This is because the evaluation of open source BI is more likely to be “bottom up” rather than “top down.” This means that the technical evaluation and even a pilot implementation may occur at minimal cost and effort before executives are involved. When and if a project becomes large enough to require executive-level approval and budgeting, the business case is based on actual results already achieved, not academic projections or third-party endorsements.

¹ Burton, Betsy, et al. Business Intelligence Focus Shifts From Tactical to Strategic. ID Number: G00139352. Gartner Group, 22 May 2006.

² Ventana Research. Open Source BI: A Ventana Research Primary Research Study. San Mateo, CA: Ventana Research, 2006.

³ Berryman, Ken. The Changing Software Landscape – Pricing and Purchasing Trends. Santa Clara, CA: McKinsey & Company, Oct 2004.

Getting the Most for Your Money

One of the biggest problems with proprietary solutions is that all the costs are born upfront by the customer before there is any reward. This dynamic is one of the key reasons for the growth and popularity of the open source movement. While open source is evangelized on many fronts, including security, flexibility, and competitive advantage, organizations adopt open source primarily because of the price/performance ratio. With open source BI, organizations can adjust spending as they go, depending on the perceived returns. Almost all of the investment is spent configuring or customizing the solution to meet the organization's needs, not on a generic system that needs to be customized just to work.

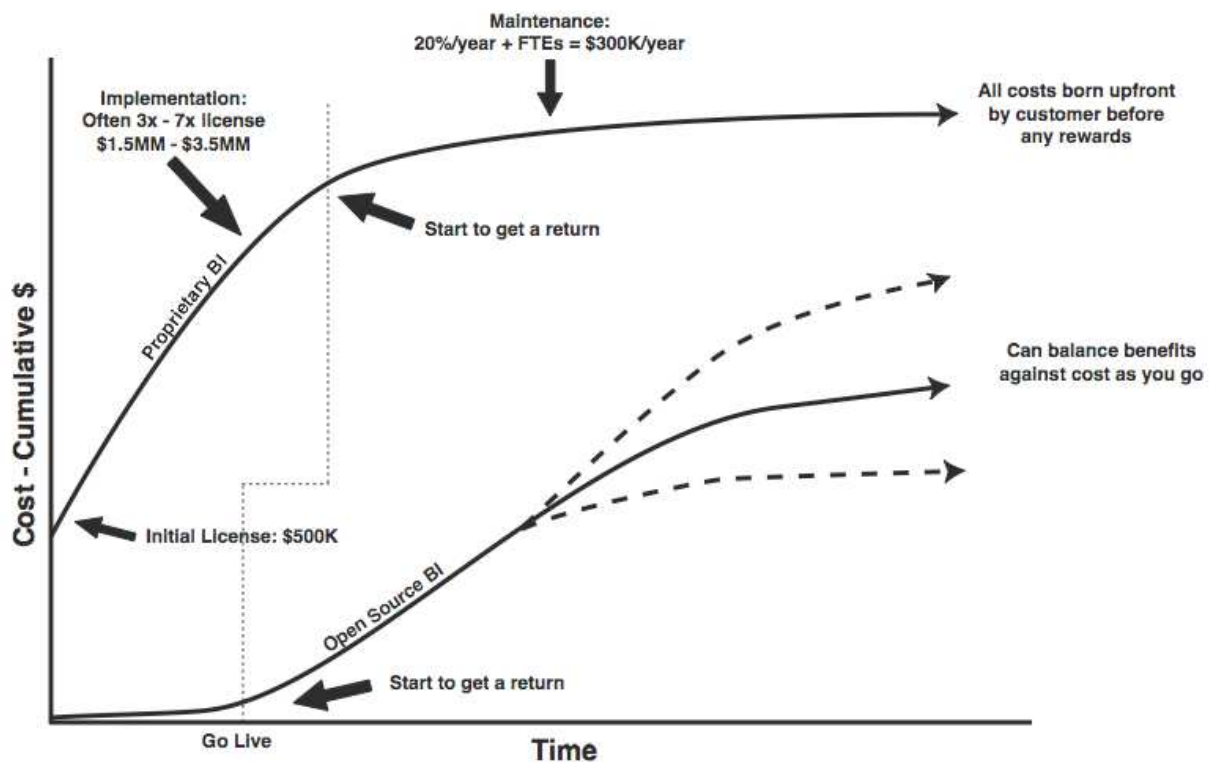


Figure 1: Open source allows faster returns on smaller investments.

