TDWI RESEARCH

FOURTH QUARTER 2014

TDWI BEST PRACTICES REPORT

Real-Time Data, BI, and Analytics

Accelerating Business to Leverage Customer Relations, Competitiveness, and Insights

By Philip Russom, David Stodder, and Fern Halper

tdwi.org

tdwi

Research Sponsors

Actian

Cloudera, Inc.

Datawatch Corporation

HP Vertica

IBM

Tableau Software

Treasure Data, Inc.

WebAction, Inc.

BEST PRACTICES REPORT

FOURTH QUARTER 2014

Real-Time Data, BI, and Analytics

Accelerating Business to Leverage Customer Relations, Competitiveness, and Insights

By Philip Russom, David Stodder, and Fern Halper

© 2014 by TDWI (The Data Warehousing Institute™), a division of 1105 Media, Inc. All rights reserved. Reproductions in whole or in part are prohibited except by written permission. E-mail requests or feedback to info@tdwi.org. Product and company names mentioned herein may be trademarks and/or registered trademarks of their respective companies.

Table of Contents

About the Authors

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 500 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. You can reach him at prussom@tdwi.org, @prussom on Twitter, and on LinkedIn at linkedin.com/in/philiprussom.

DAVID STODDER is director of TDWI Research for business intelligence. He focuses on providing research-based insight and best practices for organizations implementing BI, analytics, performance management, data discovery, data visualization, and related technologies and methods. He is the author of TDWI Best Practices Reports on mobile BI and customer analytics in the age of social media, as well as TDWI Checklist Reports on data discovery and information management. He has chaired TDWI conferences on BI agility and big data analytics. Stodder has provided thought leadership on BI, information management, and IT management for over two decades, and he was the founding chief editor of Intelligent Enterprise, where he served as editorial director for nine years. You can reach him at dstodder@tdwi.org.

FERN HALPER, Ph.D., is 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 more than 20 years of experience in data and business analysis and has published numerous articles about data mining and information technology. Halper is co-author of "Dummies" books on cloud computing, hybrid cloud, service-oriented architecture, service management, and big data. Her Ph.D. is from Texas A&M University. You can reach her at fhalper@tdwi.org, or follow her on Twitter: @fhalper.

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, data warehousing, and analytics 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 solutions. TDWI also fosters the advancement of 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, on-site courses, certification, solution provider partnerships, an awards program for best practices, live Webinars, resourceful publications, an in-depth research program, and a comprehensive website, tdwi.org.

About the TDWI Best Practices Reports Series

This series educates technical and business professionals about new 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.

To support the program, TDWI seeks vendors that collectively wish to evangelize a new approach to problems or an emerging technology discipline. By banding together, sponsors can validate a new market niche and educate organizations about alternative solutions to critical issues. To suggest a topic that meets these requirements, please contact one of the TDWI Research Directors.

Research Methodology and Demographics

Report Purpose. This report educates organizations worldwide about how and why they should accelerate their business operations and insights into real time. A number of real-world use cases are now established in which real-time techniques help grow customer accounts, achieve operational excellence, and keep pace with competitors. To assist users, many new products and technologies have arrived recently from software vendors and the open source community. This report describes all this and more.

Terminology. In this report, *real time* is defined as a number of time frames of varying lengths in which processing for data, business intelligence (BI), and analytics execute to support fast-paced business processes for operational BI, just-in-time inventory, business monitoring, customer relations, and so on.

Survey Methodology. In April 2014, TDWI sent an invitation via e-mail to the BI professionals in its database, asking them to complete an Internet-based survey. The invitation was also posted in Web pages, newsletters, and publications from TDWI and other firms. The survey drew responses from almost 400 respondents. From these, we excluded academics and vendor employees. The resulting responses of 365 respondents 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 recognized data management experts. TDWI also received briefings from vendors that offer products and services related to real-time data, BI, and analytics.

Survey Demographics. The majority of survey respondents are IT professionals (63%), whereas the others are consultants (20%) and business sponsors or users (17%). We asked consultants to fill out the survey with a recent client in mind.

The financial services industry (14%) dominates the respondent population, followed by consulting (10%), software/Internet (8%), transportation/logistics (7%), healthcare (6%), manufacturing (non-computers) (6%), and other industries. Most survey respondents reside in the U.S. (49%) or Europe (17%). Respondents are evenly distributed across all sizes of organizations.

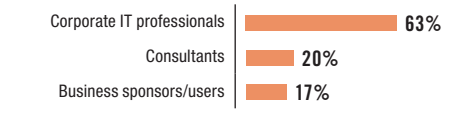
Acknowledgments

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

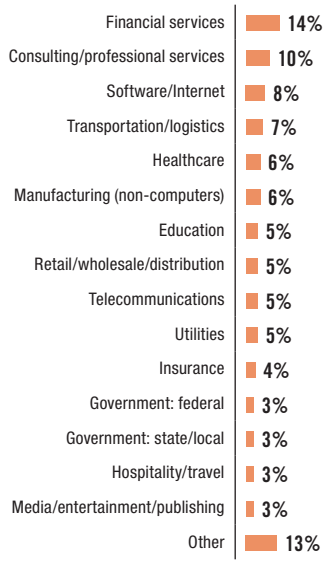
Sponsors

Actian Corporation, Cloudera, Inc., Datawatch Corporation, HP Vertica, IBM, Tableau Software, Treasure Data, Inc., and WebAction, Inc., sponsored the research and writing of this report.

Position

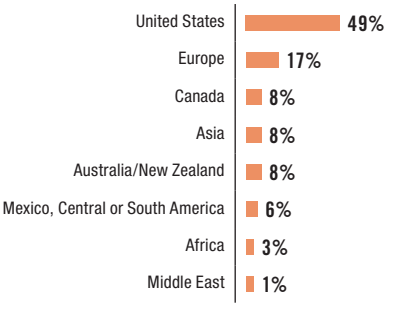


Industry

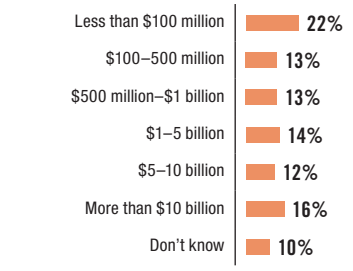


("Other" consists of multiple industries, each represented by 2% or less of respondents.)

Region



Company Size by Revenue



Based on 365 survey respondents.

Executive Summary

Real-time data, BI, and analytics are the basis of many desirable business practices.

User organizations continue to push their applications for business intelligence (BI), analytics, and data management closer to real-time operation. This is because fresh information can support fast-paced, time-sensitive business processes, such as operational BI, real-time management dashboards, just-in-time inventory, high-yield manufacturing, facility monitoring, call center information delivery, self-service information portals, recommendations in e-commerce, and so on.

According to this report's survey, the leading benefits of implementing real-time technologies for data, BI, and analytics include better actions based on BI and analytic information, improved customer service and customer experience, and automated decisions made by software, not people. The barriers to achieving such benefits include cost, difficulties in designing real-time systems, the state of data management infrastructure, and inadequate staffing or skills.

Real-time is a broad concept involving many types of processes, tools, and users.

The term *real time* has become an umbrella concept encompassing multiple time frames, speeds, and execution frequencies. BI and analytic practices are progressively rarer as we come closer to true real time (milliseconds and nanoseconds). For example, most reports and analyses today operate on data that's refreshed daily on a 24-hour cycle. Even so, BI and analytics based on hourly cycles is well established, and so-called *near-real-time* cycles (seconds and minutes) are fairly common.

A variety of technologies are in use today, enabling real-time and near-real-time data, BI, and analytics. Common ones include data federation, replication, data sync, message buses, and microbatches. A number of functions designed for high performance are increasingly applied to near-real-time uses, namely changed data capture, columnar databases, in-database analytics, in-memory processing, and solid state drives. Organizations in need of true real time are ramping up the use of complex event processing (CEP) and continuous stream mining.

Operational BI is a common manifestation of real-time data and BI.

The proliferation of operational BI is the leading driver for real-time usage in BI. Most operational BI users interact with near-real-time data via management dashboards and other types of reports. However, users are aggressively adopting visual data discovery tools because of their ease of use, strong self-service functions (which empower many user types), and the tools' ability to process large data sets with near-real-time performance.

Analytics is aggressively accelerating into real time.

As user organizations move deeper into analytics (as a complement to reporting), they also move deeper into real time. For example, this report's survey ranks operational BI and reporting as leading practices today, with 43% of respondents using these with some level of real time. However, 34% of respondents are already practicing operational analytics in real time, with an additional 34% planning to adopt the same within three years.

Other areas that should experience strong adoption within three years include real-time data warehousing (and similar practices such as active and dynamic warehousing), predictive analytics, visualization, social media analytics, text analytics, CEP, and stream mining.

A growing range of data types will soon be handled in real time.

Most of the data handled in real time today is structured (or more specifically relational), followed by application logs and semi-structured data, but survey respondents anticipate more real-time handling for social media data, Web logs and clickstreams, and unstructured data.

This report helps users understand the many available technologies and practices for real-time data, BI, and analytics, plus how these enable fast-paced business processes and yield a wide range of organizational advantages.

Introduction to Real-Time Data, BI, and Analytics

Real-time operation is diverse, including many enabling technologies executed in many time frames.

The pace of business and other organizational activity continues to accelerate, such that organizations must now react faster and more frequently to customer interactions, service-level agreements (SLAs), operational commitments, competitive pressures, and other time-sensitive issues. With real-time technologies maturing and proliferating, organizations must now evaluate the numerous options they need for the real-time delivery, access, and analysis of information, if they are to reduce or eliminate business latency.

