TDWI RESEARCH

FIRST QUARTER 2015

TDWI BEST PRACTICES REPORT

Next-Generation Analytics and Platforms

For Business Success

By Fern Halper

tdwi.org

tdwi
Advancing all things data.

Research Sponsors

Actian

Cloudera

Datawatch Corporation

Pentaho

SAP

SAS

BEST PRACTICES REPORT

FIRST QUARTER 2015

Next-Generation Analytics and Platforms

For Business Success

By Fern Halper

Table of Contents

Research Methodology and Demographics 3

Executive Summary 4

Introduction to Next-Generation Analytics Technologies 5

- What Is Next-Generation Analytics? 5
- In Their Own Words 5
- Trends Supporting Next-Generation Analytics 6
- Drivers for Next-Generation Analytics 7

Next-Generation Analytics Status 7

- BI Status 8
- The Status of More Advanced Analytics 9
- Gimme the Data! 11
- Where Is It Being Used? 12
- Who Is Using It? 14

Technologies That Support Next-Generation Analytics 15

- Platforms 15
- Platform Status 16
- The Cloud 17
- Open Source 19

Operationalizing Analytics: The Path to Action 19

- The Status of Operational Analytics 19
- Analytics Informs Action 20

Challenges and Emerging Best Practices for Next-Generation Analytics 21

- Overcoming the Challenges 22
- Acquiring Skills 23
- Other Best Practices 23

Measuring the Value of Next-Generation Analytics 24

Vendor Platforms and Tools That Support Next-Generation Analytics 25

Top Ten Best Practices 27

About the Author

FERN HALPER is well known in the analytics community, having published hundreds of articles, research reports, speeches, Webinars, and more on data mining and information technology over the past 20 years. Halper is also co-author of several "Dummies" books on cloud computing, hybrid cloud, and big data. She is the director of TDWI Research for advanced analytics, focusing on predictive analytics, social media analysis, text analytics, cloud computing, and "big data" analytics approaches. She has been a partner at industry analyst firm Hurwitz & Associates and a lead analyst for Bell Labs. Her Ph.D. is from Texas A&M University. You can reach her at fhalper@tdwi.org, on Twitter @fhalper, and on LinkedIn at www.linkedin.com/in/fbhalper.

About TDWI

TDWI is your source for in-depth education and research on all things data. For 20 years, TDWI has been helping data professionals get smarter so the companies they work for can innovate and grow faster.

TDWI provides individuals and teams with a comprehensive portfolio of business and technical education and research to acquire the knowledge and skills they need, when and where they need them. The in-depth, best-practices-based information TDWI offers can be quickly applied to develop world-class talent across your organization's business and IT functions to enhance analytical, data-driven decision making and performance.

TDWI advances the art and science of realizing business value from data by providing an objective forum where industry experts, solution providers, and practitioners can explore and enhance data competencies, practices, and technologies.

TDWI offers five major conferences, topical seminars, onsite education, a worldwide membership program, business intelligence certification, live Webinars, resourceful publications, industry news, an in-depth research program, and a comprehensive website: 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. To suggest a topic that meets these requirements, please contact TDWI Research Directors Philip Russom (prussom@tdwi.org), David Stodder (dstodder@tdwi.org), or Fern Halper (fhalper@tdwi.org).

Sponsors

Actian, Cloudera, Datawatch Corporation, Pentaho, SAP, and SAS sponsored the research for this report.

Research Methodology and Demographics

Report Scope. Analytics has become extremely important to business. Many organizations are on the cusp of moving from reporting and dashboards to newer forms of analytics. The result is that companies are looking for a way to drive insight and action using analytics without becoming mired in analytics and infrastructure issues. The purpose of this report is to accelerate understanding of the many new technologies and practices that have emerged recently around analytics.

Survey Methodology. In late July 2014, TDWI sent an invitation via e-mail to business and IT executives; VPs and directors of BI, analytics, and data warehousing; business and data analysts; data scientists; IT application managers; and other BI/DW professionals, asking them to complete an Internet-based survey. The invitation was also delivered via websites, newsletters, and publications from TDWI. The survey drew over 450 responses. From these, we excluded incomplete responses as well as some respondents who identified themselves as vendors or academics. The resulting 328 responses form the core data sample for this report.

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

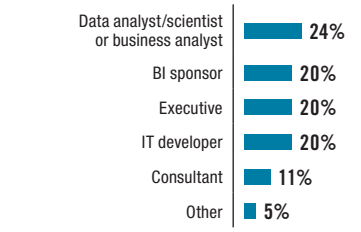
Survey Demographics. There was a good mix of survey respondents including analysts/scientists and business analysts (24%), business sponsors or users (20%), executives (20%), IT developers (20%), and consultants (11%). We asked consultants to fill out the survey with a recent client in mind.

The consulting industry (18%) dominates the respondent population, followed by software/Internet (11%), financial services (10%), and healthcare (10%). Most respondents reside in the U.S. (51%), Europe (17%), or Asia (11%). Respondents are fairly evenly distributed across all sizes of company.

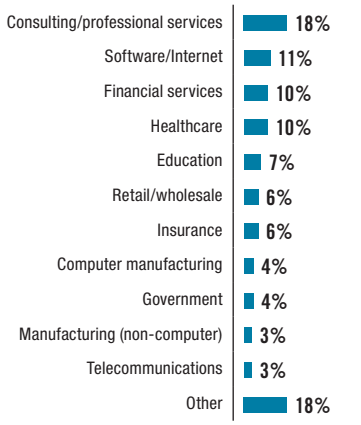
Acknowledgments

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: Michael Boyda, Roxanne Cooke, Marie Gipson, Denelle Hanlon, and James Powell.

Position

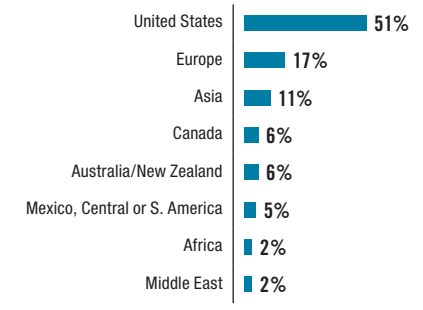


Industry

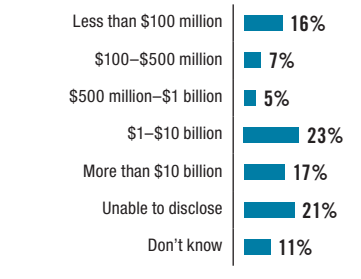


("Other" consists of multiple industries, each represented by less than 3% of respondents.)

Region



Company Size by Revenue



Based on 328 survey respondents.

Executive Summary

The market is on the cusp of moving forward.

User organizations are pushing the envelope in terms of analytics and the platforms to support analysis. These organizations realize that to be competitive, they must be predictive and proactive. However, although the phrase "next-generation platforms and analytics" can evoke images of machine learning, big data, Hadoop, and the Internet of things, most organizations are somewhere in between the technology vision and today's reality of BI and dashboards. Next-generation platforms and analytics often mean simply pushing past reports and dashboards to more advanced forms of analytics, such as predictive analytics. Next-generation analytics might move your organization from visualization to big data visualization; from slicing and dicing data to predictive analytics; or to using more than just structured data for analysis. The market is on the cusp of moving forward.

Predictive analytics, geospatial analytics, text analytics, and even in-stream analysis are poised to double in use in the next three years.

Although the majority of our survey respondents use some kind of analytics software against their data warehouse or a commercial analytics package on their server, they show great interest in moving ahead with more advanced forms of analytics and the infrastructure to support it. Technologies such as predictive analytics, geospatial analytics, text analytics, and in-stream analysis are all poised to double in use over the next three years if users stick to their plans. Additionally, more than 50% of our survey respondents are already using an analytics platform or appliance. They are looking at other platforms, too, such as in-memory databases, analytics, and in-memory computing. They are exploring (and using) the cloud. There are challenges, too: More than half of respondents cite skills as their top challenge, followed by the closely related challenge of understanding the technology.

For companies that are making use of more advanced analytics, the results are rewarding. These enterprises are monitoring and analyzing their operations and predicting and acting on behaviors of interest. They are measuring top- and bottom-line impacts. In fact, about a quarter of the survey respondents are already measuring this impact. These respondents are more likely to use advanced analytics and disparate data types. They are building a coordinated data ecosystem. There is no silver bullet to get there, but these companies are making it happen.

This TDWI Best Practices Report focuses on how organizations can and do use next-generation analytics. It provides in-depth analysis of current strategies and future trends for next-generation analytics across both organizational and technical dimensions, including organizational culture, infrastructure, data, and processes. It examines both the analytics and infrastructure necessary for next-generation analytics. This report offers recommendations and best practices for implementing analytics in an organization.

Introduction to Next-Generation Analytics Technologies

Analytics has attracted considerable market excitement. Recent research finds that companies using analytics for decision making are 6% more profitable than those that do not.[1] Companies understand the value of analytics; they want to be predictive and proactive, and advanced analytics is rapidly emerging as a big part of the analysis landscape. Additionally, as businesses realize that their traditional data warehouse can be insufficient for their analytics and performance needs, they are looking for new ways to easily access and analyze growing volumes of disparate data—often incorporating their existing data management infrastructure.

Companies that use analytics for decision making are more profitable.

Next-generation infrastructure such as analytic platforms and appliances are important in this evolving ecosystem, as are open source systems such as Hadoop. Unified information architectures are becoming a part of the picture as is the cloud. On the analytics front, businesses are looking to incorporate what they are using today as well as newer technologies (such as text analytics, stream mining, geospatial analytics, big data analytics) and more-established forms of advanced analytics (such as predictive analytics) over this infrastructure.