Opening Up Business Intelligence

While Linux, Apache, MySQL, and other open source software are routinely deployed in the enterprise, open source BI solutions are just at the point where they are sufficiently feature-rich to satisfy user requirements⁴. The major attraction of open source is to save time and money, but there's an extra

⁴ Ventana Research. Open Source BI: A Ventana Research Primary Research Study. San Mateo, CA: Ventana Research, 2006.

dimension to open source BI - open source BI allows iteration and evolution in a way that proprietary BI does not.

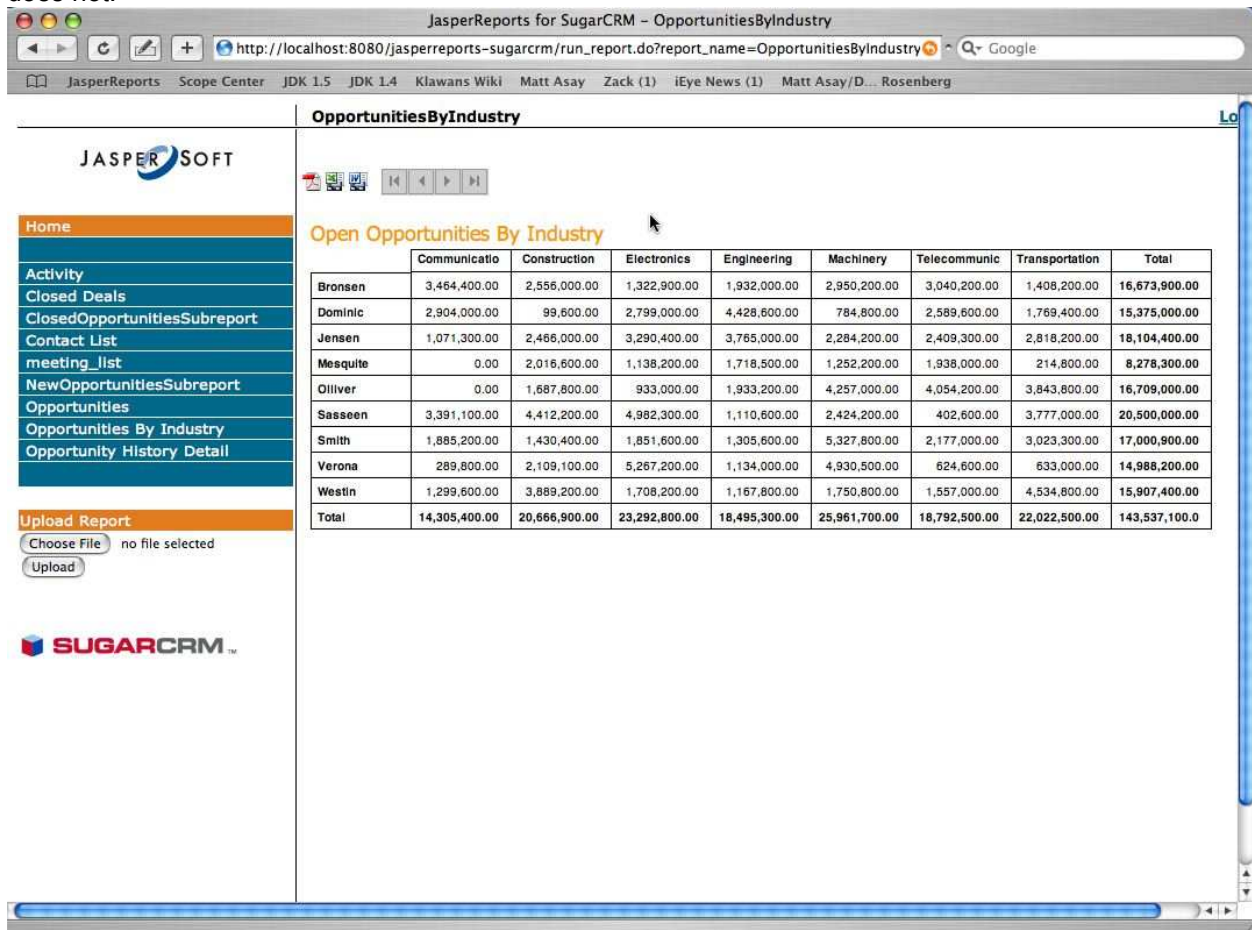


Figure 2: open source has moved up the stack to integrate applications, reporting, and analytics.