One of the insights explored in this report is that most organizations need multiple solutions to achieve real-time data, business intelligence (BI), and analytics. Two reasons stand out:

Real time has many meanings. As with most IT disciplines, real-time solutions should be designed to satisfy business requirements. This leads to tremendous diversity because business processes have different requirements for the freshness of information and hence various requirements for how fast and frequently data should be fetched, processed, and delivered. The catch is that each of the many technologies available today for real-time data, BI, and analytics has unique characteristics for performance, data types, interfaces, and processing strengths. Most organizations need multiple real-time technologies to satisfy the diverse speeds and data content of real-time business processes.

Data travels across many software systems. Even a basic technology stack for BI, data warehousing, and analytics will include multiple tool and platform types such as those for databases, data integration, reporting, and analytics. When data must travel in real time across these, each tool and platform needs some kind of real-time capability. Otherwise, one link in the chain becomes a bottleneck that hinders the level of real-time desired.

Real-time technologies tend to be specific to data, BI, or analytics.

Given that most software configurations for real-time operation involve data traveling (and being processed) across multiple systems, this report discusses three broad layers of real-time technology, plus the interoperability of the three, their integration with other systems, and common use cases and enabling technologies for each layer.

1. **Real-time data.** Real-time BI and analytics cannot operate without real-time data. Luckily, database, data integration, and other data management technologies can now handle the many new sources of real-time data that have come online recently, including new sources of machine data, which streams from sensors, robots, mobile devices, and social media. Capturing data as it is created or updated is a foundation for operational BI and real-time analytic insights into customer behaviors, competitive threats, and operational efficiencies.
2. **Real-time BI.** Traditionally limited to historical views of enterprise information, BI users today want to know what's happening *now*. In-memory computing, accelerators, and visualization are enabling real-time views via reports, OLAP, and data discovery. Organizations are revamping and automating processes for faster financial reporting, operational intelligence, and performance management. Tighter integration between BI and operational applications enables users to embed BI functions and insights into a wide variety of enterprise applications.
3. **Real-time analytics.** Increasingly, organizations need to analyze real-time data in order to rescore analytic models frequently or to correlate millisecond-old events with data from other sources and latencies. New technologies—such as CEP, stream mining, in-memory analytics, and in-database analytics—have enabled real-time analytics for customer interactions, fraud detection, and situational intelligence.

The Many Meanings of Real Time

Terminology for real time tends to be complicated because of its multiple time frames and technologies.

Let's return to the idea expressed earlier that real time has many meanings. Below are descriptions of common meanings, along with examples. The takeaway is that most terminology for real time is so fuzzy that, when you hear someone say real time or a related term, you need to ask them what time frame they have in mind.

Real time as a broad concept. Originally the term *real time* literally meant that all processing (from event reception to system response) executes within seconds, or sometimes milliseconds (or even nanoseconds.) In recent years, many of us have become sloppy in our use of the term; we sometimes say "real time" when the fetching and delivery of data takes minutes or is executed every few hours. Like it or not, real time has evolved into a broad concept that encompasses many different time frames.

Right time. TDWI regrets that the term *right time* hasn't caught on more. "Right" time reminds us that users need to understand the data requirements of a business process or technology configuration before determining the time frame for data delivery.

True real time. Given that real time can mean many different time frames, the term *true real time* has emerged recently to mean real time in the sense of milliseconds or occasionally nanoseconds. Note that not all technologies called real time are this fast, but some are, including transaction commitments, similar database updates and inserts, and some configurations of data replication, plus the latest generation of event processing and in-memory functions.

Near real time. This usually means time frames from one minute to several minutes. For example, some reporting platforms support data federation or data virtualization, so that a report can fetch fresh data from operational applications (such as ERP or CRM), then refresh report content. Even when this takes a few minutes, it's still far faster than the usual overnight refresh. Although it does not appear in this report, the term *near time* is sometimes used instead.

High performance. The speed of enterprise software and hardware has increased substantially in recent years, to the point that some functionality is now so close to true real time as to be equivalent to it. For example, TDWI regularly hears users describe in-memory data processing, in-database analytics, and columnar queries as real time.

On demand. Most real-time functionality can be configured for either push or pull data movement. Users now expect real-time performance (or close to it) with on-demand (pull) functions, such as ad hoc queries (typical of analytics) and unscheduled refreshes of reports (typical of performance management dashboards).

Analytics plus real-time decision automation. Some organizations apply complex algorithms and predictive models to drive real-time decision automation systems. Such complete real-time automation systems often include both streaming analytics and rules. To detect fraud, for example, a financial services firm might calculate complex analytics and feed the intelligence into a fraud detection system that fires off automated decisions in response to fraudulent behavior. Firms will also use this combination for real-time advertising optimization; they can recalculate thousands of customer segments continuously and apply the intelligence to serve up appropriate advertising to millions of online customers.

Definitions of real time aside, there are other concepts and practices that merit consideration in this discussion:

Real-time functions are fast and frequent, as well as constant or intermittent.

Fast versus frequent. Real-time technologies execute *fast*, whether they're running a query, delivering an event over a message bus, refreshing a report, rescoring an analytic model, and so on. An equally important characteristic of real-time operations is that they tend to execute *frequently*. For example, sensors can be configured to send data every second, minute, or other short time frame. Similarly, as organizations move closer to real-time operation of the business, they may tweak data integration and report refresh jobs (designed to run overnight in batch) so they also run a few times during the business day in so-called intraday microbatches.

Streaming data as extreme real time. Streaming data is an extreme case because some streams generate an event, record, code, or message once a second, minute, or other short time frame. A stream can come from machinery (sensors, mobile devices, GPS units) or the logs to which enterprise applications append data frequently (ranging from ERP to Web servers). Capturing and processing a stream in real time (plus correlating it with other streams and other data sources) demands very special software, typically for CEP or operational intelligence. It's worth acquiring such technology because it enables many new applications for fast-paced business monitoring, surveillance, customer service, automated responses, and so on.

Continuous versus intermittent streams. Some sensors have a power source that keeps them on and transmitting continuously, such as the digital thermometers in a chemical plant or the GPS units on trucks in logistics firms. However, most RFID chips have no power source and come to life only long enough to emit a code when hit by a transmission from an RFID transceiver; proximity to a transceiver is intermittent, so RFID codes in the stream are, too. Even though the source may not be continuously active (and, therefore, not considered real time by some definitions), receiving systems should be active at all times, ready to capture and process data at any time.

The survey for this report asked respondents to provide their own definitions of real time. Their responses corroborate that there are many definitions, viewpoints, and uses, as seen in the following representative quotes:

In your own words, what's your short definition of *real time* in the context of data, BI, and analytics?

Users have their own definitions and use cases for real-time operations.

- "We do not require immediate financial transaction accuracy, so anything fully updating within five minutes is our 'near-real-time' definition." —BI/analytics executive, insurance, U.S.
- "I would say near real time instead of real time, which [for us] means data not older than one hour." —System administrator, utilities, Europe
- "Real time is often misused as immediately updated, but should rather be considered current enough for [an] application." —Financial planner, retail, U.S.
- "Real time in this context means operational reporting and analytics to drive immediate decisions [based] on real-time activities. Transform frustrating customer-facing situations into outstanding experiences according to customer value. Increase simultaneously enterprise performance in each touch point." —BI manager, transportation and logistics, Europe
- "Represents frequent intraday updates on transnational data needed to conduct analysis and build dashboards, scorecards, reports. Ideally event driven; if not, then frequent batches." —Director of enterprise data, hospitality, U.S.

- "Getting the data at the time that you need it." —Director, professional services, Australia
- "Embedded analytics as a part of business processes and transactions." —Managing consultant, Internet firm, Asia
- "Real time is a relative and over-used term. For our company's purposes, real time equals on demand." —Head of global BI, Internet firm, U.S.
- "Ability to report on streaming data as events happen. Have triggers and alerts set up on moving data. Calculations are conducted on moving data." —BI architect, financials, Australia
- "Data that is available, in a format consumable by analytic platforms for decision making, within seconds of it posting to systems of record." —Manager of BI, non-profit, U.S.

Common Use Cases for Real-Time Technologies and Practices

Real-time software solutions monitor and analyze business activities to give a wide range of users the real-time visibility they need to see a problem or opportunity, make a fully informed decision, and then act accordingly. Numerous real-world applications are already established today:

Many use cases for real-time technology are deployed today, in many industries, locations, and department types.

- Understand customer behavior in real time across multiple channels, such as Web, mobile, social, and enterprise applications. Improve the customer experience as it's happening.
- Enhance complete views of customers with real-time data, BI, and analytics. That way, views are also up to date, not merely complete.
- Combine real-time data with historical data from data warehouses and BI systems. Judge current events more accurately in the context of performance history and seasonality.
- Evaluate sales performance in real time. Take measures now to achieve sales quotas.
- See a product recurring in abandoned shopping carts on an e-commence website. Run a promotion to close more sales of that product before interest in it wanes.
- Identify a new social media sentiment or pattern. Direct it or correct it as it evolves.
- Spot potentially fraudulent activity as it's being perpetrated. Stop it while it's in process and take action to mitigate its impact.
- Take logistics to a new level of accuracy, efficiency, and customer service. A few minutes here and there on a truck or rail freight schedule add up.
- Monitor the performance of interconnected infrastructures such as utility grids, computer networks, and manufacturing facilities. Make tactical decisions for short-term maintenance and optimization, but store and analyze the real-time data for long-term capacity planning.
- Let software take action automatically to adjust machinery, turn lights off, route energy flows on utility grids, buy/sell stocks, or send a coupon to a churning customer.