What Is Next-Generation Analytics?

What, exactly, is next-generation analytics? Is it only the newer analytics techniques? For this report, we consider analytics beyond the basics of BI reporting and dashboards. This includes exploration and discovery (often on big data) utilizing easy-to-use software and more advanced analytics (such as predictive analytics or stream mining) as well as new ways of approaching analytics, such as visualizing vast amounts of real-time data. In fact, visualization for data discovery and/or predictive analytics is often a first stepping stone to more advanced analytics. Next-generation analytics also includes the supporting infrastructure (such as Hadoop and other analytic platforms). The developing analytics ecosystem can be quite complex.

Next-generation analytics is the successor to BI reporting and dashboards.

In Their Own Words

We asked respondents to provide examples of forward-looking analytics deployments in their organizations. We wanted to understand how the respondents were thinking about "next generation." About 20% could not provide an example, including a fraction who said the example would be too proprietary to share. For those who responded, some examples were more traditional and some were newer. For instance, use cases included forecasting, predicting human (e.g., customers or patients) behavior as well as machine behavior. Some next-generation examples include:

- Healthcare: Predicting expected patient re-admittance to hospital, predicting expected visits to emergency rooms, and patient monitoring
- Insurance: Predicting future claim rates to price insurance risk
- Financial services: Fraud monitoring
- Energy: Real-time analytical processing of oil well data
- Horizontal: Market basket analysis, segmenting customers, churn analysis, predicting equipment failure
- Forecasting world events

[1] Andrew McAfee and Erik Brynjolfsson [2012]. "Big Data: The Management Revolution," *Harvard Business Review*, October.

Next-generation analytics is happening now.

Although some of these are examples that vendors in the market have been talking about for years, what is important is that companies are actually performing these kinds of analyses now.

Trends Supporting Next-Generation Analytics

Several trends have helped to both motivate and drive organizations to utilize next-generation analytics, including the following:

- **Ease of use.** In the past, building a predictive model required a scripting or programming language. Now, vendors generally provide a drag-and-drop or point-and-click interface. Some vendors have gone so far as to have their software ingest the data, determine the outcome variables, and suggest which model is best. Some automatically run a model. Many vendors have also provided collaboration features to allow a non-technically skilled user to build a model and share it with a more experienced person for feedback. Some companies require this in order to put a model into production. Such features make it easier for individuals to build models, elevating business analysts to some of the primary builders of models.
- **Democratization.** The idea behind democratization is to provide all people access to data, regardless of technical prowess, to help make more informed decisions. This is tied to the ease-of-use trend mentioned earlier. Originally, democratization focused on self-service BI for reports and dashboards. Now it also includes visualization as well as more advanced techniques.
- **Consumerization.** Consumability means either (1) BI or analytics can be utilized easily by many people—(related to democratization) or (2) that the results of BI or analytics can be consumed by the masses. In the latter case, embedding a model into a business process might be necessary. For instance, as a credit card transaction comes into a retail system, it might be scored for probability of fraud and routed to a special investigation unit. This is an example of operationalized analytics (we discuss operationalizing analytics later in this report) in which more people can make use of the analysis. In other words, someone might build a model that many people utilize.
- **Platforms.** Analytic platforms—software that provides an integrated solution for the analytics life cycle—is also gaining popularity. This next generation infrastructure can help make more advanced analytics easier to build and deploy. In a 2013 TDWI Best Practices Report[2] on predictive analytics, 83% of survey respondents stated that they would be using an analytics platform in the next three years.
- **Big data and the Internet of things.** Big data—ever-increasing amounts of disparate data at varying velocities—is important because it can drive value. Disparate data is being generated at large scale and high speed. The Internet of things (IoT) brings home the big data value proposition. As sensors and machines generate vast amounts of near- and true real-time data, organizations are beginning to use this data for applications ranging from real-time operational intelligence of a manufacturing facility to patient monitoring. Big data is also driving the use of newer infrastructure such as Hadoop and multi-platform data warehouse environments that manage, process, and analyze new forms of big data, non-structured data, and real-time data. This might include NoSQL databases, data warehouse appliances, and columnar databases. Other technologies, such as machine learning, are gaining steam because of big data from the IoT. Use cases for machine learning include predictive analytics. With big data, that might mean thousands of attributes, and organizations might use machine learning to first figure out the key variables because a predictive model with a thousand attributes might reflect more noise and error than real relationships.

[2] See the 2013 TDWI Best Practices Report *Predictive Analytics for Business Advantage*, available at tdwi.org/bpreports.

Drivers for Next-Generation Analytics

The market for next-generation platforms and analytics is growing for many reasons, but what are the drivers for user adoption of the technology? We asked respondents to score the important drivers of next-generation analytics on a five-point scale where 1 was extremely unimportant and 5 was extremely important.

Over 50% of respondents stated that using next-generation analytics was important for strategic decision making.

Decision making, understanding customers, and improving business performance ranked at the top. Companies are interested in utilizing analytics to make decisions. More often than not, they start with analytics to understand some behavior. Over 50% of the respondents (not charted) stated that using next-generation analytics is extremely important for driving strategic decision making and understanding customers. Slightly less than 50% felt it was extremely important for improving business performance and processes (not charted).

Drive new revenue: Respondents are interested in next-generation analytics to help drive new revenue opportunities, whether for sales and marketing or other business opportunities. Forty-six percent (not charted) of respondents felt this was extremely important.

Lower on the list was driving real-time actions. Analytics is useful only when acted upon. However, much of the market is not yet mature enough to implement real-time actions or take action on real-time events—a familiar situation from previous research (for instance, see the 2014 TDWI Best Practices Report *Real-Time Data, BI, and Analytics*[3]) as well. Likewise, monetizing analytics (i.e., generating revenue by actions such as selling analytics services) also ranked low. Fewer than 25% of respondents cite these drivers as extremely important.

Next-Generation Analytics Status

As pointed out earlier, the emerging analytics ecosystem consists of software, infrastructure, and methodologies. Gone are the days when the data mart or data warehouse or flat files could handle everything a company needed to do for analysis. Instead, forward-looking organizations are beginning to take an ecosystem approach to infrastructure, with different tools for different tasks. To understand what tools and techniques are being used today, and which are poised for growth, we investigated the status of BI, analytics, and technologies.

[3] Available at tdwi.org/bpreports.

BI Status

Dashboards are very common today: 83% of respondents use them.

To learn about respondents' status in terms of BI and analytics, we asked, "What kinds of BI do you perform in your company today? Three years from now?" (See Figure 1.)

What kinds of BI do you perform in your company today? Three years from now?

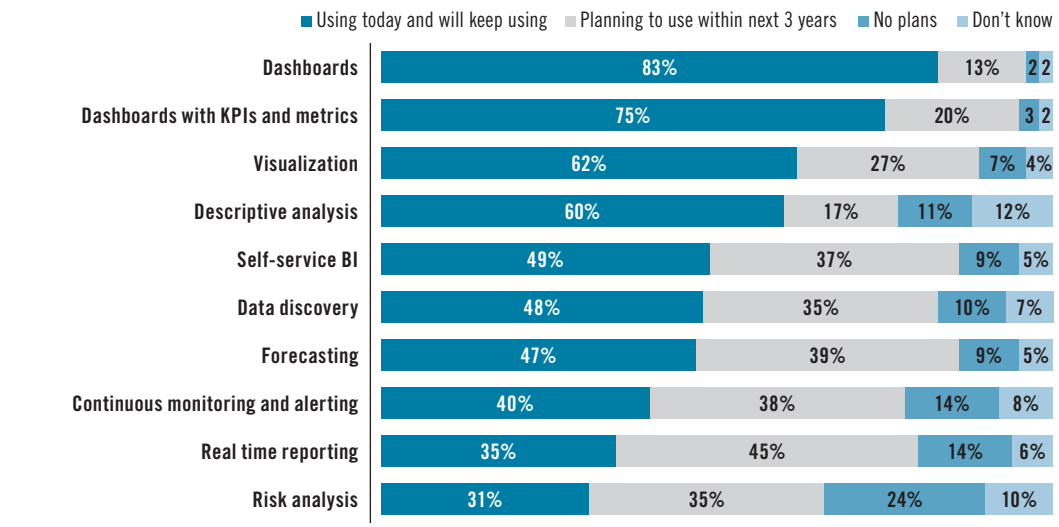


Figure 1. *Types of BI in use at respondent companies. Based on 328 respondents.*

Dashboards are the most commonly used BI technology today. Dashboards rank at the top of the list with close to 83% of respondents stating they use dashboards today. TDWI has seen similar results in other research. Seventy-five percent of respondents answered that dashboards with KPIs and metrics are being used today. Dashboards are a popular technology that helps users get an idea of what is happening, or, more likely, what has already happened in their business. Dashboards are not next-generation technology, but they can help people think analytically, which can drive next-generation approaches.

Visualization is becoming increasingly popular.

Visualization and self-service are primed for growth. Visualization has become a popular technique for data exploration and discovery; its use is exploding. Visualization can help business analysts and others in the organization to slice and dice and discover patterns in their data[4]. In our survey, 62% of respondents are already using visualization, and another 27% are planning to utilize it in the next three years. Visualization can become complex internally for pattern detection and exploration of multiple data sources, including real-time ones.

