In-Memory Analysis

Delivering Insights at the Speed of Thought

BY WAYNE ECKERSON

Director of Research, Business Applications and Architecture Group, TechTarget, December 2011

Research Background

THE PURPOSE OF THIS REPORT is to profile the capabilities of next-generation business intelligence (BI) tools with emphasis on new visual analysis tools and in-memory processing. It examines the types of BI users and capabilities they need, including casual users, power users and IT administrators. It also examines BI architectures—how front-end tools interact with back-end servers and databases—to deliver those capabilities.

The research is based on my knowledge of the BI market, interviews with BI practitioners and briefings with sponsors of this report. The research is also based on a survey of 240 BI professionals. The 10-question survey was promoted to the BI Leadership Forum, an online group of about 700 BI directors and managers and my Twitter followers (2,000-plus people) during October.

More than 240 people started the survey and 198 completed it. The respondents were primarily BI or IT professionals (64%) from large companies with more than $1 billion in annual revenues (48%). However, quite a few consultants completed the survey (21%) and individuals from medium-sized companies with between $50 million and $1 billion in annual revenues (30%) and small companies with less than $50 million in annual revenues (23%) (see **FIGURES 1** and **2**).
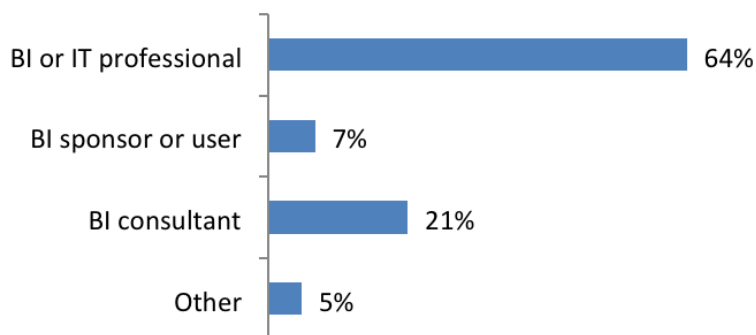

FIGURE 1:
Respondent Profile
Based on 198 respondents, BI Leadership Forum, October 2011 (www.bileadership.com).

FIGURE 2:
Company Size of Survey Respondents
Based on 198 respondents, BI Leadership Forum, October 2011 (www.bileadership.com).
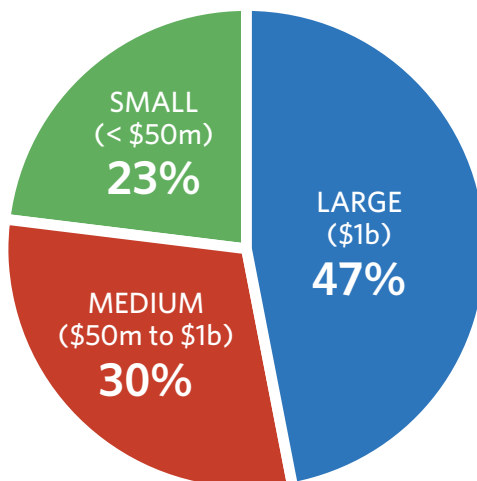
The majority of BI professionals hailed from large companies (56%), with very few (7%) from small companies. This makes sense, since most companies with less than $50 million in revenues don't have enough money to pay for a full-time BI professional, while large companies can support teams of full-time BI professionals. In contrast, business sponsors and users were evenly dispersed across small (29%), medium-sized (28%) and large companies (43%). Consultants mostly came from small companies but consulted with a range of different-sized firms. ■

Executive Summary

ALTHOUGH BI HAS BECOME a mainstream market, innovation has not stopped. In fact, by all accounts, the technology driving BI is changing more rapidly now than any other time in its 20-year history.

This avalanche of innovation is divided between front- and back-end technologies. On the front end, tools are incorporating more visual, interactive interfaces that are making it easier for users to interact with data and create their own views of information. On the back end, advances in memory, CPU and disk technology have enabled BI vendors to exploit in-memory databases and intelligent caches and specialized analytical databases and platforms that offer dramatically improved price-performance over previous generations of database technology. Never before have BI customers had so many options to store, access, analyze and consume information for decision making.

This innovation raises the question—What are the capabilities of next-generation BI tools? Certainly, the business mantra of "faster, better, cheaper" is becoming a reality. But we can also add visual, interactive, analytical, scalable, manageable, collaborative and mobile. Collectively, many of these capabilities get lumped together under the heading "self-service BI," which has been the Holy Grail of BI for nearly two decades. The more end users can interact with the data to create their own views, the more satisfied and productive they'll be with BI tools and the more corporate BI teams can focus on value-added activities, instead of creating an endless stream of custom reports and dashboards.

More specifically, next-generation BI tools blend the capabilities of top-down, metrics-driven reporting with bottom-up, ad hoc analyses seamlessly,

> *Never before have BI customers had so many options to store, access, analyze and consume information for decision making.*

making it easy for users to meet their own information needs once an IT person, power user or superuser has done the initial setup. Top-down tools expose semantic layers and widget libraries built by IT professionals and allow superusers to build ad hoc reports and dashboards (i.e., "mashboards"). Conversely, bottom-up tools, such as popular visual analysis tools, let power users and superusers explore data culled from a variety of back-end systems and build fast, highly interactive dashboards for their departmental colleagues.

Finally, you can't discuss next-generation BI tools without examining their back-end data architectures. To deliver the highest level of query performance possible, many new BI tools store data locally in an in-memory database or intelligent mid-tier cache. Others query back-end databases directly, relying on ROLAP (relational online analytical processing) SQL generation capabilities or the power of analytical platforms to handle complex queries and deliver super fast performance. And some tools give users the flexibility of caching data locally or querying back-end databases, depending on business requirements and systems availability. ∎

> *You can't discuss next-generation BI tools without examining their back-end data architectures.*

BI Mega Trends

BI FRAMEWORK 2020

At a macro level, next-generation BI capabilities are depicted in my BI Framework 2020, which I introduced in my 2011 report titled *Analytic Architectures: Approaches to Supporting Analytics Users and Workloads*. I've updated the framework since its first publication, but the contents remain much the same. The framework depicts four intelligences for delivering reporting and analysis applications: business intelligence, continuous intelligence, analytics intelligence and content intelligence, each of which brings different types of BI tools to the table (see **FIGURE 3**).
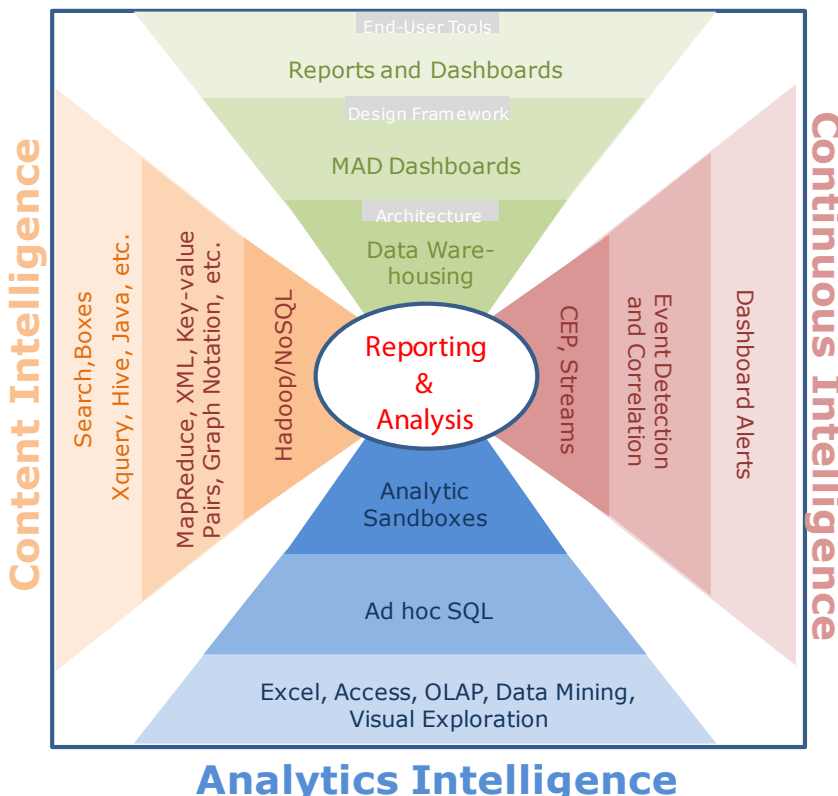

FIGURE 3:
BI Framework 2020

Business intelligence provides historical data to casual users in the form of reports and dashboards built on a data warehousing infrastructure so they can monitor and analyze routine business activity. *Continuous intelligence* accelerates the delivery of information to users, and in some cases, correlates events and triggers alerts when it's time for humans to intervene. *Analytics intelligence* gives power users a variety of desktop analysis tools to explore and analyze data in an unfettered fashion so they can answer unanticipated questions. Finally, *content intelligence* broadens the data sources that both casual and power users can access and analyze to include semi-structured and unstructured data. Casual users will use search tools to access this data, while power users will use programming and scripting languages.

Types of users and activities. There are two overlay dimensions in the BI Framework 2020. One depicts types of users, casual and power. The other depicts types of BI activities, monitoring and exploration. For the most part, these two dimensions mirror each other: Casual users generally want to monitor metrics depicted in dashboards, while power users want to explore data found in any data source. More accurately, the two dimensions have an 80/20 overlap; that is, 80% of the time casual users want to monitor predefined metrics and power users want to explore data, while 20% of the time casual users want to explore data and power users want to monitor metrics. This asymmetry is critical to understanding next-generation BI capabilities.

> *The two dimensions have an 80/20 overlap ...This asymmetry is critical to understanding next-generation BI capabilities.*

The one dimension that the BI Framework 2020 doesn't depict is delivery platforms, which have expanded significantly in the past 20 years, from mainframes and minicomputers to the client-server model, the Web and Web services. Two new important delivery platforms for BI in the coming decade are mobile devices and private and public cloud environments. Both figure in next-generation BI capabilities, although mobile is the most imminent.

REPORTING VERSUS ANALYSIS

The message behind the BI Framework 2020 is that you can't shoehorn all BI activities into a single architecture or tool set. At the highest level, the two

primary BI applications—reporting and analysis—are fundamentally differ-
ent applications with very different types users and unique workloads, design
frameworks and architectures. Not surprisingly, each application has opposite
advantages and disadvantages. Yet, they are interrelated: Analysis leads to
reports and reports trigger new questions that lead to the need for additional
analysis (see **FIGURE 4**).