This is because the evaluation period is fundamentally different. With proprietary BI, a lot of time is spent selling and buying based on anticipated results. With open source, the evaluation is tangible, productive, fast, and as close to free as an organization can get. Is there a specific problem that needs to be solved? Simply download a reporting solution, build the report, get the answer (or immediately evaluate another solution that better solves the problem), and move on.

Can this be done with proprietary software? Yes and no. Proprietary BI vendors may give organizations permission to install and evaluate software for free, but getting to that point may take much longer, and they are likely to expand the context of the problem to demonstrate greater long-term value. That's not bad, but getting fast answers for free is better. With open source BI tools and 24x7 user communities, nimble teams can solve discrete problems one after another, building knowledge and executive support along the way.

Best Practices for Cost-Effective Implementations

Implementation is where the cost benefits of open source BI versus proprietary can vary the most. The following best practices encourage an "iterate and evolve" approach that can lead to greater rewards, faster.

Categorize and Prioritize the Backlog

Organizations should categorize and prioritize the report requests backlog. If the report backlog is very large, organizations can start by working with just a small segment. Each request should be categorized into one or more of the following groups: Embed, Manage, and Analyze.

To determine how to categorize the requests, use the following criteria:

- **Embed:** the data often comes from a single application, such as a CRM, ERP, or other business application. Often the best way to present this information is in the application itself. Reports embedded within familiar business applications are easier to access and understand.
- **Manage:** the user wants a report to show up in her inbox on a regular schedule; requires a repository of canned reports; or needs a report that needs to be centrally shared, but restricted by role.
- **Analyze:** the data comes from more than one application, for which the SQL would crush the server. One indicator that a report falls into the “Analyze” category include the word BY, as in, “I want to see our sales BY state” or “deal size and discount rate BY brand.” Another indicator is when report data is exported to spreadsheets. While sometimes that means that the layout or formatting may need improvement, if data leaves a report and ends up in a spreadsheet, it’s usually because additional analysis is needed.

A single request can belong in more than one category. Precise categorization is not important; what is important is the approach, which allows iteration and evolution. The best practice is to use categorization as a tool to accelerate business benefit.

Embed BI in a Business Application First

Prioritize the backlog with the “Embed” list at the top. Very good open source reporting libraries are available that allow organizations to deliver pixel-perfect reports inside existing applications. Organizations typically need someone with Java skills to put the libraries in the right place. As for creating the reports themselves, the requirements vary. Report creators can use WYSIWYG editors or APIs. Depending on the maturity of the tool, organizations will need basic SQL skills for the query, plus scripting (such as Groovy or JavaScript) or Java skills for more sophisticated layouts.

Organizations should create a few reports and embed them in a business application. Re-creating existing reports can accelerate the learning curve. It also gives users the opportunity to express their opinions. Are the reports created using the open source tool better? In what way? Are there additional features in the open source tool that can be used to improve the reports? Could they be created more easily or more cheaply using existing tools? What else is needed to make the report better? What would that take?

While a single report from a single datasource isn’t exactly what people think about when we say “Business Intelligence,” this approach goes a long way to achieving its goal: getting the right data to the right people in the right context so they can make better, faster business decisions. The effort and cost are low and the payoff can be huge.

Extend Advanced Capabilities to One Business Unit at a Time

Managed reporting is more complex, and therefore may cost more for labor, support subscriptions, or other services. Compared to proprietary solutions, however, open source solutions allow organizations to pay only for what they need, and when they need it. Open source managed reporting solutions compare well with proprietary solutions. Prioritize the Manage list, and then build a repository for just one or two business units. For example, focus on Sales, Marketing, or Finance first. Actively solicit user feedback

and build support from users and executives within that business unit before extending the BI solution to other business units.

Allocate Special Resources for Analysis

Most complex is **Analyze**. For proprietary vendors, this phase requires big budget, big vision, and lots of time. With open source, organizations need a modest budget and a few burning questions that need to be answered. For both proprietary and open source, organizations need DBA, data warehousing, ETL, and programming skills. Some tasks are one-time tasks, such as implementing the initial cube or star schema. For proprietary vendors, the scope generally gets so large that organizations also need dedicated program/project management. With open source, organizations can start small, so they may not need as much administrative overhead.

After that, the keys to success with Analyze are the same for proprietary and open source. It is important to note that the transactional data store is not suitable for analysis. Analysis requires data to be organized into multidimensional “cubes” or “star schemas.”