Continuous alerting and monitoring is also poised for expansion. Although taking action did not score extremely high on the list of next-generation drivers, continuous alerting and monitoring is poised for growth. Forty percent of the respondents utilize it today and an additional 38% expect to do so in the next three years. This is actually a strong step toward next-generation analytics. Monitoring and alerting might not happen on a true real-time scale—for example, daily or hourly is the norm for many organizations[5]—but it is forward movement (from the thought-process perspective) in terms of automating analytics or embedding analytics into a business process or system. We discuss this in more detail in the operationalizing analytics section beginning on page 19.

[4] For more information on visualization and self-service BI, see the 2014 TDWI Best Practices Report *Business-Driven Business Intelligence and Analytics* available at tdwi.org/bpreports.

[5] For more information on real time, see the 2014 TDWI Best Practices Report *Real-Time Data, BI, and Analytics* available at tdwi.org/bpreports.

The Status of More Advanced Analytics

We also asked respondents about the status of more advanced analytics in their organizations. These technologies include:

Predictive analytics is often a first step in next-generation analytics.

- Predictive analytics: A statistical or data mining technique that can be used on both structured and unstructured data to determine outcomes such as whether a customer will "leave or stay" or "buy or not buy." Predictive analytics models provide probabilities of certain outcomes. Popular use cases include churn analysis, fraud analysis, and predictive maintenance.
- Prescriptive analytics: Whereas predictive analytics helps users determine what might happen, prescriptive analytics goes further to either suggest or automatically initiate a subsequent action to produce an optimal result. For instance, prescriptive analytics in healthcare can be used to guide clinician actions by making treatment recommendations based on models that use relevant historical intervention and outcome data. Prescriptive analytics can use both predictive analytics and optimization to do this. True prescriptive analytics often utilizes constraints.
- Geospatial analytics: This form of analytics involves manipulating and analyzing geospatial data, often called location or spatial data. This includes geocoded, remote-sensing, and GPS data. Geospatial analytics includes statistical techniques as well as techniques designed for spatial and spatial/temporal data. For instance, in industrial analytics, a large-scale, discrete manufacturer might use geospatial data to analyze manufacturing bottlenecks in real time. A retailer might use geospatial analytics to examine customer data it is already collecting to plan the location of its next store.
- Text analytics: Text analytics is the process of analyzing unstructured text, extracting relevant information, and transforming it into structured information that can be leveraged in various ways. Text analytics can be used on a range of text from e-mail messages to social media to understand the "why" behind the "what." For instance, if a customer discontinues a service, text analytics can help to understand the reasons for the action. Were they unhappy? Why?
- Operational intelligence (OI): This analytics technique involves using query analysis or other more advanced analytics against continuous, potentially real-time or near-real-time data to gain visibility into operations. Operational intelligence can go hand in hand with the Internet of things. OI could be used to evaluate complex event processing for oil well operations.

44% of respondents are already performing advanced analytics.

When we asked if respondents thought they were performing advanced analytics, 44% answered affirmatively (not charted). Another 39% said they are planning to in the next two years. Additionally, 33% of respondents said they are already performing big data analytics, and another 33% are planning to begin over the next year or two (not charted).

What are these respondents doing now and what are they planning to do? There are a range of advanced analytics being used now as well as being planned for in the future.

Time-series analysis, operational intelligence, and quality monitoring all rank high. These three technologies ranked at the top of the list of more advanced analytics (See Figure 2) in terms of current usage. More than 40% of respondents stated they use each of these technologies. Operational intelligence and quality monitoring are related in that they both analyze some kind of continuous stream of data. Of course, the data might have varying time frequencies that may not be short. Time-series analysis has been utilized for many years, and some people will use it for forecasting. However, time-series analysis is also becoming more popular in terms of real-time and near-real-time analysis of continuous data streams.

Predictive analytics adoption may double in the next three years.

Predictive analytics is poised for significant growth. Predictive analytics is rapidly gaining attention in the market, as reflected in the percentage of respondents using the technology today. It is also poised for significant growth (See Figure 2). Thirty-nine percent of respondents are currently using predictive analytics today and an additional 46% are planning to use it in the next few years. Predictive analytics is often an important first step for companies embarking on next-generation analytics. Interestingly, 23% of respondents stated that they use prescriptive analytics already, a number that seems high. It could be that many of the respondents are using an informal definition of prescriptive analytics that reflects confusion in the market around this term—or the respondents themselves confused prescriptive with predictive analytics in the survey.

Analysis in data streams is poised for growth. Tied to the earlier discussion of time-series data, in-stream analytics is also set to grow. A data stream typically arrives continuously as a sequence of instances. Such data might include sensor data, social media data, traffic feeds, or financial transactions. Often, it needs to be processed immediately as in fraud detection. Although only 20% of respondents are using it today, and it is unclear what time scale is used in their analysis of the data stream, an additional 40% are planning to use it in the next three years. (See Figure 2.)

What kinds of analytics do you use in your company today? Three years from now?

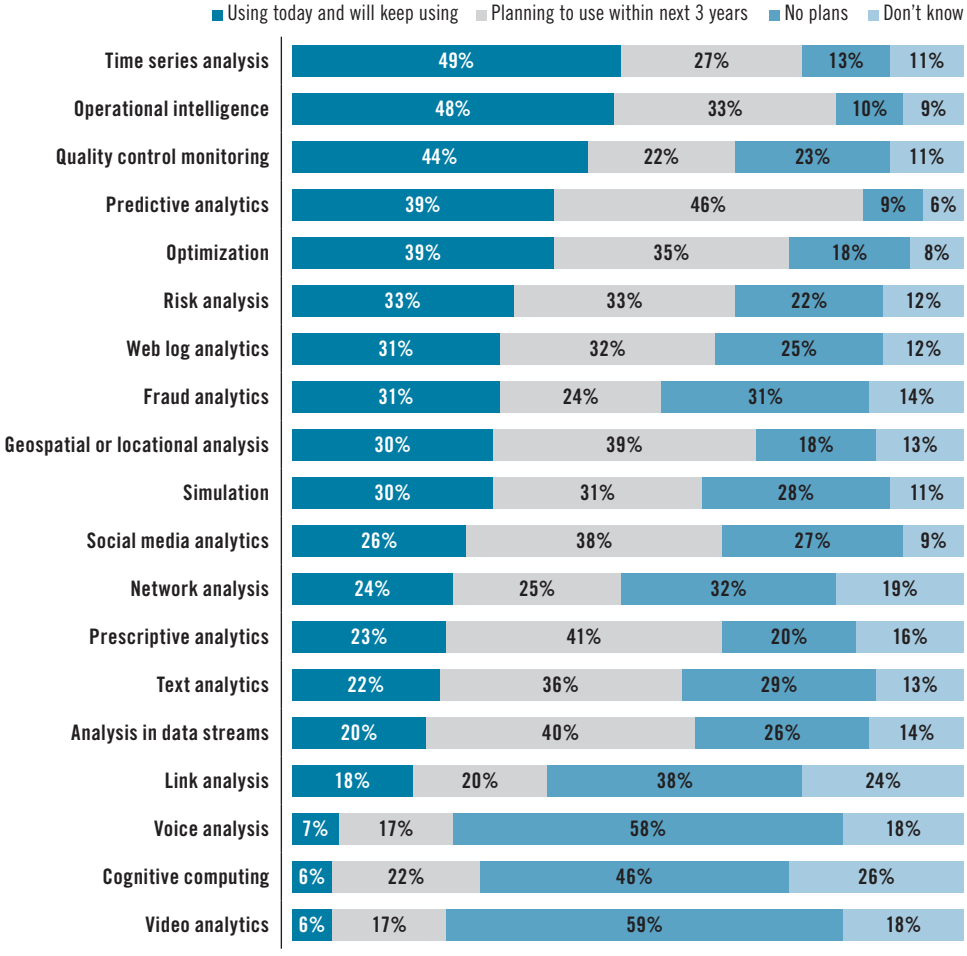


Figure 2. *Kinds of advanced analytics in use today. Based on 328 respondents.*

USER STORY

Companies often need true real-time monitoring and analysis for field operations. For example, an energy company is analyzing stream data to monitor and avoid unplanned maintenance disruptions on its equipment. It uses predictive analytics first to determine the conditions that might lead to a disruption. Then, as new data comes in, it is monitored and analyzed using complex event processing for the parameters of a potential failure, computing in-memory, in-stream and in real time. The company is expecting to minimize downtime by anticipating these issues.

Other advanced analytics techniques are also gaining in popularity. Respondents seemed interested in a range of technologies including optimization, risk analysis, Weblog analytics, and geospatial analytics. All are slated to double in usage over the next three years if users stick to their plans.

Advanced analytics such as geospatial analytics is also growing in adoption.

Gimme the Data!

An important aspect of next-generation analytics is that it should utilize disparate data. According to industry wisdom, only 20% of the data currently available is structured. That means most data is unstructured, in the form of text, audio, video, and so on. To get the most out of analytics, disparate kinds of data should be used. Next-generation platforms often support integration from multiple sources. Some vendors refer to this as data blending, others as unified information architectures, data virtualization, or the data fabric, among other terms.

Many analytic platforms are starting to support multiple data types as well. This means that a platform might support text analytics for natural language processing to extract key themes, entities, and sentiments from text data found in e-mail messages or online. This data is then treated as another data source for analysis. Some platforms will even provide voice recognition to extract words from videos as text structured for analytics. All of this can be very powerful for analysis. Churn analysis is a good example of where sentiment can help predict turnover.

The market is early for disparate data types. Currently, 94% of respondents stated they are using structured data for analytics. (See Figure 3.) Sixty-eight percent are enriching this structured data with demographic data for analysis. Slightly more than half are using time-series data. Other, more diverse data types are in use by fewer than 40% of the respondents. However, this situation is primed for change.

