



Technology Report

# **Open Source in the Business Intelligence Market**

**Crossing the Threshold to Mainstream Adoption**

**August 2009**  
**Mark R. Madsen**

Sponsored by: Infobright  
Jaspersoft

**TABLE OF CONTENTS**

About the Survey .....	1
Methodology.....	1
Respondent Profile .....	1
Executive Summary.....	2
Introduction .....	3
About Open Source.....	3
Community Versus Commercial Open Source.....	3
Open Source Offerings in the BI and Data Warehouse Market .....	4
Detailed Findings.....	5
What is the Status of Open Source Adoption in This Market?.....	5
Organization Size and Use of Open Source .....	5
Scope of Use.....	5
Size of End-User Deployments.....	6
Use by Consultants and Systems Integrators .....	7
Rationale and Benefits .....	7
What are Organizations Buying? .....	8
What Categories of Software are People Using?.....	9
How Much Data Are People Accessing? .....	11
Failed Software Evaluation and Reasons for Not Using .....	12
Profile of Survey Participants: Who is Using Open Source?.....	14
Recommendations .....	16

## About the Survey

The paper presents results of a survey on the adoption and use of open source software to deliver reporting and analytic capabilities, covering all parts of the stack from the database to end-user delivery. It is written for business and technical managers who are responsible for delivering reporting, business intelligence or analytics, whether embedded in applications, web sites or as part of a BI or data warehouse program.

This research evaluated the needs, rationale and benefits that are driving demand for and use of open source as an alternative to the traditional vendors in this market. We also looked at the types of software, the scope and status of its deployment, and some of the challenges faced by early adopters.

We conducted this survey because the business intelligence, reporting and analytics market has different drivers and requirements. Most open source studies target operating systems, development tools and application infrastructure. As the BI and data warehouse market is different, we wanted a better picture of the factors influencing IT adoption in this segment.

## Methodology

The research for this paper is based on interviews with IT managers and developers and a survey Third Nature conducted between April and June of 2009. We solicited survey participation via four open source vendors' email distribution lists and websites, several blog networks, the annual MySQL conference and the open source channel at the Business Intelligence Network. More than 200 people completed the survey, although not all respondents answered every question.

Our aim was to get a broad perspective of the perceptions, use and non-use of open source in both open source centric communities and in the broader IT market.

## Respondent Profile

The majority of survey respondents are corporate IT professionals across all sizes of company, with the next largest group being consultants. The composition or roles is shown in Figure 1. Most respondents are in North America and Europe, with 47 countries represented in the sample. Software and internet companies are the largest industry represented at 23% of the total, with the rest spread across 26 other industries.

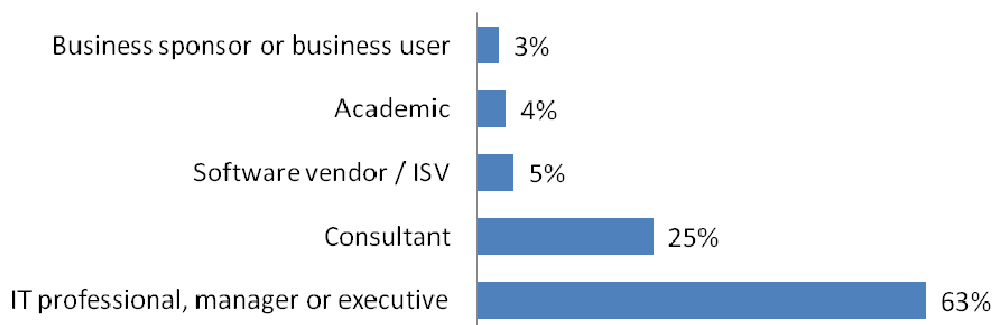


Figure 1: Roles of survey respondents.

## Executive Summary

Open source rose quickly in the data warehousing field, from almost nothing a few years ago to community and commercially-supported projects for every conceivable information delivery use or element of infrastructure. Open source has become mainstream for technical infrastructure, so it is only natural that it move from this into applications and tools.

This paper is based on a survey that looked at adoption and use of open source in the business intelligence and data warehouse market. The goal of the survey is to learn about three elements of adoption:

- What organizations are using open source in this segment of the market?
- What types of software are being deployed?
- What are the benefits and challenges?

The survey found that interest and adoption are widespread, with no specific industry having a clear lead in open source BI or data warehouse use. More than one third of respondents reporting running open source reporting, data integration or database software for analytic uses. Another third are considering or evaluating open source alternatives today. Only 8% reported no interest in or plans to use open source in this market segment.

The top reason for adopting is still cost savings, although reduced vendor dependence and ease of integration were close to the same level. The limiting of vendor technology lock-in and freedom from deployment restrictions were key elements of reducing vendor dependence. Some companies used open source deployments as a means of keeping their incumbent vendors honest.

The primary complaints about open source software or vendors are related to maturity of the software, availability of consulting services and the availability of support for community-based projects.

Open source vendors in the BI and DW market is still less mature than their proprietary counterparts. This is largely due to the relative youth of the projects and companies. In some cases, the maturity complaint was due to enterprise requirements being placed on software that was created for developer use.

Given the rapid pace of innovation in open source projects, they are quickly closing the gap in both core features and maturity between themselves and traditional vendors.

Complaints regarding open source licenses and incompatibility with corporate standards and procurement practices declined from a few years ago, showing that companies are comfortable with the ideas behind open source today.

Roughly one third of open source users are purchasing services and support from open source vendors today. Based on this pattern in what is largely an early adopter segment, expect the commercial open source vendors to continue their growth.

## Introduction

### About Open Source

Open source software (OSS) is released under a license that differs from the traditional software license. The license guarantees several freedoms: access to the source code, the ability to the code with others, and the freedom to modify or deploy it as you wish.

One misconception is that you must share any customizing or modification you make. The requirement to share only applies if you give the software to others outside your organization. If you redistribute the software, then any changes or additions you made must be provided as source code. If you don't redistribute, you are not forced to share your work.

