TDWI RESEARCH

TDWI BEST PRACTICES REPORT

FOURTH QUARTER 2011

BIG DATA ANALYTICS

By Philip Russom

tdwi.org

Research Sponsors

Cloudera

EMC Greenplum

IBM

Impetus Technologies

Kognitio

ParAccel

SAND Technology

SAP

SAS

Tableau Software

Teradata

BIG DATA ANALYTICS

By Philip Russom

Table of Contents

About the Author

PHILIP RUSSOM is director of TDWI Research for data management and oversees many of TDWI's research-oriented publications, services, and events. He is a well-known figure in data warehousing and business intelligence, having published over five hundred research reports, magazine articles, opinion columns, speeches, Webinars, and more. Before joining TDWI in 2005, Russom was an industry analyst covering BI at Forrester Research and Giga Information Group. He also ran his own business as an independent industry analyst and BI consultant and was a contributing editor with leading IT magazines. Before that, Russom worked in technical and marketing positions for various database vendors. You can reach him at prussom@tdwi.org, @prussom on Twitter, and on LinkedIn at linkedin.com/in/philiprussom.

About TDWI

TDWI, a division of 1105 Media, Inc., is the premier provider of in-depth, high-quality education and research in the business intelligence and data warehousing industry. TDWI is dedicated to educating business and information technology professionals about the best practices, strategies, techniques, and tools required to successfully design, build, maintain, and enhance business intelligence and data warehousing solutions. TDWI also fosters the advancement of business intelligence and data warehousing research and contributes to knowledge transfer and the professional development of its members. TDWI offers a worldwide membership program, five major educational conferences, topical educational seminars, role-based training, onsite courses, certification, solution provider partnerships, an awards program for best practices, live Webinars, resourceful publications, an in-depth research program, and a comprehensive Web site: tdwi.org.

About the TDWI Best Practices Reports Series

This series is designed to educate technical and business professionals about new business intelligence technologies, concepts, or approaches that address a significant problem or issue. Research for the reports is conducted via interviews with industry experts and leading-edge user companies and is supplemented by surveys of business intelligence professionals.

To support the program, TDWI seeks vendors that collectively wish to evangelize a new approach to solving business intelligence problems or an emerging technology discipline. By banding together, sponsors can validate a new market niche and educate organizations about alternative solutions to critical business intelligence issues. Please contact TDWI Research Director Philip Russom (prussom@tdwi.org) to suggest a topic that meets these requirements.

Acknowledgments

TDWI would like to thank many people who contributed to this report. First, we appreciate the many users who responded to our survey, especially those who responded to our requests for phone interviews. Second, our report sponsors, who diligently reviewed outlines, survey questions, and report drafts. Finally, we would like to recognize TDWI's production team: Jennifer Agee, Bill Grimmer, and Denelle Hanlon.

Sponsors

Cloudera, EMC Greenplum, IBM, Impetus Technologies, Kognitio, ParAccel, SAND Technology, SAP, SAS, Tableau Software, and Teradata sponsored the research for this report.

Research Methodology and Demographics

Report Scope. According to TDWI survey data, a new flood of user organizations is currently commencing or expanding solutions for analytics with big data. To supply the demand, vendors have recently released numerous new products and functions, specifically for advanced forms of analytics (beyond OLAP and reporting) and analytic databases that can manage big data. While it's good to have options, it's hard to track them and determine in which situations they are ready for use. The purpose of this report is to accelerate users' understanding of the many new tools and techniques that have emerged for analytics with big data in recent years. It will also help readers map newly available options to real-world use cases.

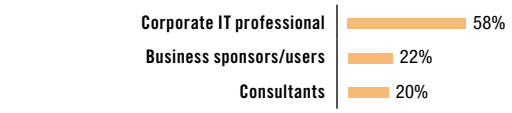
Survey Methodology. In May 2011, TDWI sent an invitation via e-mail to the data management professionals in its database, asking them to complete an Internet-based survey. The invitation was also distributed via Web sites, newsletters, and publications from TDWI and other firms. The survey drew responses from almost 360 survey respondents. From these, we excluded incomplete responses and respondents who identified themselves as academics or vendor employees. The resulting completed responses of 325 respondents form the core data sample for this report.

Survey Demographics. The majority of survey respondents are corporate IT professionals (58%), whereas the others are business sponsors or users (22%) and consultants (20%). We asked consultants to fill out the survey with a recent client in mind.

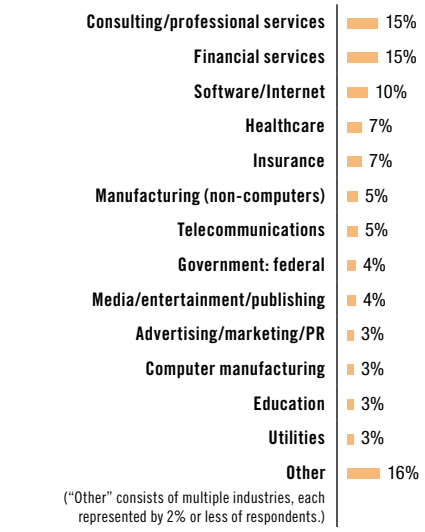
The consulting (15%) and financial services (15%) industries dominate the respondent population, followed by software (10%), healthcare (7%), insurance (7%), and other industries. Most survey respondents reside in the U.S. (56%) or Europe (17%). Respondents are fairly evenly distributed across all sizes of companies and other organizations.

Other Research Methods. In addition to the survey, TDWI Research conducted many telephone interviews with technical users, business sponsors, and recognized data management experts. TDWI also received product briefings from vendors that offer products and services related to the best practices under discussion.

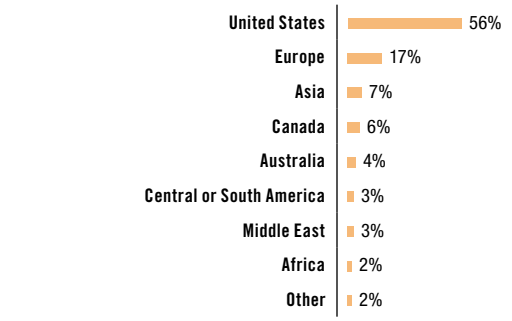
Position



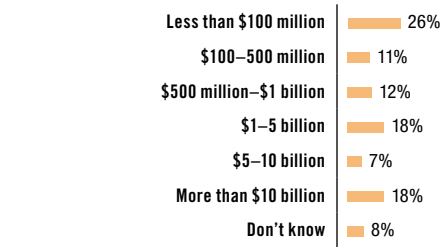
Industry



Geography



Company Size by Revenue



Based on 325 survey respondents.

Executive Summary

Big data used to be a technical problem. Now it's a business opportunity.

Oddly enough, big data was a serious problem just a few years ago. When data volumes started skyrocketing in the early 2000s, storage and CPU technologies were overwhelmed by the numerous terabytes of big data—to the point that IT faced a data scalability crisis. Then we were once again snatched from the jaws of defeat by Moore's law. Storage and CPUs not only developed greater capacity, speed, and intelligence; they also fell in price. Enterprises went from being unable to afford or manage big data to lavishing budgets on its collection and analysis.

Today, enterprises are exploring big data to discover facts they didn't know before. This is an important task right now because the recent economic recession forced deep changes into most businesses, especially those that depend on mass consumers. Using advanced analytics, businesses can study big data to understand the current state of the business and track still-evolving aspects such as customer behavior.

Big data is not just big. It's also diverse data types and streaming data.

If you really want the lowdown on what's happening in your business, you need large volumes of highly detailed data. If you truly want to see something you've never seen before, it helps to tap into data that's never been tapped for business intelligence (BI) or analytics. Some of the untapped data will be foreign to you, coming from sensors, devices, third parties, Web applications, and social media. Some big data sources feed data unceasingly in real time. Put all that together, and you see that big data is not just about giant data volumes; it's also about an extraordinary diversity of data types, delivered at various speeds and frequencies.

Big data analytics is the application of advanced analytic techniques to very big data sets.

Note that two technical entities have come together. First, there's big data for massive amounts of detailed information. Second, there's advanced analytics, which is actually a collection of different tool types, including those based on predictive analytics, data mining, statistics, artificial intelligence, natural language processing, and so on. Put them together and you get big data analytics, the hottest new practice in BI today.

Of course, businesspeople can learn a lot about the business and their customers from BI programs and data warehouses. But big data analytics explores granular details of business operations and customer interactions that seldom find their way into a data warehouse or standard report. Some organizations are already managing big data in their enterprise data warehouses (EDWs), while others have designed their DWs for the well-understood, auditable, and squeaky clean data that the average business report demands. The former tend to manage big data in the EDW and execute most analytic processing there, whereas the latter tend to distribute their efforts onto secondary analytic platforms. There are also hybrid approaches.

There are many types of vendor products to consider for big data analytics. This report discusses the types.

Regardless of approach, user organizations are currently reevaluating their analytic portfolios. In response to the demand for platforms suited to big data analytics, vendors have released a slew of new product types including analytic databases, data warehouse appliances, columnar databases, no-SQL databases, distributed file systems, and so on. There is also a new slew of analytic tools.

This report drills into all the aspects of big data analytics mentioned here to give users and their business sponsors a solid background for big data analytics, including business and technology drivers, successful business use cases, and common technology enablers. The report also uses survey data to project the future of the most common tool types, features, and functions associated with big data analytics, so users can apply this information to planning their own programs and technology stacks for big data analytics.

Introduction to Big Data Analytics

Big data analytics is where advanced analytic techniques operate on big data sets. Hence, big data analytics is really about two things—big data and analytics—plus how the two have teamed up to create one of the most profound trends in business intelligence (BI) today. Let's start by defining advanced analytics, then move on to big data and the combination of the two.

Defining Advanced Analytics as a Discovery Mission

In the last three years or so, many organizations have deployed analytics for the first time.

According to a 2009 TDWI survey, 38% of organizations surveyed reported practicing advanced analytics, whereas 85% said they would be practicing it within three years.[1] Why the rush to advanced analytics? First, change is rampant in business, as seen in the multiple "economies" we've gone through in recent years. Analytics helps us discover what has changed and how we should react. Second, as we crawl out of the recession and into the recovery, there are more and more business opportunities that should be seized. To that end, advanced analytics is the best way to discover new customer segments, identify the best suppliers, associate products of affinity, understand sales seasonality, and so on. For these reasons, TDWI has seen a steady stream of user organizations implementing analytics in recent years.

The rush to analytics means that many organizations are embracing advanced analytics for the first time, and hence are confused about how to go about it. Even if you have related experience in data warehousing, reporting, and online analytic processing (OLAP), you'll find that the business and technical requirements are different for advanced forms of analytics. To help user organizations select the right form of analytics and prepare big data for analysis, this report will discuss new options for advanced analytics and analytic databases for big data so that users can make intelligent decisions as they embrace analytics.

Note that user organizations are implementing specific forms of analytics, particularly what is sometimes called advanced analytics. This is a collection of related techniques and tool types, usually including predictive analytics, data mining, statistical analysis, and complex SQL. We might also extend the list to cover data visualization, artificial intelligence, natural language processing, and database capabilities that support analytics (such as MapReduce, in-database analytics, in-memory databases, columnar data stores).

"Discovery analytics" is a more descriptive term than "advanced analytics."

Instead of "advanced analytics," a better term would be "discovery analytics," because that's what users are trying to accomplish. (Some people call it "exploratory analytics.") In other words, with big data analytics, the user is typically a business analyst who is trying to discover new business facts that no one in the enterprise knew before. To do that, the analyst needs large volumes of data with plenty of detail. This is often data that the enterprise has not yet tapped for analytics.

For example, in the middle of the recent economic recession, companies were constantly being hit by new forms of customer churn. To discover the root cause of the newest form of churn, a business analyst would grab several terabytes of detailed data drawn from operational applications to get a view of recent customer behaviors. The analyst might mix that data with historic data from a data warehouse. Dozens of queries later, the analyst would discover a new churn behavior in a subset of the customer base. With any luck, that discovery would lead to a metric, report, analytic model, or some other product of BI, through which the company could track and predict the new form of churn.

Discovery analytics against big data can be enabled by different types of analytic tools, including those based on SQL queries, data mining, statistical analysis, fact clustering, data visualization,

[1] See the TDWI Best Practices Report *Next Generation Data Warehouse Platforms* (Q4 2009), available on tdwi.org.

natural language processing, text analytics, artificial intelligence, and so on. It's quite an arsenal of tool types, and savvy users get to know their analytic requirements before deciding which tool type is appropriate to their needs.