Structured data is still the most popular kind of data used for analysis.

Some data types set to grow. Sources such as internal text data, external Web data, and external social media data are set to double or even triple in use for analysis over the next three years. Likewise, while IoT data is used by fewer than 20% of respondents today, another 34% are expecting to use it in the next three years. Real-time streaming data, which goes hand-in-hand with IoT data, is also set to grow in use. In a separate question (not charted) about important functions for next-generation analytics platforms and tools, the ability to handle disparate data types ranked high (4.3 out of 5 possible points). As organizations start to truly understand and implement newer tools and platforms and obtain the skill set for doing so, it will become easier to use these kinds of data for analysis.

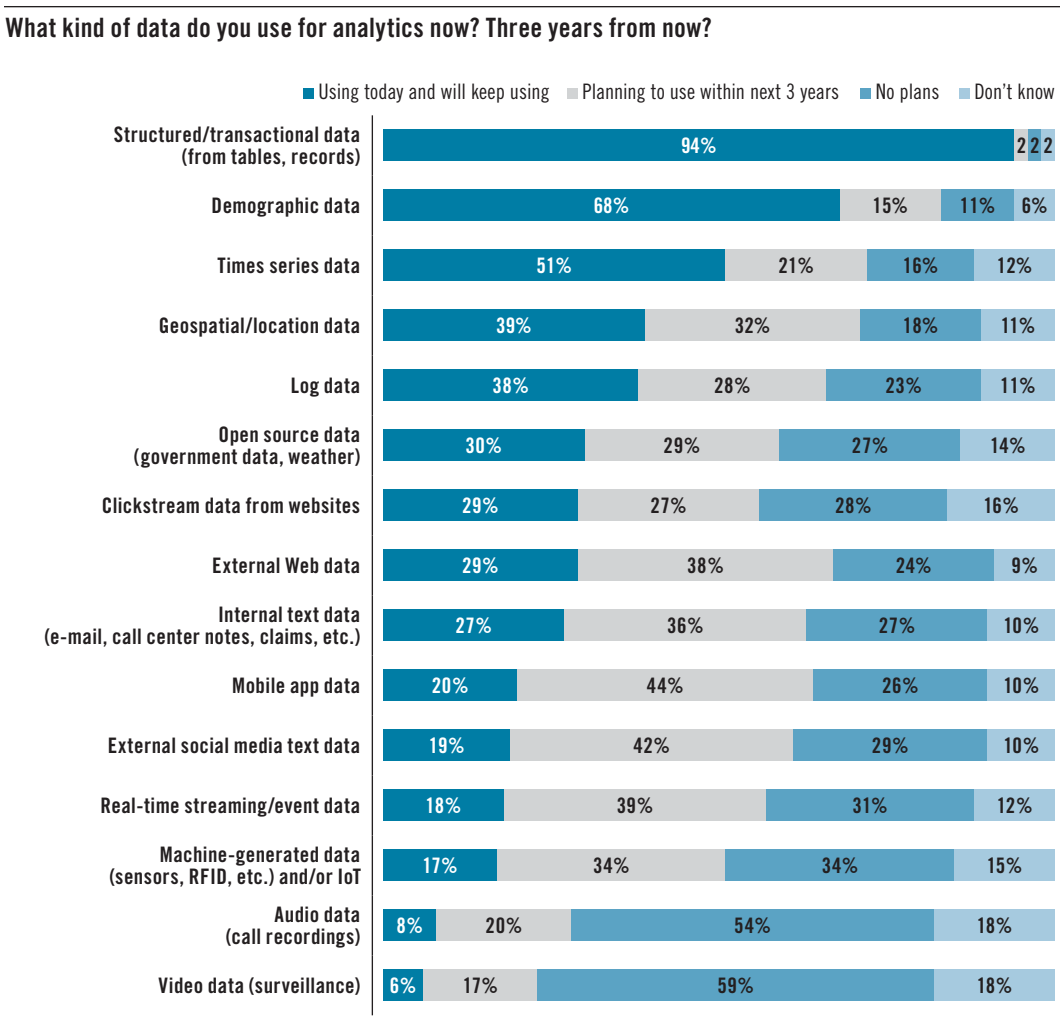


Figure 3. *Data types being used for analytics. Based on 328 respondents.*

Where Is It Being Used?

Next-generation analytics is often utilized in marketing and sales.

We asked respondents where next-generation analytics is being used in their companies. The top four answers were marketing or marketing analysis (46%), executive management (42%), sales (41%), and finance (40%) (See Figure 4.) Organizations are no doubt using forecasting in finance and sales. Additionally, one of the first places that predictive analytics takes hold is in marketing and sales. Hot on the heels of sales and marketing is customer service and support (40%), a good next step in more advanced analytics to drive impact. Both customer service and support and operations management are set to nearly double in numbers of organizations deploying these functions, if respondents stay true to their plans.

We find it interesting that when examining *only* those respondents who claim to already be using advanced analytics, the top four uses are the same, although with much higher response rates (not charted). Additionally, over half are already using analytics for customer service and support (52%), IT network or computer management (50%), and operations management (48%). These are often areas that companies attack once they become comfortable with next-generation analytics.

Another area that organizations with more advanced analytics are exploring more frequently is social media analysis (40%)—analyzing data found online in blogs, microblogs, news reports, and so on. This area is poised for growth among the wider respondent base as well. Social media analytics can be used for competitive intelligence, brand monitoring, and customer feedback analysis as well as other aspects of sales and marketing or support. Some companies utilize social media analytics on their foray into text analytics on internal data such as e-mail or call center notes. Some companies start internally and then move externally.

Where is next-generation analytics used in your company now? Three years from now?

Figure 4. *Parts of the organization where next generation analytics is being deployed. Based on 328 respondents.*

Who Is Using It?

TDWI Research points to business analysts as the new users of advanced analytics.

TDWI Research has uncovered a shift occurring in the builders of predictive models. Whereas in the past the builder was a statistician or other quantitative person, a current market move involves the business analyst in building models. It may be that models the business analyst builds have to be reviewed for quality control by a statistician or data scientist, but the move is apparent. In TDWI's 2013 Best Practices Report on predictive analytics, over 80% of active users of predictive analytics stated that business analysts would build models. This group felt that knowledge of the business, knowledge of the data, and critical-thinking abilities were enough to build a model. Seventy-nine percent stated that statisticians would be the primary builders. The assumption was that business analysts would be utilizing (that is, building) more advanced analytics models.

In this survey on next-generation analytics, we asked, "Who in your organization is involved in actually analyzing data using advanced analytics (i.e., predictive, text, Web analytics)?" In this case, respondents cited the top three users as business analysts (75%), statisticians (56%), and business users (50%) (See Figure 5.) This, too, points to the fact that business analysts, for better or worse, are more often analyzing data utilizing advanced analytics. This is an important factor in next-generation analytics.

Who in your organization analyzes data using advanced analytics (i.e., predictive, text, Web analytics)?

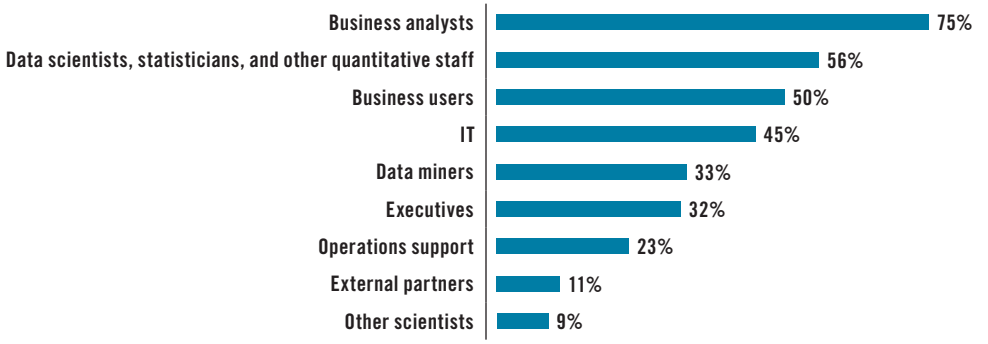


Figure 5. *Personnel using advanced analytics to analyze data. Based on 328 respondents. Multiple selections were allowed.*

EXPERT OPINION

"Predictive analytics today is the way visualization was a couple of years ago," according to Saken Kulkarni, solution principal, analytics, for Slalom Consulting. "It is on the verge of taking off with a new set of users, the data analysts. When it does take off, questions will be less about tools and methods and more about how analysts can share, collaborate, and obtain insights from these models. Analysts must be empowered to collaborate through a process similar to that which a developer uses to comment on code. Additionally, it is critical to help analysts manage models, keeping them accurate and up to date, especially as they accumulate hundreds of such models. These factors will be critical in the future as companies use both descriptive and predictive analytics."

Technologies That Support Next-Generation Analytics

Numerous platforms can support next-generation analytics.

Underlying analytics, of course, is the infrastructure to support it. Although the data warehouse still reigns supreme, most warehouses were designed and optimized for deliverables such as standardized reports and dashboards and online analytic processing (OLAP). The data warehouse might not be the best in terms of performance for all analytic workloads. An ecosystem of technologies are evolving to support analytics (see Philip Russom's Best Practices Report, *Evolving Data Warehouse Architectures in the Age of Big Data*, 2014[6]). These new technologies don't replace the data warehouse; rather, they complement it because they are optimized for workloads that manage, process, and analyze new forms of big data, non-structured data, and real-time data.