In addition to the common use cases listed above, certain departments and personnel are likely users of the technologies and practices of real-time data, BI, and analytics. (See Figure 1.)

Many types of users and departments can employ real-time functions and practices.

A wide range of business operations can use real-time operations. Operations (53%) bubbled up to the top of the survey responses, followed by more specific types of operations, such as customer service and support (51%) and supply chain/manufacturing (32%).

Sales and marketing are obvious candidates for real-time functions. Finding, closing, and retaining customers are all increasingly time sensitive, so it makes sense that sales (45%), marketing (43%), and e-commerce (28%) ranked fairly high in the survey.

Certain managers can incorporate real-time methods. Real-time dashboards, frequently refreshed reports, and alerts are commonly used today by line-of-business managers (39%), C-level executives (33%), and finance personnel (32%).

Real-time functions are useful to technical users, too. IT systems/network management (43%) typically monitors, in real time, a wide range of systems to assure availability and performance. Research and development departments (19%) increasingly depend on real-time data and analytics to design and build new products and services.

Who uses (or would use) real-time data, BI, or analytics in your organization? Select all that apply.

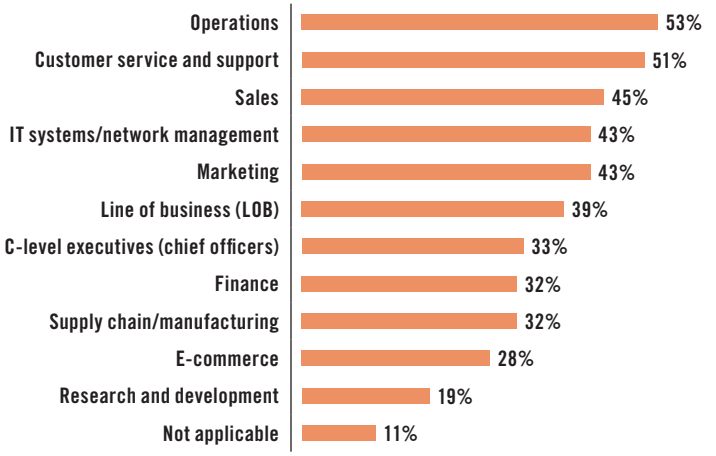


Figure 1. *Based on 1,562 responses by 365 respondents; 4.3 responses per respondent, on average.*

A number of survey respondents entered additional use cases for real-time data, BI, and analytics, including clinical decisions in healthcare, information for end customers (not just customer support), fraud detection and prevention, continuous statistical analysis of export data (for homeland security), monitoring commodity investments, using BI for auditing business processes in real time, validation for bill processing, and transaction recalls (in a retail point-of-sale system).

As you can see, real-time data, BI, and analytics (plus real-time operations in other contexts) are not mere hype or science fiction. There are many real-world applications deployed in many real-world organizations today.

Why Is Real Time Important?

The vote is in: most users consider real-time operations to be important.

Anecdotally, TDWI has seen users' needs for real-time technology increase noticeably since the new millennium began. To gauge the urgency of these needs, this report's survey asked: "How important are real-time data, BI, and analytics for the success of your organization?" (See Figure 2.)

An appreciable 24% of respondents feel that real-time operation is not a pressing issue at this time. To the contrary, however, the majority of respondents consider real-time data, BI, and analytics to be extremely important for success (33%) or moderately important (43%).

How important are real-time data, BI, and analytics for the success of your organization?

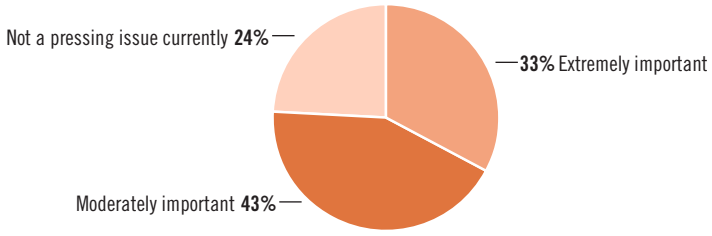


Figure 2. *Based on 352 respondents.*

Clearly, the majority of BI professionals and others who responded to the survey agree in their zeal for real-time data, BI, and analytics, but why? To get their unvarnished opinions, the survey asked the open-ended question: "Why are real-time data, BI, and analytics important (or not) to your organization?" The respondents' comments are loaded with wisdom, as seen in the representative insights that follow.

Why are real-time data, BI, and analytics important (or not) to your organization?

Users' priorities for real-time operation are quite diverse.

- "To react immediately to customer needs, differentiate from the competition, and to increase customer loyalty. Maximize the customer experience." —BI manager, transportation, Europe
- "Anticipation or early identification of problems rather than waiting on the 24-hour refresh cycle." —Senior BI analyst, manufacturing, U.S.
- "Changes in demand affect operations in real time. Also, response to unplanned outages is life-and-death critical." —Data scientist, utilities, U.S.
- "Enables monitoring and management of business performance on a near-real-time basis, allowing business users to effect changes and make informed decisions timely." —Data architect, transportation, Africa
- "Marketing automation—creating automated workflows for e-mail campaigns based on action of consumer." —Director of information strategy, education, U.S.
- "As a manager of quality assurance, we need real-time data and analytics to assess the state of the product and make release management decisions." —Quality manager, Internet, U.S.
- "Competitive environment requires innovation and deep customer understanding." —Consulting director, education, Australia

- "The top reason is for the trading desks to be able to act quickly. It's also important for sales and IT monitoring." —IT architect, financials, U.S.
- "In our industry, we need to see what our customers are doing, when they do it." —Director of planning, hospitality, U.S.
- "Improving the digital consumer experience. Improving operations execution. Proactive tactics, more than reactive." —Director of data insights, food and beverage, Europe

USER STORY SPEED AND CLIENT SERVICE HELPED QUICKEN LOANS DOUBLE ITS ONLINE MORTGAGE LENDING BUSINESS.

In 2013, Quicken Loans closed a company record $80 billion in home loan volume, breaking the previous record of $70 billion set in 2012. This catapulted the company from the eighth largest to the second largest retail mortgage lender; among online lenders, Quicken Loans is now the largest in the U.S.

According to company representatives, Quicken Loans' success is a direct result of developing and leveraging powerful performance management tools for their mortgage operations team. Those tools, in turn, enable the operations team to compete on speed and client satisfaction.

"We think in terms of shaving 'inches' off of processes to gain even more speed, as well as simplification and better service," said Dan Jones, a BI director at Quicken Loans. "Once you start looking, you can find all kinds of inches, and they really add up to make a significant impact on client satisfaction and company growth."

Quicken Loans' performance management method has a strong focus on calculating and applying metrics and KPIs in the shortest time frames possible. "The length of the time frame is relative to an individual piece of the overall process," said Nuverre Naami, the team leader for BI business analysts. "For example, more than 80% of clients who submit loan applications are called within one minute. As another example, loans are processed in 30 days on average, instead of the industry average of 60 to 90 days."

To recognize their innovation and achievement, in 2013 TDWI bestowed on Quicken Loans a TDWI Best Practices Award in the category of performance management.

Benefits and Barriers for Real-Time Data, BI, and Analytics

Real-Time: Problem or Opportunity?

Establishing and sustaining real-time operations takes time, specialized skills, coordination with the business, and the right software functionality. With these requirements in mind, this report's survey asked: "In your organization, are real-time data, BI, and analytics considered mostly a problem or mostly an opportunity?" (See Figure 3.)

Real-time operation is definitely a business opportunity; it's rarely a technology problem.

Over three-quarters (78%) of organizations surveyed consider real-time operations an opportunity. Conventional wisdom today says that real-time data, BI, and analytics enable an otherwise unobtainable speed for fact-based decision making and analysis based on fresh data.

Less than a quarter (22%) consider real time a problem. Real-time data, BI, and analytics pose a number of technical and organizational challenges, as noted in the next section of this report. Relatively few organizations find that the challenges outweigh the benefits.

In your organization, are real-time data, BI, and analytics considered mostly a problem or mostly an opportunity?

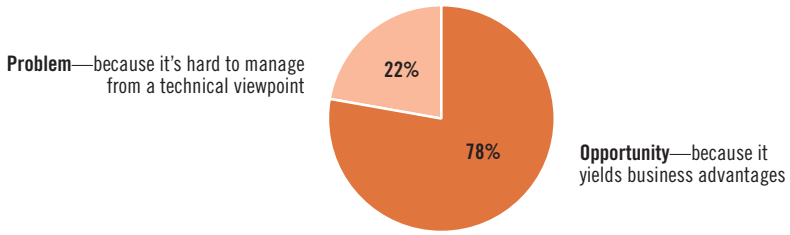


Figure 3. *Based on 317 respondents.*

Benefits of Effective Real-Time Practices

In the perceptions of survey respondents, real-time data, BI, and analytics offer a number of potential benefits. (See Figure 4.)

Real-time functions can improve reactions, customer touches, BI and data in general, and analytic applications.

Real-time technology can make BI and analytics more actionable. Real-time implementations—as applied to data, BI, and analytics—are all about taking action in a time-sensitive business scenario. By their nature, such implementations enhance the ability to take action on data analysis (53%) and employees' ability to be proactive (30%). As more real-time implementations encode business rules and similar logic, taking action rises to the level of automated decisions for real-time processes (e.g., approvals, customer service) (47%).

Anything involving customers tends to gain from real-time technology. After all, speed and personalization have become hallmarks of quality customer experience and service (49%), which for 21% of organizations includes e-commerce customer interactions.