All these techniques have been around for years, many of them appearing in the 1990s. The difference today is that far more user organizations are actually using them. That's because most of these techniques adapt well to very large, multi-terabyte data sets with minimal data preparation. That brings us to big data.

Defining Big Data Via the Three Vs

Big data isn't just about data volume.

Most definitions of big data focus on the size of data in storage. Size matters, but there are other important attributes of big data, namely data *variety* and data *velocity*. The three Vs of big data (volume, variety, and velocity) constitute a comprehensive definition, and they bust the myth that big data is only about data volume. In addition, each of the three Vs has its own ramifications for analytics.[2] (See Figure 1.)

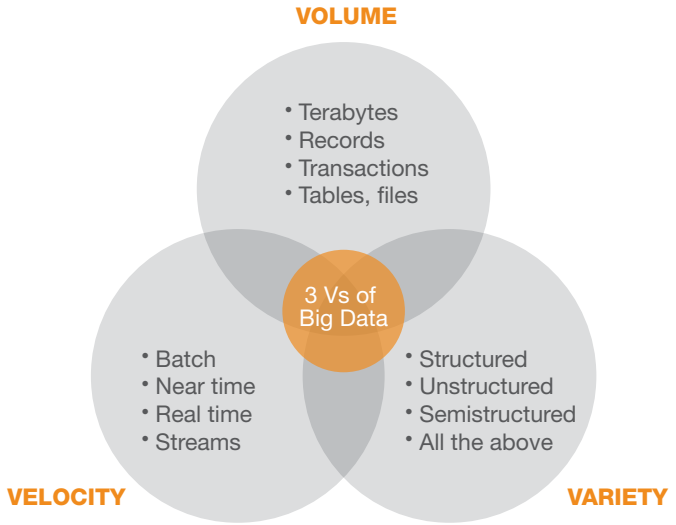


Figure 1. The three Vs of big data

Data volume as a defining attribute of big data.

It's obvious that data volume is the primary attribute of big data. With that in mind, most people define big data in terabytes—sometimes petabytes. For example, a number of users interviewed by TDWI are managing 3 to 10 terabytes (TB) of data for analytics. Yet, big data can also be quantified by counting records, transactions, tables, or files. Some organizations find it more useful to quantify big data in terms of time. For example, due to the seven-year statute of limitations in the U.S., many firms prefer to keep seven years of data available for risk, compliance, and legal analysis.

The scope of big data varies widely.

The scope of big data affects its quantification, too. For example, in many organizations, the data collected for general data warehousing differs from data collected specifically for analytics. Different forms of analytics may have different data sets. Some analytic practices lead a business analyst or similar user to create ad hoc analytic data sets per analytic project. Then, there's the entire enterprise, which *in toto* has its own, even larger scope of big data. Furthermore, each of these

[2] These definitions of big data were originally developed in TDWI blog posts, available at tdwi.org/blogs/philip-russom.

quantifications of big data grows continuously. All this makes big data for analytics a moving target that's tough to quantify.

USER STORY THERE ARE VARIOUS WAYS TO QUANTIFY BIG DATA.

TDWI asked a user how many terabytes he's managing for analytics, and he said: "I don't know, because I don't have to worry about storage. IT provides it generously, and I tap it like crazy." Another user said: "We don't count terabytes. We count records. My analytic database for quality assurance alone has 3 billion records. There's another 3 billion in other analytic databases."

Data type variety as a defining attribute of big data.

Big data is remarkably diverse in terms of sources, data types, and entities represented.

One of the things that makes big data really big is that it's coming from a greater variety of sources than ever before. Many of the newer ones are Web sources, including logs, clickstreams, and social media. Sure, user organizations have been collecting Web data for years. But, for most organizations, it's been a kind of hoarding. We've seen similar untapped big data collected and hoarded, such as RFID data from supply chain applications, text data from call center applications, semistructured data from various business-to-business processes, and geospatial data in logistics. What's changed is that far more users are now analyzing big data instead of merely hoarding it. The few organizations that have been analyzing this data now do so at a more complex and sophisticated level. Big data isn't new, but the effective analytical leveraging of big data is.

The recent tapping of these sources for analytics means that so-called structured data (which previously held unchallenged hegemony in analytics) is now joined by unstructured data (text and human language) and semistructured data (XML, RSS feeds). There's also data that's hard to categorize, as it comes from audio, video, and other devices. Plus, multidimensional data can be drawn from a data warehouse to add historic context to big data. That's a far more eclectic mix of data types than analytics has ever seen. So, with big data, variety is just as big as volume. In addition, variety and volume tend to fuel each other.

USER STORY HADOOP IS ABOUT DATA VARIETY, NOT JUST DATA VOLUME.

TDWI found a couple of users who have employed Hadoop as an analytic platform. Both said the same thing: Hadoop's scalability for big data volumes is impressive, but the real reason they're working with Hadoop is its ability to manage a very broad range of data types in its file system, plus process analytic queries via MapReduce across numerous eccentric data types. It's not just Hadoop; TDWI has heard users make similar comments about other analytic platforms.

Data feed velocity as a defining attribute of big data.

The leading edge of big data is streaming data.

Big data can be described by its velocity or speed. You may prefer to think of it as the frequency of data generation or the frequency of data delivery. For example, think of the stream of data coming off of any kind of device or sensor, say robotic manufacturing machines, thermometers sensing temperature, microphones listening for movement in a secure area, or video cameras scanning for a specific face in a crowd. The collection of big data in real time isn't new; many firms have been collecting clickstream data from Web sites for years, using streaming data to make purchase recommendations to Web visitors. With sensor and Web data flying at you relentlessly in real time, data volumes get big in a hurry. Even more challenging, the analytics that go with streaming data have to make sense of the data and possibly take action—all in real time.

USER STORY **PROCESSING STREAMING BIG DATA ENABLES NEW ANALYTIC APPLICATIONS.**

A consultant who specializes in streaming data told TDWI about the video and audio analytic applications he's looking into: "Think about the algorithms that enable us to parse text and perform sentiment analysis, sometimes in real time. Very similar algorithms can parse video images to document and analyze changes in the thing that's being imaged. For example, satellite images could monitor and analyze troop movements, a flood plane, cloud patterns, or grass fires. Or a video analysis system could monitor a sensitive or valuable facility, watching for possible intruders, then alert authorities in real time.

"You can implement similar applications with sound monitoring. One of my analytic applications involves 2,000 underground microphones that listen for movement in geologic formations. I hope that the big data the application is collecting can eventually help predict earthquakes."

Defining Big Data Analytics

Most users are familiar with big data analytics but don't use the term.

Again, *big data analytics* is where advanced analytic techniques operate on big data. The definition is easy to understand, but do users actually use the term? To quantify this question, the survey for this report asked: "Which of the following best characterizes your familiarity with big data analytics and how you name it?" (See Figure 2.) The survey results show that most users understand the concept of big data analytics, whether they have a name for it or not:

Few respondents are unfamiliar with the concept. Only 7% report that they "haven't seen or heard of anything resembling big data analytics."

Most users surveyed don't have a name for big data analytics. Even so, they understand the definition (65% of respondents).

Roughly a quarter of respondents have a name for big data analytics. Twenty-eight percent both understand the concept and have named it.

Which of the following best characterizes your familiarity with big data analytics and how you name it?

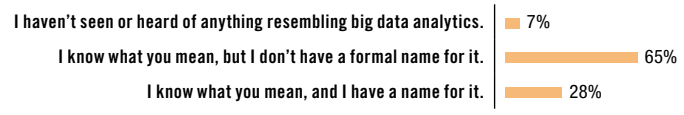


Figure 2. Based on 325 respondents.

When users have a term, it's most often "big data analytics."

Most of the survey respondents who report having a name for big data analytics typed the name they use into the survey software. The name entered most often is the term used in this report: "big data analytics" (18% in Figure 3). Similar terms appeared, such as large-volume or large-data-set analytics (7%). Many use the popular term advanced analytics (12%), or they simply call it analytics (12%). A few common terms were entered, such as data warehousing (4%), data mining (2%), and predictive analytics (2%). A whopping 43% entered a unique name, showing that names for analytic methods are amazingly diverse.

Finally, a few survey respondents entered humorous but revealing terms such as honking big data, my day job, pain in the neck, and we-need-to-buy-more-hardware analytics.

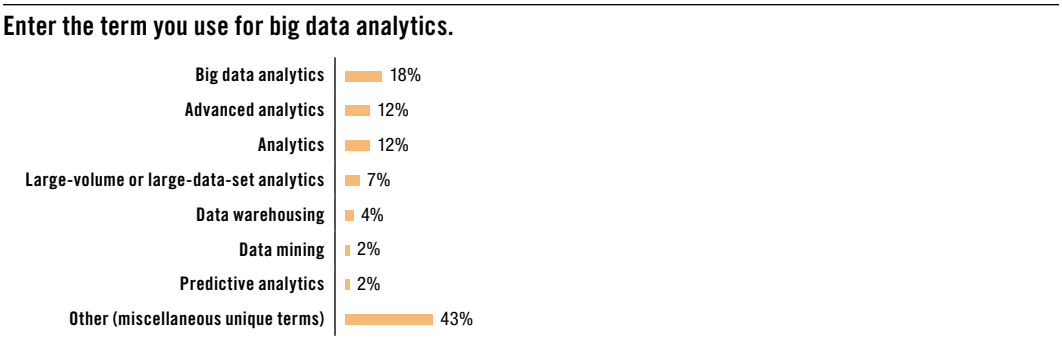


Figure 3. Based on 92 respondents who report having a name for big data analytics.

Why Put Big Data and Analytics Together Now?

Analytic platforms today handle big data better than ever.

Big data provides gigantic statistical samples, which enhance analytic tool results. Most tools designed for data mining or statistical analysis tend to be optimized for large data sets. In fact, the general rule is that the larger the data sample, the more accurate are the statistics and other products of the analysis. Instead of using mining and statistical tools, many users generate or hand-code complex SQL, which parses big data in search of just the right customer segment, churn profile, or excessive operational cost. The newest generation of data visualization tools and in-database analytic functions likewise operate on big data.

Analytic tools and databases can now handle big data. They can also execute big queries and parse tables in record time. Recent generations of vendor tools and platforms have lifted us onto a new plateau of performance that is very compelling for applications involving big data.

The economics of analytics is now more embraceable than ever. This is due to a precipitous drop in the cost of data storage and processing bandwidth. The fact that tools and platforms for big data analytics are relatively affordable is significant because big data is not just for big business. Many small-to-midsize businesses (especially those deep into digital processes for sales, customer interactions, or supply chain) also need to manage and leverage big data.

Big data is an enterprise asset that yields actionable business insights.

There's a lot to learn from messy data, as long as it's big. Most modern tools and techniques for advanced analytics and big data are very tolerant of raw source data, with its transactional schema, non-standard data, and poor-quality data. That's a good thing, because discovery and predictive analytics depend on lots of details—even questionable data. For example, analytic applications for fraud detection often depend on outliers and non-standard data as indications of fraud. So, be careful: If you apply ETL and data quality processes to big data as you do for a data warehouse, you run the risk of stripping out the very nuggets that make big data a treasure trove for advanced analytics.[3]

Big data is a special asset that merits leverage. That's the real point of big data analytics. The new technologies and new best practices are fascinating, even mesmerizing, and there's a certain macho coolness to working with dozens of terabytes. But don't do it for the technology. Put big data and discovery analytics together for the new insights they give the business.

Analytics based on large data samples reveals and leverages business change. The recession has accelerated the already quickening pace of business. The recovery, though welcome, brings even more change. In fact, the average business has changed beyond all recognition because of the recent economic recession and recovery. The change has not gone unnoticed. Businesspeople now share a wholesale recognition that they must explore change just to understand the new state of the business.

[3] The preparation of big data for advanced analytics rarely follows the same best practices we associate with mainstream data warehousing, reporting, and OLAP. To understand the differences, see the TDWI Checklist Report *Data Requirements for Advanced Analytics*, available on tdwi.org.

Even more compelling, however, is the prospect of discovering problems that need fixing (such as new forms of customer churn and competitive pressure) and opportunities that merit leverage (such as new customer segments and sales prospects).

The State of Big Data Analytics

Big Data Analytics Adoption

Advanced analytics is common, and big data analytics has a good presence.

Big data analytics is a fast-growing and influential practice. But how many user organizations are actually doing it? To find out, this report's survey asked respondents: "Does your organization execute advanced analytics against big data today?" (See Figure 4.)

Advanced analytics is fairly common today. Roughly three quarters (74%) of organizations surveyed have adopted some form of analytics today, regardless of the analytic method or tool type, whether with big data or not. This reveals a strong adoption of advanced analytics, which isn't a surprise, given that it's been around for at least 15 years. (Later, Figure 16 will reveal which analytic methods are the most common today.)