Top down. Reporting is a "top-down" activity that monitors business activity
using metrics that are aligned with strategic goals and objectives. To design
reports and dashboards—visual exception reports—you need to know in
advance the questions casual users are going to ask, which will differ depend-
ing on their roles in the organization. To create reports, the typical organiza-
tion first builds a data warehouse or data mart that contains a model of how

FIGURE 4:
Reporting Versus Analysis: Distinct Workloads, Users and Architectures

Pros:
- Alignment
- Consistency

Cons:
- Hard to build
- Politically charged
- Hard to change
- Expensive
- "Schema Heavy"

Pros:
- Quick to build
- Politically uncharged
- Easy to change
- Low cost

Cons:
- Alignment
- Consistency
- "Schema Light"
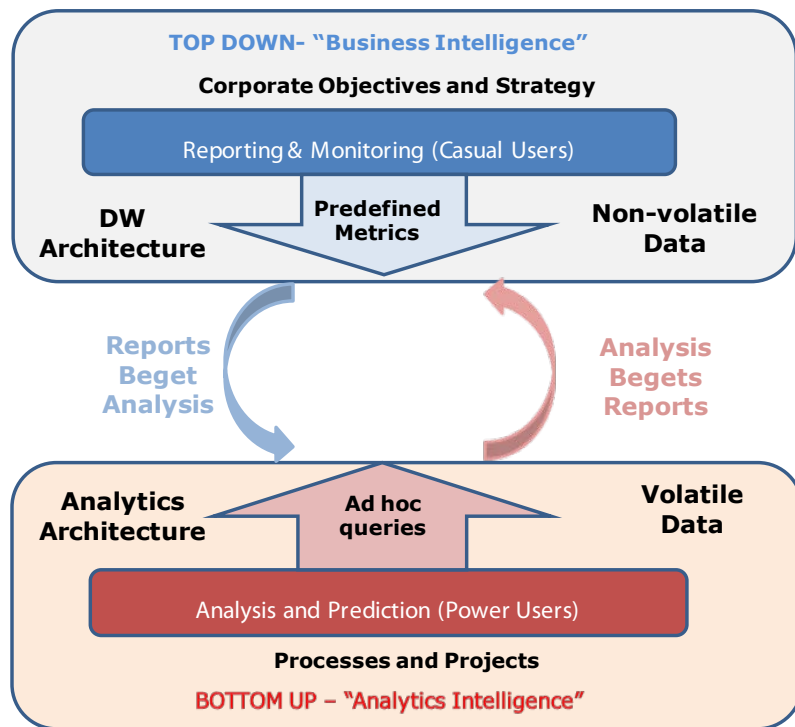
TOP DOWN- "Business Intelligence"
Corporate Objectives and Strategy
Reporting & Monitoring (Casual Users)
DW Architecture
Predefined Metrics
Non-volatile Data

Reports Beget Analysis
Analysis Begets Reports

Analytics Architecture
Ad hoc queries
Volatile Data
Analysis and Prediction (Power Users)
Processes and Projects
BOTTOM UP – "Analytics Intelligence"

its business operates. Thus, reports and dashboards built on a data ware-house take a lot of time and money to create and are hard to change, but they ensure a consistent mapping of enterprise information, which is a key business requirement among executives who want to run their businesses on a single set of numbers.

Bottom up. In contrast, analysis is a "bottom-up" activity in which analytical experts (a.k.a. business analysts or power users) use a variety of tools to answer unanticipated questions from business managers and executives who need to create or refine a strategy. To answer such questions, power users often need to access a variety of data sources, explore and merge the data, analyze the results and present their findings in a concise and comprehensive way to business executives and managers. This type of exploratory analysis is great for answering unanticipated business questions but is a highly inefficient use of analysts' time and often undermines information consistency since each power user defines business rules independently.

SYNERGISTIC COMBINATION
Misplaced expectations. The challenges that most organizations experience with BI tools often have less to do with vendor products than with customer expectations about the products. Reporting and dashboard tools are primarily designed to monitor predefined metrics aligned with role-based objectives, not explore and analyze data. Customers that expect reporting and dashboard tools to support ad hoc analysis will become frustrated. Conversely, companies that expect analysis tools to deliver pixel-perfect reports or enforce business rules across the enterprise will be sorely disappointed.

Recognize that you need both top-down and bottom-up BI tools and that these tools need to work together, not against each other.

As a result of misplaced expectations, many organizations flip-flop between top-down and bottom-up BI initiatives. When data warehouse-driven reports and dashboards become too slow and costly to deliver, they embrace light-weight analysis tools. When the bottom-up tools hit the wall because they can't scale or sort out complex and dirty source data, they revert to enterprise tools rather than buy lots of additional hardware and write tons of script to make the tools work.

The solution to this dynamic is to recognize that you need both top-down and bottom-up BI tools, and that these tools need to work together, not against each other. Buying reporting and analysis tools from a single enterprise BI vendor does not guarantee the tools will interoperate, but it is a step in the right direction. Also, it is important to establish BI tool standards for each type of user in your organization. At the highest level, casual users should get one type of tool and power users another, although casual users may end up using the Web-based published output of the power-user tool. At a finer-grained level, there are many different types of casual and power users, and you may need a tool standard for each type.

Multiple modalities. Muddying up the waters even more, some users play multiple roles and thus should have different types of tools for each role. Unfortunately, few BI sponsors want to purchase multiple BI tools, and most casual users don't want to have to learn to use multiple tools. So we're back to the quest for a single BI tool that offers multiple modalities geared to different types of BI activity. We are starting to see tools that offer such functionality.

For example, a prototypical casual user spends 80% of his time viewing, navigating and interacting with data but then adds a new metric, dimension, attribute or prompt that is not in the data set. The user crosses the 80/20 line and becomes a power user who wants to create new data. Most BI products don't handle this pivot point gracefully, but some are starting to provide optional "do more" modalities that expose new functionality on the fly, when users need it and are ready for it.

Summary. The BI Framework 2020 provides a macro view of the BI environment of the future. It shows that there is no one BI environment that can adequately support the various BI activities involved in building and using reporting and analysis applications. Rather than search for a single BI architecture, tool set and design environment, BI professionals need to think about creating a BI ecosystem in which multiple reporting and analysis environments interoperate to help businesspeople use information to make smarter decisions. The BI team may not oversee all these distinct environments, but they need to be aware that they exist. They must work to propagate standards for shared information entities used across these environments and eliminate overlaps and inefficiencies in the information delivery workflow. ■

Next-Generation BI Capabilities

A NEXT-GENERATION BI TOOL tries to blend the best of top-down and bottom-up BI while minimizing the downsides of each approach. Reconciling these contradictory requirements is challenging. However, many BI vendors are making progress toward bridging the gulf.

Top-down capabilities. Top-down BI tools, which often are referred to as "enterprise BI" tools, address the needs of casual users. IT developers use these tools to create metrics-driven reports and dashboards that deliver a single version of truth in each business domain. Superusers use the tools to create ad hoc reports and dashboards around the edges of the standard departmental and enterprise reports.

Top-down BI is more IT-oriented because standard reports and dashboards normally run against a data warehouse or data mart via a semantic layer that converts back-end data elements into business-friendly query objects. The hallmark of a top-down BI tool is that it scales to the enterprise, delivers clean, consistent, accurate data and displays metrics tailored to each individual's role in the organization.

Bottom-up capabilities. In contrast, bottom-up BI meets the needs of power users by supporting ad hoc analyses against any structured data source, including local files and external data sets. As such, bottom-up BI is more business-oriented, because the pace and volatility of many businesses today require an ad hoc approach to asking questions and getting answers.

Bottom-up BI tools, which vendors claim deliver "agile BI," try to deliver "faster, better, cheaper" BI. These tools are purchased by department heads frustrated by the IT backlog and unwilling to wait for months or years for a custom dashboard or data-centric solution. Many of these tools originated as desktop tools for power users that can be used to create interactive, departmental dashboards for casual users. They are fast, highly visual, interactive,

analytical and low-cost (at least to start). Many use in-memory databases to speed query performance.

Today, given the challenges of melding these two environments, many user organizations are simply standardizing on a top-down tool for reporting and a bottom-up tool for ad hoc analysis. The problem here is that there is little interoperability between these two tools unless they come from the same vendor (and even then, interoperability is not guaranteed). So, although standardization is a reasonable option for blending the two worlds today, the race is on to see which category of BI tools can subsume the other's capabilities.

The race is on to see which category of BI tools can subsume the other's capabilities.

END-USER CHARACTERISTICS

If you are looking for a next-generation BI tool, here are some of the key end-user characteristics the tool should support:

Self-service. Next-generation BI tools need to offer more than a static report or dashboard. They need to enable users to change what they see on the screen without IT or power-user involvement. This ranges from giving users the ability to navigate to more detailed views of data to letting them create new reports and dashboards from scratch. In many respects, self-service is an umbrella term that represents nearly all the user-oriented capabilities listed here. Because of the importance of self-service to next-generation BI tools, I've devoted an entire section to the topic later in this report (see "Self-Service BI.")

Interactive. An interactive BI tool lets users interact with the display rather than just view the data. There are many ways to make a BI tool interactive. Users may want to personalize the look and feel of the display or select the metrics, charts and tables to display from a library of such objects. Or they may want to navigate the data by drilling down along predefined paths or interact with the objects themselves, applying filters, toggling between chart types or printing, exporting or snapshotting views of the data. A menu bar of icons or right-click should expose a list of context-sensitive actions that administrators should be able to hide or expose, depending on a user's role, task or experience.

Visual. Visualizing data makes it easier for users to quickly see patterns, trends and anomalies in the data. Next-generation BI tools support a rich array of interactive visualizations and suggest appropriate visualizations based on the type of data users want to display. Rather than render static images, BI tools should let users interact with the graphical displays. For instance, a user should be able to mouse over a chart element to view its underlying data properties or click on the chart to drill down and view more detail. On a scatter plot, a user should be able to lasso data points and use them to create a new group or as a filter for another chart or table. And on a heat map, a user should be able to right-click to display a time-series chart, among other things.

Flexible. Next-generation BI tools need to support the kinds of displays and analyses that users want to create without forcing them to dump data into Excel and do their work there. Casual users may want to apply filters to view data from multiple dimensions or perhaps sort or rank the data, and then display it in various chart types. Power users may want to create new calculated or parameter-driven fields, build what-if simulations or apply various algorithms to identify segments, associations or correlations within a data set.

Analytical. A BI tool needs to go beyond just displaying metrics, although this is what most casual users primarily want. If a metric exceeds a threshold, users need to analyze the root cause of the problem by exploring the data dimensionally and visually and by examining its details. Power users, on the other hand, may want to create custom groups and then compare these groups to others using a range of mathematical and statistical formulas. The tools need to come with a wide variety of functions and invoke server-side or database functions that run against enterprise data and return the results.

Predictive. Next-generation BI tools can't just look at data in the rear-view mirror; they need to extrapolate from the past about what the future holds. This may be as simple as running a regression algorithm against this morning's sales data to predict sales by the end of day. It may also entail running more sophisticated pricing and demand forecasts using dozens of input variables. It may also mean creating an analytical model and running it against existing records in a data warehouse or new records as they get created in a transaction system. The output score is then adding to each record and used to guide future interactions with a customer, supplier or activity.