Real-time technology takes BI to a grander level. Long-term and latent BI processes aren't going away, but they're being extended or complemented with real-time ones. Hence, real time can extend and improve business decisions (45%), operational BI (45%), and dashboards and performance management (42%).

Real-time technology can enhance business operations and their goals. These are seen in the survey as business performance and execution (39%), the efficiency of business operations (39%), operational excellence (28%), and cross-unit business processes (18%).

Some data-driven processes can benefit from real-time technologies. Clearly, real-time functionality increases data freshness or timeliness (32%). It can also contribute to the business leverage of data assets (17%), data-driven corporate objectives (17%), data governance and stewardship (16%), data sharing across business units (16%), and views of the business via data (14%).

A wide variety of analytic applications may benefit. These include various forms of situation awareness (29%), fraud detection (28%), social media monitoring (23%), and customer churn detection (16%). A few respondents added other types of analytic applications, namely forecasting, predictive analytics, and risk assessment.

Which of the following would improve in your organization if you implemented some form of real-time data, BI, or analytics? (Select 10 or fewer.)

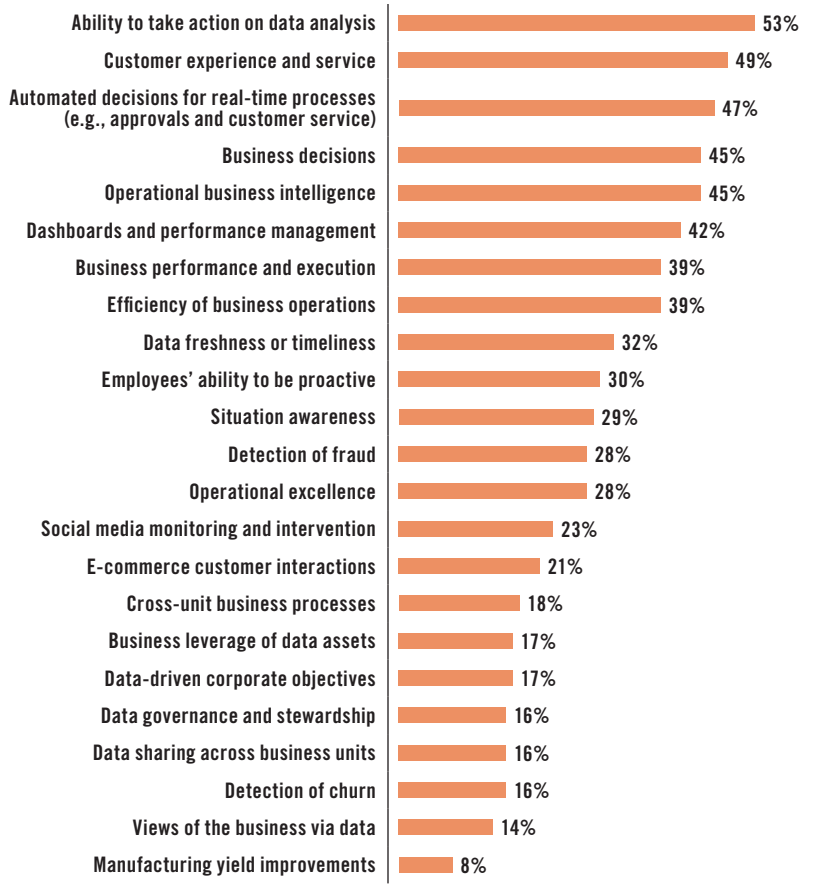


Figure 4. *Based on 2,130 responses from 317 respondents; 6.7 responses per respondent, on average.*

Barriers to Success with Real-Time Practices

Real-time data, BI, and analytics all have benefits, as we just saw. Yet, they also have barriers. (See Figure 5.)

Leading barriers involve cost, architecture, and the state of data and its management.

The cost of real-time technology is the leading barrier (56%). Costs arise from acquiring new tools, using more functions of existing tools (which may increase license fees), and hiring or training employees. As with any technology program, proper funding may be inhibited by the lack of business sponsorship (31%) or the lack of a compelling business case (31%).

Architecting a complex real-time system can be daunting (40%). This is especially true when real-time data, BI, and analytics are new to an organization. Successful real-time architecture can be inhibited by inadequate staffing or skills (33%) and concerns about the impact on operational systems (31%), as well as when current data, BI, and analytic systems are designed for historical data (32%) or they can't handle data in real time (28%). One way to side-step these barriers is to use a cloud-based, real-time solution that's easy to integrate, is relatively inexpensive, and provides value without a lengthy implementation or training period.

The current state of data and its management is critical. Data quality is a top concern (37%). Almost as critical are data management infrastructure (25%), governance or stewardship (24%), and metadata (20%). Because real-time data travels through many interfaces, application integration middleware (17%) can be an issue. Furthermore, data, BI, and analytic systems are not real time unless they are also highly available (16%).

A few respondents added their own barriers, such as: IT wants to "own" this instead of the business; legal concerns about using personal data; massive data volumes; and unique business rules. On the upside, a number of respondents said there are "no barriers," and one added that it's "only a matter of time to implement, test, and deploy."

In your organization, what are the top potential barriers to implementing real-time data, BI, and analytics? (Select 10 or fewer.)

Figure 5. *Based on 1,421 responses from 317 respondents; 4.5 responses per respondent, on average.*

USER STORY IF YOU DON'T HAVE A COMPELLING BUSINESS CASE, DON'T IMPLEMENT REAL-TIME TECHNOLOGIES.

"We rely on daily reporting almost exclusively," said a BI professional at an educational services firm. "The only exceptions would be a few reports we run a few times intraday, so we can check the information that's coming in through our website and the performance of the site. For most of our business processes, the firm couldn't act quickly enough to justify real-time or near-real-time technologies, and we don't have any customer service, operational, or competitive reasons that would push us into real-time operations. Even so, things could change as we grow and develop new services, so we'll keep an eye on real-time BI and implement it—if and when a business case materializes."

Real-Time Data

In many ways, real-time data is a foundation for building real-time solutions for BI and analytics. After all, without tools for databases, data integration, and data management fetching, preparing, and delivering data at the right time, BI and analytic solutions wouldn't have any information to work on. Let's start our drill-down into real-time technologies by looking specifically at some of the issues and best practices for real-time data.

Data Refresh Rates for Data Warehouses

24 hours is as "real" as most data needs to be.

Daily refresh is the norm for data warehouse (DW) data. TDWI regularly encounters organizations that refresh most data in data warehouses and similar databases daily. There are other names for this, including "nightly update," "batch window," "warehouse loads," or "24-hour cycle." As the names imply, refreshing DW data once a day usually involves running batch processing (typically ETL) at night during off-peak business and IT times.

For example, 83% of users surveyed work in an organization that supports data refreshes with a frequency of one day. (See Figure 6.) Most of them support other frequencies, too; only 10% of respondents do nothing but daily refreshes (not charted). From a different viewpoint, 51% of DW data is refreshed daily (averaged across survey respondents). Either way, once a day is the norm for refreshing and loading DWs.

Executing data refreshes more frequently than daily is relatively rare. For example, a small minority of user organizations refresh a DW in true real time (9%) with an even smaller minority of DW data (2%). This isn't surprising because true real time is an extreme scenario that few organizations need today. In near-real-time scenarios, the need and adoption have progressed further. For example, once-an-hour refreshes and continuous trickle feeds are more common (21% of users, each), although still with small amounts of data (3% and 4%, respectively).

Although true real time is rare, near-real-time data refresh is now common with data warehouses.

Multiple data refreshes a day are fairly common. These have become required for certain popular BI practices, namely operational BI, performance management, and dashboarding. The time frame is about four hours. For example, data for metrics and key performance indicators (KPIs)—which these practices depend on—are updated in the usual nightly batch runs. In addition, however, a time-sensitive subset of this data is also updated mid-morning and mid-afternoon during the business day (sometimes more frequently), often with so-called microbatches. The near-real-time data gives managers a view into corporate performance yesterday, this morning, and this afternoon. The cadence of this practice works well with a wide range of organizations, and it's an easy but rewarding first step into real-time data, BI, and analytics.

Real time is for time-sensitive data in the context of a fast-paced business process. That combination is rarer than the hype around real time might admit. The survey data charted in the bottom half of Figure 6 reveals that small percentages of enterprise data are involved. Yet, respectable percentages of organizations are already using some kind of real-time data refresh that's run more than once a day. This indicates that small amounts of timely information can merit special treatment from many organizations if they are to get full business value before the time to value expires. However, it's important to set proper expectations: real-time technology is never for all data or all aspects of all business processes.

For the data in your organization's primary data warehouse, what approximate percentage is currently captured and processed in each of the following time frames? Enter integers that sum to 100.