Analytics doesn't require big data. The two get jammed into the same sentence so much lately that we forget that they don't have to go together. In fact, 40% of survey respondents practice advanced analytics without big data.

One-third of organizations (34%) do big data analytics today, although it's new. In other words, they practice some form of advanced analytics, and they apply it to big data. This is a respectable presence for big data analytics, given the newness of the combination of advanced analytics and big data.

Does your organization execute advanced analytics against big data today?

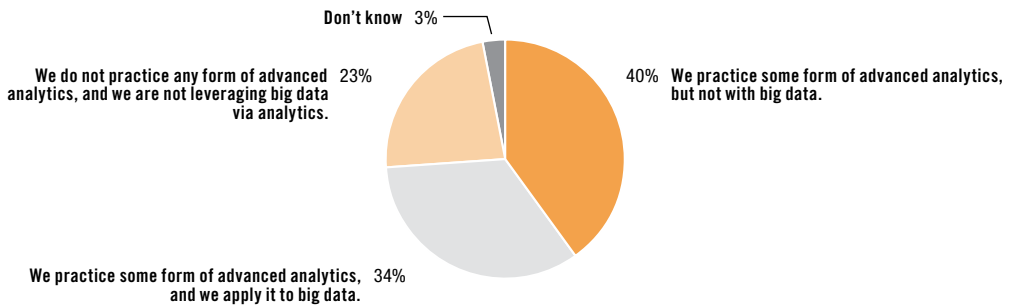


Figure 4. Based on 325 respondents.

Benefits of Big Data Analytics

Big data analytics can benefit customer relations, business intelligence, and many analytic applications.

We just saw that user organizations have adopted big data analytics in appreciable numbers. To determine the potential benefits that are driving the adoption, TDWI's survey asked: "Which of the following benefits would ensue if your organization implemented some form of big data analytics?" The most likely benefits (seen at the top of Figure 5) are those most often selected by survey respondents, and the likelihood of a benefit declines as the list proceeds downward.

Anything involving customers could benefit from big data analytics. Near the top of the list (in Figure 5), this includes better-targeted social-influencer marketing (61%), customer-base segmentation (41%), and recognition of sales and market opportunities (38%). Recent economic changes worldwide have changed consumer behaviors. Big data analytics can help develop definitions of churn and other customer behaviors (35%), as well as an understanding of consumer behavior from clickstreams (27%).

Business intelligence in general can benefit from big data analytics. This could result in more numerous and accurate business insights (45%), an understanding of business change (30%), better planning and forecasting (29%), and the identification of root causes of cost (29%).

Specific analytic applications are likely beneficiaries of big data analytics. For example, consider analytic applications for the detection of fraud (33%), the quantification of risks (30%), or market sentiment trending (30%). At the leading edge, big data analytics might help automate decisions for real-time business processes such as loan approvals or fraud detection (37%).

Potential benefits entered by survey respondents selecting "other" include customer loyalty, service experience optimization, healthcare delivery optimization, and supplier performance based on cost and quality.

Which of the following benefits would ensue if your organization implemented some form of big data analytics? (Select five or fewer.)

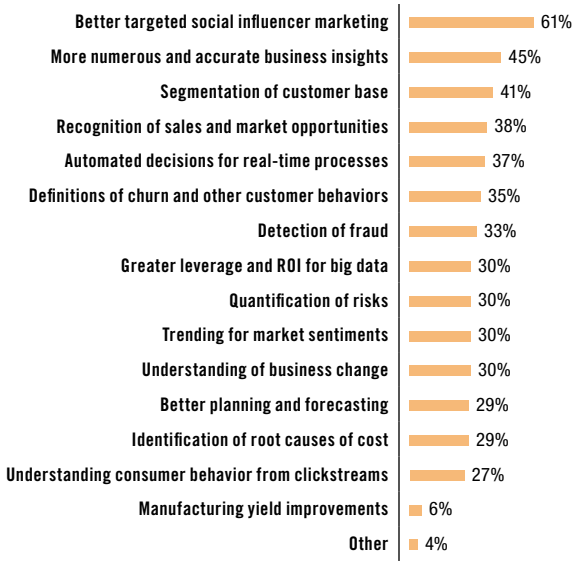


Figure 5. Based on 1,635 responses from 325 respondents; 5 responses per respondent, on average.

Barriers to Big Data Analytics

Problems with skills, sponsors, and software are the leading barriers.

Big data analytics has its benefits, as we just saw. Yet, it also has barriers. To get a sense of which barriers are more likely than others, this report's survey asked: "In your organization, what are the top potential barriers to implementing big data analytics?" The most likely barriers (seen at the top of Figure 6) are those most often selected by survey respondents, and the likelihood of a barrier declines as the list proceeds downward.

Inadequate staffing and skills are the leading barriers to big data analytics (46%). After all, many organizations are still new to big data analytics. And its skill set is not quite the same as that for business intelligence and data warehousing, for which most organizations have developed their skills. Other skill-related barriers include the difficulty of architecting a big data analytic system (33%) and problems with making big data usable for end users (22%).

A lack of business support can hinder a big data analytics program. Survey respondents pointed to a lack of business sponsorship (38%) and a lack of a compelling business case (28%), plus the related issue of overall cost (42%).

Problems with database software can be barriers to big data analytics. Issues arise when the current database software lacks in-database analytics (32%), has scalability problems with big data (23%), can't process analytic queries fast enough (22%), or cannot load data fast enough (21%). In a related issue, managing big data in a data warehouse is challenging when that warehouse is modeled for reports and OLAP only (22%).

Possible barriers entered by survey respondents selecting "other" include competing with other initiatives, lack of test and control rigor, and sourcing and rationalizing big data from multiple systems.

In your organization, what are the top potential barriers to implementing big data analytics? (Select five or fewer.)

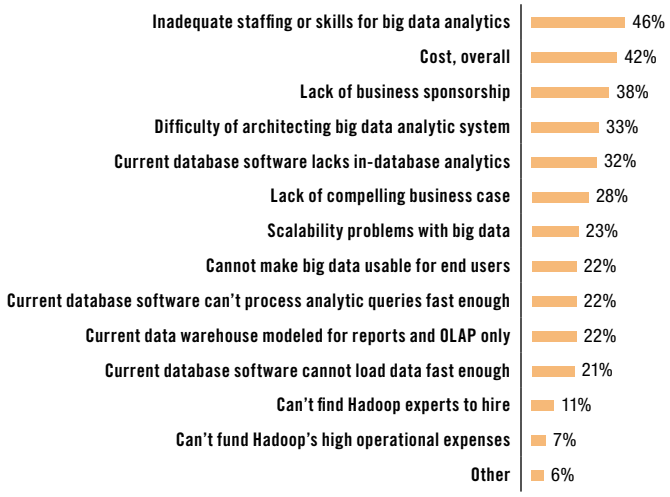


Figure 6. Based on 1,153 responses from 325 respondents; 3.5 responses per respondent, on average.

Big Data: Problem or Opportunity?

Big data is mostly an opportunity, not a problem.

TDWI has seen many user organizations emerge only recently from a scalability crisis where big data was more of a curse than a blessing. With that in mind, we asked: "In your organization, is big data considered mostly a problem or mostly an opportunity?" (See Figure 7.)

Only 30% consider big data a problem. There's no doubt that big data presents technical challenges due to its volume, variety, and velocity. Data volume alone is a showstopper for some organizations.

The vast majority (70%) considers big data an opportunity. Through exploratory, detailed analyses of big data, a user organization can discover new facts about their customers, markets, partners, costs, and operations—then use that information for business advantage.

In your organization, is big data considered mostly a problem or mostly an opportunity?

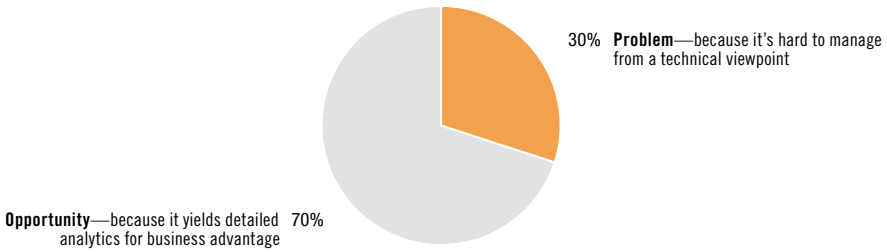


Figure 7. Based on 325 responses

USER STORY ADOPTION OF BIG DATA ANALYTICS IS DRIVEN BY A PERFECT STORM OF TECHNOLOGIES, BUSINESS MANAGEMENT, AND ECONOMICS.

"It's been kind of like a perfect storm, where several things came together to make big data analytics practical and affordable," said a BI director interviewed by TDWI. "For one thing, there are more analytic products to consider now. The new analytic databases are pretty affordable. And they're designed specifically for analytics. For another thing, firms have started thawing budgets—which the recession had frozen—so there's more corporate money to spend on analytics. And there's been an attitude shift that's hard to pinpoint; it seems like more managers today are convinced of the management power of analytics.

"I think that an even more beneficial change has been that key infrastructure pieces—like data storage, CPUs, memory, and network bandwidth—have both gotten better and come down in price. Oddly enough, scaling up to big data was a serious problem just a few years ago. We survived the scalability crisis by buying cheap infrastructure, such that big data is now a good thing instead of a problem. Now, if we could just solve the real-time data processing crisis, we'd be all set!"

Organizational Issues

Ownership and Control of Big Data Analytics

Analytics is usually owned by a BI/DW team; less often by a department.

The most common owner of big data analytics is the BI/DW team (41% in Figure 8). This is no surprise, since the majority of organizations centralize as many business intelligence (BI) and data warehouse (DW) functions as possible through a single team. Furthermore, several survey respondents selected "other" and explained that they have a separate analytics team, although it's managed through the BI/DW team.

Occasionally a department owns and controls big data analytics (21%). Although this flies in the face of centralization best practices for enterprise business intelligence (EBI) and the enterprise data warehouse (EDW), there are good reasons why some departments go rogue with their own team and platforms for analytics, as explained in the next section of this report.

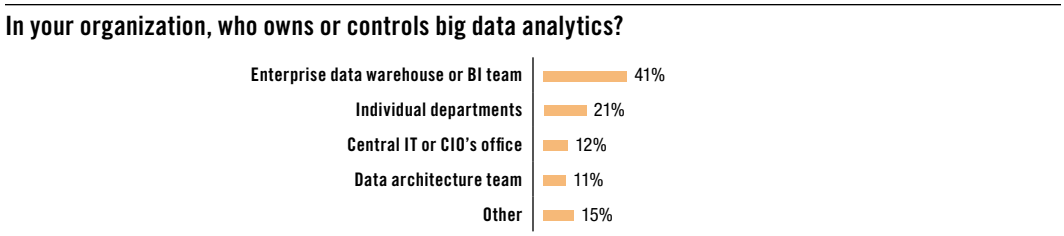


Figure 8. Based on 109 respondents who report practicing big data analytics.

Big Data Analytics Can Have a Departmental Focus

Much of the action in big data analytics is at the departmental level.

Analytic applications are departmental by nature. Just about any analytic application you think of is focused on tasks, data domains, and business opportunities that are associated with specific departments. For example, customer base segmentation should be owned and executed by marketing and sales departments. The actuarial department does risk analysis. The procurement department does supply and supplier analysis. Hence, the average analytic application satisfies departmental requirements, not enterprise ones, even if implemented by an enterprise team.

Not all BI/DW technology stacks are designed for advanced analytics. In most organizations, users have designed and optimized their stacks for reporting, performance management, and OLAP. This is natural, since these are the most common deliverables that must come out of a BI/DW solution. Furthermore, this optimization is invaluable for "big picture" reports and analyses that span enterprisewide processes (especially financial ones). Such stacks are also capable of satisfying most departmental requirements for reporting and OLAP. But in many organizations, the BI/DW technology stack is simply not designed to satisfy departmental requirements for advanced analytics and big data. Note that this limitation is due to a conscious design decision that users made, not the failure of a vendor product.

Some departments deploy their own platforms for big data and analytics. They do this when the department has a strong business need for analytics with big data, plus the budget and sponsorship to back it up. In summary, in some organizations (approximately 21%), big data analytics is a departmental affair, implemented by the department's team on a departmentally owned platform.

Job Titles for Big Data Analytics

With big data analytics, the most common job titles are business analyst and data analyst.

Analysts design and execute analytics. Of course, that's why they're called analysts. This includes the popular job title business analyst (14% in Figure 9) and the more-or-less equivalent title data analyst (6%). Note that a relatively large number of survey respondents entered director and manager titles for analytics (14%), indicating that analytic teams exist and that these teams have managers.