Platforms

Platforms come in many shapes and sizes, and typically provide an environment to process, store, and analyze disparate data. Examples include the following.

- **Hadoop file system:** Hadoop is an open source project managed by the Apache Software Foundation. It includes the Hadoop distributed file system (HDFS), which is a reliable, high-bandwidth, low-cost data storage cluster that facilitates the management of related files across machines. It also includes the MapReduce engine, a high-performance, distributed/parallel processing implementation of the MapReduce algorithm. It helps to break data into manageable chunks and then makes it available for either consumption or additional processing. More recently, YARN (yet another resource negotiator) has become an important part of the Hadoop ecosystem. YARN is a framework for job scheduling and cluster resource management. There are many other components of the Hadoop ecosystem, including Spark for stream processing and Impala for interactive SQL.

  The power of Hadoop is that it utilizes schema on read. With a data warehouse, you often have to know what the tables look like before loading data. With Hadoop, you can pull in data from any source or of any type and then determine how to organize it. One common use case for Hadoop is as a staging area for disparate data types. Some organizations store data in Hadoop and call it a "data lake." Other organizations are looking to process information and query the data in Hadoop. Some enterprises use it for analytics, especially big data analytics against semi-structured and unstructured data. Vendors are also providing commercial distributions of Hadoop to make it easier to use[7].

- **Analytic databases:** The following DBMS/platforms are designed specifically to support analytics[8]:

  - **In-memory databases:** These systems load data into memory, meaning that analytics engines process data and perform computations in RAM rather than on disk, thus avoiding time-consuming I/O. Theoretically, in-memory processing can be thousands of times faster than data access from disk, making it appealing for computationally intensive analytics such as complex querying or data discovery.
  - **Columnar databases:** These store data in columns, not rows, which makes querying faster.
  - **Massively parallel processing (MPP) databases:** These databases spread workloads over clusters of machines, enabling the servers to share workloads.
  - **Appliances:** Appliances integrate hardware and software (e.g., operating systems, DBMSs) tuned for data warehousing.

[6] Available at tdwi.org/bpreports.

[7] For more information about Hadoop, see the 20103 TDWI Best Practices Report *Managing Big Data*, available at tdwi.org/bpreports.

[8] For more information about columnar databases and other analytic databases, see the 2012 TDWI Checklist Report *Analytic Databases for Big Data*, available at tdwi.org/checklists.

- **Streaming/complex event processing engines:** This platform is used to track, analyze, and process data as an event happens. Typically these platforms combine data from multiple sources and utilize rules to identify important events.
- **In-database analytics:** This technology executes data management and analytics inside the database where the data resides, theoretically decreasing the cost of moving data from one environment to another for analysis. It will typically sit on top of a data warehouse.

Platform Status

Organizations are beginning to step up their use of some of these next-generation platforms, although the tried-and-true methods still prevail.

Databases, data warehouses, and flat files are still the platform of choice for analytics.

Databases, data warehouses, and flat files are still the platform of choice for analysis. The majority of respondents still use these three technologies for at least some portion of their infrastructure to support next-generation analytics. (See Figure 6.)

Interest in new and pre-packaged platforms is picking up steam. Usage is also increasing. More than half of all respondents (56%) are using some sort of platform or appliance for analytics. An additional 26% of respondents plan to use platforms in the next three years, a finding that is fairly consistent with results from our 2013 TDWI Best Practices Report on predictive analytics. The numbers in 2013 were higher, perhaps because the respondents were already using predictive analytics and therefore might be more advanced.

In-memory computing is set to double in the next three years.

Respondents also see the benefit of in-memory computing. Over 30% of respondents are using in-memory computing now, and usage is set to double in the next three years. Advanced techniques being used on big data can benefit greatly from in-memory computing, which is fast, thus reducing the time to interact with large amounts of data. The technology is also helpful when building especially complex models that may require iteration.

Mobile platforms are also growing. Mobile platform use is poised to more than double in the next three years. More than 20% of respondents are using this technology now. Mobile is especially useful for certain kinds of people in an organization, such as sales and operations staff. For example, mobile devices are being used more frequently on manufacturing floors to monitor operations.

Various DBMS are growing too. This includes MPP DBMSs (19%), in-memory DBMSs (15%), and NoSQL DBMSs (14%). Other technologies such as in-database analytics are also increasing in use.

Stream mining and CEP are relatively immature.

Stream computing and complex event processing platforms are still relatively immature. Fewer than 10% of respondents reported using either of these technologies. However, usage is expected to double in the next three years if users stick to their plans. Given respondents' expectations to be performing in-stream analysis and using stream data, these platforms might prove useful.

Of the following tools, infrastructure, and techniques, which is your organization using for next-generation analytics?

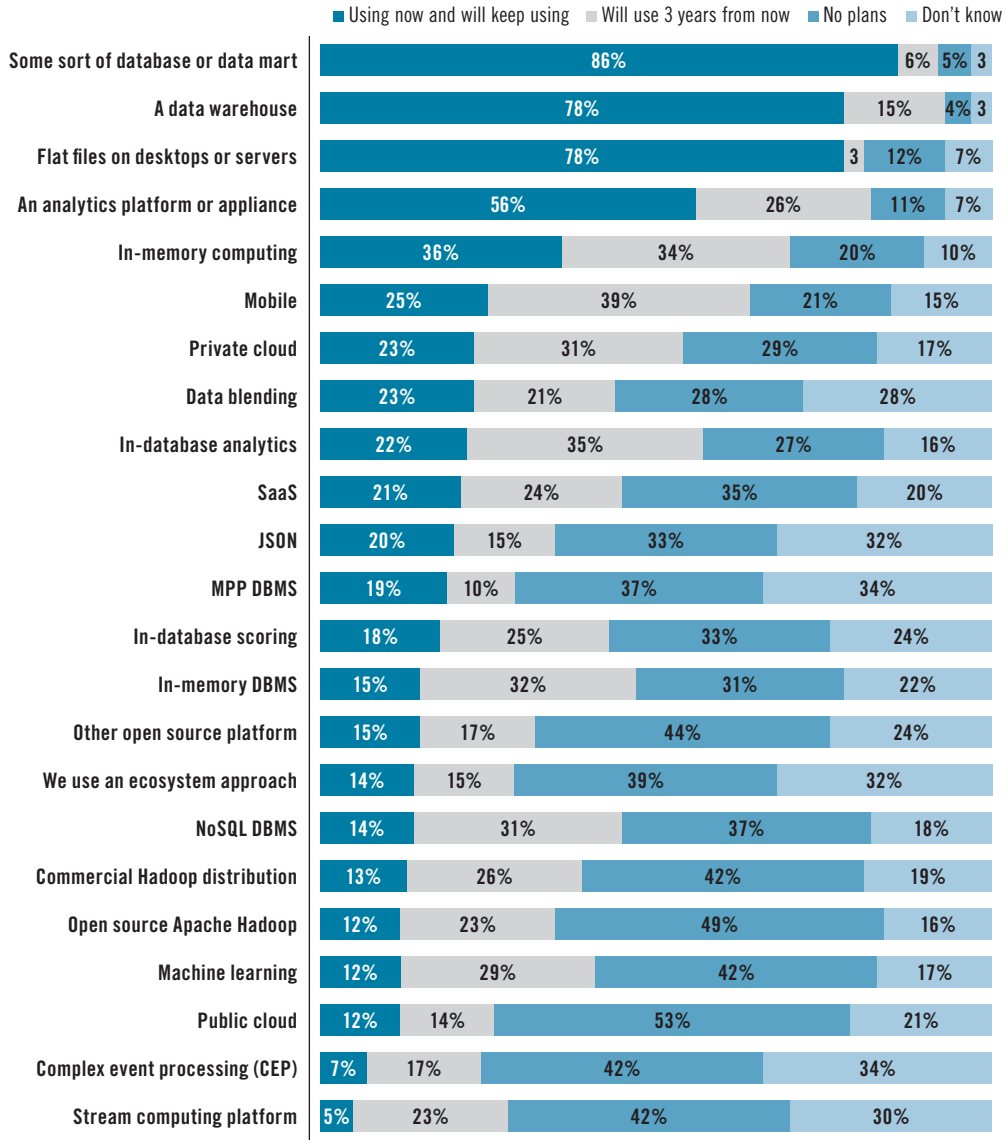


Figure 6. *Platforms in use for next-generation analytics. Based on 328 respondents.*

The Cloud

Cloud analytics is not just about software-as-a-service. There are many deployment and delivery models for using the cloud with BI and analytics, including public, private, and hybrid deployment models. Companies are making use of infrastructure-as-a-service (IaaS) and platform-as-a-service (PaaS) on public and private clouds as well. In Figure 6, more than 20% of respondents reported using a private cloud for next-generation analytics. About 12% reported using the public cloud.

Respondents who are more advanced in their analytics deployments are more likely to use the cloud in an intentional way.

In a separate question, we asked respondents if they are using the cloud now for BI or next-generation analytics activities (Figure 7). Fourteen percent said they would never use the cloud for BI or next-generation analytics; 40% are thinking about it. The rest are already using the cloud in some way or aren't sure. TDWI typically sees about 20% of survey respondents say they would never use the cloud for BI or analytics, so this group's responses are a bit of an improvement.