The software is available as a project which is maintained by a community of people who write the code and documentation, provide QA and help to manage distribution. Depending on the project, these people may be independent or they may all work for a software company maintaining the software.

Vendors of open source use this license to enable a means of software production and distribution which provides cost reductions and other benefits back to the vendor.

### Community Versus Commercial Open Source

There are two open source models for sharing development and distribution costs. One is community-based open source, often call "free and open source" or FOSS for short. The other is commercial open source software, usually abbreviated as COSS.

Most people are familiar with the free and open source model because it's been given the most press. In this model volunteers contribute their efforts to development and maintenance. In some cases they may be full-time employees of a non-profit organization owning the software, but the project does not operate like a traditional software company.

Many of the myths about commercial use of open source stem from the model's origins in the shareware and FOSS world. In the early days open source was often designed and built for personal use rather than as part of an organization's IT infrastructure.

Most people are familiar with the ideas of the FOSS model, so they are applied to the COSS model, which came about for a different reason. Commercial open source vendors aim to make money by filling the gaps in the FOSS model.

Enterprise use of open source was challenging in the early days because the focus on personal use meant the products were less polished. Documentation in FOSS projects is often weak, quality varies, and regular fixes, software releases and support are often lacking.

Commercial open source evolved with recognition that companies are willing to pay for support, service, and other less tangible items like indemnification or certifying interoperability. A commercial open source vendor operates just like a traditional software vendor, except that the source code is not shrouded in secrecy. This enables

more and deeper interaction between the community of customers and developers, making the open source model more user-focused than the traditional model.

In contrast to the majority of FOSS projects, commercial open source vendors employ most of the core developers and expect to make a profit while doing so. They provide the same services and support that traditional vendors do, and frequently with more flexibility and lower cost. COSS vendors use elements of the proprietary model such as providing support contracts or selling non-open source components that can be purchased in addition to, or in place of, the free version of the software.

This split between community and paid versions causes confusion because some companies release binaries not covered under an open source license yet call the software open source. If there is no source code but the software executable is provided free then they are really offering a free trial version. Unless the vendor delivers software with source code that is under an OSI-certified license, it is not open source.

### Open Source Offerings in the BI and Data Warehouse Market

There are open source projects available for every element of the business intelligence and data warehouse stack. This includes core products like relational databases as well as more modern analytic database platforms.

The availability of advanced technology like analytic databases surprises many because the assumption has been that open source only applies to the commodity market rather than emerging technologies. COSS changes the dynamics of software development and brings these technologies to market as open source before they have a chance to become a commodity.

Software is available for every type of user-facing application, from traditional reporting and OLAP tools to advanced data mining and statistics software, as well as more esoteric items like data visualization and web-based geographic information systems.

Data integration software is a more recent entrant in the developer tools market, with several data integration, ETL, data quality and federation options available. These are being applied equally to analytic and transaction processing applications.

Regardless of what you are seeking, it's likely that there is an open source project to fill that need. Enterprise caliber software is more readily available. In the case of core software like databases, BI and ETL tools.

Software developed as open source is no different from traditional commercial software. The difference lies in a license that gives you more freedom with the code than a proprietary license. This means you should evaluate it as you would any other software, by asking whether it meets your requirements at a reasonable price.

## Detailed Findings

### What is the Status of Open Source Adoption in This Market?

#### Organization Size and Use of Open Source

One persistent myth is that small companies are the primary users of open source. While there are more small organizations evaluating and using open source than mid-sized or large, as show in Figure 2, the data also shows that both small and large organizations are leading adoption over the mid-sized organizations.

There are clear signs that the momentum for smaller organizations is going to continue. As we'll see later, this is a challenge for commercial open source vendors because the biggest adopting category is also the least likely to pay for support or services.

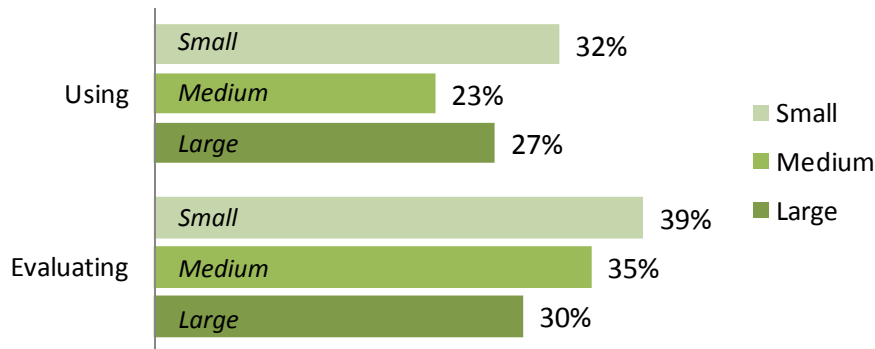


Figure 2: Open source use and evaluation by size of organization.

#### Scope of Use

One common belief about analytics and BI projects is that open source is more likely to be used by departments in large organizations and across the company in smaller organizations.

Figure3 shows that small organizations are more likely than medium and large to do company-wide deployments, and large organizations are doing smaller deployments, supporting this belief.

Small companies and departments of large organizations share similar characteristics: they are often constrained by budget, they have a smaller user base and their usage is more uniform, making smaller deployments easier.