Architects are regularly involved in analytics. We can see this from survey responses for data architect (10%) and the related title data scientist (4%).

Engineers and researchers use analytics. People who develop products need analytics, as seen in the job titles engineer (10%), research analyst (3%), and R&D specialist (2%).

BI professionals do analytics. This includes BI directors (5%) and BI specialists (3%).

A wide range of people are involved in the design and execution of analytics. Almost one-third of survey responses (29%) were a mixed bag, describing marketers, consultants, statisticians, data governors, risk managers, and so on. This breadth of job titles is significant because it shows that analytics is not

just for an analytic specialist. On the contrary, analytics is becoming a standard competency for a wide range of business and technology people.

Enter the job titles of people who design and execute advanced analytics in your organization.

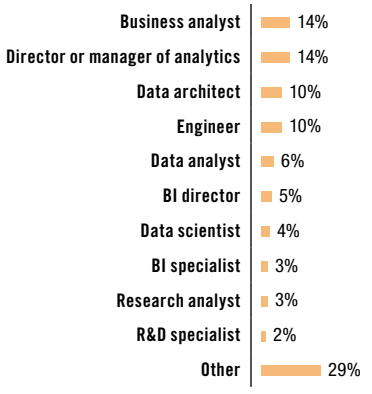


Figure 9. Based on 113 responses from 109 respondents who report practicing big data analytics.

Best Practices in Big Data Analytics

Although measured in terabytes today, big data for analytics is on the cusp of petabytes.

Volume Growth of Analytic Big Data

In recent years, TDWI has observed a strong trend toward the application of advanced analytics to very large data sets. To help quantify the perceived trend, the survey asked: What's the approximate total data volume that your organization manages only for analytics, both today and in three years? (See Figure 10.) Due to branching in the survey, the question was posed only to survey respondents who reported (in their responses to earlier survey questions) that they work for an organization that is already practicing big data analytics. In other words, these people speak from experience, not mere opinion. Note that they are describing big data only for analytics, not for BI, DW, or other enterprise applications.

One-third of organizations surveyed have already broken the 10 TB barrier (37%). In fact, the 10- to 100-terabyte range received more responses than any other, making it the norm for today's volume of big data specifically applied to advanced analytics.

Smaller analytic data sets will become less common as they grow into larger ones. In forecasting analytic data volumes for three years from now, survey respondents project considerably fewer analytic data sets in the 1 TB, 1–3 TB, and 3–10 TB ranges. A TDWI Technology Survey run in August 2011 returned almost identical results.

Large analytic data sets will become more common. The number of very large analytic data sets will triple or quadruple in the 100–500 TB and 500+ TB ranges. Clearly, users conduct advanced analytics with ever-larger analytic data sets.

Today almost all user organizations quantify analytic big data in terabytes. On its current trajectory, however, big data for analytics will soon cross into petabytes, which will become the prevailing unit of measure. Growth is the nature of big data, and user organizations should conduct periodic capacity planning exercises to ensure they can continue to scale up and perform with big data analytics.

What's the approximate total data volume that your organization manages *only* for analytics, both today and in three years?

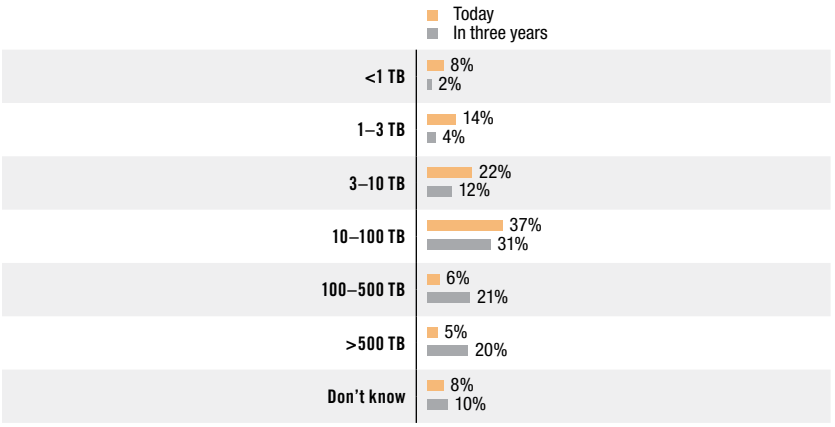


Figure 10. Based on 109 respondents who report practicing big data analytics.

USER STORY **BIG DATA IS OFTEN FILTERED DOWN TO SMALL, SIGNIFICANT DATA SETS.**

A BI professional at a prominent Internet-based business told TDWI: "We load 200 GB a day into our data warehouse. But that's processed down from several terabytes of Web log and clickstream data. We mix this big data with data about our customers drawn from other touch points, then analyze it. Although the Web data is streaming, we collect the stream on disk, then process it down and analyze it overnight. Our next step is to process and analyze streaming big data in real time in hopes of influencing customer behavior in mid-process. We're definitely a customer-oriented business, so understanding customers and serving them better is the goal of analytics. We just need to do it both after the fact in batch and—eventually—in real time."

Managing Analytic Big Data

Manage big data in the EDW? Buy a vendor's analytic database? Adopt a new platform?

One of the big questions for big data is: Where should you manage it and operate on it? After all, there are many vendor-built database platforms available today that can handle large analytic data sets. And user organizations have diverse business and technology requirements that lead them to equally diverse designs, models, and architectures.

To get a sense of what users think about these issues, this report's survey asked a pair of related questions: Today, where is big data for advanced analytics managed and operated on? Where would you prefer that big data for advanced analytics be managed and operated on? (See Figure 11.)

The EDW is a much-used and much-preferred platform for analytics. Roughly two-thirds of users surveyed report using an EDW today (64%), and two-thirds say it's a preference (63%). As pointed out elsewhere in this report, some EDWs were originally designed by users for reporting, performance management, and OLAP. Because of this, some EDW designs can also handle advanced analytics—in terms of scalability and query performance—and some cannot. Again, this is a issue with users' designs, not a problem with vendor products. Then there are questions of sponsorship, funding, and enterprise versus departmental analytic requirements. Whether to manage analytic big data in a shared, centralized EDW or in a separate-but-related database is one of the basic architectural decisions users face when starting or expanding a program for big data analytics.

The tradition of analytic databases outside the EDW proper continues, but with a twist. On the one hand, survey respondents anticipate reducing the number of data marts and operational data stores (ODSs). This is no surprise, since we're all trained to keep the proliferation of these in check. On the other hand, the survey shows that users are already using vendor databases designed for big data analytics (28%), and these are considered preferable (30%). Users interviewed by TDWI described their use of new vendor-built analytic databases as being akin to an old-fashioned data mart, but with far greater data volume, detailed data, and data type diversity. Only time will tell whether new analytic databases will lead to the proliferation abuses we associate with data marts.

New types of analytic platforms are coming. A few users report using cloud-based analytic platforms today, and many more users would prefer them. TDWI expects various types of clouds to become common platforms for analytics within a few years. The survey also shows that Hadoop is already in use by 24% of respondents, which is a respectable presence. Given the open-source nature of Hadoop, TDWI suspects that many of these are simply downloads that are in experimental use. Hadoop is so heavily hyped at the moment that it is difficult to say whether its current experimental use will evolve into a permanent presence in IT.

Other analytic platform choices abound. Some of these were entered by survey responses who selected "other," such as a separate sandbox attached to the EDW, a hierarchy of databases within the EDW environment, an analytic data warehouse, a farm of data warehouse appliances, and hand-coded software.

Today, where is big data for advanced analytics managed and operated on?
Where would you prefer that big data for advanced analytics be managed and operated on?

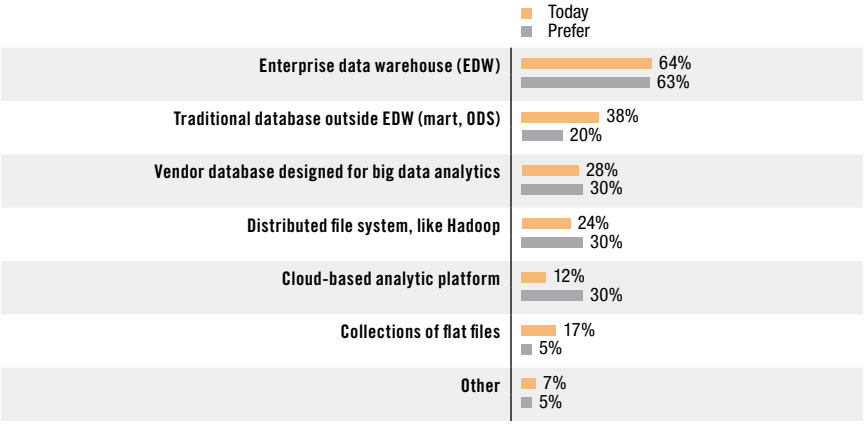


Figure 11. Based on 207 responses from 109 respondents who report practicing big data analytics; 1.9 responses per respondent, on average.

Data Types for Big Data

Remember the three Vs of big data? The second V is data variety, and it's manifested by a growing number of data types that are being managed and analyzed with increasing rapidity.

Structured data still rules, but is slowly joined by many other data types.

Structured data maintains hegemony over other data types. The majority of data handled via analytic platforms today falls under the rubric *structured* data. This is primarily about the tables and other data structures of relational databases. But other sources yield predictable structures, such as the record formats of most applications and the character-delimited rows of many flat files. In our survey, a whopping 92% of respondents report handling structured data today. (See Figure 12.)

Semistructured and complex data are coming on strong. The hegemony of structured data types will eventually be challenged by a wider range of data types. In particular, today 54% of respondents report handling some form of semistructured data (XML and similar standards) or complex data (hierarchical or legacy sources). These data types are driven up by increased use of industry standards (SWIFT, ACORD, HL7) and XML applied to business-to-business data exchange (which tends to be modeled in hierarchies).

Unstructured data (mostly text expressing human language) continues to gain (35%). TDWI has interviewed many of its members who use text mining or text analytics tools to convert facts (discovered in textual documents) into structured data (typically a record or table row per discovered fact). For example, insurance companies regularly extract facts from text gathered in the claims process, then use that data to extend their analytic data sets for risk management and fraud detection.

Web data is finally getting the attention it deserves. Users interviewed by TDWI talked about how for years they didn't have the skills or proper IT platforms for analyzing Web data. Now that they have the skills and the platforms, they're aggressively exploring social media data (blogs, tweets, social networks; 34%) and Web logs and clickstreams (31%).

Real-time data types lag at the moment. But they stand a good chance of becoming more common as real-time technologies continue to improve and to be adopted by user organizations. This includes event data (45%), spatial data (GPS output; 29%), and machine-generated data (from sensors, RFID chips, robots, and various devices; 28%).

Which of the following data types are you collecting as big data and/or using with advanced analytics today? Select all that apply.

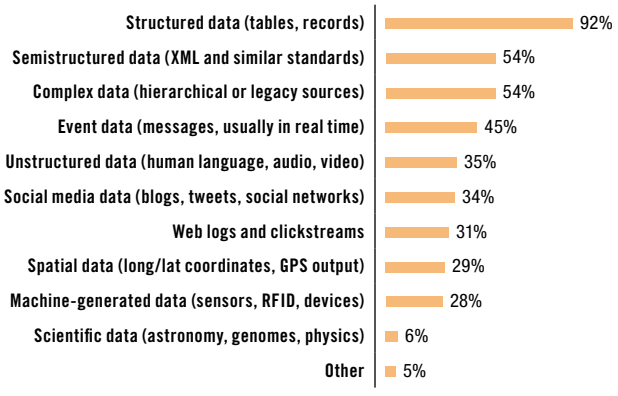


Figure 12. Based on 450 responses from 109 respondents who report practicing big data analytics; 4.1 responses per respondent, on average.

USER STORY MIXING DATA TYPES AND DATA VELOCITIES ENABLES NEW ANALYTIC APPLICATIONS.

Here's the kind of streaming big data application that many people are dreaming of, as described in a TDWI interview: "Imagine I'm a consumer walking around downtown in a city, and I'm shopping. Now imagine letting a shopping service know the kinds of goods I'm looking for. As I walk, my GPS coordinates could stream to the shopping service, its analytic application could match my interests with goods available locally, then point me to the appropriate stores."

Refresh Rates for Analytic Data

Real-time analytics are relatively rare today, but will soon be common.

Data velocity is the third V in the three Vs of big data. Data velocity concerns both frequency and speed. In other words, data velocity is about how frequently data is generated by an application. But it can also be about the speed of data delivery into an analytic data set once data is generated. For example, streaming data is an extreme case where the generation and delivery of data is continuous; the analysis of that data may also be continuous.