It does appear that respondents having more advanced analytics deployments are more likely to use the cloud in an intentional way. We compared two groups: those who were already using advanced analytics and those who had no plans to do so. Those respondents already using advanced analytics are more likely to use a hybrid cloud approach. In other words, they were using both public and private clouds. Organizations are making use of various types of cloud deployment and delivery options for analytics. TDWI research suggests that companies that tend to use the cloud for analytics are often more advanced in terms of their analytics. The hybrid model (public and private) is typical for these companies. For example, if data is generated in the public cloud, it is often analyzed there as well. This analysis might be basic or complex. More often, companies are capturing big data in the cloud and then experimenting with it there. Based on the analysis, certain data is brought on-premises to the data warehouse for reporting or dashboards. As one user put it, "We intend to use the cloud for quite a lot. We already have our CRM in the cloud. Our next-generation infrastructure might be cloud-based. We would use IaaS and store data and run applications in the cloud."

Do you use the cloud for any of your BI or next-generation analytics activities?

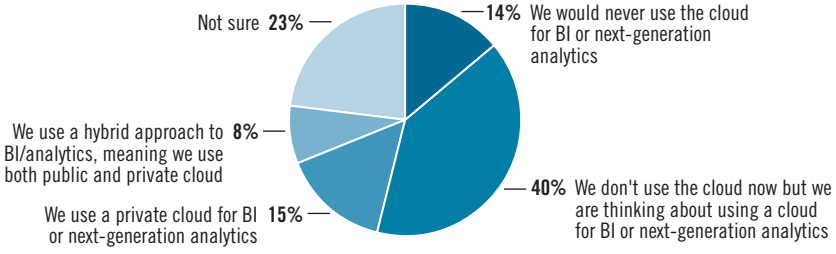


Figure 7. *Cloud usage for next-generation analytics. Based on 328 respondents.*

USER STORY

"Like many groups, our BI team focus started with finance, operations, and sales, areas which all continue to grow their analytical demands," said a senior IT director at a software firm. "Also, we have seen interest in analytics increase significantly in other parts of the business such as marketing, engineering, and product development. In order to support the increased demand and the intensified focus on data, we have been working to get a hybrid cloud architecture in place so that we have the necessary flexibility and scalability. As this comes into play, we are fostering the pockets of analytics that are developing and working with our chief data officer to extend consumption of our data and to foster data governance and stewardship."

Open Source

Aside from Hadoop and on the analytics front, the emergence of the R language is also evidence of the growing popularity of open source. Many analytics vendors are already incorporating support for R into their packages. The open source Python programming language is also increasingly popular for analytics. Open source is important because it enables a community to innovate, and that is what is happening around the analytics ecosystem.

Slightly more than 10% of respondents are using open source technologies to analyze or manage data.

Figure 6 shows that only about 12% of respondents reported using open source Apache Hadoop. About the same percentage reported using a commercial distribution of Hadoop. In a separate question, we asked respondents about their use of open source technologies, and 44% were not using it at all (not charted). Twenty percent (also not charted) are using open source technologies in conjunction with other commercial technologies in their organizations. Slightly more than 10% are using open source to analyze or manage data. Often, IT does not allow companies to use open source. Some people use it for proof of concepts only. When some organizations use open source for advanced analysis, they use it for prototyping, rather than putting a model into production.

Operationalizing Analytics: The Path to Action

Analytics must be actionable to be useful, and this is a big push with the next generation of analytics. Operationalizing refers to making analytics part of a business process; i.e., deploying analytics into production. In this way, the output of analytics can be acted upon. Operationalizing occurs in different ways. It may be as simple as manually routing all claims that seem to have a high probability of fraud to a special investigation unit, or it might be as complex as embedding analytics in a system that automatically takes action based on the analytics.

In order to make analytics useful it must be actionable.

Operationalizing also goes by many names, depending on your point of reference, including operationalizing analytics, operational intelligence, embedded analytics, real-time monitoring, or alerting. Some terms refer to operationalizing analytics as combining predictive analytics and decision rules together to enable decision management. Operationalizing might include event processing or stream mining. It might include in-database scoring, where a model is put into a database system and scored there as data comes into the database. Others may be talking about embedding models in a system that takes action. As data from the IoT, social media, and other big data sources becomes more common, organizations will have to determine how to deliver the right information to the right person at the right time. Analytics will become invisibly embedded, even if this simply means embedding analytics into a dashboard. The market is still relatively new to this concept.

The Status of Operational Analytics

We asked respondents which statement best describes how they have operationalized analytics. The results indicate many ways to operationalize analytics (Figure 8).

About 25% have not operationalized analytics. Eighteen percent of respondents clearly stated that they have not operationalized analytics at all. Another 7% did not know. This group probably uses analytics for insight and not for action or doesn't use it at all.

Organizations still operationalize analytics manually. An additional 15% stated that they take the output of analysis and manually fold it into a business process. For example, a marketing department might provide a list of names that a call center should contact for a promotion. The list is manually handed off to the call center.

Others have built custom integrations or use vendor software. About 34% of respondents are using either homegrown solutions or vendor solutions to integrate analytics into their operations.

A small percent embed analytics. Seven percent stated that they embed analytics into system processes. An additional 1% score data in motion in event streams.

Still others are running batch jobs to update their systems. About 11% of respondents are using batch updates as a form of operationalizing analytics. For example, they may update an analytics dashboard daily.

Which statement best describes how you have operationalized analytics?

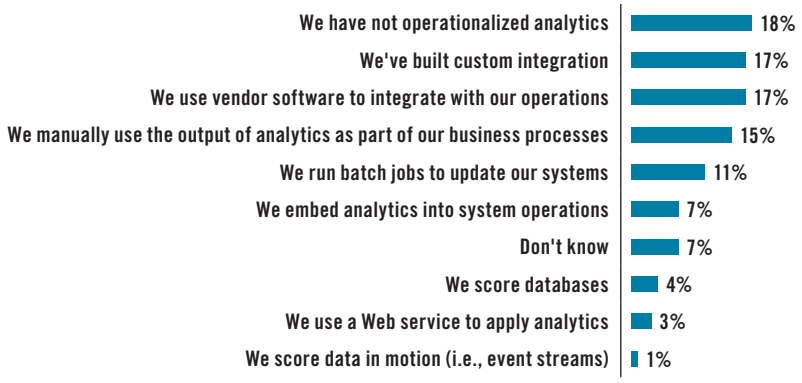


Figure 8. *How respondents have operationalized analytics. Based on 328 respondents.*

Analytics Informs Action

We asked respondents about how they take action on analytics. About 25% stated they had not taken action on analytics yet (not charted). However, the rest are using a variety of methods. Many of the respondents who were performing analytics used the manual approach for taking action, at least for some of their analytics. Fully 70% of respondents (not charted) stated that they "look at the results of our analysis and decide what action to take." In other words, most organizations still use analytics to help make manual decisions, at least in some part of the business, but they are starting to operationalize their analytics in a more automated way to act on the results.

However, organizations are advancing in how they make analytics actionable. For example, over a quarter of respondents were using some sort of alerting system (not charted). Fifteen percent claimed to have embedded analytics somehow into a business process that automatically takes action. This is slightly different from the 7% above who claimed to have embedded analytics into a system. Twenty-three percent were using interactive analytics systems that suggest a course of action to the organization. Although these are probably primarily dashboards, the point is that the market is moving forward in terms of taking action. Most organizations still use analytics to help make manual decisions but they may be starting to operationalize their analytics in a more automated way to act on the results.

EXPERT OPINION

"The value of any analytic lies in its ability to improve an organization's decision making," said James Taylor, CEO of Decision Management Inc. "As customers and employees become mobile, as organizations must operate worldwide 24x7, as customers continue to demand self-service and personalized service, the reality is that this decision making must be real time. Organizations can no longer rely on operationalizing analytics by having someone look at them or by handing off static deliverables to someone else in the process. The future belongs to organizations that can deeply embed increasingly sophisticated analytics into their decision making so that data-driven, analytic decisions can be taken in real time. The days of batch calculation of analytics are surely numbered because the right response to an event or stimulus requires analytics calculated right then using every data point available. Real-time execution of analytics in line, as a process executes and a decision must be made, is the future. Because these analytics must evolve quickly to capture narrow windows of opportunity and change to react to evolving market conditions, the operationalization of these analytics must become more automated also. Organizations are learning that analytics are not "done" when they have the insight; they are done *only* when the business has been improved. In a real-time, mobile, fast-changing world, operationalizing analytics is critical to deliver long-term value from data and analytic investments."

"Organizations are learning that analytics are not 'done' when they have the insight; they are done only when the business has been improved."

Challenges and Emerging Best Practices for Next-Generation Analytics

Organizations clearly must make many, many technology choices as they move their analytics programs forward. Gone are the days of simply deploying a database or data warehouse for analytics. Is it becoming overwhelming? What are some of the challenges that organizations are encountering with next-generation analytics and platforms and how are they overcoming these challenges? Figure 9 illustrates these challenges.

Lack of skills is a key challenge for next-generation analytics.

What are your top next-generation analytics challenges? Select 3 or fewer.

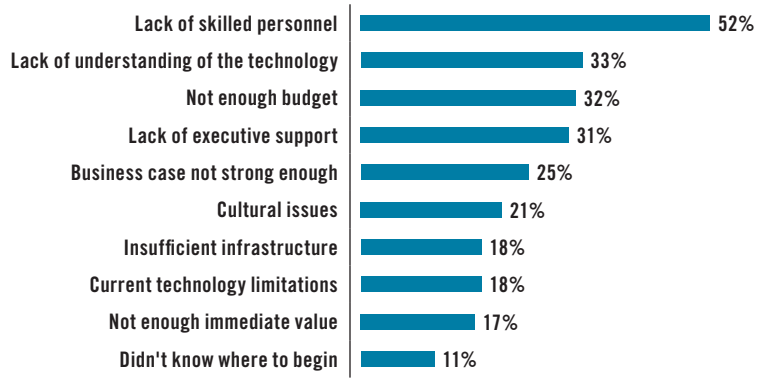


Figure 9. *Next-generation analytics challenges faced by respondents. Based on 328 respondents. Multiple selections were allowed.*

Lack of skills ranks at the top. Not surprisingly, companies are concerned about the lack of skills that they have in place to deal with new data and new analytics. In fact, over half (52%) cite this as a top challenge. This is a new skill set that requires people with knowledge of Hadoop, newer kinds of databases, and big data and advanced analytics. Thirty-three percent of respondents cite the lack of understanding of the technology as a big challenge. Skills and understanding technology, of course, go hand in hand.