Critical to success is knowing the questions that need to be answered—this comes from the report requests backlog. Organizations must always keep in mind that the objective is not to “build a data mart” or “roll out BI,” but to solve business problems. With the right questions, data architects can model the cube or star schema properly. (Remember the criteria for the Analyze bucket: “I want to see deal size and discount rate BY product”? “Deal size” and “discount rate” are facts and “product” is a dimension.) (See Figure 3: Typical star schema for analysis).

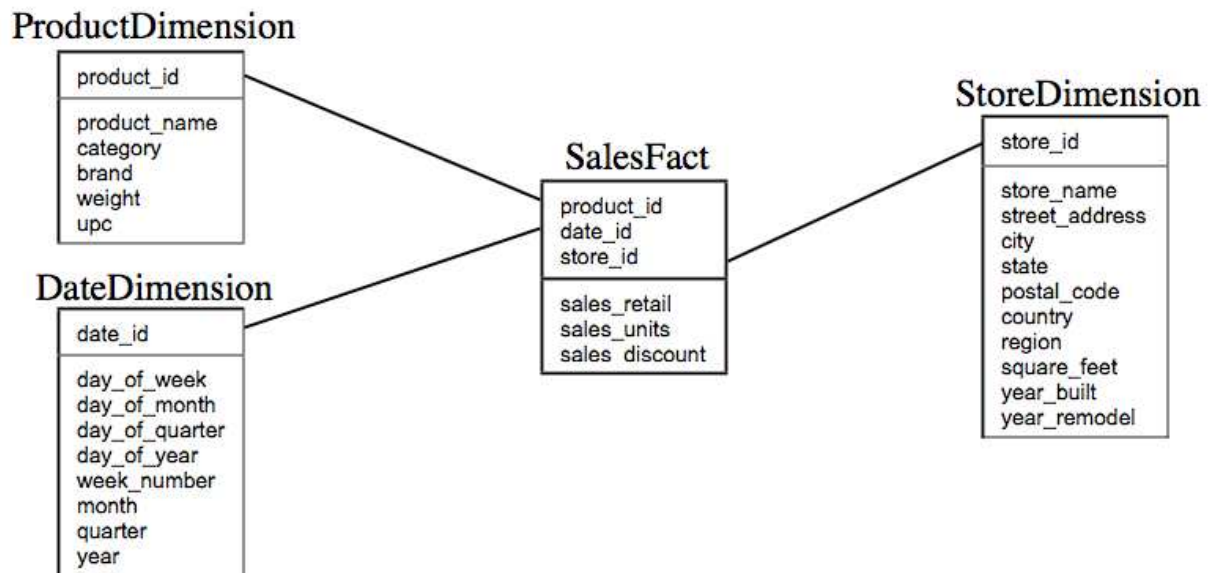


Figure 3: Typical star schema for analysis

A few reminders for star schemas: facts are the core. The Fact table will be “tall” – that is, it will have many rows, but each row is small. Fact tables contain the most detailed information, and the actual size depends on the facts that are being modeled. A “sales” Fact table, for example, will have one row for each order. Put another way, the Fact table holds numerical data for analysis, plus indices. Dimensions are smaller, “wide” de-

normalized tables that hold structured information about the facts. Dimensions are the “BY” tables.

Organizations should design the Fact and Dimension tables before starting to increase the likelihood of success. However, it should be acknowledged that the first implementation will not be “right,” however, and that users will change requirements. BI is iterative, and it’s one of the key reasons that organizations should NOT invest huge amounts of time or money in a big project until they deliver real value.

Here are a few tips on maintaining the database:

- Organizations can use a proprietary or open source database for the data mart.
- Tune the database correctly. Be sure to create indices on the dimensions used for searching, and on all the columns used for the primary key/foreign key relations.
- Don’t worry about the impact of indices on inserts – there won’t be any.
- Take advantage of any data warehousing extensions the RDBMS provider has created. For example, if using MySQL, use the MERGE engine to build Fact tables from subtables, preferably striped across multiple disks.
- Structure queries to search on Dimensions first, then pull in only the Fact entries needed. For example, in the star schema in Figure 3, identify the Product first and then only pull in the sales facts related to that product.
- Data Mart tables are BIG - avoid full table scans
- Run “ANALYZE TABLE ...” or its equivalent after updating the database so the RDBMS updates its performance optimization plans.
- Learn how to use the tools the RDBMS vendor provides to understand how the database is going to perform query execution.

Also, be prepared to iterate the ETL process many times to move the data from production systems to the data mart. As organizations set up the ETL process, new questions will be raised that influence which facts and dimensions really need to be included in the data mart.