Most analyses today, however, aren't that continuous. A deployed solution for advanced analytics will rerun analyses as data and business situations change. For example, predictive models are rescored and analytic databases are updated. To get a sense of how often this occurs, the survey asked: In your organization, what percent of analyses are rerun and/or rescored at the following intervals? (See Figure 13.)

Based on survey responses, today most analytic updates and rescores occur daily, weekly, monthly, and/or annually. In other words, the refresh of most deployed analyses is latent, with intraday refresh (every few hours, hourly, or in real time) being rare. This puts analytics behind the times, as compared to reporting, where real-time refresh is the norm for some report types.

With other surveys run by TDWI, respondents claim to refresh standard reports many times intraday, with dashboard-style reports being refreshed the most frequently.[4] This is due to operational BI—a popular management technique that requires frequently refreshed reports. Note that reporting took many years to cross the line from overnight refresh to frequent intraday refreshes. The march toward real time is affecting many enterprise applications types; no doubt, analytics will soon come closer to real time as part of this trend.

In your organization, what percent of analyses are rerun and/or rescored at the following intervals?

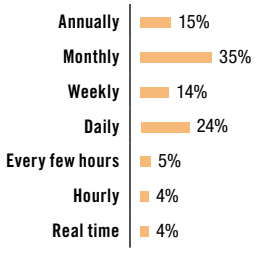


Figure 13. Based on 96 respondents who report practicing big data analytics.

USER STORY REAL-TIME PROCESSING OF STREAMING BIG DATA IS CRUCIAL TO FINDING MEANINGFUL DATA AND REACTING TO IT.

A consultant who specializes in streaming data told TDWI recently: "You don't need all of the streaming data. You just need the interesting pieces or just the one piece that identifies what you're looking for. We've all seen video footage from the U.S. military's unmanned jet drones. A drone is processing several frames of video per second looking for shapes or light signatures that match its programming. For example, it might be looking for shapes that look like tanks or sun reflections that could come from metallic weapons. The drone deletes almost all of the frames, because they're not of interest. And that helps to avoid data glut that could choke the system."

[4] For example, see Figures 8, 10, and 12 in the 2007 TDWI report *Best Practices in Operational BI*, available on tdwi.org.

Replacing Analytics Platforms

One-third of users will replace their analytics platforms within three years.

An analytics platform can take many forms. To some users, it's the analytic tool where they create analytic models or fashion complex queries. To others, it's the database management system where analytic big data is managed and operated on. It could be both. And everyone longs for heftier hardware for the analytics platform. Regardless of the definition, some users are contemplating a replacement of their analytics platforms.

Roughly half of user organizations will keep their analytics platforms. They have no plans to replace the current analytics platforms (47%). No doubt, some would like a replacement, but don't have the budget or approval to do so. TDWI suspects that many get what they need from the current platform, so they are not compelled to swap it for another.

A tenth (9%) have already replaced their analytics platform. The survey results aside, a few users interviewed talked about swapping data mining tools for statistical packages or replacing old data marts with new data warehouse appliances or analytic databases.

One-third anticipate replacing their analytics platform within three years. Thirty-three percent say the replacement will occur in 2011, 2012, or 2013. Another 11% anticipate a replacement in the four years after that.

When do you anticipate replacing your current analytics platforms?

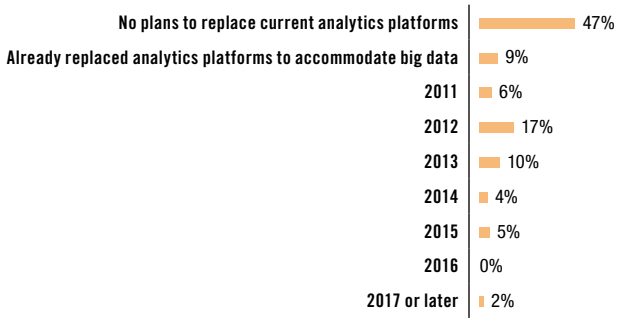


Figure 14. Based on 325 respondents.

Users replace analytic platforms to get better performance, productivity, and modern capabilities.

Ripping out and replacing an analytics platform is rather expensive for IT budgets and intrusive for business users. Yet, this is exactly what more than half of the users surveyed by TDWI say they are contemplating—or have just done. To find out what dire circumstances would lead so many people down such a drastic path, our survey asked: What problems will eventually drive you to replace your analytics platforms?

Big data analytics requires massive performance and scalability. Not just any platform can live up to such stringent demands. Common problems voiced by users are that old platforms can't scale to big data volumes (42%), load data too slowly (29%), respond to queries too slowly (24%), lack processing (CPU) capacity for analytics (17%), and can't handle concurrent mixed workloads (11%).

Sometimes the old analytics platform is a mismatch to today's requirements. This is revealed by users who complained that their system can't support the analytic modeling they need (32%) and the current platform is OLAP-only whereas they need advanced analytics (28%).

Some users need a tool that works the way they want to work. For example, some platforms are poorly suited to self-service for end users (18%) or are poorly suited to visual analytics (11%).

Sometimes users need a new platform that supports modern capabilities. This is evident when the old platform is considered poorly suited to real-time analytics (24%), can't support in-database analytics (20%), has inadequate support for Web services and service-oriented architecture (SOA) (14%), lacks support for cloud or virtualization (10%), or has inadequate support for in-memory processing (7%).

What problems will eventually drive you to replace your analytics platforms? Select five or fewer.

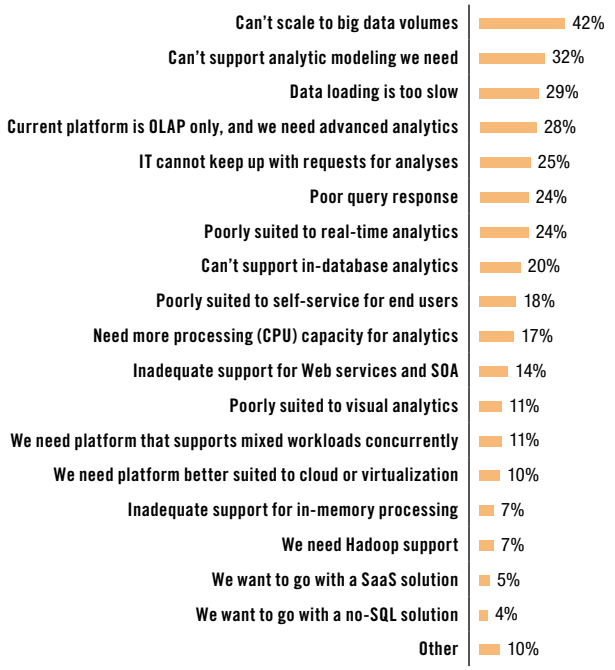


Figure 15. Based on 1,098 responses from 325 respondents; 3.4 responses per respondent, on average.

Tools, Techniques, and Trends for Big Data Analytics

Good news: There are many options for big data analytics.

Bad news: it's hard to know them all and select the best one.

By now, you've probably noticed that there are many different options that you can select for your big data analytics program. Options include vendor tool types and tool features, users' techniques and methodologies, and team or organizational structures. The list is long and complex, and it includes a few items you probably haven't considered seriously. Regardless of what project stage you're in with big data analytics, knowing the available options is foundational to making good decisions about approaches to take and software or hardware products to evaluate.

To quantify these and other issues, TDWI presented survey respondents with a long list of options for big data analytics. (See Figure 16.) The list includes options that have arrived fairly recently (clouds, MapReduce, complex event processing), have been around for a few years but are just now experiencing broad adoption (data visualization, predictive analytics), or have been around for years and are firmly established (statistical analysis, hand-coded SQL). The list is a catalog of available options for big data analytics, and responses to survey questions indicate what combinations of analytic functions, platforms, and tool types users are employing today, as well as which they anticipate using in a few years. From this information, we can deduce priorities that can guide users in planning. We can also quantify trends and project future directions for advanced analytics and big data.

Survey responses reveal which options for big data analytics are in common use today.

Concerning the list of big data analytics options, the survey asked: "What kinds of techniques and tool types is your organization using for advanced analytics and big data, both today and in three years?" Survey responses for these two questions are charted as pairs of bars on the left side of Figure 16. Within each pair of bars, the value for Using Today is the percentage of survey respondents who claim to use that option now. Similarly, the value of Using in Three Years is the percentage of survey respondents who anticipate using that option in coming years.

Survey responses indicate how usage of options for big data analytics will increase or decline.

The pairs of bars on the right side of Figure 16 portray a slightly different view of option usage. The Potential Growth bars calculate the per-option difference between responses for Using Today and Using in Three Years; this delta provides an indication of how much the usage of a big data analytics option will increase or decrease. An option's Commitment value is the percentage of survey respondents who are committed to using that option, whether today, in three years, or both. Note that no option will be used by all survey respondents in all time frames, which is why none of the values in Figure 16 tally to 100%.

What kinds of techniques and tool types is your organization using for advanced analytics and big data, both today and in three years? (Checking nothing on a row means you have no plans for that technique or tool.)

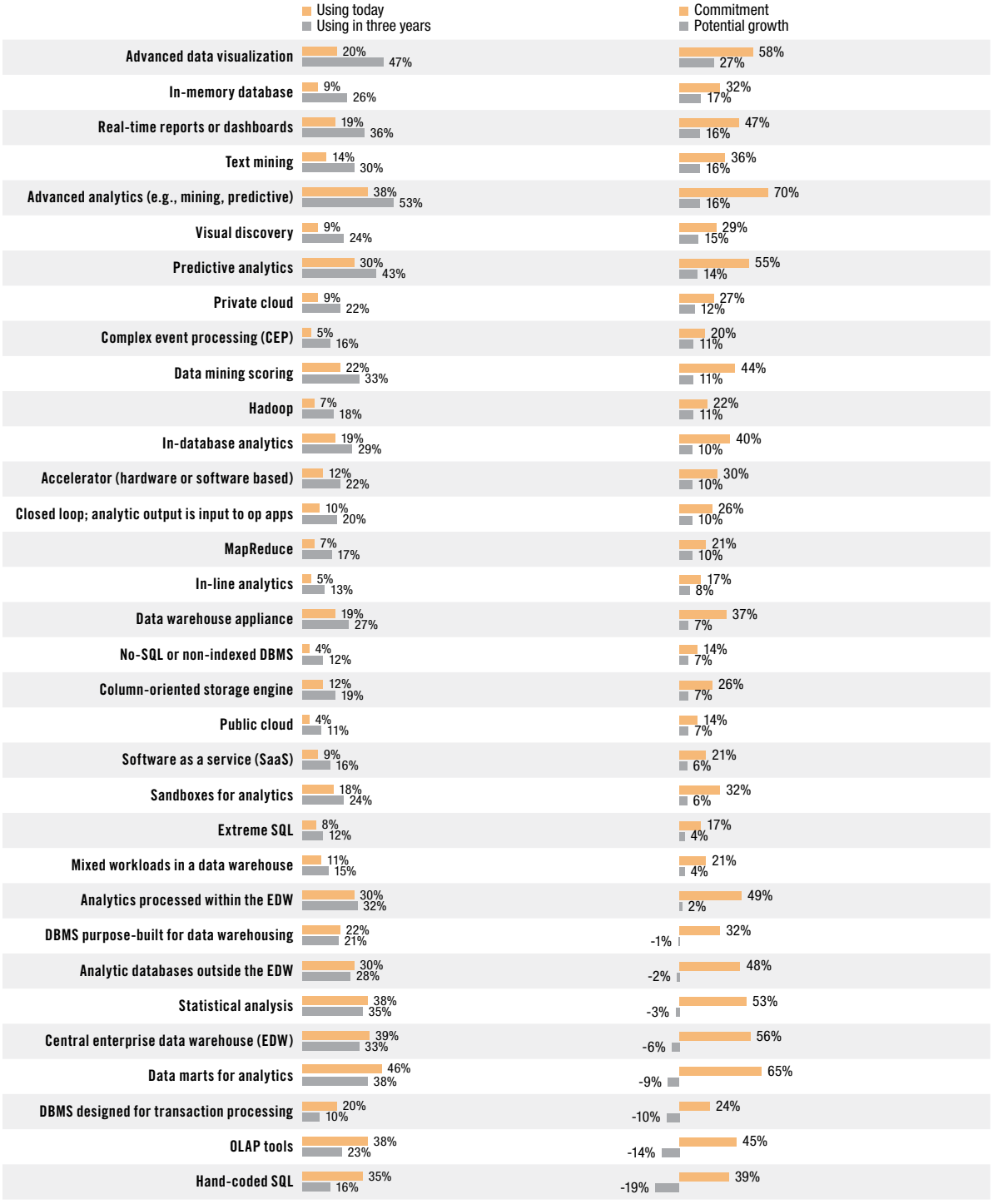


Figure 16. Based on varying numbers of responses from 325 respondents. The charts are sorted by the "Potential Growth" column of values.