Budget approval, a clearly identified business case, and executive support are also issues. Respondents also cite these interrelated factors as challenges to successful next-generation analytics. To get funding, you often need to build the business case—at least to move past the experimentation phase. Additionally, it often helps to have an executive championing your effort because it can help to get initial funding for an effort.

Cultural and organizational issues are also challenges for next-generation analytics.

Cultural issues can hinder progress. Twenty-one percent of respondents said cultural issues are a challenge for practicing next generation analytics. As one respondent put it, "We do not have an analytics culture; we have pockets of analytics. There are different issues that come up. Sometimes people will want to stick with what they have because they are afraid they will lose headcount, even if a new technology is 10 times faster." Another respondent said, "We really need the right person to help us evangelize analytics because otherwise no one will get excited about it. We had [one group] and that didn't work. We're trying something different now and we'll know soon enough."

Overcoming the Challenges

What are respondents doing to overcome the challenges? Their responses naturally fell into several groups.

"Training and patience is how we're overcoming challenges."

- **Skills.** Skills ranked at the top of the list of next-generation analytics and platform challenges. To overcome this challenge, some respondents talked about hiring fewer but more skilled personnel such as data analysts and data scientists. Others talked about training from within because current employees understand the business. Our survey revealed that many organizations are doing both (see next section). Additionally, some organizations are building competency centers where they can train from within. They also include workshops for business analysts. When organizations don't have an adequate training budget, many are doing self-study. Some are taking a phased approach to training. As one respondent put it, "Training and patience is how we're overcoming challenges."
- **Education.** In addition to training staff to understand technologies, many respondents cited the need for educating their organizations about the value of next-generation analytics. Often this involves socializing and evangelizing the concepts, especially if an executive is not yet on board. As one respondent explained, "You need to get out in front of people. We are getting out to various departments and managers and sitting down with them to understand what their business metrics are and then we're figuring out a better way to deliver these to them." Another respondent stated that he is really now in the "business development" business. Still another respondent spoke about "unrelenting advocacy and politicking."
- **Proof of concept (POC).** A number of respondents talked about the value of a POC, or several variations of one. For example, some respondents talked about slowly making good POCs on real business problems, or taking small use cases with definitive value and scaling up—i.e., "Smaller chunks to prove value." They talked about trial and error and continuous reinforcement and using what they have on hand to show value. Some organizations are taking a bolder approach and developing disruptive use cases to get people's attention. The success of this approach depends on the culture.

- **Funding.** Many respondents talked about finding the right business case and presenting the ROI for it. Some are monitoring opportunities to find the strong business case. Others are looking at competitor examples in their peer community to help drive the business case for the funding. Still others were fortunate enough to have an executive sponsor willing to fund even small efforts.

Unfortunately, there is no silver bullet when it comes to overcoming these challenges.

Acquiring Skills

Securing the necessary analytics skills is a big part of success for next-generation analytics and platforms. We asked respondents how they were learning these skills and their answers are shown in Figure 10. As discussed, over 40% of respondents are using a combination of approaches to obtain the skills they need for next-generation analytics, including training from within and hiring from outside. Only a small group (3%) claim to already have the needed skills. Only a tiny percentage (1.2%) are hiring from universities.

Over 40% of respondents are using a combination of approaches to obtain needed skills.

Which statement best describes how you find the skills to deal with next-generation analytics and platforms?

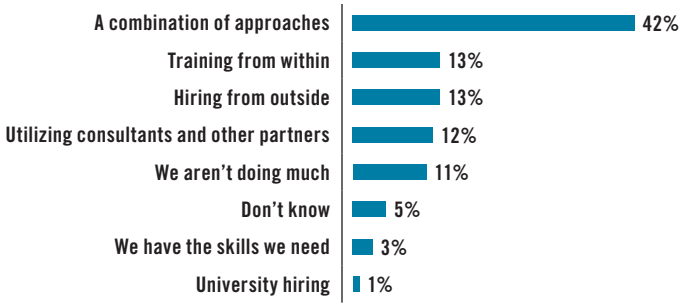


Figure 10. *Respondents' sources for acquiring skills. Based on 328 respondents.*

Other Best Practices

We asked respondents about other best practices related to next-generation analytics. In particular, we asked about whether they use a center of excellence and how they measure results over time.

Center of excellence. A CoE typically consists of a cross-functional team that provides leadership in analytics. The team might be responsible for working on analytics problems, training, and disseminating best practices. It evangelizes analytics. Thirty-four percent of respondents already have a CoE in place (not charted); an additional 30% are planning to in the next year or two. These teams can be useful because they help provide consistency and innovation across the organization.

Measuring results. An analytics culture is necessary when dealing with next-generation analytics. This culture is data-driven, which means results-driven. Results inform actions, a best practice for analytics, whether it is simply tracking key performance indicators or monitoring the results of more advanced models that get stale. How often you update a model depends on the model. It can be a period as long as six months or as short as daily (or shorter, depending on how advanced the model is).

About 40% of respondents measure and adjust analytics monthly.

We asked respondents how they measure results and improve over time. The results are presented in Figure 11. Twenty-eight percent did not measure results; an additional 9% did not adjust analytics, even if they did measure results. That means more than a third of respondents may not be data- or analytics-driven. About 40% of respondents measure and adjust analytics monthly. Fewer than 25% measure and adjust metrics more often than monthly.

How do you measure results and improve over time?

Figure 11. *How respondents measure results and improve over time. Based on 328 respondents.*

Measuring the Value of Next-Generation Analytics

We have been citing best practices for next-generation analytics throughout this report. However, in order to explore best practices further, we examined the characteristics of companies that are monitoring and obtaining measurable value from next-generation analytics. We also wanted to see if those companies differed from companies that were not reporting measurable value.

Those respondents who measured a top- or bottom-line impact were more likely to be using advanced analytics for several years.

We separated those using the technology into two groups—those who measured either a top- or bottom-line impact (or both) from next-generation analytics (81 respondents) and those who could not measure an impact (181 respondents). (The latter group stated they thought they might gain efficiencies or insights but could not measure anything.) There were some interesting results.

- **Advanced analytics:** Those who measured a top- or bottom-line impact were more likely to have been using advanced analytics for several years versus those who did not. For example, those who have utilized operational intelligence or optimization for more than two years were more likely to measure a top- or bottom-line impact than those who have not.
- **Analyzing big data.** Those who did not measure a top- or bottom-line impact are more likely not to have any short-term plans for big data.
- **Taking action.** Not surprisingly, those respondents who take action on their data are more likely to measure value than those who do not.
- **Analytics ecosystem.** Those who measured a top- or bottom-line impact were more likely to already have a coordinated analytics ecosystem in place.
- **Technology infrastructure.** Those who measured a top- or bottom-line impact were also more likely to be using mobile technology or the public cloud.

Correlation does not mean causation.

Of course, correlation does not imply causation. However, it appears that those respondents who are more mature in their analytics efforts are more likely to measure value than those who are less mature. These organizations are already incorporating more advanced analytics, running on both old and new technologies. It takes time to build an analytics-driven organization. With time seems to come reward.

Vendor Platforms and Tools That Support Next-Generation Analytics

The firms that sponsored this report are all good examples of software vendors that offer either analytics and/or platforms as well as professional services that support next-generation analytics. These sponsors form a representative sample of the vendor community, yet their offerings illustrate different approaches to integrating Hadoop with business intelligence, data warehousing, data integration, and analytics.[9]

Actian

Actian provides a Hadoop-based analytics platform for big data that is designed to accelerate time to value, even for sophisticated analytics. The platform offering includes a visual data flow framework for data blending, enrichment, and analytics, so minimal coding is required. More advanced analytics on Hadoop is provided via partners. For example, Actian provides pre-built data access, quality, and transformation capabilities for use with KNIME. In addition, high performance (approaching real-time speed) is made possible by deploying the Actian X100 vector processing engine across every node of the Hadoop architecture, with YARN-certified administration. Actian's Extreme Performance Edition enables real-time and low-latency use cases, including fraud detection, trading surveillance, network intrusion detection, digital ad optimization, and campaign optimization.

Cloudera

Cloudera was founded in 2008 to develop the first enterprise-class implementation of Apache Hadoop. The company now provides a family of products and services that complement open source Apache Hadoop software products for big data. Cloudera refers to this as the Enterprise Data Hub. The idea is to use this hub to store, process, and analyze big data. The hub can also connect back out to other data stores, such as a data warehouse. CDH5 is the latest release of Cloudera's distribution of the Apache Hadoop Distributed File System (HDFS) that supports the hub. Cloudera Manager makes HDFS viable for serious enterprise users by providing expert technical support, upgrades, and administrative tools for Hadoop clusters as well as professional services to assist implementations, training, and certification. Cloudera is a major contributor to open source Hadoop technologies and it has recently shown leadership in the design of Impala, an SQL engine for real-time, ad hoc queries in Apache Hadoop. It is also providing commercial distributions of Spark for real-time streaming and in-memory computing.