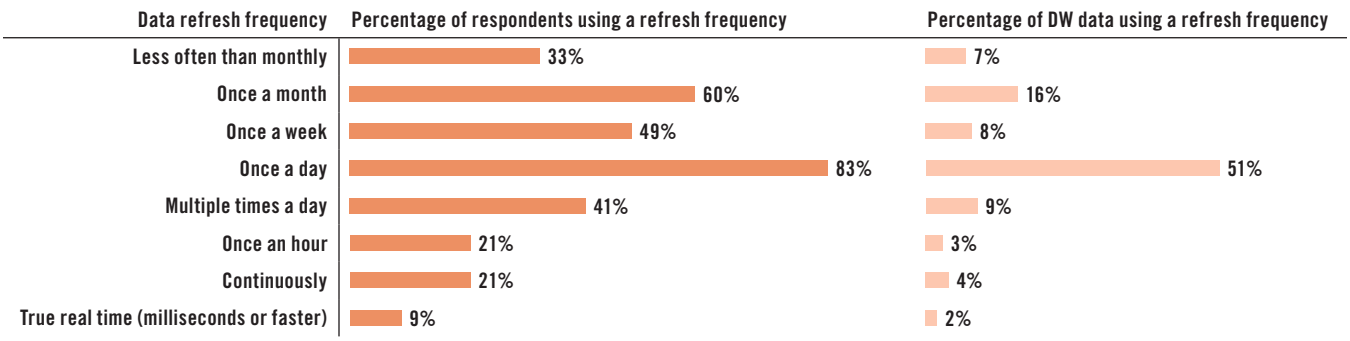


Figure 6. *Based on 259 respondents.*

Common Technologies for Real-Time Data

Available in the marketplace today are many types of tools and data platforms (plus functions within these) that are conducive to handling data in true real time, near real time, on demand, and with high performance. Following is a brief inventory with a few use case examples.

Data federation and virtualization. Data federation fetches fresh data in near real time from source systems, but only when an application, report, or user asks for it. This reduces the overhead of continuous data capture, and it makes federation a viable strategy for collecting small amounts of time-sensitive data that don't require much processing prior to presentation in a report or analysis. Federation has been around for years in low-end forms such as distributed queries and materialized views. Modern tools, however, provide superior design and maintenance functions, plus high performance and virtualization features. Federation and virtualization are supported in tools and platforms for reporting, analytics, visualization, data integration, and databases.

Replication and data sync are the unsung heroes of data management.

Data replication and data synchronization. According to a TDWI report, almost half of data integration specialists are using data replication today or its bidirectional form known as data synchronization.[1] Yet, we seldom hear much about replication and data sync in the IT press. Both can be configured in many ways, including configurations that move time-sensitive data in time frames ranging from true-real-time trickles to daily batch runs. TDWI's position is that given their real-time performance and flexibility, every real-time data strategy should include replication and data sync. These are available in standalone tools, data integration tools, and all mature relational databases.

Application integration technologies. This involves a bus, supporting messages, events, and services. Although not designed for data integration, a bus can carry data in its messages and data processing instructions via services. For organizations with a bus implementation in place, this infrastructure is quite effective for some data integration functions, especially those that must reach operational applications and/or operate in near real time. Data services are now supported by almost all tools and platforms involved with data integration, reporting, and analytics.

Batch and microbatch. Used by 95% of surveyed organizations, ETL is without doubt the preferred data integration technique for data warehousing, BI, and analytics.[2] Users usually design ETL jobs to run in highly latent, nightly batches, but it doesn't have to be that way. ETL tools are now optimized to run in microbatches that are fast and non-intrusive for source systems. Nowadays, a savvy ETL developer creates jobs and transformational logic that can be deployed to run in multiple time frames, from big, latent batches to small, fast microbatches.

[1] See Figure 1 in the 2011 TDWI Best Practices Report *Next Generation Data Integration*, available at tdwi.org/bpreports.

[2] According to the 2011 TDWI Best Practices Report *Next Generation Data Integration*, available at tdwi.org/bpreports.

Columnar DBMS. This is a relational database management system that relies on a column-oriented data store, unlike the row-oriented stores of most relational DBMSs. In a column store, data is stored by table columns. The close proximity of related data in storage speeds up the DBMS's retrieval of data for a specific column. Because most BI and analytic queries involve columnar data, most queries execute quickly on a columnar DBMS. Performance is so good with a columnar DBMS that many users say it's as close to real time as they need for SQL-based lookups and analyses of time-sensitive data.[3]

Hadoop is inherently latent, but its real-time capability is poised to improve.

Hadoop. The Hadoop distributed file system (HDFS) is inherently batch oriented, as are most Hadoop tools (especially MapReduce). This imposes fundamental limitations on the speed and frequency with which data is loaded into or retrieved from HDFS. As with any batch-oriented system, some Hadoop jobs can be run multiple times intraday—not just overnight—say, to load very recent Web log appends or blocks of machine data. For example, 25% of Hadoop users surveyed load HDFS every 90 minutes or so.[4] Furthermore, near-real-time performance for queries, analytics, and streams is coming soon to Hadoop thanks to open source incubator projects such as Spark, Stinger, and Storm.

Complex event processing. As more organizations move deeper into monitoring operations in real time, there's a growing need to quickly capture and process data expressed as messages or events in a data stream. At the same time, the number of data streams is increasing because many new forms of big data are communicated via streams (e.g., sensor and machine data, plus log and Web data). Furthermore, the greatest analytic insights result from correlating data from multiple streams, as well as from traditional enterprise sources. Examples of applications include financial trading systems, business activity monitoring, utility grid monitoring, e-commerce product recommendations, and facility monitoring and surveillance.

Complex event processing has arisen in recent years as the preferred method for leveraging streaming data. Technologies for simple event processing have been available for years, but most are designed to monitor only one stream of events at a time. Even if users monitor multiple streams, they end up with multiple, siloed views into real-time business operations. The newer practice of CEP can monitor multiple streams at once while correlating across multiple streams, correlating streaming data with data of other vintages, and continuously analyzing the results. Single-purpose, standalone CEP tools are available from a handful of vendors today.

In-database analytics saves time by taking processing to the data instead of vice versa.

In-database analytics. Analytic processing based on data mining, statistical analysis, and other non-relational approaches has for many years required that the data be located in a special database or a flat file. Moving large volumes of data for advanced forms of analytics is time-consuming—and therefore antithetical to real time. As data volumes soar, moving data becomes ever less tenable. Hence, a new trend takes analytic algorithms to the data instead of vice versa. That's exactly what in-database analytics does. It assists speed and scale for large-volume analytics, and—depending on how you set it up—it also reduces architectural complexity. In-database analytics is fully enabled in leading brands of relational databases, and it is one of the more compelling reasons to consider the adoption of Hadoop version 2.0 products.

Reducing disk I/O increases performance for data-intense processing.

In-memory databases. One way to get real-time responses from a database is to manage it in server memory, thereby eliminating disk input/output (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. In-memory databases can support many use cases. In BI they usually support operational BI or performance management. The "database" is usually just a table or OLAP cube that stores metrics and KPIs for the frequent on-demand refresh of management dashboards. We're now seeing similar growth among use cases for advanced analytics, typically to accelerate access to and the scoring of analytic models.

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

[4] According to the 2013 TDWI Best Practices Report *Integrating Hadoop into Business Intelligence and Data Warehousing*, available at tdwi.org/bpreports.

Massively parallel processing (MPP). Variations of MPP and other parallel computing architectures are critical for the speed and scale of all the real-time and high-performance technologies mentioned here.

Cloud-based real-time solutions. There's a growing number of these, with a focus on handling streaming data types and providing analytic applications for streaming data. Cloud-based solutions make sense for organizations that can't rationalize developing a real-time competency, need a real-time solution soon, and want to minimize the costs of real-time technologies.

Solid-state drives. Leading vendors now offer data warehouse appliances and similar configurations with solid-state drives (SSDs) or other uses of Flash memory. I/O-bound operations with SSDs are very fast compared to traditional drives. Workload management software on some platforms can sense which data is accessed most, then automatically move that "hot" data from traditional drives into solid-state drives or in-memory caches for improved performance.

Best Practices for Real-Time Data

A few practices apply to many real-time data scenarios.

Use real-time operational data stores (RT ODSs). Many homegrown solutions for real-time data are based around a specialized database called the operational data store (ODS). Bill Inmon's book on the subject has guided users in the construction of class I or class II ODSs.[5] The former operates in real time, the latter in near real time. Typically using replication, the database of an operational system of record sends a copy of each insert and update to the replica database that constitutes the ODS. Most ODSs and similar replicas serve multiple purposes, including high availability (in case the system-of-record database fails), intraday operational reporting, near-real-time analytics, and data staging for a data warehouse. These functions work on ODS data, which prevents additional real-time loads on already strapped operational databases and data warehouses.

Use real-time data warehouses. These can take many different forms. However, a generic definition says that a real-time DW employs many real-time, near-real-time, and high-performance technologies and practices (potentially, all those discussed in this section of this report) to provide real-time and time-sensitive data for real-time BI and analytics, plus the fast-paced business processes that depend on them. For this purpose, some database vendors have beefed up the real-time capabilities and performance of their DBMSs, but users can get a similar result by adding tools with real-time functions to their DW environments.

Implement changed data capture (CDC). This best practice can be implemented with a variety of technologies, both homegrown and vendor built. Most are based on replication, ETL, and/or hand coded SQL. CDC identifies data that is new or updated since the last extraction from a given source. This greatly reduces the amount of source data that needs to be parsed and extracted, which in turn accelerates downstream processing and delivery. CDC is not a real-time technology per se, but the high performance that results comes close. CDC often loads real-time ODSs or DWs.

Present real-time data and historic data side by side. In BI and analytic applications, the latest value of a parameter culled from real-time data should be compared to previous values of that parameter at meaningful time periods (say, the same time of day yesterday, last week, and last month). Likewise, the latest value should be compared to the average, adjusted for seasonality. Such comparisons apply to many use cases, from tailoring a purchase recommendation (based on prior purchases) to detecting potential fraud (as when an insured motorist is involved in multiple, similar losses over time). This way, the end user (or an automatic algorithm) is fully informed about the tracked entity's performance, which is fundamental to a good decision at any speed.

[5] See W. H. Inmon,1999. *Building the Operational Data Store*, from John Wiley & Sons, Inc.