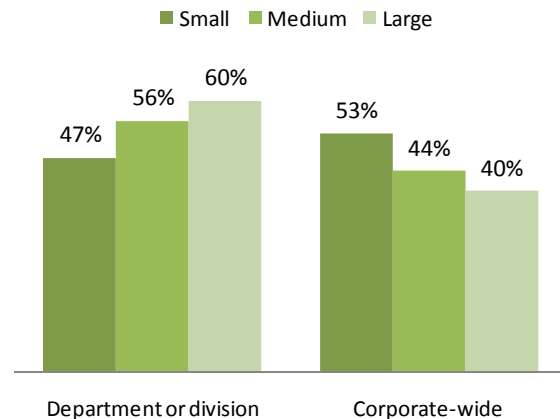


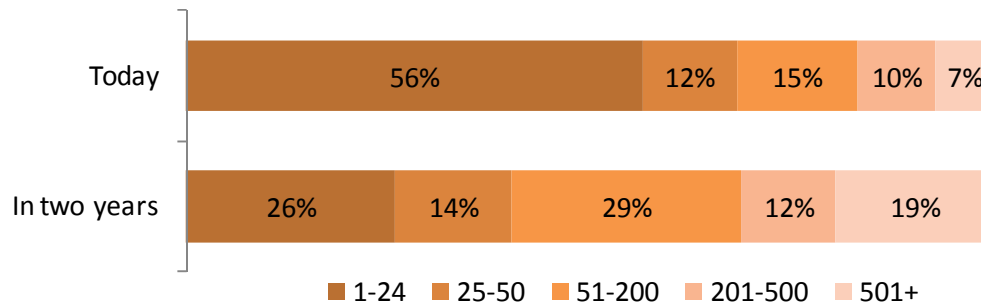
Figure 3: Scope of open source use.

Despite this general pattern, there are enterprise-wide deployments of open source in large organizations. 40% of large organizations plan to or have deployed a BI or DW application corporate-wide with some open source components, demonstrating a level of software maturity.

### Size of End-User Deployments

The number of users in environments with open source is similar to what is reported for actual usage in proprietary data warehouse environments. Actual usage in these environments can be significantly less than the number of licenses purchased.

Open source has an advantage over proprietary solutions regarding scope because there is more deployment flexibility. People often buy more software than they need from traditional vendors because the high license cost makes obtaining funds for more seats a challenge, and because of the way software is discounted in volume purchases. Open source can be less expensive, and FOSS versions carry no penalty for increasing or decreasing use. Figure 4 shows the numbers of users reported in the survey.



**Figure 4: Number of users of the system today and expected in two years**

The numbers of users also tells us something about what stage organizations are at in their production deployments of open source. Most are still in the early phases of deployment, with 56% rolling out to less than 25 users.

This low number of users does not mirror the size distribution of the organizations. Many companies also reported division (21%) or corporate-wide (47%) deployments which would imply larger numbers of users. Contact with respondents showed that many were in the pilot phase of their projects and were planning to add users in the near future, accounting for their low user counts.

Examining the number of users by size of organization reveals that small and medium institutions expect to deploy in the 50-200 user range in the future (14% and 17% respectively), while large organizations are much more likely to plan in excess of 500 users (34%).

The increase in number of users of the systems in two years suggests that scalability on the user concurrency axis will be a bigger requirement than it is today for open source vendors who provide business intelligence, reporting or database software.

### Use by Consultants and Systems Integrators

Another common belief is that consulting firms and systems integrators are more likely to use open source because it allows them to be more competitive by saving customers money on software as well as freeing budget that could be used for services. The survey data disproves that belief.

Some interesting patterns emerge when we look at the details of open source use. Consultants are generally less likely than IT professionals to be using open source tools in this space (10% for consultants to 36% for IT). The usage by respondent role is shown in Figure 5.

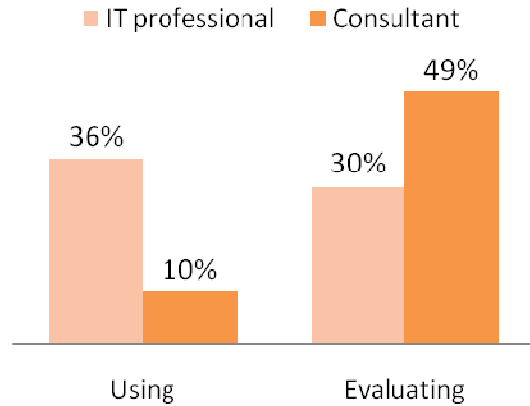


Figure 5: Open source use by IT and consultants

It is notable that 49% of the consultants and systems integrators are evaluating open source software today, signaling a possible shift in their use.

Even with the sudden rise in evaluation, consultants and SIs significantly trail IT departments. If you are in an IT organization relying on consulting services then this is an important element to factor in to consideration.

### Rationale and Benefits

The reasons for using open source in this market are not different from those in the infrastructure and development tools markets, with cost savings as the number one stated reason. Fortunately for the companies involved, cost savings is also the number one reported benefit after deployment. Figure 6 shows the top reasons given for exploring the use of open source and the benefits reported after deployment.

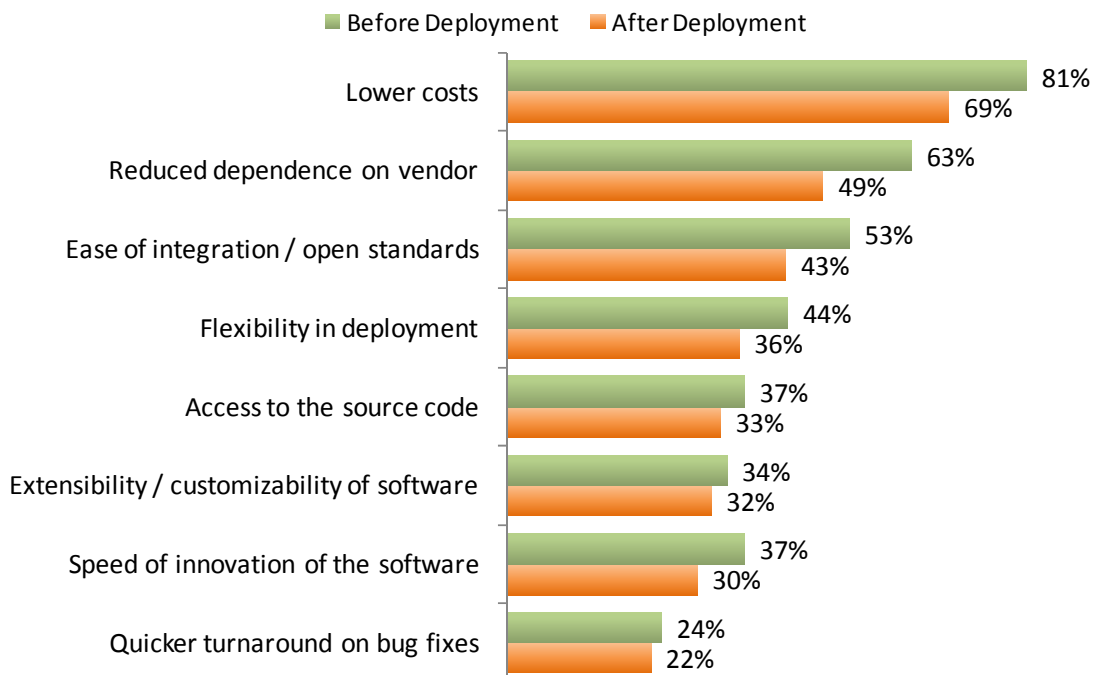


Figure 6: Top 10 reasons for wanting to use open source software.