Potential Growth versus Commitment for Big Data Analytics Options

Figure 16 reveals a number of interesting things about the use of tools and techniques for big data analytics. For example, Figure 16 is sorted by the Potential Growth column in descending order. In this sort order, "advanced data visualization" (ADV) appears at the top of the chart, because—with a delta of 27%—this option has the greatest potential growth.

However, not all organizations plan to use the ADV option. In the Commitment column, we see that 58% of survey respondents have committed to implement ADV at some point. Conversely, 42% of respondents have no plans to implement it. By scanning the Commitment column in Figure 16, you can see that 58% is a relatively high level of commitment for a big data analytics option. Given ADV's strong Potential Growth and strong Commitment, it's likely that most organizations will include some form of advanced data visualization in their arsenal of big data analytics options.

Commitment and Potential Growth are two different metrics for the future of big data analytics.

From this, we see that there are two forces at work in Figure 16, as well as in the planning processes of user organizations.

- **Potential growth.** The Potential Growth chart subtracts Using Now from Using in Three Years, and the delta provides a rough indicator for the growth or decline of usage of options for big data analytics over the next three years. The charted numbers are positive or negative. Note that a negative number indicates that the use of an option may decline or remain flat instead of grow. A positive number indicates growth, and that growth can be good or strong.
- **Commitment.** Collected during the survey process, the numbers in the Commitment column represent the percentage of survey respondents (based on a total of 325 respondents) who selected Using Today and/or Using in Three Years. Note that the measure of commitment is cumulative, in that the commitment may be realized today, in the near future, or both.

  Furthermore, the survey question told respondents: "Checking nothing on a row means you have no plans for that technique or tool." This way, if a survey respondent has no plans for an option, he/she could leave it unchecked.
- **Balance of commitment and potential growth.** To get a complete picture, it's important to look at the metrics for both growth and commitment. For example, some features or techniques may have significant growth rates, but within a weakly committed segment of the user community (clouds, SaaS, no-SQL databases). Or they could have low growth rates but be strongly committed through common use today (analytic data marts, OLAP tools). Options seeing the greatest activity in the near future will most likely be those with strong ratings for both growth and commitment (ADV, advanced analytics, predictive analytics).

  To visualize the balance of growth and commitment, Figure 17 includes the Potential Growth and Commitment numbers from Figure 16 as opposing axes of a single chart. Big data analytics options are plotted in terms of growing or declining usage (*x*-axis) and narrow or broad commitment (*y*-axis).

Options for Big Data Analytics Plotted by Potential Growth and Commitment

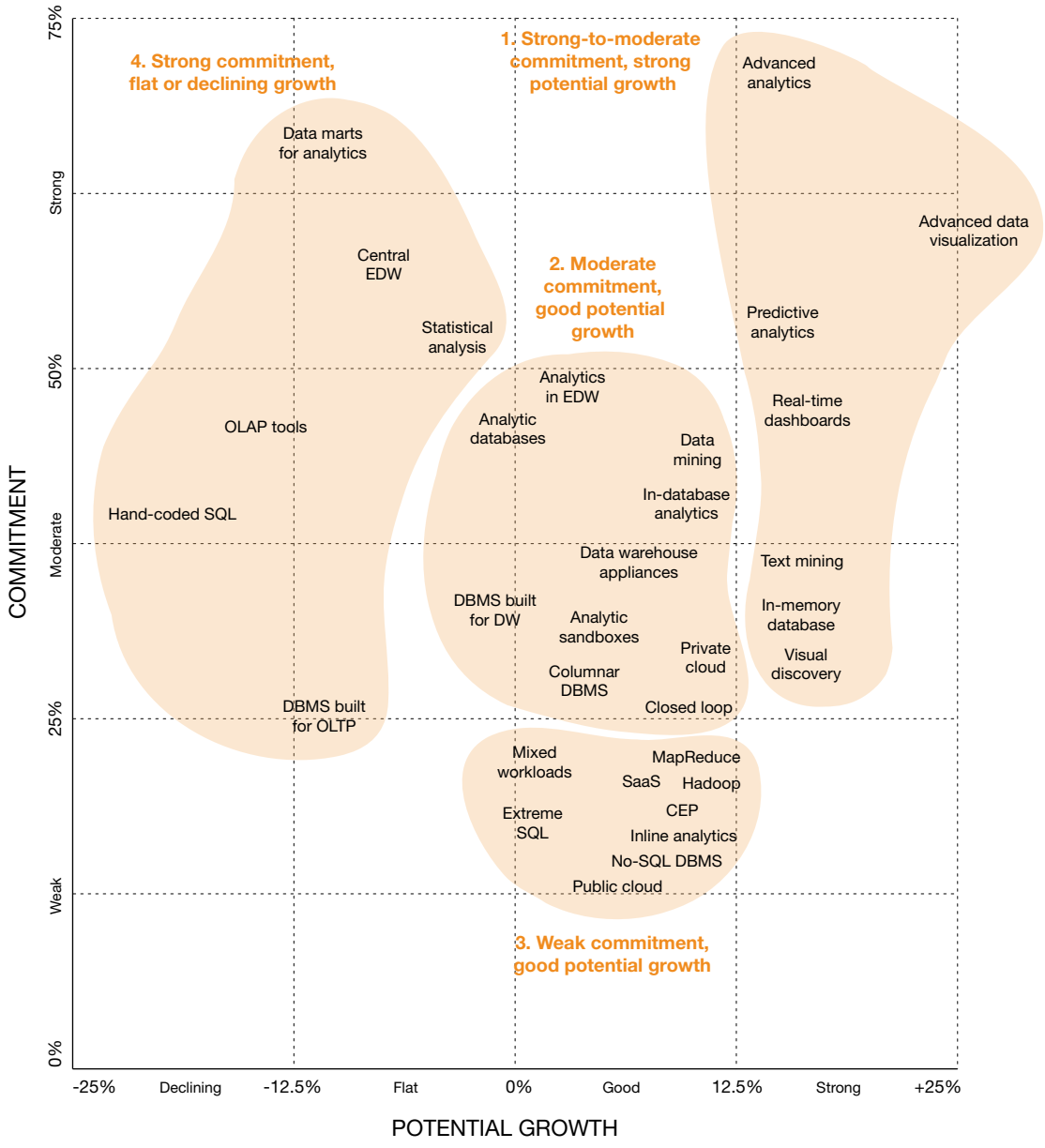


Figure 17. Plots are approximate, based on values from Figure 16.

Trends for Big Data Analytics Options

Rates of growth and commitment identify four groups of options for big data analytics.

Figures 16 and 17 reveal that most big data analytics options will experience some level of growth in the near future. The figures also indicate which options will grow the most, as well as those that will stagnate or decline. In particular, four groups of options stand out based on combinations of growth and commitment. (See the groups circled, numbered, and labeled in Figure 17.) The groups are indicative of trends in advanced analytics and big data.

Group 1 – Strong-to-moderate commitment, strong potential growth.

Options that have the highest probability of altering best practices for big data analytics are those with strong potential growth (according to survey results) coupled with a moderate or strong organizational commitment. Group 1 meets both of those requirements, and it includes tool types and techniques that TDWI has seen adopted aggressively in recent years. Furthermore, today's strongest trends in BI, data warehousing, and analytics are apparent in Group 1:

Visualization and advanced analytics are poised for aggressive adoption.

Advanced analytics. The strongest commitment among options for big data analytics is to advanced analytics (in the upper, right-hand corner of Figure 17). Closely related options such as predictive analytics, data mining, and statistical analysis have a similar commitment. Recall that advanced analytics is a collection of techniques and tool types, including predictive analytics, data mining, statistical analysis, complex SQL, data visualization, artificial intelligence, natural language processing, and database methods that support analytics. The move into advanced analytics (beyond reporting and OLAP) is one of the strongest trends in BI today, for all the reasons explained early in this report. Given the strong corporate commitment that advanced analytics has, it will no doubt be a leading growth area for users and vendors alike for several years to come.

Visualization. The strongest potential growth among options for big data analytics is projected for advanced data visualization (ADV). ADV is the rightmost option plotted on Figure 17, and the closely related option visual discovery is also in Group 1. ADV is a natural fit for big data analytics. ADV can scale its visualizations to represent thousands or millions of data points—unlike standard pie, bar, and line charts. ADV can handle diverse data types and then present analytic data structures that aren't easily flattened onto a computer screen (such as hierarchies and neural nets). Most ADV tools and functions today support interfaces to all leading data sources so that business analyst users can explore data widely in search of just the right analytic data set—and usually do so in real time. Furthermore, most ADV tools have evolved for ease of use and self-service, in response to growing constituencies of analytic users. Anecdotally speaking, TDWI has seen many organizations adopt ADV and visual discovery, as both standalone analytic tools and general-purpose BI platforms, on both departmental and enterprise levels.

Real time is the strongest BI trend, yet it hasn't hit big data analytics much as yet.

Real time. Operational business intelligence is a business practice that measures and monitors the performance of business operations frequently. It is enabled by BI technologies, especially dashboard-style reports. Although the definition of "frequently" varies, most operational BI implementations fetch data in real time (or close to it) to refresh real-time management dashboards (which are poised for growth in Figure 17). Users' aggressive adoption of operational BI in recent years has (among other things) pushed BI technologies into real-time operation, as seen in management dashboards. As users evolve operational BI to be more analytic (not merely reporting based on metrics), analytics are likewise being pushed into real time.

The third V in the three Vs of big data stands for velocity. As numerous examples in this report have shown, there are many real-world applications of analytics to streaming big data available today, plus more coming. But real-time analytic applications are still new, and they are practiced today by relatively few organizations. Perhaps this explains why real-time and streaming data didn't fare well in this report's survey. (For example, Figure 13 shows low use of real-time data among survey respondents.) Even so, given that real-time is the strongest trend in BI today, it will no doubt transform analytics soon, just as it has transformed reporting.

In-memory databases. One way to get real-time response from a database is to manage it in server memory, thereby eliminating disk I/O and other speed bumps. For several years now, TDWI has seen consistent adoption of in-memory databases among its members and other organizations. An in-memory database can serve many purposes, but in BI they usually support real-time dashboards for operational BI, and the database usually stores metrics, key performance indicators (KPIs), and sometimes OLAP cubes. We're now seeing a similar growth among users in the adoption of in-memory databases for advanced analytics, typically to speed access to and the scoring of analytic models. Leading vendors now offer data warehouse appliances with Flash memory or solid state drives, to which in-memory databases will soon move.

Unstructured data. We all give lip service to the fact that there's valuable, actionable information in natural language text and other unstructured data. Yet, organizations haven't tapped that information much until very recently. Tools for text mining and text analytics have slowly gained usage because they can find facts about key business entities in text and turn those facts into usable, structured data. The resulting data can be applied to customer sentiment analysis, but it has many other applications, too. For example, many insurance companies use text analytics to parse the mountains of text that result from the claims process, turn text into structured records, then add that data to the samples studied via data mining or statistical tools for risk, fraud, and actuarial analyses.

Group 2 – Moderate commitment, good potential growth.

Different types of analytic database platforms dominate Group 2. Thanks to recent innovations by vendor firms, we now have more analytic database platforms to choose from, including data warehouse appliances, dedicated analytic DBMSs, columnar data stores, and sandboxes—plus older options. And thanks to user adoption, the newer analytic database platforms have achieved moderate commitment and good potential growth.[5]

Gating factor: manage and operate on analytic big data in an EDW? If not, where?

For most user organizations facing a new analytics program (or a renovation of an established one), one issue is a determining factor: Can the current or planned enterprise data warehouse (EDW) handle big data and advanced analytics without degrading performance of other workloads for reporting and online analytic processing (OLAP)? A simpler question is: Can our EDW perform and scale with concurrent mixed workloads? The answer to this question will determine whether analytic data is managed and operated on in the EDW proper or in a separate platform (which is usually integrated with the EDW).

[5] For a complete list and discussion of vendor analytic database platforms, replay the TDWI Webinar "Data Warehouse Appliances: An Update on the State of the Art," online at tdwi.org.