Define acceptable thresholds and business rules for all entities tracked in real time. For example, a modern chemical manufacturing plant is monitored online via CEP technologies, and most adjustments to the manufacturing process are made via software. If the temperature reading from a sensor is outside a prescribed threshold, the software automatically executes a script that adjusts the machinery. As another example, when a streaming event says that a customer deactivated service a moment ago, a business rule should automatically look up that customer's profile, which contains metrics for profitability, lifetime spend, loyalty, and so on. The business rule can then calculate whether to make an incentive offer, asking the customer to reinstate service.

Best Practices for Streaming Data

Although streaming data is a new frontier, best practices are already codified.

Continuous data stream mining. This form of mining uses filters (queries, matches, clusters, and so on) to extract relevant information from the data elements that fly by in a data stream, where each filter captures a specific type of information. It's common to have several filters per stream (multiplied by several streams) and to correlate filtered data with other filters and other data sources. Some applications include analytic models and/or data models that are rescored and refreshed continuously, based on filtered data. In turn, BI and analytic tools access these models for very fresh data and results.[6]

Managing a stream as if it were a queue. Old-fashioned queue management techniques can be applied to the processing of new data streams. For example, in near-real-time applications where a few seconds or minutes of latency are tolerable, it is efficient to group many events, then process each group as a microbatch. Other queue management techniques apply to stream processing, such as promoting high-priority events to the head of the queue and sorting events into types for greater processing efficiency.

Time windows. Many reports and analytic applications require time-variant data. To that end, events can be grouped every minute or some other right-time period to form a so-called time window for data collection. As each time window expires, it is closed and replaced by a new window. The data of each closed window is immediately processed for time-variant calculations such as subtotals, averages, dimensions, and time-series slices. A window may be persisted to disk for offline analytics. Time windows are common in both homegrown and vendor-built solutions.

Lambda data architecture. Common with streaming data (especially in Web environments), the lambda architecture basically "forks" a stream into two identical streams, each with different targets and eventual processing. One stream feeds the "speed layer," which filters for data and events that demand a real-time reaction. The other stream feeds the "persistent layer," which takes a more traditional approach to repurposing stream data for storage in a database or file system for near-real-time or long-term reporting and analysis. The lambda architecture is simple but highly effective with single-stream applications, which are common in e-commerce and facility monitoring.

Enrich streaming data with other enterprise data. Although visibility into streaming data by itself is extremely valuable to the business, its value can be enhanced even more by combining it with data that exists in structured databases and data warehouses. For example, an insurance claim ID in streaming data enables you to fetch a customer profile from a customer master. This helps you understand real-time claims processing analytics in the context of known customer attributes and proclivities.

[6] For a more detailed discussion, see the 2014 TDWI Checklist Report *Using Streaming Analytics for Continuous Operational Intelligence*, available at tdwi.org/checklists.

Store stream data to build up a useful history. Although it is broadcast in real time, streaming data is also valuable in other time frames if captured and stored appropriately. As a historical record, streaming data is rarely updated or deleted; this characteristic makes large stores of unchanging stream data ideal for offline analytics. Stored stream data provides a historical context and seasonality for the most recent stream data generated in real time.[7]

USER STORY REAL-TIME DATA ENABLES HEALTHCARE FIRMS TO IMPROVE EFFICIENCY AND QUALITY.

Healthcare providers are in the midst of dramatic changes driven by new regulations and policies, exacting demands for cost effectiveness and accountability, and pressures to deliver higher quality care. From executives to clinicians, doctors, and nurses, the need for timely information is acute. Flexible and easy-to-use BI and visual analytics tools can play vital roles in enabling healthcare organizations to gain actionable insight into day-to-day concerns, including tracking resources and personnel, monitoring the security and safety of healthcare facilities, keeping tabs on patients' whereabouts, and other unheralded but essential efficiency, quality, and safety processes.

For most of these matters, the closer to real time that data can be gathered and delivered, the better. Stanley Healthcare, a division of Stanley Black and Decker, offers real-time location services (RTLS) over Wi-Fi networks "so that medical facilities have visibility into both the location of things and people," said Joel Cook, director of healthcare solutions at Stanley Healthcare. The RTLS technology was developed by AeroScout, which Stanley acquired in 2012. The technology has been applied in a variety of industries; in healthcare, examples of RTLS deployment include patient and staff workflow; infant tracking, to protect against mistaken identity or abduction; environmental monitoring of temperature and humidity; asset management of items such as infusion pumps; and patient and staff safety monitoring for falls, duress, and other dangers.

The RTLS system "constantly sends data from tags and other sensors into the database," said Cook, to provide status and location updates. To enable users to view and interact with the data, Stanley provides customers with a software application called MobileView. This application, integrated with Tableau Server for data visualization and analytics, brings together data alerting and interaction functionality with business metrics and process logic. Stanley's complex event processing engine manages real-time alerting to both users and healthcare systems and applications by automating business status changes that drive the Tableau dashboards.

With MobileView, "you can know when an event really occurred; you are not limited to knowledge based on documentation," said Lauran Hazan, director of software product management. "Hospitals, under pressure to improve efficiency without sacrificing patient care, are excited about what they could do with this knowledge," she said. "Slow room turnover, for example, wastes huge amounts of money. More up-to-date tracking has the potential to save hospitals very significant money each year and improve the patient experience."

Real-Time and Right-Time BI

Driving out data latency and shortening query response times have long been key goals of business intelligence technology and practices. For BI, latency is the period between when an event or transaction occurs and when it becomes data that can be accessed and consumed for reporting, analysis, and presentation. For many types of BI applications, including operational BI and performance monitoring, the freshest data is the most valuable; the longer the data is unavailable for analysis, the more of its potential business value is lost. In addition, if queries take too long to return results, opportunities could be lost to take timely action on the data-driven insights.

[7] For more information about emerging best practices for streaming data, see the 2013 TDWI Checklist Report *Operational Intelligence: Real-Time Business Analytics from Big Data,* available at tdwi.org/checklists.

As BI applications have matured in terms of functionality and optimization, and users have come to depend on them for data access, analysis, and presentation necessary for daily decisions, BI applications have become increasingly mission-critical. The applications and their underlying systems must be reliable and available for continuous data updates and refreshes. Unfortunately, no BI application is fully in control of its technology environment. They depend on data entry applications or transaction processing systems that record the data, as well as the stack of underlying software, networks, storage, data integration, and other systems that must come together to send data to (or materialize data views for) BI applications. When there is a problem in the flow of data, it can be difficult to determine which component is the cause or whether the problem is due to a flawed data model, a poorly written query, or a processing challenge encountered by the DBMS.

Most queries for BI execute in near real time, in minutes and seconds.

Query response time in BI and analytics applications also depends on the organization's prevailing development methods, SLAs, and the cost management priorities of the business and IT groups that fund and support the applications. BI application and DBMS optimization—not to mention the performance of underlying networks, systems, and storage—also greatly impact query response time. Given this complexity, TDWI Research finds that the query response rate for most organizations in our sampling is fairly respectable. (See Figure 7.) For the majority of queries submitted to their organizations' primary data warehouses, about one-third (32%) of research participants said the SLA or expected response time on average is within minutes, and almost as many (28%) said it was within seconds. A smaller percentage (16%) said it was within hours.

For the majority of queries submitted to your organization's primary data warehouse, what is the approximate service-level agreement (SLA) or expected response time, on average?

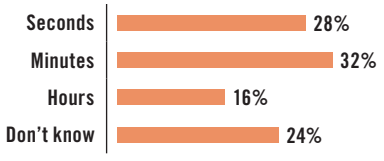


Figure 7. *Based on 282 respondents.*

Business Drivers for Real-Time BI

Pressure is mounting on organizations to reduce data latency and accelerate query performance in BI applications, particularly as they become more integral to meeting business and regulatory data access, analysis, and presentation requirements. As IT moves to reduce data latency, it must also raise—and guarantee—the availability of BI systems, data warehouses, and the processes that feed data into them. Often availability must approach continuous, 24/7 levels. Most organizations need to choose carefully which BI and analytics applications need data at or near real time because it may be prohibitive in terms of costs and resources for all applications.

Some departments build shadow IT systems to satisfy real-time needs.

Demand for timelier data and quicker response times are proving to be disruptive forces in IT. In many organizations, if IT cannot meet data demands, lines of business (LOBs) and departments will develop and deploy "shadow" IT systems, complete with their own data silos, which often later become data integration, consolidation, and governance headaches for IT. LOBs and business departments are also addressing dynamic business needs by deploying cloud computing and shifting how they account for both cloud and on-premises technology purchases from capital IT expenditures to more flexible subscription, leasing, or other operating expenditure arrangements.

Among the biggest business drivers behind demands to accelerate BI applications today are the following:

- **Business agility.** Organizations see value in gaining access to more timely data and realizing faster query response so that they can sense and respond to events in their environments sooner and coordinate and control activities internally and across business partner networks. Organizations need timelier data so that they can rapidly analyze what has happened, why it is happening, and using predictive analytics, determine what could happen given the presence of certain variables. BI applications can enable organizations to track and analyze the effectiveness of decisions and actions to improve them over time. Some data visualization tools enable users to view real-time live and streaming "data in motion," which they can combine with historical data for fuller analysis.
- **Financial performance and risk control.** Many organizations seek to use BI applications to increase transparency in financial and operational reporting and planning for auditing, controlling risk, and developing strategy. If forecasting, planning, budgeting, and risk analysis processes are using old or inconsistent data, transparency will suffer. BI systems that can give users a single view of "the truth" will reduce the time wasted on deciphering who has the best data. Although financial performance management may not always require true real-time data, with timelier data, organizations will know sooner whether targets are being met and be able to take informed corrective action if not.
- **Business optimization.** More timely data plus advanced, real-time analytics can dramatically improve how organizations plan and execute complex business processes. Organizations can use BI and analytics applications to support both human and automated decisions in operations. Firms that are employing sensors to track resources and are using CEP engines and other analytic applications to detect changes, anomalies, and patterns often require access to streams of data in real time or near it. BI and data visualization tools often play key roles in presenting data and analytics so that users can see what's important and react appropriately.
- **Customer intelligence, interaction, and engagement.** Organizations are at a competitive disadvantage if they do not have the most current data about customer behavior, preferences, and records of engagement across all channels. The faster they can collect and analyze data on current or potential customers' activity, the better they can tune cross-sell and upsell offers. Organizations need the most current information so that they can resolve customers' products and services issues and analyze customer interaction data to discover if there are patterns and trends in the reporting of such issues. Sources such as clickstreams and Web application logs can be consumed by some advanced BI and analytics tools to enable analysis and visual presentation of real-time data.