Collaborative. Few decisions are made in a vacuum, but most BI tools today look like they were designed to be used by an isolated employee in a cubicle. New BI tools let power users publish live, interactive reports and dashboards to internal and external Web pages where others with permission can access the information. Beyond publishing, casual users need BI tools that enable them to associate comments with an entire dashboard, individual charts or tables or even cells within a chart or table. Ideally, the comments form an ongoing discussion among team members about what's happened and what to do about it. The discussion culminates when team members evaluate the effectiveness of the decision or action taken by the team.

Collaboration is perhaps even more important for power users.

But collaboration is perhaps even more important for power users, who always work in isolation and often re-create the work of other analysts without knowing it. A BI tool should allow power users to share their analyses with others and track who's doing what. Power users should be given the ability to rate analyses, follow other power users and add comments and links to any published analysis. This type of power user collaboration can greatly enhance productivity through reuse and make work more interesting and dynamic.

Mobile. Users should be able to access interactive reports and dashboards through any device. These reports and dashboards should look, feel and act like the normal applications that users access through the Web or a desktop tool, yet they should also exploit the unique characteristics of the underlying device, giving users the best of both worlds. The applications should also work offline.

IT CHARACTERISTICS

The above list focused on end-user characteristics of a next-generation BI tool set. The following list focuses on IT characteristics.

Fast. Because of the "Google effect," users now expect subsecond response times for complex queries. After all, if Google can locate highly relevant documents out of billions on the Internet, why can't a BI tool locate and display the right data from a corporate data warehouse and other systems?

To meet this demand, most visual analysis vendors, including Advizor Solutions, Tableau Software and Tibco Spotfire, load data into an in-memory database. Some enterprise BI vendors, including SAS and Oracle, have also recently shipped in-memory analytics systems (i.e., SAS High Performance Analytics and Oracle Exalytics). Other BI vendors, such as Pentaho, now integrate with third-party distributed caching systems to get the benefits of in-memory-based processing without having to move all data into memory. Still other BI vendors, such as MicroStrategy and Arcplan have built intelligent caches into their mid-tier servers to speed processing on previously run queries. But there is more to delivering fast performance than managing memory (see the "Architecting for BI" section below).

Deploys quickly. Businesspeople don't like to wait for a system to be built. So next-generation BI tools dramatically shorten the time to value by turning power users into developers. BI tools need to install quickly, connect effort-lessly to any source, display relevant metadata, drag and drop data objects onto a visual canvas and begin interacting with them. Without scripting or coding, power users should be able to configure the metadata to suit their needs. This includes relabeling objects, creating hierarchies, adding calcula-tions, creating transformations and so on. Users should be able to publish these models on a shared server so others can reuse them.

Any data source. Next-generation BI tools should be able to connect natively to any data source without much or any IT intervention. This includes rela-tional databases regardless of the schema as well as local files, external data sets, NoSQL databases and even text-based files and documents. Users should be able to mash or blend these sources together without scripting or coding, although some sources may be too complex for even the average power user to handle. In this case, a good BI tool will support a workflow that lets a power user pass work to an IT professional and work collaboratively to create the new data environment.

Scalable. Next-generation BI tools scale linearly as the number of concur-rent users and amount of data grow. That's because the BI software runs in a server environment that supports load balancing and clustering, makes effective use of memory and threads and exploits the inherent parallelism of multi-core processors. On the data side, next-generation BI tools run against

enterprise-caliber database servers that can hold and process large volumes. These servers include massively parallel processing (MPP) databases, database appliances and NoSQL systems that typically run on a grid of commodity servers.

Reusable. Next-generation BI tools facilitate reuse by sharing business objects in a server environment. In a top-down environment, these objects refer to semantic layers and mashboard libraries (see "Self-Service BI" for more information on these BI artifacts). In a bottom-up environment, these objects are user-generated metadata within the dashboards (i.e., metrics, dimensions, attributes, hierarchies). When faced with complex source environments, organizations should lean toward top-down tools, which are better designed to handle complexity than bottom-up environments.

No BI tool, no matter how beloved by business users, will last long if it is not easy to maintain and modify.

Maintainable. No BI tool, no matter how beloved by business users, will last long if it is not easy to maintain and modify. Otherwise, the costs of supporting the tool skyrocket, making it an untenable option. Total cost of ownership for BI tools is best measured by the amount of time BI developers have to spend setting up and modifying the tool's metadata and creating reports for end users. Top-down tools take longer to deploy and cost more to maintain because IT developers must do the work. But in complex environments, they can save a lot of time in the long run. Bottom-up tools work well until they hit complex data sets that require substantial knowledge of SQL and the application to sort out.

Manageable. Next-generation BI tools offer strong administrative tools to track usage and optimize performance, queries and other facets of a BI environment. They show which tables and fields are used most, make suggestions for adding aggregate tables and indexes and modifying schema. They also provide query governors that prevent runaway queries from bogging down performance for everyone, among other things. These tools also interoperate with third-party security frameworks, like Lightweight Directory Access Protocol and Active Directory, to manage users and permissions.

Comprehensive. Since it's unlikely that one BI tool can support the needs of all users, the best BI tools offer tools or modules to support the complete BI stack, from pixel-perfect reporting and online analytical processing (OLAP) to visual analysis and data mining. Some even go beyond the BI stack and offer data integration and data quality tools as well as databases and even the hardware that everything runs on. Some of the bigger enterprise BI vendors will soon package all these components into a single BI appliance or online service (i.e., cloud offering).

Portable. Next-generation BI tools need to be able to publish live, interactive output to any environment, including Windows and Macintosh desktops; HTML browsers; Flash, Asynchronous JavaScript and XML, or Ajax; mobile devices and various portals. This includes the entire report, dashboard or component parts packaged as widgets, components or mini-applications. Ideally, these parts are managed in an online library that users can select from to build larger applications.

SUMMARY

Next-generation BI tools merge top-down and bottom-up capabilities into a single-user environment. Top-down BI vendors are working to accelerate the time required to install, configure and model a BI semantic layer and data model and are adding new modules to compete with in-memory visual analysis tools. Bottom-up BI vendors are working to increase the scalability, reliability, reusability and manageability of their BI tool environments to avoid creating departmental silos.

Casual users benefit because they will get more visual, interactive and easier-to-use tools while superusers will get tools that make it easier to build ad hoc, interactive reports and dashboards for casual users. Power users also benefit because they get more flexible BI tools that attach effortlessly to almost any data source and provide the analytical flexibility they need to carry out their analyses without having to dump data into Excel.

However, until vendors ship an uber BI tool, many organizations standardize on top-down BI tools for metrics-driven reports and bottom-up BI tools for ad hoc analysis. ∎

Self-Service BI

The Holy Grail of BI. For years, self-service BI has been the Holy Grail for BI professionals and a key characteristic of next-generation BI tools. Self-service BI promises to provide business users with easy-to-use tools that enable them to get the information they want, when and how they want it without IT or power-user intervention. And self-service BI liberates the BI team from having to fulfill endless requests for custom reports, each of which takes weeks or months to deliver and rarely contains all the data users want in the form or shape they desire. Thus, self-service BI provides a win-win situation that makes everyone more productive.

Self-service gone awry. Or so the theory goes. In reality, self-service BI has fallen short of the mark. Self-service BI tools have proven too complex for most casual users who simply want to monitor key metrics and perform a few drill-downs. And some power users (mainly superusers) exploit self-service BI to create thousands of reports, most of which are variations on the same theme. So, self-service BI has led to either no use or overuse, while the BI team still wades through a backlog of report requests.

APPROACHES TO SELF-SERVICE BI

For years, BI professionals have asked vendors to deliver easy-to-use self-service BI tools that enable users to create their own reports and dashboards. And vendors have responded. Top-down vendors have offered semantic layers and, more recently, mashboards that enable casual users to modify an existing report or dashboard to fit their needs or create a new one from scratch. Conversely, bottom-up vendors offer visual analysis tools that let power users publish analyses to a managed server environment so authorized casual users can access and interact with the output (see **FIGURE 5**).

Both top-down and bottom-up approaches to self-service BI have strengths and drawbacks. The top-down tools require casual users to think like an IT

person who has query and design knowledge when creating ad hoc reports and dashboards. And bottom-up tools require business analysts to dumb down their work—eliminating filters, simplifying parameters, cleaning up screens—before publishing them for casual users. Nonetheless, both types of tools are closing the gap between top-down and bottom-up BI and delivering true self-service capabilities.

Top-Down Approaches

Semantic layers. The traditional top-down approach to self-service BI involves creating a semantic layer that defines back-end data entities as a set of business-oriented objects. Users then drag and drop these objects from a folder structure into a query panel and submit the query. They then use a point-and-click design tool to format the output into tables or charts with appropriate fonts, colors and other design elements. This is a lot of work for a casual user, who can barely remember how to log in to the BI tool. However, it's a perfect environment for superusers, who can now create ad hoc reports without writing a line of code or learning SQL.

FIGURE 5.
Top-Down and Bottom-Up Approaches to Self-Service BI
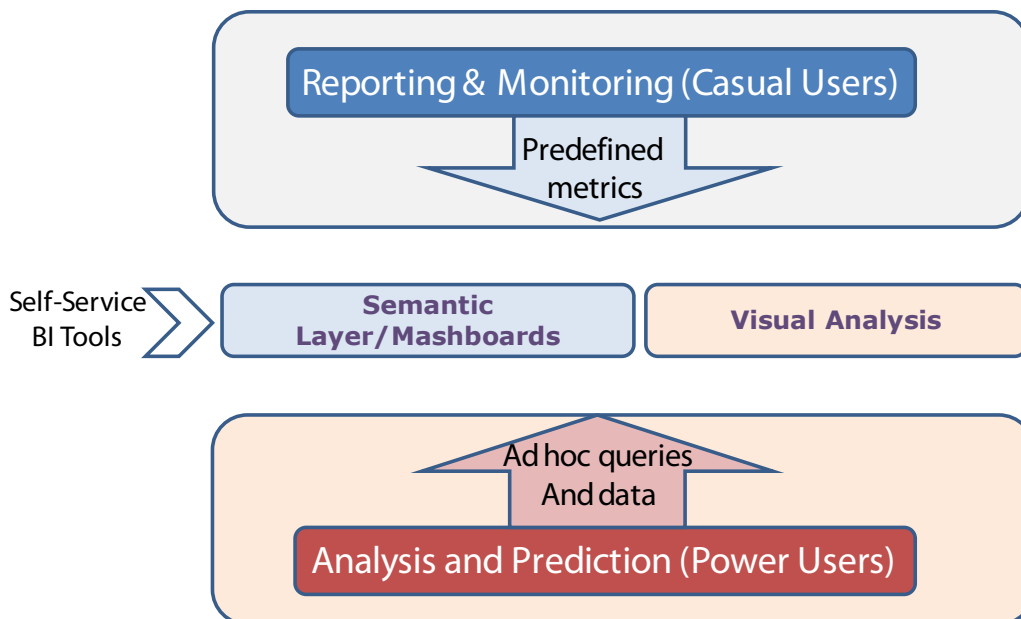
Mashboards. The newest type of top-down self-service tool is a dashboard, which is a visual exception report. Casual users love dashboards because with a quick glance, they can tell what's going well and what's not and with a few clicks get all the information they need to understand and deal with a problem. Unlike traditional dashboards, which are built by the IT department, mashboards let superusers create ad hoc dashboards without direct IT involvement.