As we saw in Figure 11, the EDW is a much used and much preferred platform for analytics, which proves that many EDWs can, indeed, scale and perform with mixed workloads. In fact, in-database analytics has very recently become common, showing that EDWs can handle advanced analytic workloads. Yet, not everyone is willing to host analytics on an EDW. That's because the management of big data and the processing workloads of advanced analytics make stringent demands of server resources, such that (depending on the EDW platform you've assembled) they can rob server resources from other data warehouse workloads, resulting in slow queries and report refreshes. To avoid performance degradation due to mixed workloads, some BI professionals prefer to isolate big data and analytic workloads on separate platforms outside the EDW. Performance aside, separate analytic database platforms make sense when analytics is funded or controlled by a department instead of the EDW's sponsor (as seen in Figure 8).

There are now more choices for analytic database platforms.

Although two-thirds of organizations tend toward analytics on a central EDW (according to Figure 11), there's enough demand for dedicated analytic database platforms that these have become permanent fixtures in data warehouse programs worldwide. The movement toward these began in 2003, when the first data warehouse appliances appeared and the IT centralization frenzy of the early 2000s subsided. (Despite the name, a data warehouse appliance is almost always used for advanced analytics with big data, not an EDW.) After that came new vendor-built databases with columnar data stores, which inherently accelerate column-oriented analytic queries. More recently, vendors have brought out analytic platforms based on MapReduce, distributed file systems, and no-SQL indexing. Today is a great time to be selecting a data warehouse platform or analytic database platform. There are more choices than ever, and most of the choices are built specifically for data warehousing and/or analytics.

Group 3 – Weak commitment, good growth.

Organizational commitment to the options of Group 3 is weak because they are all relatively new. Even so, potential growth is good within committed organizations, so we should expect these options to be in use by more organizations soon.

Interest is high in distributed file systems and distributed analytic processing.

Hadoop Distributed File System (HDFS). At the moment, users' interest in the HDFS is extremely high (hence, the good potential growth in Figure 17), although rarely adopted (hence, the weak commitment). Interest is high because big data tends to be diverse in terms of data types, and a data-type-agnostic file system could be a good fit for that diversity. Also, many of the complex data types we associate with big data originate in files, examples being Web logs and XML documents. Transforming these into data structures suited to storage via a traditional database management system (DBMS) is a problem when dealing with big data because of time-consuming processes for data modeling, data integration, and bulk data load. Plus, data transformation could potentially lose the data details and anomalies that fuel some forms of analytics. Some users would prefer to simply copy files into a file system without preparing the data much (if at all), as long as the big data strewn across a million or more files is accessible for analytics.

MapReduce. This relatively new analytic option is also of great interest today, similar to the interest in Hadoop. In fact, the two are closely related, in that MapReduce makes a distributed file system like the HDFS addressable through analytic logic. For example, with MapReduce, a user defines a data operation—such as a query or analysis—and the platform "maps" the operation across all relevant nodes for distributed parallel processing and data collection. The mapping and analytic processing work despite diverse data types strewn across many distributed files. The platform then consolidates

and reduces the responses that come back. Distributed file systems aside, MapReduce can also work well in a database management system with a relational store, as it does in the Aster Data database. Due to the distributed processing of MapReduce, analytics against very big data is possible—and with good performance.

Complex event processing (CEP). This option arrived very recently, yet it is currently experiencing rapid adoption. For example, a recent TDWI report discovered that 20% of survey respondents have incorporated some form of event processing into their data integration solutions; that is significant given the newness of this practice.[6] Although it doesn't have to, CEP often operates in real time, so its adoption is driven partially by the real-time trend. CEP can be used in conjunction with analytics, which is another driver. CEP technologies are evolving to handle streaming big data.

SQL. Trends in BI are sometimes at odds, almost cancelling each other out. That's currently the case with SQL, as some organizations deepen their use of SQL while others do the opposite.

Some users want more SQL for analytics. Others want less.

On the one hand, many organizations rely heavily on SQL as the primary approach to advanced analytics. After all, BI professionals know SQL, and almost all tools and databases support it. An experienced BI professional can create complex SQL programs (plotted as "Extreme SQL" on Figure 17), and these work well with big data that's SQL-addressable. Extreme SQL is typically applied to highly detailed source data, still in its original schema (or lightly transformed). The SQL is "extreme" because it's creating multi-dimensional structures and other complex data models on the fly, without remodeling and transforming the data ahead of time.

On the other hand, a small minority of organizations are embracing so-called no-SQL databases. This makes sense when the majority of data types analyzed aren't relational and converting them to tabular structures (or other SQL-addressable structures) isn't practical. Given that the second V in the three Vs of big data stands for variety, it's possible that no-SQL databases will gain traction. No-SQL databases also tend to appeal to application developers, who don't have the BI professional's attachment to SQL.

Clouds. TDWI Technology Surveys about clouds have consistently shown that BI professionals prefer private clouds over public ones, especially for BI, DW, and analytic purposes. This helps explain why the public cloud has the weakest commitment in Figure 17. The preference for private clouds is mostly due to paranoia over data security and governance. Even so, some organizations experiment with analytic tools and databases on a public cloud, then move them onto a private cloud once they decide analytics is mission critical. In a related issue, software-as-a-service (SaaS) doesn't necessarily require a cloud, but most SaaS-based analytic applications or analytic database platforms are on a tightly secured public cloud.

As with other IT systems, analytic tools and databases are heading into the clouds.

[6] As explained in the TDWI Best Practices Report, *Next Generation Data Integration*, available on tdwi.org.

Group 4 – Strong commitment, flat or declining growth.

Group 4 includes essential options such as data marts for analytics, centralized EDWs, OLAP tools, hand-coded SQL, and DBMSs built for OLTP. In fact, these are some of the most common options in use today for BI, data warehousing, and analytics. If these are so popular, why does the survey show them in decline? There are two reasons:

Some mature tools and errant practices will decline in use and priority.

Users are maintaining mature investments while shifting new investments to more modern options. For example, almost all organizations with a BI program have developed solutions for online analytic processing (OLAP). But the current trend is to implement forms of advanced analytics, which are new to many organizations. OLAP won't go away. In fact, OLAP is today the most common form of analytics, and it will remain so for years. No doubt, users' spend for OLAP will grow, albeit modestly compared to other analytic options.

Databases designed for online transaction processing (OLTP) are in a similar situation. As we saw in the discussion of Group 2, many users have come to the conclusion that their organizations would be better served by an analytic database platform built specifically for data warehousing and analytics. They will maintain their investments in older relational databases (designed for OLTP, although used for DW) as they shift new investments to databases purpose-built for data warehousing or analytics.

Users are correcting problems with their designs or best practices. Due to recent requirements for compliance and data sharing, data marts are even more problematic than ever. Although data marts regularly host analytic data sets, they are typically on older platforms that include an OLTP database and an SMP hardware architecture. Whether to rein in proliferated marts or to get a better analytic database platform, many user organizations are aggressively decommissioning analytic data marts.

Hand-coded SQL is a natural option to base analytics on. The catch is that hand-coding tends to be non-productive and anti-collaborative. SQL cannot go away, because (as the leading language for data) it's supported by almost every tool and platform in IT, plus the skill sets of most data management professionals. In fact, analytics is driving up the use of hand-coded SQL. Instead of hand-coding SQL, most organizations should consider tools that generate SQL based on analytic applications developed in a user-friendly GUI. This needs to happen to make developers more productive, as well as to make analytic tools more palatable to business people and mildly technical personnel.

Vendor Products for Big Data Analytics

Since the firms that sponsored this report are all good examples of software and hardware vendors that offer tools, platforms, and services for big data analytics, let's take a brief look at the product and service portfolio of each. The sponsors form a representative sample of the vendor community, yet their offerings illustrate different approaches to big data analytics.[7]

Cloudera

Cloudera makes a business by distributing open source software based on Apache Hadoop. IT personnel demand a number of features and services that Hadoop lacks. To help organizations reliably use Hadoop in production, Cloudera Enterprise is specifically designed to improve the manageability of Hadoop deployments. Cloudera makes Hadoop viable for serious enterprise users by providing technical support, upgrades, administrative tools for Hadoop clusters, professional services, training, and certification. Hence, Cloudera collects and develops additional components to strengthen and extend Hadoop, while still retaining Hadoop's open-source affordability, big data scalability, and flexibility across a wide range of data types.

EMC Greenplum

EMC Corporation is the world's leading provider of data storage platforms and other information infrastructure solutions. In 2010, EMC acquired Greenplum and has since built it up as the EMC Data Computing Division, which has become a leading platform for big data analytics. Greenplum customers are some of the largest firms in the world, and they regularly deploy Greenplum products on grids or clouds to scale up to very big data. EMC Greenplum Database is known for its shared-nothing massively parallel processing (MPP) architecture, high-performance parallel dataflow engine, and gNet software interconnect technology. Recently, EMC Greenplum has released Greenplum HD (an enterprise-ready Hadoop distribution), EMC Greenplum Data Computing Appliance Product Family (purpose-built for big data analytics), and Greenplum Chorus (software for collaboration over analytics).

IBM

IBM has one of the largest product portfolios of any software vendor, with analytics as a significant focus in support of IBM's global campaign for Business Analytics and Optimization (BAO). Within this massive portfolio, three products stand out because of their recent contributions to enabling big data analytics. First, IBM's acquisition of Netezza in 2010 adds to the portfolio the product that invented the data warehouse appliance and defined the modern analytic database platform. Second, just announced in 2011, IBM InfoSphere BigInsights is IBM's Hadoop-based offering that combines the power of Hadoop with IBM-unique code to address enterprise requirements. Enterprise features include built-in text analytics, a spreadsheet-style data discovery and exploration tool, enterprise-grade security, and administrative tools. Third, IBM InfoSphere Streams is a platform for real-time analytic processing (RTAP), which uniquely provides velocity for big streaming data analytics on structured and unstructured data.

Impetus Technologies

Impetus Technologies offers product engineering and technology R&D services for software product development. In the area of big data analytics, Impetus offers consulting, advisory, and professional services. Impetus' customers are large corporations that manage big data as part of operating a business, but most clients also leverage big data with analytics. Impetus helps such firms evaluate and embrace new technologies and business practices that are related to big data analytics. Impetus provides advisory consulting (to assess big data and analytic opportunities), implementation consulting (to design and develop big data analytic infrastructure and applications), and long-term support (to help clients evolve as new practices and technologies for big data analytics evolve). Impetus Technologies provides end-to-end, vendor- and technology-agnostic advice and engineering to objectively determine what's best for the client's business goals and how to achieve the goals with new technologies and practices.

[7] The vendors and products mentioned here are representative, and the list is not intended to be comprehensive.

Kognitio Kognitio offers WX2, an analytic database platform that can be deployed in one of three ways: as a software-only license, as a fully configured data warehouse appliance running on industry-standard hardware, or on-demand via Kognitio's affordable cloud-based data-warehousing-as-a service (DaaS) solution. Since its founding in 2005, Kognitio has rolled out many innovations, namely in-memory big data analytics, data warehouse appliance configurations, MPP shared-nothing database architecture, database high availability, software-as-a-service (SaaS), and clouds as viable platforms for analytic databases. Kognitio customers are known for their SQL coding prowess applied to discovery analytics with big data. To complement this, Kognitio recently released Pablo, with which multi-terabyte virtual cubes for online analytic processing (OLAP) can be created, deployed, and repopulated within seconds.

ParAccel ParAccel Analytic Database (PADB) is a columnar, massively parallel processing (MPP) analytic database platform with strong features for query optimization and compilation, compression, and network interconnect. Users of PADB interviewed by TDWI report high performance with complex SQL workloads executed against big data, as well as with a variety of other analytic workloads. Because PADB is schema-neutral, its users employ agile load-and-go analytic methodologies. PADB's secret sauce is the Omne optimizer, which can optimize any SQL code, no matter how long, complex, or poorly structured it is. Using ParAccel's Extensibility Framework, users can develop routines for parallelized in-database execution. Through On Demand Integration modules, users can integrate PADB with other platforms, including Teradata and Hadoop. PADB performs well on commodity hardware, amounting to a favorable cost-to-performance ratio. PADB is deployable in all enterprise environments, even in clouds and other virtualized standard operating environments.

SAND Technology The SAND Analytic Platform is a columnar analytic database platform that achieves linear data scalability through massively parallel processing (MPP), breaking the constraints of shared-nothing architectures with fully distributed processing and dynamic allocation of resources. SAND supports thousands of concurrent users with mixed workloads, infinite query optimization (requiring no tuning once data is loaded), in-memory analytics, full text search, and SANDboxing for immediate data testing. The SAND Analytic Platform focuses on complex analytics tasks, including customer loyalty marketing, churn analytics, and financial analytics.