USER STORY HOWARD HUGHES MEDICAL INSTITUTE PLOTS HOW IT WILL MOVE OUT OF BATCH

Howard Hughes Medical Institute (HHMI) is a driving force for innovation in biomedical research and science education. Headquartered in Chevy Chase, Maryland, the nonprofit science philanthropy mostly funds researchers at host institutions around the U.S., although the Institute built its own research facility in 2008 in Ashburn, Virginia, that is devoted to research on how the brain works and advanced imaging technologies. Brad Sheridan, the Institute's manager of business intelligence, focuses primarily on BI for complicated finance aspects of budgeting, purchasing, expenses, and funding to support the organization's researchers and administrative staff.

Sheridan said the Institute is evolving its BI to enable shorter decision cycles and better decision making overall. The Institute depends on its ERP system for data updates, which occur overnight in batch. The Institute's data warehouse is therefore also limited to overnight updates, meaning that the data for BI is always a day behind. "It's frustrating to do something 20 minutes before leaving the office knowing you can't pick up the results of your query until the next morning," he said.

Fortunately, exciting changes are coming that will help the Institute's personnel, including its scientists, use data more effectively. The Institute is in the process of evaluating the replacement of its aging ERP system with a more modern one, which will create an opportunity to make big changes to the BI platform, data warehouse, and data marts as well. For finance concerns, Sheridan said his group is planning for intra-day refreshes for its data warehouse environment and BI reports. The Institute is planning for a new data warehouse running on a columnar database that will be focused on supporting easier-to-use visual data discovery and analytics of financial data for the scientists and administrative personnel. Sheridan called the strategy an improvement over simply trying to scale out existing BI reporting, which would require "an army of IT personnel."

BI Technology: Progressing toward Real Time

Since BI's beginnings, users have dreamed of having real-time data in their applications. Executives and other business users of early BI, executive information, and decision support systems were often disappointed to find that, after considerable technology investment, they would not be getting real-time data in their applications. Traditionally, BI applications have served up historical data that could be days, weeks, or even months old. BI reports delivered data that was loaded into data warehouses, data marts, or cubes in batch at set intervals; ad hoc queries for further analysis would have to be scheduled for off hours so that dreaded "queries from hell" would not unexpectedly bring systems to their knees. Requests to add new data would also be subject to considerable delay as the data moved through various data management, quality, and profiling processes, which might also need to run in batch.

Fortunately, many of the technology innovations mentioned in this report—including changed data capture, data virtualization (or federation), analytic databases, in-memory computing, and extract, load, and transform (ELT) alternatives to ETL—have expanded the options for reducing data latency and improving query performance in BI applications. Many organizations are also deploying dedicated analytic platforms, data accelerators, data appliances, and cloud-based solutions to both speed up BI performance and take pressure off of existing systems. These technologies may still not deliver true real-time data to BI users, but they are enabling organizations to update or refresh data far more frequently and deliver answers to queries sooner. With easier-to-use visualizations, users can also prepare to consume the data much sooner.

Big Data Technologies for BI Applications

Technology changes are enabling organizations that have invested in Hadoop and related open source technologies to move beyond batch processing to perform fast, interactive queries against what are often massive Hadoop data files. Technologies that implement YARN (Yet Another Resource Negotiator), the recent 2.x overhaul of MapReduce, can offer improved concurrency over the previous version and provide other services for multiple applications and types of workloads. These would include both open source and commercial applications for BI-style interactive querying and streaming-data analysis. As YARN-based technologies mature, organizations will be able to realize greater value from Hadoop resources and integrate BI and analytic application access to them alongside other resources, such as enterprise data warehouses.

Operational BI: Major Focus for Real Time

The proliferation of operational BI is the leading driver for real-time usage in BI.

Many organizations today want to provide BI applications—particularly dashboards—not just to their corporate and LOB executives, but also to managers and frontline users responsible for operational processes. Call and contact centers, for example, want to use dashboards to provide managers and agents with timely information about customers, marketing campaigns, issues with products, and more while they are in contact with customers. Supply chain and inventory managers need the ability to access and analyze current data about products, suppliers, and more.

Operational BI typically requires highly frequent data updates; not surprisingly, as we saw in Figure 1, our research shows that the operations department is the top current and top potential user of real-time data. Operational BI is also one of the top areas where research participants see potential improvement if they had real-time data, BI, or analytics. Customer experience and service operations are also cited by research participants as potentially improved by real-time data, BI, or analytics. However, most operational BI applications work with data that is not quite real time; the objective is usually to provision the applications with "right-time" data that is timely enough to support operational decisions and actions. Data is typically refreshed at least once a day, with some providing hourly or more frequent updates into operational data stores or data marts, or capturing snapshots of live data.

Most operational BI users interact with data via dashboards. Dashboards provide a single, visual interface that can consolidate views of many types of information and analysis, including both historical and near-real-time data feeds for monitoring events and situations. Dashboards can include KPIs, scorecards, alerts, and functionality for drill-down analysis into deeper levels of data from visualizations such as graphs and charts. Some advanced operational dashboards allow users to essentially push a button to invoke processes in business applications or to pull real-time data snapshots or feeds from them. Although organizations often aim to consolidate access to multiple applications through one dashboard, in practice, many dashboards are specific to one application or business process. The clear downside is that users have to switch from one dashboard to another as they manage multiple applications or processes, which can increase business decision latency.

Updating Data for OLAP, Data Marts, and Analytic Platforms

For most BI deliverables, daily data refresh is the norm.

Integrated with many BI applications or suites are tools for online analytical processing (OLAP). These tools provide classic slice-and-dice, drill-down, and consolidation capabilities, enabling users to aggregate data, look at data from different perspectives, or view different dimensions. Organizations are also using various types of data marts and analytic platforms that exist outside the enterprise data

warehouse to support BI applications. We asked research participants to break out how often data is currently captured and processed for these data resources. (See Figure 8.)

The highest percentage of data is captured once a day. TDWI Research finds that for OLAP, data marts, and other analytic platforms, most of the data (47%) is captured once a day. Summing up the responses from the research participants, we find that much of their data for these platforms is captured no more than a week, with 18% of the data captured once a month and 9% captured less often than monthly. OLAP in particular is most often used for historical analysis, so it is not entirely surprising that a smaller percentage of data is captured more often, if not continuously. Plus, it can prove challenging to update particularly large OLAP cubes from data stored on disk at a continuously fast rate.

Self-service, visual data discovery tools may drive demand for higher refresh rates. The "self-service" or self-directed wave in BI, analytics, and data discovery technology is strong; in previous research, TDWI has found that increasing users' self-reliance and reducing their dependence on IT are very important goals for most organizations. The tools tend to be easier to use than OLAP, enabling business users who are not expert in manipulating data to perform discovery analysis and explore data on their own. Many of these tools can work with in-memory computing to make more data available to users who wish to build data visualizations or run complex queries without having to go through I/O bottlenecks to access data in disk storage. For some queries or reports, however, it is still better for performance to access data on disk-based analytic platforms. Whether held in memory or on disk, TDWI expects that over time users will request that data be refreshed far more often than has been typical with traditional OLAP, data marts, and analytic platforms.

Regarding data in your organization's OLAP cubes, data marts, or other analytic platforms outside of the data warehouse, what approximate percentage of data is currently captured and processed in each of the following time frames? Enter integers that sum to 100.

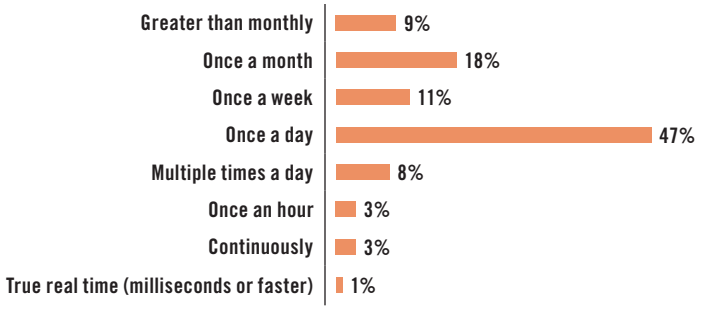


Figure 8. *Based on 243 respondents.*

Refreshing Data in Standard BI Reports

Reports continue to be the meat and potatoes of BI. Increasingly, users receive and interact with reports through their dashboards, including on mobile platforms. Snapshot reports are typically scheduled rather than requested on demand, although some users create snapshots manually. The results are often stored in a cache or database as a "snapshot" of a certain point in time. Usually, reports must be consistent; then, trends identified and comparisons made over time are more easily made, and remain valid.