Evaluating Open Source BI Solutions

Despite all its advantages, open source is not risk-free. The major hurdles are skill set, support, product maturity, and the glut of products to choose from.

Skill Set

Smaller organizations that are priced out of proprietary BI solutions might think that open source BI is also beyond their reach. This is because they don’t have programmers on staff to evaluate and deploy open source solutions. For example, a \$100M company may have only a handful of people in IT. These folks are dedicated to maintaining email servers, backup systems, security management, and provisioning employees. They have neither the time nor the required skills to set up and maintain a data mart.

The solution? Work with a preferred VAR or systems integrator. There are over 30,000 regional VARs in North America alone. Low barrier to entry and community support make open source solutions very attractive to the smaller VARs, which increase the buyer’s ability to get a great implementation at a great price. Organizations that do not have a preferred VAR can check the open source provider’s “Partners” web page, or post to a community site. A significant benefit is that the money spent with a VAR or systems integrator will be for limited-scope integration or customization services, not license fees.

Support

Many open source projects are part-time projects that are poorly supported or abandoned. This is a huge risk when building mission-critical systems and applications. Over the last five years, proprietary vendors have sponsored many top-tier open source solutions. These vendors offer pay-as-you-go support subscriptions and services, receive and post bug fixes, receive and implement new features, and publish a public roadmap developed with community input.

Product Maturity and Selection

Organizations such as O'Reilly's CodeZoo, Open BRR, SpikeSource, and Freshmeat offer information and services that make it easier to evaluate and implement open source solutions. Technology industry analysts such as Gartner, Forrester, and the 451group also track popular open source solutions alongside their proprietary counterparts.

Finally, open source should also not be confused with free or not-for-profit. For a comparison, review the Free Software Foundation and open source Initiative (OSI) websites at www.fsf.org and www.opensource.org. These sites include essays that explain the differences between the free software and open source movements, provide important information about the GPL (GNU public license), LGPL (lesser GNU public license), and much more.

Conclusion

Business Intelligence has been a "high investment/high return" solution for a long time now. However, the way it was evaluated, sold, and implemented priced proprietary BI out of the reach of most organizations. Cost consideration is driving interest in open source BI, but organizations should adopt new evaluation and implementation best practices to maximize cost savings and accelerate business benefits.

These best practices take advantage of the fact that open source software is freely available, which allows organizations to quickly prototype business solutions. This allows smaller teams to approach executive management sooner rather than later with a value proposition that applies to their organizations. With open source BI, organizations also have the flexibility to iterate and evolve. Rather than assembling large teams to formalize an ROI on proprietary software, nimble teams can solve discrete business problems one after another, building knowledge and delivering value along the way. And the best part is--little to no "investment" is needed.

For More Information

Business Readiness Rating (BRR)	Allows community ratings of open source software in an open and standardized way.	www.openbrr.org
Free Software Foundation (FSF)	Promotes the development and use of free software. Provides information that compares free software with open source.	www.fsf.org
freshmeat.com	Resources and information on Linux downloads and Linux training.	www.freshmeat.com
JasperForge.org	Development portal for open source Business Intelligence solutions.	www.jasperforge.org

O'Reilly CodeZoo	Provides metrics and information to help developers find open source components that are actively maintained. Ratings are provided by O'Reilly.	www.codezoo.com
open source Initiative (OSI)	Non-profit corporation dedicated to managing and promoting the open source Definition.	www.opensource.org
SourceForge.net	Provides free hosting to open source software development projects. Has the largest repository of open source code and applications, but can be difficult to assess project maturity.	sourceforge.net
SpikeSource	For-profit organization that distributes, integrates, manages and supports open source solutions.	www.spikesource.com

About Jaspersoft

Jaspersoft is the market leader in open source business intelligence (BI) and the most widely deployed BI product in the world. With more than 3 million downloads worldwide and more than 8,000 commercial customers in 96 countries, Jaspersoft offers the most widely used open source business intelligence software in the world. The Jaspersoft Business Intelligence Suite is comprised of an interactive reporting server, graphical and ad hoc report design interfaces, OLAP analysis, an ETL tool for data integration, and a Java reporting library for use with either stand-alone or embedded business intelligence applications. Jaspersoft is based in San Francisco and is backed by leading venture capital firms Morgenthaler Ventures, DCM, Partech International, Scale Venture Partners and SAP Ventures. More information is available at www.jaspersoft.com and www.jasperforge.org.