SAP Over the years, SAP has deepened its commitment to BI/DW and analytics through the internal development of SAP BW, SAP BEx, and SAP NetWeaver, plus the external acquisition of Business Objects and Sybase. In this direction, SAP took a giant step forward in early 2011, when it released SAP In-Memory Appliance (also known as SAP HANA). HANA is an enterprise software architecture that enables analytic queries to run against detailed source data—and run fast in real time—without need for transforming the data into data models optimized for analysis. To achieve this, HANA implements a variant of MapReduce. That means that the user needn't define analytic queries months in advance, then wait for IT to model data for them. HANA gives logical data modeling a new twist, so that the analyst user can run queries as fast as he/she thinks them up, and without being limited by data models. Eventually, SAP BW and all SAP analytic applications will run atop HANA, giving them scalability for big data and speed for discovery analytics.

SAS

SAS is famous for its predictive analytics capabilities, which include data management, data visualization tools, and prepackaged business solutions. SAS High Performance Computing is specifically designed to support big data initiatives, and it includes in-memory, in-database, and grid computing support. SAS On Demand provides support for private and public clouds, including the ability to deploy any SAS solution on SAS-hosted infrastructure. The SAS Data Integration Studio provides support for Hadoop, allowing integration specialists to design integration jobs using a graphical interface that generates Pig code. SAS includes canned Hadoop transforms, and plans to support in-database capabilities for Hadoop. By the end of 2011, SAS will release a high-performance analytics solution in the form of Teradata and EMC Greenplum appliances, which will provide another option for supporting big data analytics.

Tableau Software

Tableau is known for its strong visualization features, which can support exploratory or discovery analytics, where the point is to explore and discover things the enterprise didn't already know. Analytics aside, Tableau is also used as an all-purpose BI platform, applied to either enterprise or departmental needs. The visual approach seen in Tableau enables high ease of use so that—with simple drag-and-drop methods—an analyst or other user can interact directly with the visualization and other visual controls to form queries, reports, and analyses. If the user knows the basics of enterprise data, he or she doesn't need to wait for assistance from IT. With a few mouse-clicks, a user can access a database, identify data structures of interest, and bring big data into server memory for reporting or analysis—all in a self-service manner.

Teradata

Teradata Database is famous for supporting large and mostly centralized EDWs that yield scalability and fast performance, despite the fact that they're supporting concurrent mixed workloads, such as those for standard reports, performance management, OLAP, advanced analytics, and real-time or streaming data. Furthermore, Teradata's support for third normal form and in-database analytic processing makes it a good platform for managing and analyzing detailed big data. The centralized EDW has distinct advantages. Yet, some Teradata customers need analytic databases outside the main Teradata System. In response, Teradata introduced a line of data warehouse appliances and acquired Aster Data. Since then, Aster Data has received a patent on its native SQL integration with MapReduce called SQL-MapReduce (which Hadoop lacks). And Teradata continues to improve support for partnering analytic tools and platforms. Today, the Teradata portfolio is amazingly diverse, including products and services for just about any BI, DW, or analytic configuration.

Recommendations

Leverage big data via advanced analytics.

Explore big data so you can discover business facts you never knew. This is how you understand what has changed in your business, as well as where the opportunities are for new customer segments or cost reductions.

Put advanced analytics and big data together. The exploratory- and discovery-oriented methods of advanced analytics are appropriate to learning from big data. And these analytics methods benefit from the massive data samples produced from big data. But the main point is that big data is a special enterprise asset that merits leverage, and advanced analytics provides that leverage.

Think of big data as an opportunity, not a problem. Seventy percent of your peers do. Sure, the management of big data presents technical challenges. But its analytic insights can lead to cost reductions and revenue lift.

Know the three Vs of big data and the many types of advanced analytics.

Remember the three Vs of big data. They are data volume, data type variety, and data feed velocity. Scaling up to big data's volume is a challenge, but there's more to it. To get the most out of big data, you need tools and platforms that can analyze diverse data types, and you may need tools that can handle the velocity of streaming data in real time.

Know the types of advanced analytics so you can make informed choices. Advanced analytics is a collection of related techniques and tool types, including predictive analytics, data mining, statistical analysis, complex SQL, data visualization, artificial intelligence, natural language processing, and database methods that support analytics.

Don't expect OLAP to go away. OLAP is by far the most common analytic approach today, and it will remain so. Expect to maintain your OLAP solution as you implement other analytics.

Go for the benefits. Avoid the barriers.

Embrace big data analytics for the benefits. Don't be seduced by the sexy technologies. Big data analytics benefits customer relations, all things BI, and many types of analytic applications.

Beware the barriers to big data analytics. These include inadequate staffing or skills, a lack of business support, and problems with database software.

Design your big data analytics solutions for capacity, architecture, scale, speed, and governance.

Plan capacity for hundreds of terabytes of big data. And that's only for analytics, not your entire DW or enterprise. Twenty percent of users will manage half a petabyte of big data just for analytics in three years.

Choose carefully where you will manage and operate on analytic data. The decision will impact data warehouse architecture, scalability limits, query speed, and sponsorship.

Control your analytic databases. Don't let them proliferate like gargantuan data marts. Analytic data requires governance, privacy, and security, just like any enterprise data.

Analyze non-structured data, even if it means acquiring a new analytic platform.

Question the hegemony of structured data. Non-structured data types are daunting for the uninitiated, but they are the final frontier—the data your enterprise hasn't tapped for analytics.

Reevaluate your current portfolio of analytic databases and tools. Over half of organizations are contemplating platform replacements to get a platform that performs well, handles diverse big data, or satisfies modern requirements for ease-of-use or self-service.

Give priority to visualization, in-memory databases, and SQL.

Take another look at data visualization. According to survey responses, use of this function will grow faster than any other big data analytics option. Perhaps you need it as much as your peers do.

Rely on in-memory databases for speed. Use of these has skyrocketed in recent years, because they yield blistering fast query responses and real-time analytics.

Rely on SQL for analytics—but not too much. As the language of data, SQL is a natural fit for analytics, even more so than competing analytic methods such as data mining, statistics, artificial intelligence, or natural language processing. But if you hand-code SQL, it can be non-productive and non-collaborative. Instead, consider an analytic tool that generates good-quality SQL.

Seriously consider the newest analytic platforms. These include Hadoop, MapReduce, no-SQL databases, public and private clouds, SaaS, and complex event processing, which uniquely satisfy new requirements for highly diverse data types, unrestricted indexing and querying, outsourcing of analytics, and real-time analytics.

Cloudera
www.cloudera.com

Cloudera, the leader in Apache Hadoop–based software and services, enables data-driven enterprises to easily derive business value from all their data. Cloudera's Distribution Including Apache Hadoop (CDH)—www.cloudera.com/downloads—is the most comprehensive and widely deployed Hadoop distribution. To quickly, reliably use this open source technology in production, organizations subscribe to Cloudera Enterprise, comprised of Cloudera Support and software including Cloudera Management Suite. Cloudera also offers consulting services, training, and certification. As the top contributor to the Apache community and tens of thousands of nodes under management, Cloudera's sharing of expertise is unrivaled.

EMC Greenplum
www.greenplum.com

EMC Greenplum is driving the future of data warehousing and analytics with breakthrough products including the Greenplum Data Computing Appliance, Greenplum Database, Greenplum HD enterprise-ready Apache Hadoop, and Greenplum Chorus—the industry's first Enterprise Data Cloud platform. EMC Corporation (NYSE: EMC) is the world's leading developer and provider of information infrastructure technology and solutions.

IBM
www.ibm.com

IBM has one of the largest big data product portfolios of any software vendor. IBM InfoSphere BigInsights combines the power of Hadoop with IBM-unique code to address enterprise requirements such as built-in text analytics, a spreadsheet-style data discovery and exploration tool, enterprise-grade security, and administrative tools. IBM InfoSphere Streams is a platform for real-time analytic processing (RTAP), which uniquely provides velocity for big data streaming analytics on structured and unstructured data. IBM Netezza, Smart Analytics System, and InfoSphere Warehouse also provide a broad platform base for scalable, workload-optimized data warehousing appliances, flexible systems, and software.

Impetus Technologies
www.impetus.com

Impetus is a leading technology and R&D services provider offering advisory and professional services in the area of big data analytics and high-performance computing. Impetus has built top-tier expertise and employs thought leaders in technologies such as Hadoop, numerous no-SQL DBs including Cassandra, Mongo DB, CouchBase, a range of commercial massively parallel processing database products, and ETL, BI, and analytics tools. Impetus consultants operating both on client premises and/or offshore help enterprises and technology companies evaluate and implement big data and analytics solutions tailored to their specific contexts.

Kognitio
www.kognitio.com

As the data explosion continues, organizations count on Kognitio and its analytical in-memory database solution—Kognitio WX2—to get comprehensive answers to their business questions, fast. Where other solutions cannot cope with overwhelming data volumes and force the user to rely on data samples, Kognitio has helped numerous customers around the globe survive the Big Data phenomenon and analyze whatever, whenever and get answers to complex questions in subsecond response times. Kognitio WX2 is available as a software license, analytical appliance, or on-demand via cloud-based data warehousing as a service (DaaS). Learn more at www.kognitio.com.

ParAccel
paraccel.com

In today's analytics-driven environment, gaining fast and accurate business insights from massive volumes of data provides significant strategic advantage. As the leader in the high-performance Analytics Platform, ParAccel enables organizations to address their most dynamic and complex analytic challenges and rapidly gain ultra-fast, deep insights from very large data sets. ParAccel's *Fortune* 1000 customers include companies in the financial services, retail and healthcare industries, as well as government agencies. Each organization uses ParAccel to address their business-critical data issues outside the scope of conventional data warehouses and existing analytic tools.

SAND Technology
www.sand.com

SAND Technology delivers a columnar analytic database platform achieving linear data scalability through agile massively parallel processing (AMPP), breaking the constraints of shared-nothing architectures with fully distributed processing and dynamic allocation of resources. SAND supports thousands of concurrent users with mixed workloads, infinite query optimization (requiring no tuning once data is loaded), in-memory analytics, full text search, and SANDboxing for immediate data testing. The SAND Analytic Platform focuses on complex analytics tasks, including customer loyalty marketing, churn analytics, and financial analytics.

SAP
www.sap.com

As market leader in enterprise application software, SAP (NYSE: SAP) helps companies of all sizes and industries run better. From back office to boardroom, warehouse to storefront, desktop to mobile device—SAP empowers people and organizations to work together more efficiently and use business insight more effectively to stay ahead of the competition. SAP applications and services enable more than 109,000 customers to operate profitably, adapt continuously, and grow sustainably.

SAS
www.sas.com

As data volume, complexity, and velocity grow, so does the need for organizations to uncover timely insights and drive value for competitive differentiation. The leader in business analytics, SAS has been managing “big data” before the phrase existed. Whether it’s customer intelligence, fraud detection, risk management, social media engagements, or any number of unique business challenges, SAS has been at the forefront of solving the most complex problems using big data analytics for more than 35 years. Since 1976 SAS has been giving customers around the world THE POWER TO KNOW®.

Tableau Software
www.tableausoftware.com

Tableau Software, a privately held company in Seattle, WA, builds software that delivers fast analytics and rapid-fire business intelligence to everyday businesspeople. Our mission is simple: help people see and understand data. Tableau’s award-winning products integrate data exploration and visualization to make analytics fast, easy, and fun. They include Tableau Desktop, Tableau Server, Tableau Digital, and the free Tableau Public. We understand the needs of businesspeople, non-technical and technical alike, when it comes to retrieving and analyzing large volumes of data. As a result, Tableau has already attracted over 65,000 licensed users in companies from one-person businesses to the world’s largest organizations.

Teradata
teradata.com

Teradata is the world’s largest company solely focused on data warehousing and integrated marketing management through database software, enterprise data warehousing, data warehouse appliances, and analytics. Teradata provides the best database for analytics with the architectural flexibility to address any technology and business need for companies of all sizes. Supported by active technology for unmatched performance and scalability, Teradata’s experienced professionals and analytic solutions empower leaders and innovators to create visibility, cutting through the complexities of business to make smarter, faster decisions. Simply put, Teradata solutions give companies the agility to outperform and outmaneuver for the competitive edge. Visit teradata.com.

TDWI RESEARCH

TDWI Research provides research and advice for business intelligence and data warehousing professionals worldwide. TDWI Research focuses exclusively on BI/DW issues and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of the business and technical challenges surrounding the deployment and use of business intelligence and data warehousing solutions. TDWI Research offers in-depth research reports, commentary, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.

1201 Monster Road SW
Suite 250
Renton, WA 98057-2996

T 425.277.9126
F 425.687.2842
E info@tdwi.org

tdwi.org