However, to create a mashboard, the IT department must first create a bunch of reports. Then, they "widgetize" the artifacts of those reports—tables, charts, controls—and store them in an online catalog. From there, authorized superusers can drag and drop the widgets from the catalog onto a dashboard canvas to create a personalized dashboard, or "data portal." Well-designed widgets "recognize" each other when dragged onto a canvas and can automatically synchronize their displays in response to user inputs. Mashboards are a great way for organizations to get more value out of their existing report content.

Bottom-Up Approach
Visual analysis. The primary bottom-up approach to self-service BI uses visual desktop analysis tools to publish dashboards to a managed server. Here, power users conduct an analysis and publish their output for casual users to use as a live, interactive dashboard. Conversely, superusers can use visual analysis tools explicitly to create departmental dashboards for their colleagues. In many respects, visual analysis tools offer the best of both worlds. They are powerful analysis tools that provide easy access to any structured data source and speed-of-thought analysis, yet their publishing capabilities turn them into highly interactive departmental dashboards.

> *In many respects, visual analysis tools offer the best of both worlds.*

Many in-memory visual analysis tools can hold 50 million records or more, depending on the memory footprint of the server hardware. In effect, these tools serve as departmental data marts (or small data warehouses) tailored to a specific audience. Although most of these tools update their in-memory databases once a day, administrators can schedule the data to refresh on an incremental basis, as frequently as every five minutes. Some even offer the option to store data locally in memory or query the data directly in the source

systems. This lets users opt for greater speed (i.e., local in-memory system) or current data (i.e., direct access), depending on their requirements and the capabilities of source systems.

Offloading the data warehouse. One architectural benefit of in-memory visual analysis tools is that they offload query processing from the data warehouse so they can more efficiently handle other analytical workloads. For instance, one large retailer plans to implement a BI tool with a mid-tier intelligent cache to support dashboards with predefined queries. Early each morning, they'll execute the dashboard queries to put the previous day's data into the cache, which essentially becomes an in-memory data mart and frees up the data warehouse to handle ad hoc queries and other workloads without adding more hardware.

There is a downside to visual analysis tools: They are largely departmental in scope and can have difficulty if source systems have complex schema or contain lots of dirty data. Although these tools have scripting and transformation capabilities, they work best against local files, flat external data sets and existing data warehouses and data marts.

Also, if the tool only runs against an in-memory database, there is a chance that the volume of data and number of concurrent users may exceed available memory, which would give users an error message. As a result, administrators have to carefully estimate hardware capacity to meet application requirements. To minimize the chance of exceeding memory limits, some visual analysis vendors, such as QlikTech, recommend building many smaller, purpose-built applications rather than one general-purpose BI application. In many cases, users prefer this approach, since it mimics on a larger scale the purpose-built applications many people now use on their smartphones. However, it may also confuse or frustrate users who have to sort through dozens of applications to find specific information or functionality. Thus, a danger of visual analysis tools is they can create a fragmented view of enterprise data. Another downside is these tools don't create formatted reports, which is still a requirement for many users, although they can export the data into Excel or a database.

FUNCTIONALITY ON DEMAND

So far, we've defined self-service BI as giving users the ability to create new reports and dashboards without IT's assistance. This is important, but it's only one dimension of self-service BI. A more refined view of self-service BI

involves takes into account other types of func-
tionality and users' readiness to use those func-
tions. Matching users with the right functional-
ity at the right time is the key to getting users to
adopt BI tools and deliver organizational value.

The challenge here is that users' needs
change based on their roles or tasks at any
given hour or day. In addition, users often seek
additional functionality as they gain experience
with the tools in the context of their business
data and processes. In other words, delivering
BI functionality to users is a moving target. So,
BI tools that expose functionality to users as
they need new capabilities are a critical element of self-service BI. Next-gen-
eration BI tools expose functionality on demand, increasing user adoption.

> *Matching users with the right functionality at the right time is the key to getting users to adopt BI tools and deliver organizational value.*

Two Types of Self-Service

One reason self-service BI has been so elusive is most vendors and BI profes-
sionals don't realize that there are two types of self-service BI they must sup-
port: one for information consumers and another for information producers.

An information consumer uses information created by an information
producer. Simple enough. But an information consumer isn't equivalent to a
casual user, and an information producer doesn't equate to a power user. Let
me explain.

An information consumer is anyone who consumes information. This can
be a casual user or a power user. An information producer is anyone who pro-
duces information. This is typically a power user or IT professional but can
occasionally be a casual user.

For example, casual users consume information 80% of the time and produce
it 20% of the time (or at least try to produce it until they give up and ask for
help from a superuser or IT professional). The reverse is true for power users.
They produce information 80% of the time and consume it 20% of the time.
(Ideally, power users should do less producing and more consuming or analyz-
ing.) IT professionals are almost entirely information producers (see **FIGURE 6**).

Functional Hierarchies

Next, it's important to understand that information consumers and producers

have separate sets of functional requirements. These requirements generally flow in a functional hierarchy, as depicted in **Figure 6**. In other words, information consumers and producers want additional functionality as they become more experienced with the BI tool and more knowledgeable about their data and business processes. They may also need different functionality if their roles or tasks change frequently as part of their jobs.

Consumer hierarchy. The functional hierarchy for information consumers is the following:

1. **View.** Users simply view information online or on paper. Users study the data but do not interact with it in any way, although some may copy and paste numbers into a spreadsheet.
2. **Navigate.** Users navigate an existing data set by drilling down along pre-defined paths to view detailed data or by selecting predefined filters to narrow or expand views.
3. **Modify.** Users change an existing data set by sorting or ranking data, toggling among charts and tables, adding columns using calculations and pivoting axes to change data views.
4. **Explore.** Users add new data to an existing data set by accessing additional predefined data sources and manipulating metadata.
5. **Model.** Users create what-if scenarios and predictive models by manipulating independent and dependent variables.
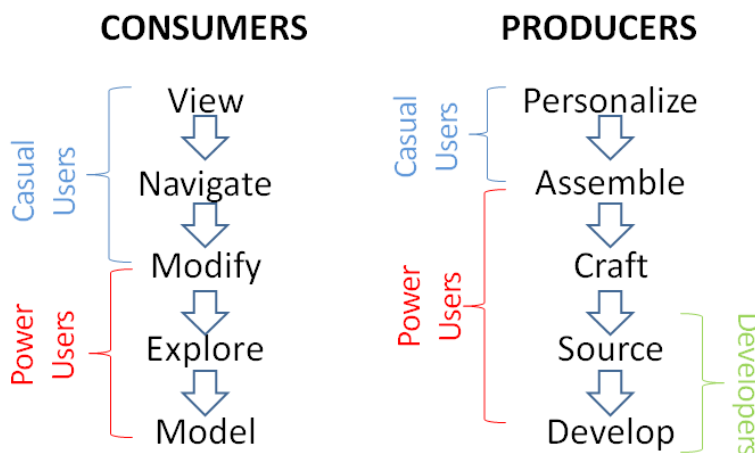

Figure 6:
Mapping Types of Users to Self-Service Hierarchies

Both casual users and power users both consume and produce reports and analyses, but power users exploit more advanced features.

Producer hierarchy. The functional hierarchy for information producers is the following:

1. **Personalize.** Users select which information objects to display on their screens and customize the look and feel to suit their tastes.
2. **Assemble.** Users create new reports and dashboards from widgets created from existing report parts, such as charts and tables.
3. **Craft.** Users create new reports or dashboards from scratch using a semantic layer of predefined information objects.
4. **Source.** Users query new data sources, including local files and external data sets, using SQL or built-in query functions, and integrate the result sets using point-and-click logic or Excel.
5. **Develop.** Users write scripts or programming code to query or manipulate data to support consumer requirements.

Mapping Users to Functional Hierarchies

Casual users. As **Figure 6** shows, casual users generally *consume* information by viewing and navigating data—the first two levels in the consumer functional hierarchy. On the producer side, casual users might only personalize their dashboards or portals, if that functionality is available to them. This bounded set of functionality aligns with my dashboard framework, called MAD, which stands for monitor, analyze and drill to detail.[1] In a MAD dashboard, casual users devote the majority of their dashboard time to monitoring predefined metrics and navigating to details to understand root causes and impacts.

Power users. Conversely, power users *consume* information in more sophisticated ways than casual users. They spend most of their time modifying, exploring and modeling data. On the producer side, power users typically assemble, craft and source information and may occasionally write code or scripts. (Actually, superusers primarily assemble and craft reports and dashboards, and business analysts, modelers and data scientists source data and write scripts and code to manipulate data.) IT professionals source and develop data-centric applications.

---

1. See the second edition of my book, *Performance Dashboards: Measuring, Monitoring, and Managing Your Business* (Hoboken, NJ: John Wiley & Sons, 2010), available at most online bookstores.

Evolving Requirements

When casual users first use a BI tool to consume information, they may only view data. After a while, as they become more familiar with the tool, they may want to navigate to the details. Later, they may want to modify data in the report or dashboard by adding a calculated column. Conversely, information producers may first want to assemble data from pre-existing report parts or craft them using a semantic layer. But after a while they may want to source data independently and mix it with other data.

Next-generation BI tools need to offer the full spectrum of functionality for both information consumers and producers. But, more important, they need to expose this functionality on demand, as users need them and are capable of using them. Most tools enable administrators to control what functionality users can access using fine-grained access control lists. But this approach alone is cumbersome, since it's hard to know exactly when users are ready for more. But exposing all functionality at once can overwhelm even the most experienced users, undermining their productivity. (Think of Microsoft Office 2007.) So the best BI tools expose functionality discreetly by means of icons displayed in the menu bar, ribbon or content frame, but only to users who are most likely to use those functions.

SUMMARY

Self-service BI is a key feature of next-generation BI tools. Top-down and bottom-up BI tools both offer self-service capabilities. Top-down tools are designed to help casual users perform ad hoc tasks, while bottom-up tools help power users publish interactive dashboards. Today, mashboards and visual analysis tools hold the most promise for delivering on the promise of self-service BI.