Reduced vendor dependence is surprisingly high in the list of reasons and benefits. The benefits anticipated are more than the obvious avoiding of a vendor's technology lock-in, for example the requirement that one run Windows and SQLServer in order to use Microsoft's BI tools. Also mentioned were more options to resolve problems, community support reducing the requirement for vendor aid, and using open source to offset vendor acquisitions.

The business intelligence and data warehousing market has seen several years of steady consolidation across all software categories. This consolidation makes it increasingly likely that a formerly multi-vendor installation is now entirely dependent on a single vendor. Many managers view having all of their technology decisions in the hands of a single vendor as a risk.

In light of recent price increases and restrictions imposed by vendors, using open source is proving to be a way to reduce dependence and balance the risk of more vendor acquisitions or unilateral actions like raising prices on a captive base.

Other advantages of open source software are the ease of adjusting deployment and customizing or extending it to fit specific project circumstances. Neither of these is as simple with traditional software because of licensing and code access restrictions.

Across all of the items listed, the percentage of respondents reporting benefits relative to the reasons for using open source dropped. This indicates that some expected benefits did not materialize. This is not a surprise since people often underestimate the effort or overestimate the benefits of software projects.

One caveat to the reported benefits is that most organizations do not have formal mechanisms in place to track returns, financial or otherwise, once a system has gone into production. This is no different for open source, so measurement is based on perception rather than reliable metrics.

### What are Organizations Buying?

Confirming a commonly held belief, small companies are the least likely to purchase anything from an open source vendor. They prefer the pure FOSS option. This is not a surprise since they generally have the lowest ability to pay, even at commercial open source prices.

Mid-sized and large companies are more than twice as likely to purchase subscriptions. 32% and 31% purchased subscriptions respectively, compared to just 12% of smaller organizations.

Half of all respondents bought nothing, preferring the FOSS versions of the software and relying on internal resources or the community for support. The purchase data is shown in Figure 7.

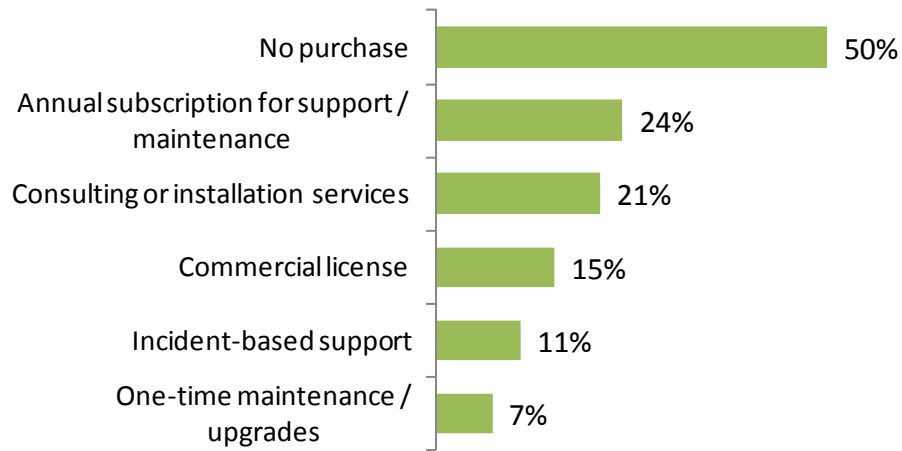


Figure 7: Purchases related to open source

### What Categories of Software are People Using?

Use is distributed across all categories and follows a pattern aligned with maturity of the software and the most common uses. Fewer people have a need for advanced analytics or embedding reports into an application so we should expect these to show somewhat lower use, evaluations and interest.

For this survey we divided software into categories based on the layers in the traditional business intelligence or data warehouse stack: database or warehouse platform, data integration, and information delivery tools. Because there are so many different front-end tools, we divided them into reporting and OLAP, embedded and application reporting, and advanced analytics. The latter comprises data mining, statistics, visualization and GIS software. Software use by category is shown in figure 8.