Datawatch Corporation

Datawatch provides visual data discovery software based on its 2013 acquisition of Panopticon Software. Datawatch Desktop and Datawatch Server enable users to interactively explore data through a range of visualizations. Datawatch's visual discovery software can consume streaming data from CEP engines, message brokers, and other sources. The software can visualize data "in motion" in real time and use algorithms to identify anomalies, including data that is part of the IoT and the industrial Internet (such as true real-time data from sensors). Views can show live, not just static, data. Users can build ad hoc aggregations to support OLAP-style slicing and dicing and drill-down. Users can run continuous queries and set up alerts to enable notification of trends and events. Through dashboards, users can visually filter real-time feeds and use algorithms to identify anomalies and outliers from a variety of "data in motion" sources. For operational BI and analytics, users can discover and monitor key performance indicators in real time.

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

Pentaho

Pentaho, founded in 2004, provides a business analytics platform that combines data integration and business analytics. It provides an enterprise data integration component based on the Kettle open source ETL project. The platform also includes a big data layer that provides integration across Hadoop, NoSQL, and analytic database systems and enables data blending across these and traditional data stores. Pentaho ensures portability across Hadoop distributions from Cloudera, Hortonworks, MapR, Intel, Cassandra, MongoDB, and Splunk. In addition to traditional reporting and dashboards, the company's business analytics tools include visualization tools, data mining, and predictive algorithms along with an analytic modeling workbench. Its predictive analytics features include visualizations and integration with R and Weka. Pentaho also offers a cloud-based solution for both data integration and analytics. Pentaho Labs is constantly working to bring new innovations to market.

SAP

Software leader SAP has recently put a great deal of focus into its analytics portfolio. It acquired predictive analytics vendor KXEN in 2012. Since then it has been integrating products and providing new next-generation capabilities, such as support for geospatial data, for both its visualization product (SAP Lumira) and its predictive analytics suite (Infinite Insight 7.0). Infinite Insight has also added support for Apache Hive. SAP has integrated Lumira with its core BI suite. Perhaps more important, the company has set forth a strategy to make its analytics easy to use by employing mobile design principles. It is also simplifying its technology stack and putting SAP HANA as its core. SAP HANA has been re-launched by the company as an in-memory platform for databases, data management, and applications. It provides libraries for more advanced analytics such as predictive analytics, text analytics, and geospatial analytics. SAP enables its customers to integrate Hadoop into their existing BI and data warehousing environments in multiple ways, giving customers the ability to tailor the integration to their needs. The company also offers its analytics in the cloud.

SAS

SAS offers software for the analytics life cycle including products for data management and analytics. As a leader in advanced analytics, machine learning, and data mining, the company offers software for advanced statistical analysis, data mining, text analysis, optimization, and more in SAS Enterprise Miner. SAS ETS provides support for forecasting and time-series analysis. SAS High Performance Computing is specifically designed to support initiatives for big data analytics, and it includes in-memory, in-database, and grid computing in a shared-nothing hardware environment. It supports various kinds of analysis, including predictive analytics, as well. SAS has shown leadership by promoting in-database analytics, i.e., where SAS analytic functionality is pushed down for execution in another vendor's database. In-database analytics is now supported for Hadoop. SAS also provides a cloud solution.

Top Ten Best Practices

1. **Realize there is no silver bullet, but don't do nothing.** Building an analytics culture using next-generation analytics and putting the ecosystem together takes time. It is important not to try to boil the ocean. However, it is also important not to ignore the work and simply hope that success will magically happen. Companies that are measuring value with analytics are taking risks, experimenting, and finding success. They are evangelizing and communicating. It may take time, but they are certainly getting there.
2. **Consider new infrastructure technology.** Companies succeeding with next-generation analytics are putting an ecosystem together that consists of multiple technology types. Yes, this can include the data warehouse (don't expect the new stuff to replace the old), but it should also include the right tools for the jobs, including in-memory computing for highly iterative analysis or the cloud to deal with vast amounts of data that might be generated in the public cloud and on premises.
3. **Consider more advanced analytics.** Companies measuring value are using more advanced analytics. Although this requires skills and training, the upside is clear. Often a good first step into the world of advanced analytics is predictive analytics. Vendors are making the tools easier to use, and with the right controls in place (see below), this can be a good place to start.
4. **Start with a proof of concept.** Companies succeeding with predictive analytics often start with a metric they are already measuring, so they can demonstrate that they can predict that metric and they know it is valuable and will get attention.
5. **Utilize disparate data.** Although structured data and demographic data are the mainstay of analysts and modelers, disparate data types can enrich a data set and provide lift to models. Think about incorporating data beyond the traditional types that you might have in your data warehouse or on your servers. Good starting points include geospatial data and text.
6. **Take training seriously.** The democratization of analytics is moving ahead. However, you need to think about the skills you will require for data management as well as the skills to build your models and deal with your data. With statisticians and other quants in short supply, think about what skills you will need for the kinds of models you want to build. Part of the process is balancing the costs and benefits of the models you are considering. Allocate your resources wisely. Training will become an important part of your next-generation strategy.
7. **Put controls in place.** Democratization means that business analysts will try to use more advanced technology. Make sure controls are in place before a model is put into production. This might include confirming the validity of a model.
8. **Act on your data.** Analytics without action won't yield measurable impact. Even if you aren't ready to operationalize your analysis, it makes sense to start to implement a process to take action, even if it is manual action. You will be building a more analytically driven culture for when you want to build more operational intelligence.
9. **Build a center of excellence.** A CoE can be a great way to make sure that the infrastructure and analytics you implement are coherent. CoEs can help you disseminate information, provide training, and establish or maintain governance.
10. **Remember to monitor your analysis.** Data can get stale. Models can get stale. It is important to revisit any kind of analysis where action is taking place on a periodic basis to make sure that your data is still relevant and that your model still makes sense.

Actian

www.actian.com

Actian transforms big data into business value for any organization—not just the privileged few. Actian drives revenue growth while mitigating risk with high-performance, low-latency in-database analytics, extensive connectivity, and data preparation. Using off-the-shelf hardware, the Actian Analytics Platform empowers users to connect, analyze, and take action on big data with actionable insights delivered in seconds or minutes. Actian makes Hadoop enterprise-grade, providing high-performance data enrichment, visual design, and SQL analytics with no need for MapReduce. Tens of thousands of organizations apply Actian analytics in financial services, telecommunications, digital media, healthcare, and retail. Follow Actian at www.actian.com, Facebook, Twitter, and LinkedIn.

Cloudera, Inc.

www.cloudera.com

Cloudera is revolutionizing enterprise data management by offering the first unified platform for big data, an enterprise data hub built on Apache Hadoop. Cloudera offers enterprises one place to store, process, and analyze all their data, empowering them to extend the value of existing investments while enabling fundamental new ways to derive value from their data. Only Cloudera offers everything needed on a journey to an enterprise data hub, including software for business-critical data challenges such as storage, access, management, analysis, security, and search. As the leading educator of Hadoop professionals, Cloudera has trained over 22,000 individuals worldwide. Over 1,000 partners and a seasoned professional services team help deliver better time to value. Finally, only Cloudera provides proactive and predictive support to run an enterprise data hub with confidence. Leading organizations in every industry plus top public sector organizations globally run Cloudera in production. www.cloudera.com

Datawatch Corporation

www.datawatch.com

Datawatch provides visual data discovery software that optimizes any data—regardless of its variety, volume, or velocity—delivering next generation analytics to reveal valuable insights for improving business. Its unique ability to integrate structured, unstructured, and semi-structured sources like reports, PDF files, and EDI streams with real-time streaming data into visually rich analytic applications allows users to dynamically discover key factors that impact any operational aspect of their business. This ability to perform visual discovery against any data sets Datawatch apart in the big data and visualization markets. Organizations of every size, worldwide, use Datawatch products, including 93 of the *Fortune* 100.

Pentaho

www.pentaho.com

Pentaho is building the future of business analytics. Pentaho's open source heritage drives our continued innovation in a modern, integrated, embeddable platform built for accessing all data sources, including diverse and big data. Pentaho tightly couples data integration with business analytics to bring together IT, data scientists, and business users to easily access, mine, visualize, and explore all data that impacts business results. For more information visit pentaho.com or call +1 866-660-7555.

SAP

www.sap.com

SAP is at the center of today's technology revolution, developing innovations that not only help businesses run like never before, but also improve the lives of people everywhere. As the market leader in enterprise application software, we help 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 248,500 customers to operate profitably, adapt continuously, and grow sustainably.

SAS

www.sas.com

SAS is the leader in advanced analytics software and services, and the largest independent vendor in the business intelligence market. Through innovative solutions, SAS helps customers at more than 70,000 sites worldwide benefit from advanced analytics—whether for revealing relationships between pricing and demographics, understanding how social media influences your customers, or predicting risk and fraud. With an unwavering focus on analytics since 1976, SAS offers the breadth and depth of advanced algorithms and techniques such as machine learning, text analytics, and a broad set of predictive techniques. Learn more about SAS advanced analytics at http://www.sas.com/en_us/software/analytics.html

TDWI RESEARCH

TDWI Research provides research and advice for data professionals worldwide. TDWI Research focuses exclusively on business intelligence, data warehousing, and analytics 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, data warehousing, and analytics 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.

tdwi

555 S Renton Village Place, Ste. 700
Renton, WA 98057-3295

T 425.277.9126
F 425.687.2842
E info@tdwi.org

tdwi.org