Besides tools that bridge the gulf between metrics-driven reporting and ad hoc analysis, we also need BI tools that expose functionality on demand to information consumers and producers. The two types of users traverse different functional hierarchies. And individual users descend to deeper levels in the hierarchy as they gain more experience with a tool. Thus, it's imperative that tools support the full range of functionality for each type of user but expose functionality as users need it to maximize adoption and usage. This is another critical, but often overlooked, dimension of self-service BI. ∎

Architecting for BI

IT'S FUN TO TALK ABOUT BI tool functionality, past, present and future. But, BI tools don't work unless there is a back-end data management system that they can query. The nature of that system—how it stores data and runs queries—plays a significant role in the success of any BI tool. A well-designed BI tool running against a slow data management system inevitably fails.

To ensure adequate performance, BI professionals need to think long and hard about how to architect their BI environments. The key decision they need to make is where to house data and how to process queries. There are three basic options to perform this work: database, application server or client. In more complex scenarios, architects spread data and query processing across two or more of these layers to optimize performance.

At the same time, many BI tools now come with in-memory databases, which, in effect, create self-contained analytical environments that may or may not interoperate much with the larger BI environment. This is a blessing and a curse. On one end, a self-contained BI environment lets business departments and power users access and analyze data without IT involvement. On the other, it has the potential to create redundant silos of analytical information, unless the tools are implemented under the purview of the IT department, which can point them to existing data warehouses and data marts to preserve unity of information.

A well-designed BI tool running against a slow data management system inevitably fails.

DATABASE PROCESSING

The most basic BI architecture is one in which a BI tool queries a remote database. Here, the BI tool generates the queries, displays the results and enables users to view and manipulate the delivered data set. The database holds all the data, optimizes and processes the queries and generates the result sets,

which it feeds to the BI tool. With the database providing so much of the processing, there are several ways to architect a BI system.

Data warehouse. For simplicity, most BI architects store all data in a data warehouse, a single repository of integrated data optimized for query processing. Data warehouses work well until the number of concurrent users submitting queries, both easy and complex, starts to bog down performance. At this point, architects can scale up the data warehouse by adding more indexes, aggregations or hardware; throttle or kill offending queries; or offload data and users to other machines.

Data marts. One way to offload data is to create a data mart on a separate machine. A data mart is an application-specific instance of the data warehouse geared to a particular workgroup or department (e.g., finance, sales or marketing). Some data marts are logical sets of tables inside a data warehouse designed to simplify queries and partition workloads, but others are physically independent systems, such as in-memory visual analysis tools and OLAP cubes that extract and model a data set for local consumption. Visual analysis tools often use in-memory databases to ensure fast performance, creating an outboard data mart. The same is true for multidimensional (i.e., OLAP) cubes, which aggregate data dimensionally, either dynamically or in advance or some combination of the two.

One way to offload data is to create a data mart on a separate machine.

The problem with data marts in general, and cubes and in-memory databases in particular, is that they often replicate data, which can make it harder to maintain a single version of truth if IT is not heavily involved in the process. Data marts can also increase overhead costs by duplicating systems and staffing.

Multidimensional cubes. Multidimensional cubes, or OLAP engines, are specialized databases that aggregate data dimensionally, either dynamically or in advance or some combination of the two. These cubes provide fast performance for dimensional queries, enabling users to slice and dice data at the speed of thought. The problem with cubes is that they replicate data and systems, creating an administrative headache and increasing costs.

BI PROCESSING

The second most basic architecture puts more of the onus for storing data and processing queries on the BI tool. There are several options here as well:

In-memory database. Here, the BI tool comes with its own in-memory database to store and process all data used by the BI tool. Typically, these tools pull data from a data warehouse or directly from transaction systems in a single nightly batch load. However, most can be updated at any frequency, applying only changes in source systems to the target files. The benefit here is that queries run extraordinarily fast when data is held in memory (i.e., random-access memory [RAM]) versus disk. Plus, an in-memory database functions as a data mart, freeing up processing cycles in the data warehouse to handle ad hoc queries and other workloads. The downside is that you may not be able to fit all your data into RAM and support large numbers of concurrent users. And although the cost of RAM has dropped considerably, if you have lots of data and users, it can still get expensive. Plus, you again replicate data and systems, which can lead to analytical silos.

Some tools now give users the option to hold data in memory or query it directly in the data warehouse or other system. Users may decide to download data to a local in-memory database to get consistently fast performance on a frequently used data set or when back-end systems are bogged down. They'll choose to query data sources directly when they want the most current data possible or if they don't own the data and can't download it.

Mid-tier cache. To gain the benefits of memory without the constraints, some BI vendors create an intelligent, mid-tier cache on the BI server that stores query result sets temporarily (see **Figure 7**). Here, BI tools run queries against the cached data when possible, instead of the remote database. When the cache fills up, it bumps the least used data back to disk. Intelligent caches apply permissions to data to keep users from viewing unauthorized data. In many respects, mid-tier caches deliver the same results as in-memory databases, except the data is not permanently stored in the cache. The problem with mid-tier caches is that the first user to run a query each day doesn't get the benefit of in-memory processing. To circumvent this problem, administrators often run certain queries early in the morning to ensure that frequently used data is already in the cache.

FIGURE 7:
In-Memory Databases Versus Intelligent Caches
Intelligent caches speed response times for pre-run queries while still giving users access to data stored in remote databases (of any size). In-memory databases speed response times for all queries running against a downloaded data set but don't provide direct access to remote data. Some tools use a hybrid approach users can choose to query remote data or download data to a local server or desktop to improve query performance.
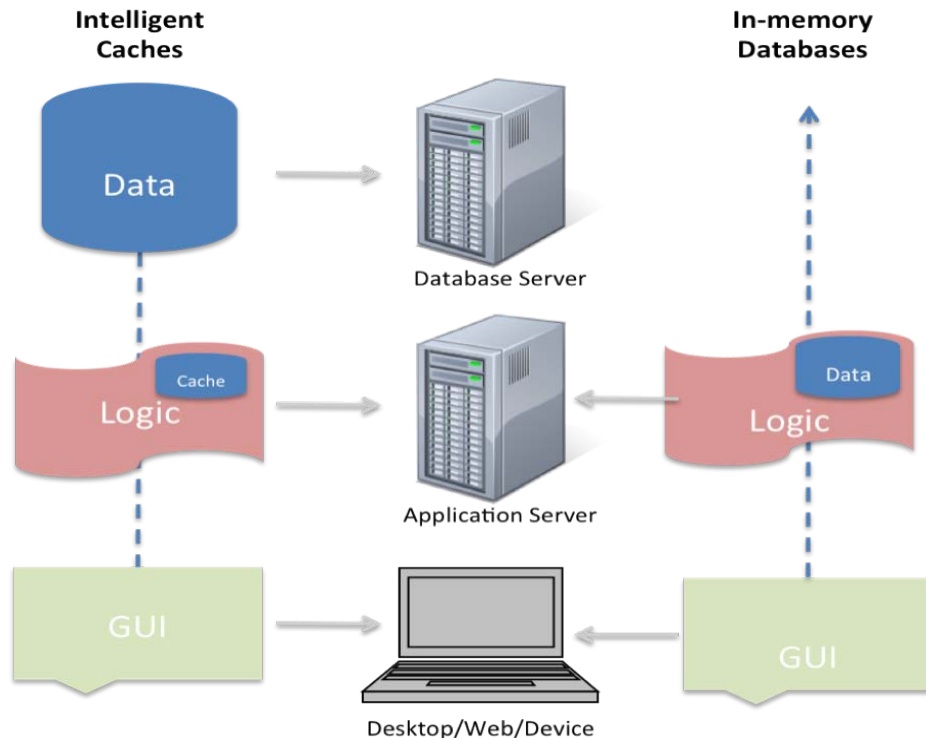
ROLAP servers. To get the benefits of cubes without replicating data, some BI tools use a relational OLAP, or ROLAP, architecture to generate dimensional result sets on the fly. These tools generate complex SQL that runs against data warehouses or data marts. The downside is performance, since generating dimensional aggregations on the fly can be slow. To circumvent this problem, some ROLAP vendors now split query processing between a database server, which handles basic SQL operations, and the BI server. To help ROLAP performance, administrators often create aggregate tables, which function like multidimensional cubes but without the downsides of duplicating or moving data, although aggregate tables can suck up IT development and administration time.

Federated queries. Some BI tools assume responsibility not just for generating queries but optimizing them as well. BI virtualization tools—such as Oracle's Common Enterprise Information Model, part of Oracle BI Enterprise Edition—create a global semantic model that makes multiple remote databases appear as a single local database. Behind the scenes, the BI virtualization tool, which understands the profile of each remote database, optimizes queries to run against those systems, deciding which database to query first, then moves data to the next database or brings all data back to the BI server to finish an operation.

Client cache. Some BI tools also move data to the client machine to speed performance. For example, Flash-based dashboards require users to download data and animations to their client machines before they can view and interact with the data. Although Flash now can query back-end systems, Flash-based designers need to be careful to minimize the amount of data displayed on any given dashboard page to avoid lengthy downloads that can irritate users. Also, some mobile BI applications, namely, those designed for Apple's iPad and iPhone, download data to those devices to enhance performance.

DATABASE OPERATIONS

Database management vendors continually add new features and techniques to their products to improve query performance and keep up with growing workloads from both top-down and bottom-up BI. To display a simple dashboard may require a database to execute multiple, complex queries and return results in a split second. Power users can generate hundreds of complex, ad hoc queries that may require scanning a huge fact table and joining it with other tables. Here are some of the more salient features that BI architects need to understand to deliver fast performance:

Parallelization. Databases now try to parallelize all operations to remove bottlenecks. This includes loading, streaming, aggregating, scanning, joining, summing, sorting and merging. This parallelization happens at the processor level to exploit multi-core processors as well as the node level, if the database runs on a massively parallel processing system.

Columnar. Many databases now store data in columns as well as rows. Columnar storage reduces total disk space since columnar data can often be

compressed 10 to 1. In addition, since many queries only retrieve a fraction of the columns in a record, this reduces the amount of data that a database must retrieve from disk—the slowest part of the database operation—and processes it in memory.

Storage-level filtering. Many databases now run some or all SQL functions at the storage layer so data can be processed at the binary level, further minimizing the data that must be retrieved from disk.

Solid-state disks. Database vendors are supplementing traditional mechanical disk drives with solid-state drives, which are infinitely faster but much more expensive. To get the biggest bang for the buck, some vendors, like Teradata, have figured out how to automatically move frequently used, or "hot," data to solid state while keeping less frequently used, or "cold," data on less-expensive hard disks.

Indexes and aggregates. To deliver the performance of OLAP cubes or in-memory databases, database administrators often create aggregate tables and apply indexes to tables that are frequently joined, using functions such as Teradata's Aggregate Join Index, for example. Of course, it takes time to create and maintain these aggregates and indexes, and BI tools need to understand how to exploit them to achieve highest levels of performance.

Native SQL. Similarly, most databases have native SQL dialects that contain functions which can dramatically speed performance if the BI tools know the dialect and can exploit it within SQL queries.