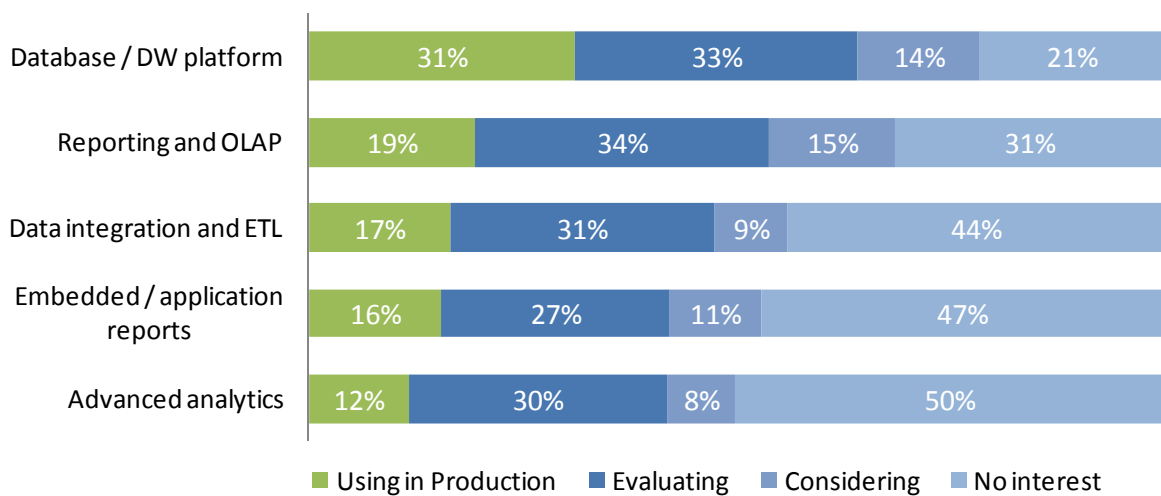
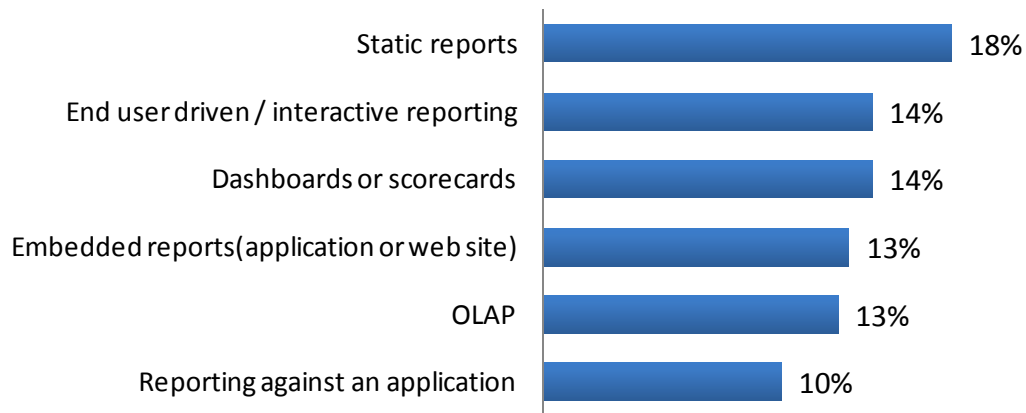


Figure 8: Interest and use of open source by software type.

Production use of open source databases for BI is lower than one would expect, even though they show the greatest use in this market. Open source databases have been around for many years, while many of the most prominent projects in other categories are less than five years old. Given the youth of these projects, it's surprising that they are seeing such a rapid uptake.

One source of problems holding back open source databases is the nature of analytic workloads. Most engineering effort for OSS databases focused on transaction processing and horizontal scaling. Analytic use requires better handling of complex queries, large single query data volumes, and lower user concurrency. Database use would be higher if it weren't for poor query performance and analytic support.

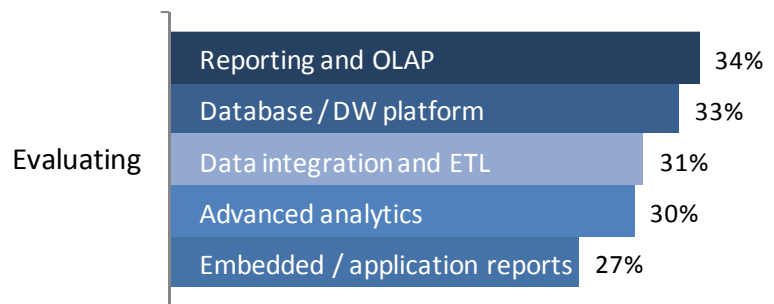
A detailed breakdown of use in the reporting tools space is outlined in Figure 9. What we see is that static reports are the most common, with interactive reporting and dashboards tied for second. This mirrors what we seen in the traditional BI market.



**Figure 9: Breakdown of production use for the BI and reporting tools category.**

The survey asked about evaluation and interest to understand where the momentum is in the market. Evaluations are the leading indicator of what will be installed in the future. About one third of the respondents reported active evaluations across any given software category, as shown in Figure 10.

Evaluations and interest are highest in the reporting and OLAP space. This is logical since many companies have data management and integration infrastructure in place, but are less likely to have reporting and analysis tools. This is particularly true of smaller and mid-sized organizations where these products have frequently been unavailable due to the high cost of enterprise business intelligence products.

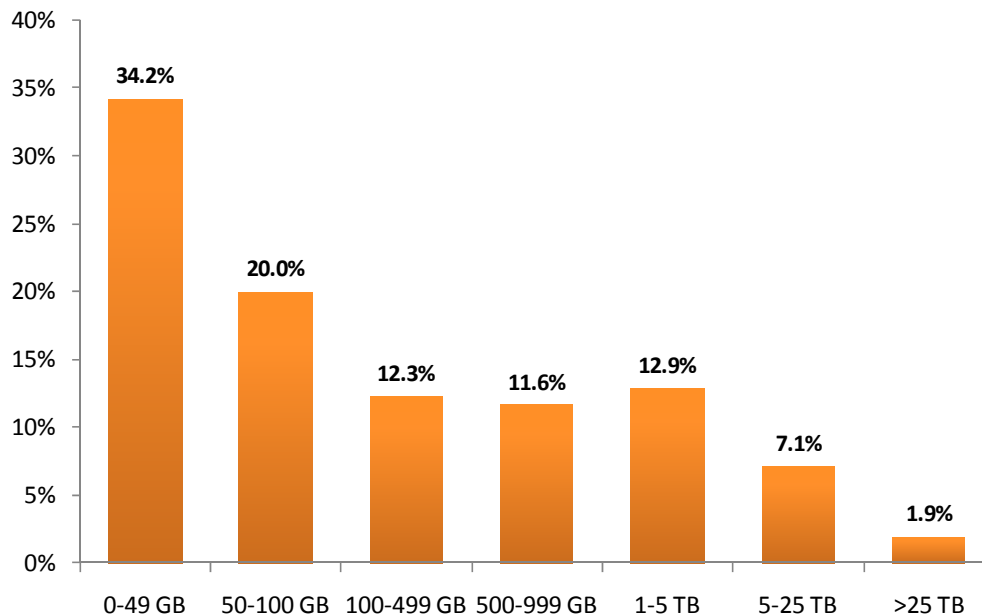


**Figure 10: Respondents evaluating by software category.**

Interest in all BI categories is strong and growing. One element gleaned from interviews is that experience with open source often leads to increased adoption of related open source tools. This means we should expect to see more open source in the BI and DW stack as more companies gain experience.