In-database functions. Database vendors are adding many analytical functions to their systems that previously required users to download data from the database into a BI tool or specialized analytical system to run. For example, SAS is pushing a lot of its data management and analytics functions inside various databases to reduce data movement and leverage parallel processing. Many also publish application programming interfaces, or APIs, so users can write their own database functions. By running these functions in the database instead of in a BI or analytics tool, companies can speed up processing significantly. More important, they won't have to move data out of the database to an application server and then back again.

Special data types. Similarly, databases now support unique data types, such as geospatial data, along with special SQL functions to manipulate this data. Without built-in support for such functions, users would need to download data to a specialized application server and carry out the processing there.

PURPOSE-BUILT SANDBOXES

Finally, BI architects need to understand that database vendors now deliver purpose-built machines, databases or data services geared to specific types of analytical workloads. In most cases, these purpose-built environments support the needs of small groups of highly specialized business analysts or data scientists. Here are some examples:

Software-only analytical databases. Some database vendors offer software-only products geared to supporting the most complex analytical queries with super fast performance. These specialized databases, such as ParAccel's Analytic Platform, typically use an MPP architecture, exploit columnar storage and come with dozens of built-in analytical functions. As such, they are perfect sandboxes for small groups of highly skilled analysts who want to run complex queries.

Analytical appliances. Other vendors offer software-hardware combinations tuned to handle a variety of workloads. For example, Teradata offers its Extreme Performance Appliance, which uses solid-state disks to provide fast performance, and Extreme Data Appliance to let users query petabytes of data. SAS also offers an analytical appliance, SAS High Performance Analytics, built on configured hardware from Teradata and EMC Greenplum and designed to analyze large volumes of data.

Cloud-based services. And some vendors, like ParAccel, are putting their databases in the private or public cloud and allowing users to quickly provision new instances, upload their own data and select from preloaded data sets.

Hadoop. Many companies are storing large volumes of clickstream and other unstructured or semi-structured data in Hadoop, which runs queries against an open source distributed file system. Specialized analysts, known as data

scientists, write programs in Java and other programming languages using the MapReduce framework to process data stored in Hadoop. Higher-level languages, such as Hive and Pig, are emerging to make it easier for data scientists to query Hadoop data stores.

SUMMARY

There are many options for architecting an analytics environment. BI architects need to select the right option or combination of options to ensure adequate performance and scalability for their BI applications. If the IT department controls the data management architecture, it will need to adapt to what already exists unless it wants to create and manage its own data architecture. Most vendors are quick to adapt new technologies and often leapfrog each other in capability. New advances in the use of memory, solid-state disks, columnar storage, storage-level processing and in-database analytics are rapidly advancing the state of the art, giving BI applications a badly needed turbo boost. ∎

User Input on Next-Generation BI

THIS FINAL SECTION PRESENTS the results of our user survey. To keep the survey short, I focused the survey on casual users and the BI tool capabilities needed to support them. I also asked a question on BI architectures.

END-USER CAPABILITIES

To identify the most important BI tool features for this set of users, I asked BI professionals a delicious question: "If next month you had to buy a new BI tool for your casual users, what end-user capabilities would be most important?"

"Ease of use" grabbed the top spot by a wide margin. Almost all respondents (95%) rated ease of use as "very high" or "high" in importance. In fact, an astounding 75% rated it "very high" in importance, almost double for any other feature in our list (see **FIGURE 8**).

This raises the question—What constitutes "ease of use"? Examining responses to the survey's only open-ended question shows that "ease of use" is a catch-all category to address all the problems that casual users experience with BI tools.

Almost all survey respondents (95%) rated ease of use as "very high" or "high" in importance.

In response to the question "What are the biggest drawbacks to your current BI tool set for casual users?" many respondents said their tools were too "clunky" or "cumbersome" or "complicated." Others cited more specific issues, such as "difficulty drilling" or "inability to insert user calculations" or "No Google-like, easy, metadata-driven search." Others said self-service capabilities of the tools simply expect too much of casual users.

- "[The tool] assumes users spend most of their work time dinking with data when they don't."

FIGURE 8:
End-User Capabilities of BI Tools Desired by BI Professionals
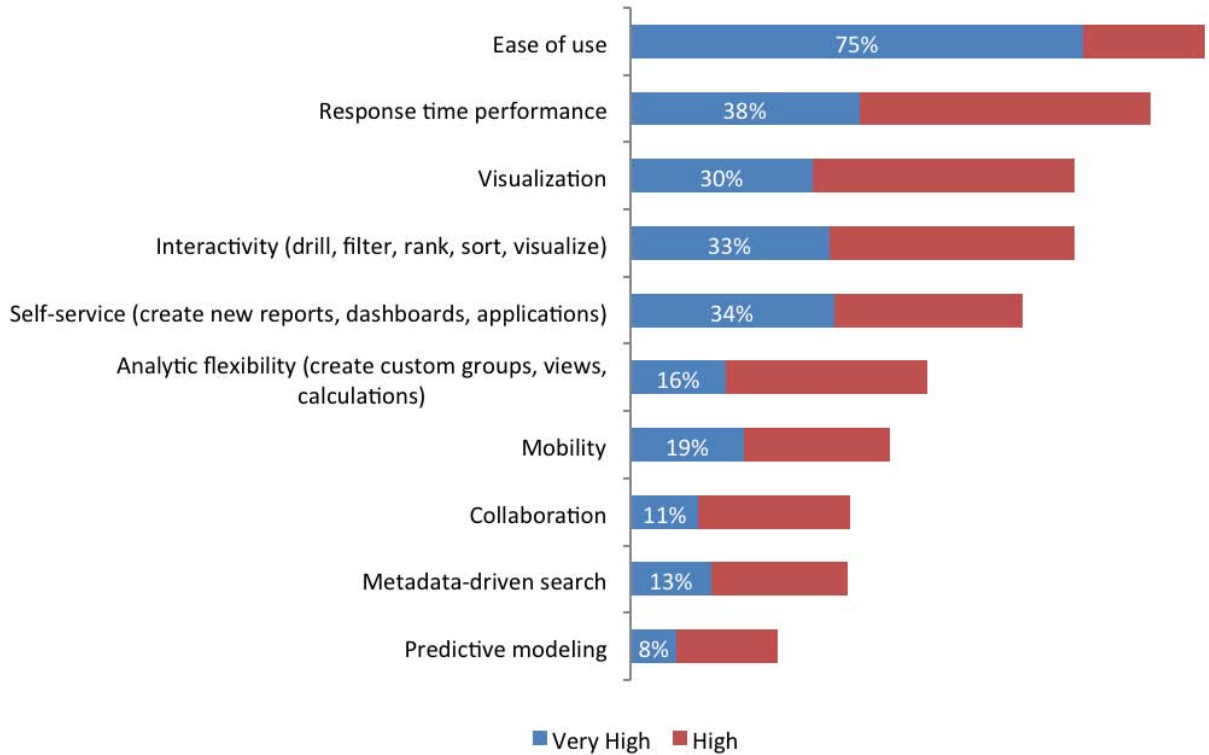
- "Not intuitive enough. Users shouldn't have to think, except about the business problem they are trying to solve."

Others shifted the onus from the tool to power users who design interactive reports and dashboards for casual users or provide mashboards for ad hoc dashboard creation:

- "Power users have made it too complicated to use for the casual users because we don't have a varied tool set."
- "The creation of an ad hoc mashboard is a little too complex for casual users if it requires new parameters not already in the data."

HIERARCHY OF NEEDS
As the results in **FIGURE 8** show, there is a four-tiered hierarchy of end-user capabilities that BI professionals expect BI tools to offer.

At the top of the heap is "ease of use" and "performance." Unless a tool is intuitive and fast, users won't bother. Next in importance are "visualization" and "interactivity," which now seem to go hand in hand: Users want to view data graphically and change those views with a simple mouse click or two. Third in the hierarchy is "self-service" and "analytic flexibility," both of which empower casual users to add, change or delete data in reports or dashboards, although this is still beyond the capabilities of many casual users today, as evidenced by respondent comments cited above. Last, BI professionals value emerging capabilities, such as mobile BI, collaboration and metadata-driven search.

Satisfaction ratings. Despite the criticism, BI professionals gave decent scores to the end-user capabilities of their current BI tools for casual users. More than a third of respondents (39%) rated their satisfaction with their current BI tools for casual users as "very high" or "high." Another 37% were "moderately" satisfied with their BI tools, while 17% gave the tools a "fair" or "low" rating (see **FIGURE 9**).

Interestingly, small companies (15%) are twice as likely as medium-sized companies (7%) to give "very high" satisfaction ratings to their current tools, and medium-sized companies are twice as likely as large companies (3%) to

FIGURE 9:
BI Professionals' Satisfaction with End-User Capabilities of Current BI Tools

Rating	Percentage
Very high	7%
High	32%
Moderate	37%
Fair	12%
Low	5%
Doesn't apply	8%
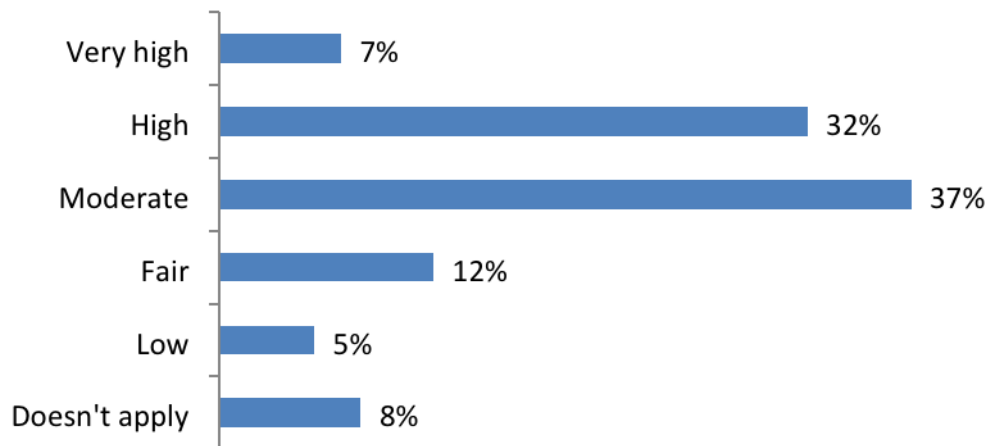
give a "very high" satisfaction rating. This raises the question: What type of BI system do small companies use that delivers such high degrees of satisfaction? While we didn't ask about specific tools, we can guarantee that small companies can benefit most from the avalanche of new low-cost, BI tools built on the latest technologies. Most other rating degrees were comparable across organizations of different sizes (see **FIGURE 10**).

FIGURE 10:
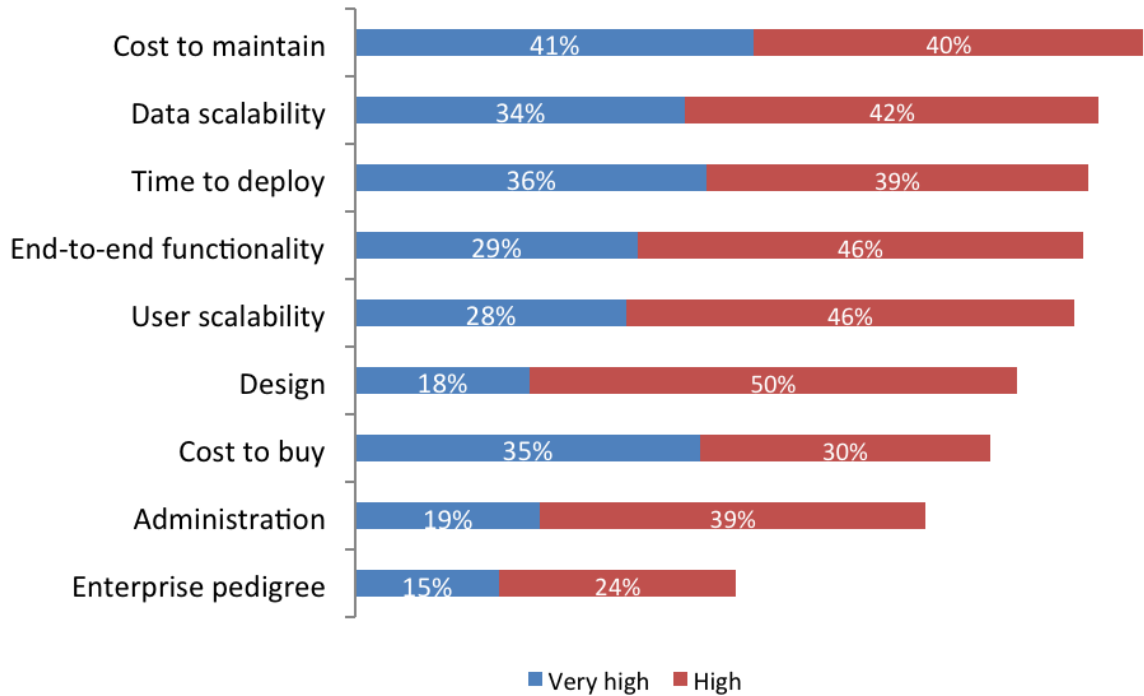*Satisfaction with Existing BI Tools by Company Size*
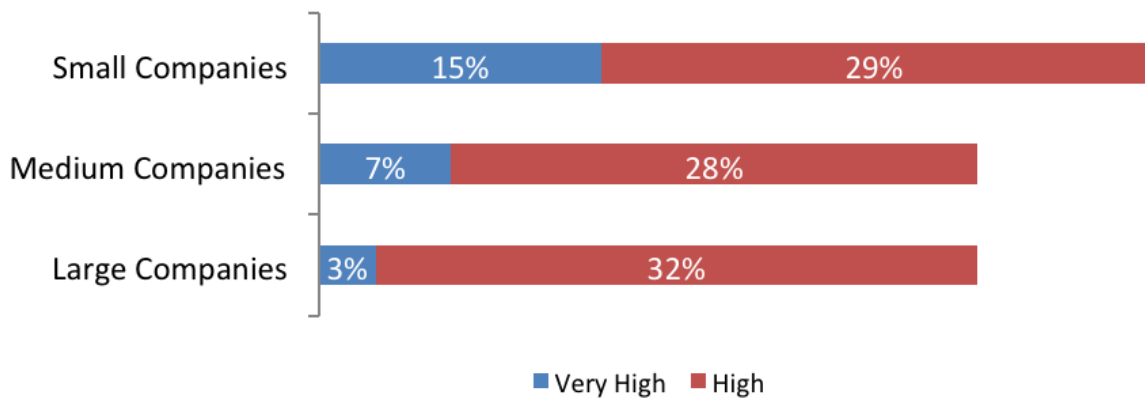
IT CAPABILITIES

Respondents had varied answers when asked to rate the importance of IT capabilities for casual users in a BI tool they might purchase next month. All but one of the listed IT capabilities, "enterprise pedigree," received a "very high" or "high" importance rating from 50% or more of the respondents. And five of the nine selections garnered a "very high" or "high" rating from 75% or more of the respondents. In other words, BI professionals universally favor high levels of IT capabilities in their BI tools. No surprise there (see **FIGURE 11**).

"Cost to maintain" a BI tool was deemed the most important overall, with 81% of respondents citing its importance as "very high" or "high." Obviously,

FIGURE 11:

IT Capabilities of Casual User BI Tools Desired by BI Professionals

Capability	Very high	High
Cost to maintain	41%	40%
Data scalability	34%	42%
Time to deploy	36%	39%
End-to-end functionality	29%	46%
User scalability	28%	46%
Design	18%	50%
Cost to buy	35%	30%
Administration	19%	39%
Enterprise pedigree	15%	24%

■ Very high ■ High

as IT teams are asked to do more with less, keeping a close watch on maintenance budgets is key to delivering more value to the business. BI tools that require developers to create and modify reports and semantic layers are cumbersome to administer and add to expense. And they carry annual software maintenance fees. Presumably, a tool that offers greater self-service capabilities reduces the tool's overall cost of ownership.

In addition, BI professionals who responded to the survey were almost equally enamored with other IT capabilities, including "data scalability" (76%), "time to deploy" (75%), "end-to-end functionality" (75%), and "user scalability" (74%). There was a slight drop-off to the next set of desired IT capabilities, including "design" (68%) and "cost to buy" (65%), and another drop-off to "administration" (58%) and "enterprise pedigree" (38%).

It's clear that BI professionals want their BI tools to offer strong IT capabilities. And they don't care whether it's an "enterprise" BI tool or not. A closer examination, however, shows they want enterprise scalable BI tools (both

user and data scalability) that don't cost a lot to buy or maintain. In essence, they want the best of both worlds.

Satisfaction with current IT capabilities. Interestingly, most BI professionals seem relatively satisfied with the IT capabilities of their current BI tools for casual users. The results are nearly identical to the results for end-user capabilities depicted in **Figure 12**. Almost half (42%) rated the IT capabilities of their current BI tools for casual users as "very high" or "high," and 37% gave them a "moderate" rating.

Figure 12:
BI Professionals' Satisfaction with IT Capabilities of Current BI Tools
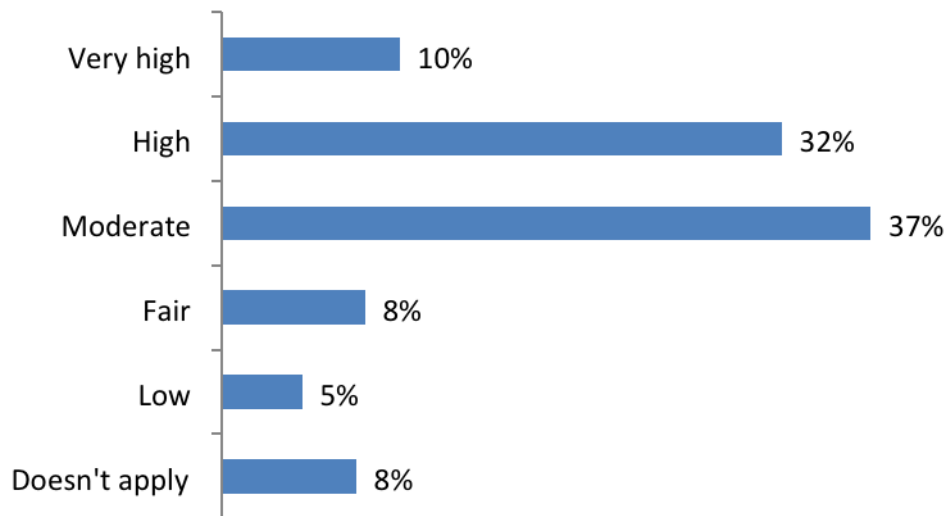
Survey respondents did not comment nearly as much about IT capabilities as they did about end-user capabilities. The most prevalent comment was "cost to maintain" followed by "amount of IT support required." Some mentioned scalability and performance, while one mentioned "longer learning curve for developers." Another user cited "weak user administration tools," another "lack of scheduled report delivery," and a third "needs better metadata."
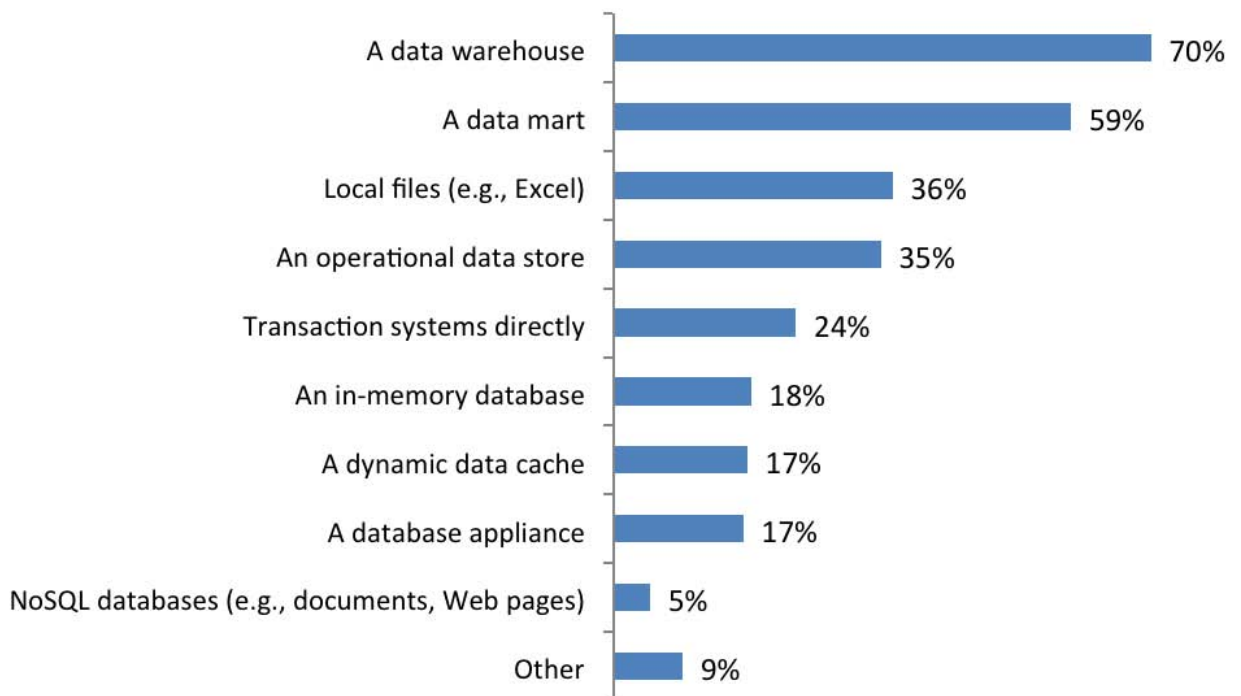
BI ARCHITECTURES

When it comes to architecting a BI environment, most organizations point their BI tools at a data warehouse (70%) or data mart (59%) (see **FIGURE 13**). After that, there is a significant drop-off, but the next two most popular data sources are a surprise: local files (36%) and operational data stores, or ODSes, (35%). The local files represent the unplanned, ad hoc nature of many BI activities, while the ODS represents operational reporting, still a major part of most BI endeavors.