### How Much Data Are People Accessing?

The chart in Figure 11 shows the responses to the question "How much raw data is being stored or accessed?" This chart shows the size regardless of database type - the databases in use are not all open source. The constraint is that people are using open source in some part of the data warehouse stack, so an open source BI tool accessing a proprietary database would be included. Even so, 23% of respondents are using open source databases like MySQL, Postgres and Ingres.



**Figure 11: Data volume distribution (includes open source and proprietary databases).**

The distribution is the same as what we see in the overall data warehouse market where the estimate is that approximately 65% of databases are less than one terabyte in size. This survey shows most (78%) of respondents are accessing less than a terabyte of data and 54% less than 100GB.

Open source data warehouses are smaller in size than they are in proprietary database market. In this sample, 90% of the open source databases running in production are less than a terabyte. Even so, 27% are more than 500 GB in size, so the sizes are not insignificant.

Data size is a critical factor affecting performance, but query response time at the sub-terabyte scale is a challenge for many organizations. The difficulty of getting good query performance is one of the major factors driving people to look at columnar databases and other specialized data warehouse platforms. This is particularly true if they are using standard FOSS databases like MySQL because of the lack of data warehouse-specific scaling and performance features.

## Failed Software Evaluation and Reasons for Not Using

While most respondents provided favorable information regarding open source, it is not without problems. Figure 12 shows that 47% of respondents reported at least one failed open source evaluation.



Figure 12: Responses to " Did any open source software fail your evaluation?"

While there were many different reasons cited for the failed evaluations, reasons clustered around several key issues shown in Figure 13.

Software maturity is the primary limitation. This includes problems like scalability and reliability, with "missing required features " at the top the list. The single biggest write-in complaint was poor documentation, something FOSS projects often struggle with. This is one of the gaps COSS vendors are trying to fill to make the software more enterprise-friendly.

Market issues include problems finding software or support, and lack of consulting, training or other services.

As open source grows in popularity the complaints about availability of services should diminish. It also highlights the fact that companies are looking for services, particularly consulting, and not finding them.

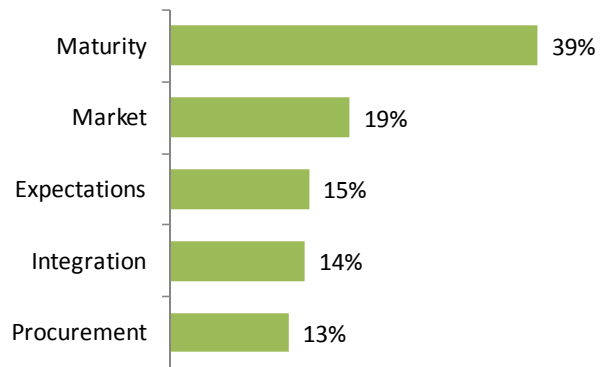


Figure 13: Reasons open source software failed evaluations.

While maturity of the software is the top rated complaint regardless of organization size, the size matters in other areas. The biggest disparity between small, medium and large organizations is complaints regarding market issues like service availability or challenges finding software.

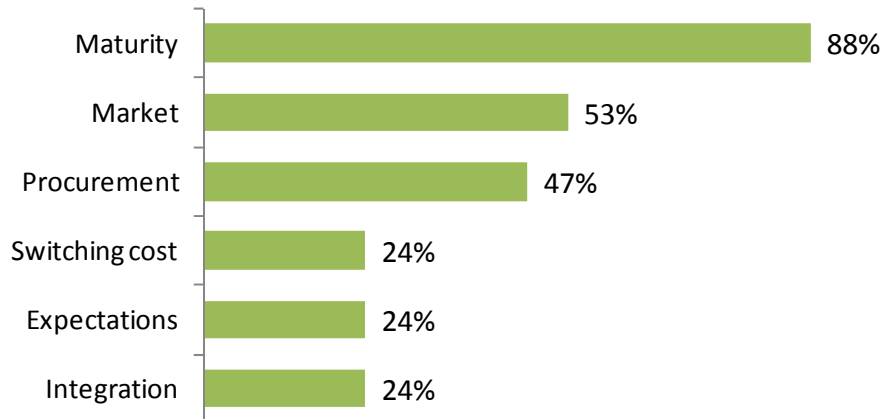
Mid-sized companies are more concerned about the availability of consulting than their larger or smaller peers. They are also less likely to have trouble finding software to meet their needs. This is likely because they don't need the out of box simplicity required by smaller companies, nor the features and flexibility required by larger companies.

Complaints about lack of service and support are directly related to the size of an organization, with large organizations least likely to mention and small organization most likely. This may be because the larger an organization is, the more likely they are to have staff with appropriate skills in house.

In general, procurement issues (corporate standards, IT resistance, license and legal difficulties) are less of a problem today as organizations have become more familiar with open source. Another factor which reduces this complain is that many companies (39%) are choosing to purchase commercially licensed or subscription versions.

9% of organizations reported that they are not using open source for any area of data warehousing. We can't draw firm conclusions from their responses, just indicators of what the reasons are.

Of the organizations that chose to not use any open source, maturity of the software was the biggest roadblock at 88%, followed by market issues like lack of service or support. Figure 14 lists reasons people gave for not using open source.



**Figure 14: Reasons for not using any open source in this market segment.**

One common write-in answer was switching costs. For companies with an existing solution in place, the cost savings on licenses or support did not exceed the estimated costs duplicate reports or rebuild ELT jobs in new software. There was no compelling reason to change given equivalent features between OSS and proprietary software.

## Profile of Survey Participants: Who is Using Open Source?

Industry representation is broad, with 204 respondents from 26 different industries. The list of industries and respondent distribution is shown in Figure 15.

Internet and software companies make up the largest portion at 23% of the total. Financial services is the next largest at 9% which is not a surprise - financial services has been an early adopter of other categories of open source software.

Software companies are adding reporting and business intelligence features to their packages either because it is expected by new customers or to add differentiation. Interviews provide anecdotal evidence that these companies are turning to open source because of lower embedding costs, but more importantly because it is easier to incorporate than traditionally licensed software.

Several companies mentioned that they were able to add missing features to their software and turn the code over to the vendor for inclusion in the open source package, thus saving the vendor from having to maintain the code. This is far less likely to happen with traditional vendors.

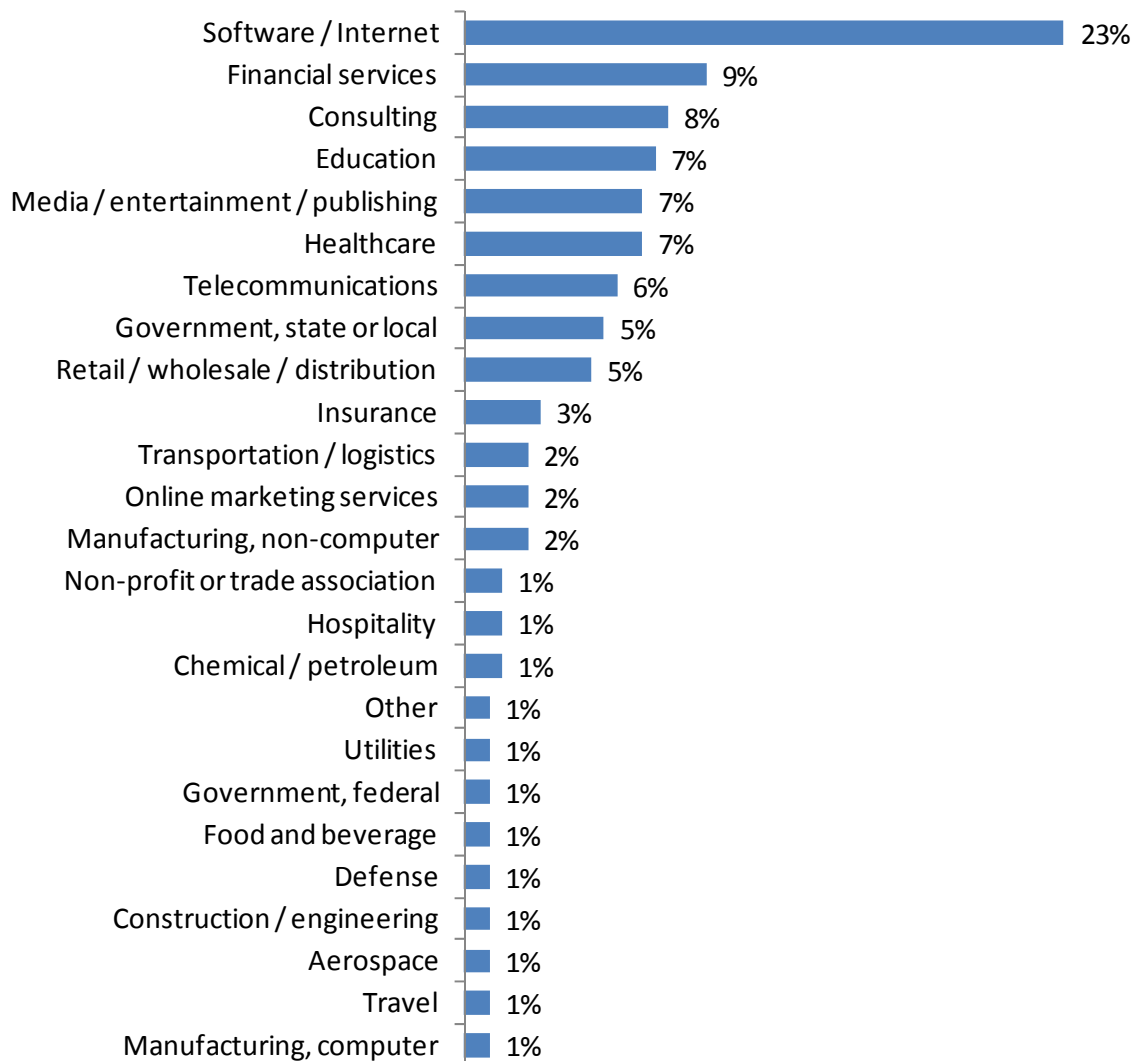


Figure 15: Industries and distribution of survey respondents.

The bulk of the respondents are from the US and Europe, representing a total of 47 different countries worldwide. The distribution across regions is shown in Figure 16. This reflects a similar distribution in other areas of open source software

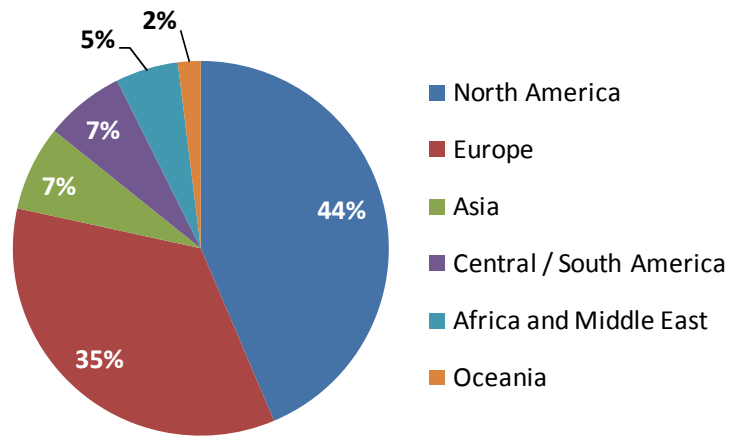


Figure 16: Geographic distribution of survey respondents.

The size of organizations represented in the survey is widely distributed, from Global 1000 to very small companies.