Interestingly, we've given a lot of attention to in-memory databases, database appliances and dynamic data caches in this report, but these sources only account for 18% of sources among surveyed companies. NoSQL databases barely register at 5%.

Data sources by company size. When we filter the same data by company, a few interesting things pop out. First, large companies are most likely to

FIGURE 13:
Data Sources for BI Tools

| Data Source	Percentage
A data warehouse	70%
A data mart	59%
Local files (e.g., Excel)	36%
An operational data store	35%
Transaction systems directly	24%
An in-memory database	18%
A dynamic data cache	17%
A database appliance	17%
NoSQL databases (e.g., documents, Web pages)	5%
Other	9%

employ data warehouses (78%) or data marts (69%) with their BI tools than either small or medium-sized companies (**Figure 14**). This is no surprise, since large companies have been doing BI longer, and BI has traditionally gone hand in hand with data warehouse and data mart deployments. In addition, larger companies largely have a greater need for enterprise views of data, hence their investment in data warehouses and BI tools that run against them.

Next, small companies are more likely to employ local files (44%), such as Excel spreadsheets, as the source of data for their BI tools than medium-sized or large companies are. Again, that's no surprise, since small companies largely run on spreadsheets and are less advanced in their adoption of BI tools.

In addition, medium-sized companies are more likely to run their BI tools against transaction systems (29%) or an in-memory database (23%) than small or large companies are. My speculation is that medium-sized companies are just beginning to invest in BI and, with limited capital budgets, are most likely to purchase departmentally based in-memory BI tools, which have been popular in recent years.

Figure 14:
Data Sources by Company Size
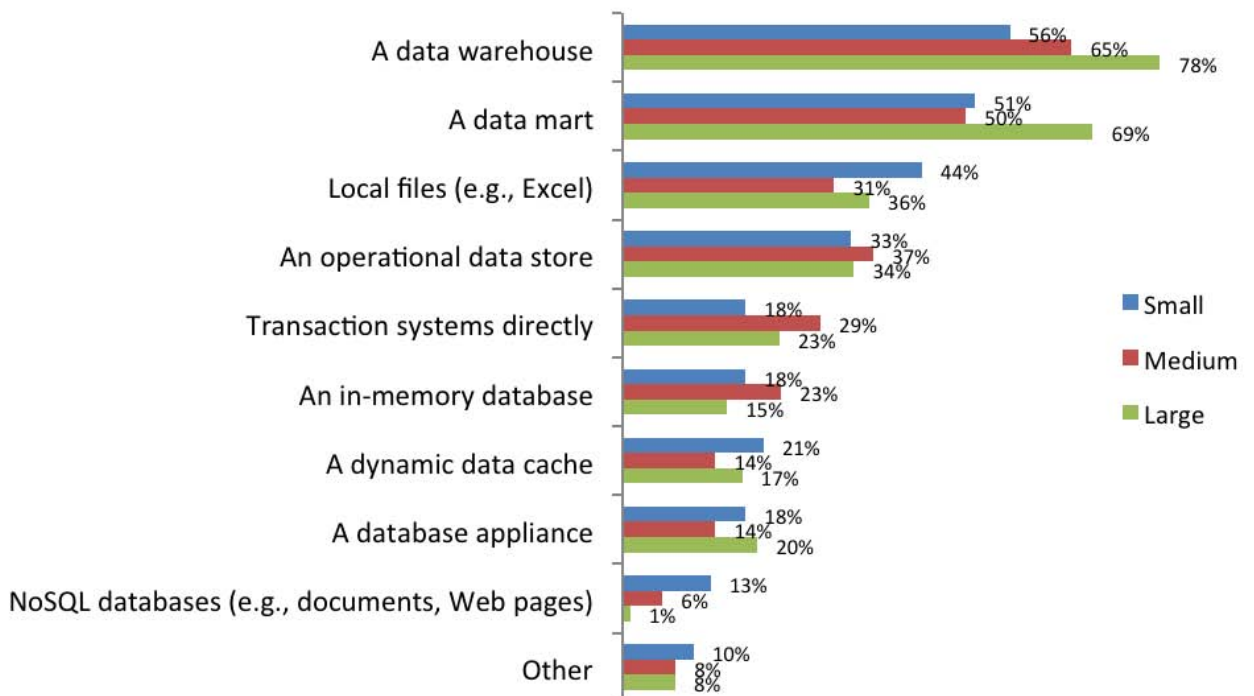
Finally, large companies are slightly more likely to purchase analytical appliances (20%) than either small companies (18%) or medium-sized companies (14%) are. These appliances typically aren't cheap. Large companies use them as analytical sandboxes to complement a data warehouse, while small and medium-sized companies use them as their data warehouse platforms.

SUMMARY

If BI professionals could purchase a new BI tool today, they would look for ones that are faster, better and easier to use. They also want tools that are more visual, interactive and analytical. From an administrative perspective, they want tools that are scalable and easier to maintain. Despite these and other emphatic wishes, about 40% of BI professionals are satisfied with their current BI tools. Small companies seem happier with their tools than large companies, perhaps because they jumped into the BI game later and purchased more modern, lower-cost, departmentally oriented BI tools. ■

If BI professionals could purchase a new BI tool today, they would look for ones that are faster, better and easier to use.

Recommendations

GIVEN THE CURRENT STATE OF THE ART in BI technology, both front and back end, here are some recommendations to guide your next BI purchase:

1. **Give your existing tools a performance boost.** If BI tool adoption lags because query performance is not optimal, consider investing in a visual analysis tool with an in-memory database or new specialized databases and systems geared to analytical processing. Or invest in both. Visual analysis tools can hold up to 50 million records in memory or more, depending on the tool's memory footprint, and function as a physically independent data mart optimized to meet the needs of an individual workgroup or department. Specialized analytical platforms include database appliances, MPP databases, columnar databases and NoSQL databases. Although these systems are not inexpensive, the outsized performance they deliver can kickstart a BI initiative and create an avalanche of new requests for data-intensive applications.

2. **Standardize on top-down and bottom-up tools.** Today, you need both top-down and bottom-up BI tools to meet the needs of casual users and power users. Top-down tools built on data warehouses deliver enterprise-caliber standard reports and dashboards for casual users, while bottom-up tools give power users the ability to perform ad hoc analyses against any data source and create highly interactive, departmental dashboards for their colleagues. In the future, one tool set may be sufficient to meet the needs of all users. Bottom-up BI vendors, in particular, are working to deliver the scalability, reliability and maintainability espoused by their enterprise brethren. In contrast, enterprise BI vendors are delivering in-memory visual analysis tools to compete with the upstart bottom-up folks. Today, it's unclear which set of vendors will dominate the market or even whether it's possible to meet all needs in a single tool set.

3. **Embrace self-service BI with governance.** Self-service BI promises to liberate end users, giving them the information they want, when and how they want it. Reporting tools with semantic layers and mashboard services and visual analysis tools with in-memory databases have made it easier for power users to create ad hoc reports and analyses. Visual analysis tools, in particular, have the potential to finally deliver on the promise of self-service BI. However, self-service BI must be accompanied by a strong governance program to ensure that superusers and power users don't create a plethora of conflicting reports that overwhelm and confuse casual users.

4. **Recognize that there are two types of self-service.** Casual users need self-service 20% of the time, while power users need self-service 80% of the time. Casual users need BI tools that allow them to change the views in a report or dashboard with simple mouse clicks to suit their needs. Power users need highly flexible tools that can attach to any source. They also need tools that let them model and manipulate the data without scripting and collaborate with other power users to reuse data workflows and analyses instead of reinventing them each time. Don't give casual users self-service tools designed for power users or vice versa. This doesn't work. Casual users will find the tools too hard, while power users will find them inflexible. ■

ABOUT THE AUTHOR

Wayne Eckerson has been a thought leader in the data warehousing, business intelligence (BI) and performance management fields since 1995. He has conducted numerous in-depth research studies and is the author of the best-selling book *Performance Dashboards: Measuring, Monitoring, and Managing Your Business*. He is a noted keynote speaker and blogger and he consults and conducts workshops on business analytics, performance dashboards and BI, among other topics. For many years, Eckerson served as director of education and research at The Data Warehousing Institute, where he oversaw the company's content and training programs and chaired its BI Executive Summit.

Eckerson is currently director of research at TechTarget, where he writes a popular weekly blog called Wayne's World, which focuses on industry trends and examines best practices in the application of BI (www.b-eye-network.com/blogs/eckerson). Eckerson is also president of BI Leader Consulting (www.bi leader.com) and founder of BI Leadership Forum (www.bileadership.com), a network of BI directors who meet regularly to exchange ideas about best practices in BI and educating the larger BI community. He can be reached at weckerson@techtarget.com.

ADVIZOR SOLUTIONS, INC.

- **Agile BI...What is it? (Webinar)**

- **Visual Reporting and Analysis - Use Case Examples (Webinar and TDWI Report)**

- **Predictive Analytics is for YOU!! (Webinar and TDWI Report)**

ORACLE®

- **Oracle Exalytics In-Memory Machine**

- **Customers Discuss the Benefits of Oracle Exalytics**

PARACCEL™

- **Learn more about the key criteria for selecting an analytic platform**

- **Learn more about offloading analytics and reporting from the data warehouse**

- **Learn more about how to deliver unconstrained analytics with an analytic platform**

- **Learn about Pentaho in a 5 minute video**

- **Download a 30-day trial of Pentaho with free technical support**

- **Learn how Delta Dental usues Pentaho to become a data driven company**

- **Levelling the Playing Field: How Companies Use Data for Competitive Advantage**

- **The Intelligence Future: Simple, Seamless, Social**

- **SAP HANA: Helping Businesses Run Better - In Real-Time**

- **Big Data: Harnessing a game-changing asset**

- **AllAnalytics.com - The Community for Data Management, Business Intelligence and Analytics**

- **SAS In-Memory Analytics**

- **Analytics At The Speed Of Thought**

- **Selecting a Visual Analytics Application**

- **A New Approach to Business Intelligence: Rapid-fire BI**

- **Teradata Purpose-Built Platform Family**

- **Teradata Aggregate Designer**

- **Teradata Enterprise Active Data Warehouse**

- **Whitepaper: In-memory Analytical Systems: Perspective, Trade-offs and Implementation**

- **Whitepaper: Dos and Don'ts of Collaborative Meetings**

- **Cloud-based Analytics and Dashboards - Free Trial**