The size of organizations can be measured by revenue or number of employees. Rather than use company revenue, this report uses employee count for the size metric because it is applicable across both private, public and non-profit institutions, and because it is a more reliable gauge of the scale of an organization than the revenues.

For this survey, small organizations are considered to be those with less than 100 employees, mid-sized organizations are between 100 and 2,000 employees, and large organizations are those with more than 2,000 employees.

The size distribution of organizations in this survey is shown in Figure 17 by both revenues and employee count.

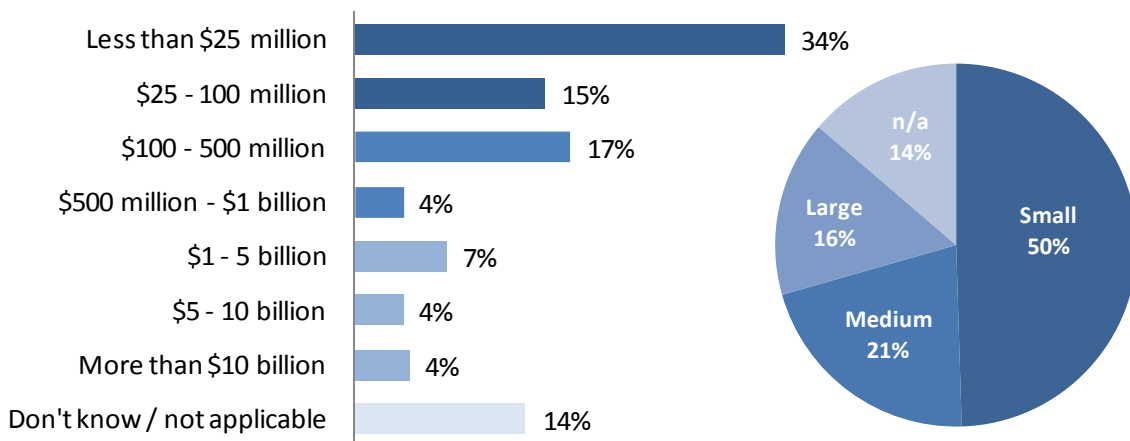


Figure 17: Revenue and employee size of organizations in the survey sample.

## Recommendations

Open source in the business intelligence and data warehousing field is beyond the early adoption stage but there are still some challenges. People responsible for evaluating BI and data warehousing tools can benefit from the following guidelines.

- **Don't focus solely on cost savings.** While cost is important, it's only one factor. The other top-ranked benefits are reduced dependence on vendors and ease of integration. If you are building a case to justify open source use, it will help to include other factors such as risk reduction and deployment flexibility. People often did not mention these as reasons for consideration, but as benefits they discovered later.
- **Make open source the default option.** When in an environment with few or no tools, open source should be the preferred alternative. It is the simplest, fastest and likely the least expensive route when compared to hand-coding or proprietary products. Look to proprietary tools when open source tools can't do the job, or when you have products in house already and expanding licenses is not as big an obstruction.
- **Plan to augment, not replace, existing software with open source.** One of the obstacles encountered while evaluating open source BI and ETL tools was the high cost of redeveloping reports or integration jobs. Rather than look at saving money by replacing software, look at gaps in the BI portfolio or data warehouse stack and use open source to supplement your systems.
- **Consider developing open source policies.** Most organizations are adopting open source in an ad-hoc fashion, project by project. While this works, it can also reduce cost savings by duplicating effort. Open source can bypass the procurement process, leading to situations where departments deploy their own tools, unaware that someone else has already done evaluations or deployed different software. There are also some differences with open source licenses that can put your organization at risk if your legal department hasn't done proper review.
- **Evaluate open source like any other software.** It doesn't matter if the software is free if it takes longer to build, manage and deploy solutions to end users, if it is unstable, or if it is missing a key feature. Open source is still software, and should be evaluated against the same set of requirements you would use with any other application. Keep in mind that while it may not always be as feature rich as traditional software, there are other potential benefits to quantify like time to market, deployment flexibility and customizability which are due to the nature of open source.

Open source use is growing in the BI and data warehousing field. As we move into the early mainstream stage, the software will become more polished and lose its rough edges. We're already seeing signs of a shift to mainstream adoption as consulting companies and systems integrators begin to evaluate open source for themselves and their clients.

## About the Sponsors



**Infobright**, the open source data warehousing company, delivers a self-managing, low cost analytic database that enables companies with limited staff and budget to implement a highly scalable data warehouse. Infobright combines a column-oriented database with unique Knowledge Grid technology that reduces ongoing administration by up to 90 percent and cost by more than 50%. Infobright is being used by companies in online marketing, financial services, telecommunications and other industries to provide rapid access to critical business data with unmatched operational simplicity. For more information, please visit [www.infobright.com](http://www.infobright.com) and <http://www.infobright.org>.



**Jaspersoft's [open source business intelligence suite](http://www.jaspersoft.com)** is the world's most widely used BI software, with nearly 9 million total downloads worldwide and more than 10,000 commercial customers in 96 countries. The company's Jaspersoft Business Intelligence Suite provides a web-based, open and modular approach to the evolving business intelligence needs of the enterprise. Jaspersoft's software is rapidly updated by a community of more than 100,000 registered members working on more than 350 projects, which represents the world's largest business intelligence community. More information is available at [www.jaspersoft.com](http://www.jaspersoft.com) and [www.jasperforge.org](http://www.jasperforge.org).

## About the Author

**MARK MADSEN** is president of Third Nature, a consulting and technology research firm focused on information management. Mark is an award-winning architect and former CTO whose work has been featured in numerous industry publications. He is an international speaker, a contributing editor at Intelligent Enterprise, and manages the open source channel at the Business Intelligence Network. For more information or to contact Mark, visit <http://ThirdNature.net>.



Third Nature is a research and consulting firm focused on new practices and emerging technology for business intelligence, data integration and information management.

Our goal is to help companies learn how to take advantage of new information-driven management practices and applications. We offer consulting, education and research services to support business and IT organizations as well as technology vendors.