Most reports are refreshed on 24-hour cycles or longer. TDWI Research finds that currently, the majority of research participants' standard reports are refreshed only once a day, if not less often. (See Figure 9.) Nearly one-fifth (19%) said that 100% of their reports are refreshed in cycles of 24 hours or longer; more than half (59%) said that 60% or more of their reports are refreshed at that rate. Because the processes that support standard reports—which could include ETL and data quality and profiling routines—can be difficult to rewrite, users who would prefer more frequently updated reports may be waiting for IT developers to build new ones that meet those requirements. Self-service BI and data discovery tools plus analytic database platforms could offer a solution; with these technologies, users can craft and run their own reports with less IT involvement, which relieves pressure on the IT backlog. Users do not have to wait for IT to build and run reports; they can respond to immediate business needs, customize visualizations, and do exploratory analytics as needed on their own.

For your organization's standard reports, what approximate percent is refreshed on a 24-hour cycle or longer?

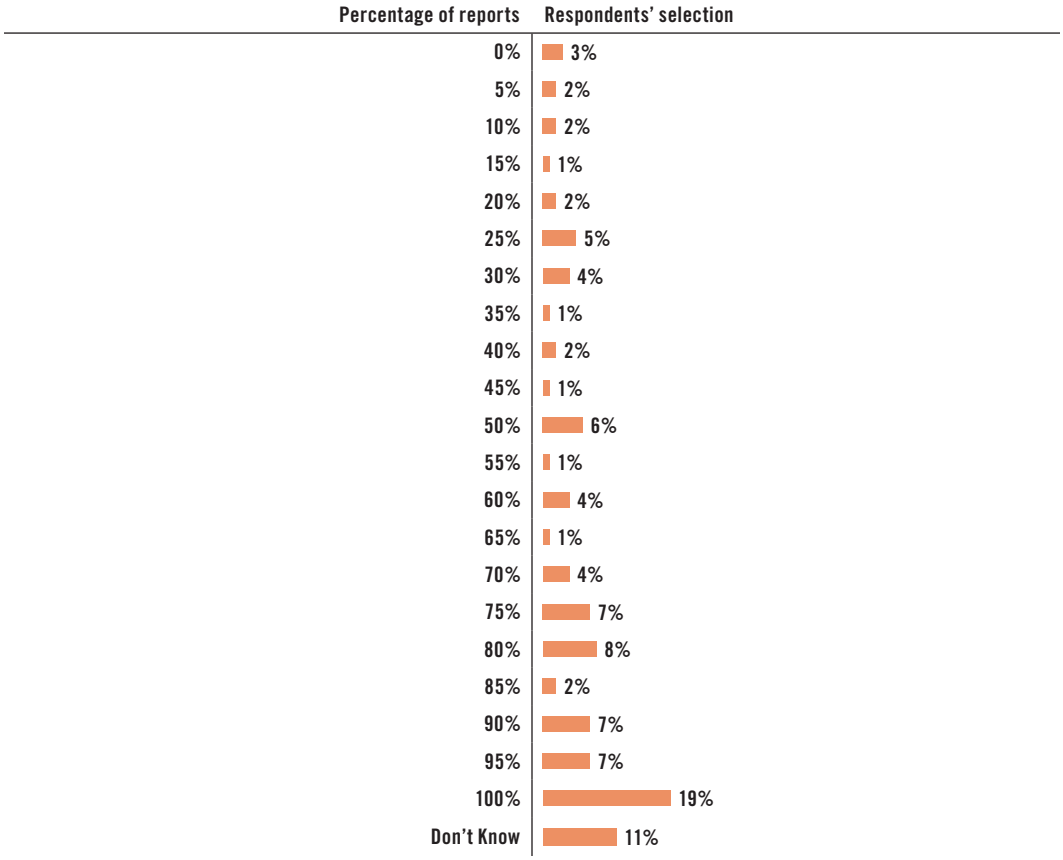


Figure 9. *Based on 282 respondents.*

Refresh Rates for Common BI Applications

In Figure 10, we see a comparative view of the frequency of data updates for a variety of common BI and other applications. As we saw in Figure 4, reports and scorecards, which are often deployed to support performance management, are most commonly updated once a day (47%). The largest percentage of research participants (39%) said their operational monitoring and alerting applications are also refreshed once a day, but 20% said these are updated multiple times a day and 24% said they were updated even more frequently. This is not surprising because the intention behind most operational monitoring and alerting applications is to keep operational personnel informed about situations or events that demand immediate attention.

Most visual data discovery and analysis applications are updated once a day. Just as we found with OLAP, TDWI Research finds that the largest percentage of research participants (47%) update data for data discovery and analysis applications once a day. However, a fairly large percentage (37%) updates data for these applications less often, certainly less frequently than applications for operational monitoring and alerting. Most research participants (40%) also update other operationalized analytics applications once a day, with 44% updating them less frequently. The majority of research participants (50%) update their BI dashboards once a day.

What is the most frequent rate of data update for the following types of BI applications?

	Less often than monthly	Once a month	Once a week	Once a day	Multiple times a day	Once an hour	Continuously	True real time (ms or faster)
Reports and scorecards	7%	19%	13%	47%	9%	1%	3%	1%
Operational monitoring/ alerting	5%	5%	7%	39%	20%	5%	16%	3%
Visual data discovery and analysis	10%	12%	15%	47%	9%	3%	3%	1%
BI dashboards	6%	16%	12%	50%	8%	4%	3%	1%
Scoring analytic models	18%	21%	15%	38%	4%	2%	1%	1%
Other operationalized analytics	12%	19%	13%	40%	7%	2%	6%	1%

Figure 10. *Based on 277 respondents.*

Most BI users appear to be reasonably satisfied with the frequency of data updates. The largest percentage of research participants (50% in Figure 11) reports that their users are "somewhat" satisfied with how frequently data is updated in their BI applications, with 16% saying their users are "very" satisfied. This suggests that for BI applications, research participants' organizations appear to be fairly satisfied with current strategies and practices for keeping data updated and refreshed and may not be among those pushing hardest to raise update frequency. Just over one-quarter (26%), however, are not as satisfied.

In general, how satisfied are business users in your organization with how frequently data is updated in their BI applications?

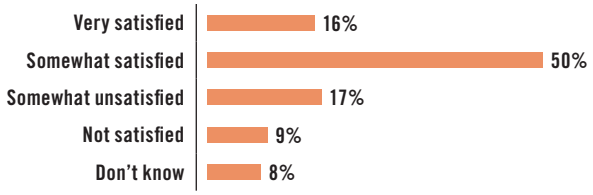


Figure 11. *Based on 277 respondents.*

USER STORY RAKUTEN LINKSHARE SERVES AFFILIATE MARKETING INTELLIGENCE IN NEAR REAL TIME.

"It's all about data and analytics," said Michael Brandt, director of BI and analytics at Rakuten LinkShare, a leading provider of online marketing services. Founded in 1996, LinkShare was one of the originators of the affiliate marketing network concept and continues to have a dominant offering as a division of Rakuten. Its services support millions of business partnerships generating tens of millions of transactions every day. Affiliate marketing is about establishing relationships between publishers who create content and advertisers who seek to attract customers to goods and services they offer. It is inherently a data-driven business, with both sides clamoring to know as much as they can as soon as they can about traffic, leads, and sales generated by customers who click on banners and links. Rakuten LinkShare also provides payment and marketing services for its networks.

The key element in affiliate marketing is the product feed; advertisers will create product files that typically feature a URL that links to product lists on their sites along with attributes such as pricing, commissions, and other internal information. Product feeds are made available to publishers, who can query product feeds and upload information onto their websites. Rakuten LinkShare can provide near-real-time reporting of aggregated data about customer clicks, including about whether customers are using PCs, mobile tablets, smartphones, or other devices. Rakuten LinkShare can enrich the up-to-the-minute statistics with solutions that calculate earnings for the links, sales and commission trends, and correlations with other data feeds such as geographic location.

Data integration, synchronization, and availability are three of the biggest challenges that Brandt has had to address in delivering near-real-time insight and intelligence. Along with BI reporting and analytics, the company's data warehouses are used to capture data first and then process and pass it on to operational systems. To ensure availability, two identical data warehouses are loaded independently, with automatic failover. The company recently implemented MongoDB and Apache Storm for enhanced real-time data processing. With high-performance requirements to be met for analytical applications, Rakuten LinkShare performs, 24/7, near-real-time and actual real-time data synchronization between multiple systems. To decrease lag time and reduce overhead on its data warehouses, the company plans to move canned reporting to a separate data mart geared to serve up fast, visual reports. It is also looking into alternatives to its current heavy on-premises technology stack.

Real-Time Analytics

What about the combination of more advanced analytics with real-time data? Real-time analytics can involve complex analysis against some sort of continuous, in-motion, (potentially) high-frequency data stream. For instance, analytics can be used against data of varying frequencies. This might involve analyzing data that is generated several times a day or multiple times a second in a stream. The analysis of the stream might involve querying it. More advanced analytics, such as predictive analytics or text analytics, can be used against this data, as well. For the purposes of this report, the focus on real-time analytics will be on three emerging and overlapping areas: operationalizing analytics, embedded analytics, and stream mining. In these cases, there is generally a continuous flow of data to be analyzed.

The Business Case for Real-Time Analytics. The value of analyzing data in motion can be significant, depending on your business. Data in motion can provide insight about what is happening now, which can help reduce costs by gaining insight and acting on issues such as operations maintenance, supply chain, manufacturing, or service delivery. It can help drive top-line benefits by understanding and acting on customer behavior. It can help to save lives by analyzing continuous data streams on patients. The use cases are varied and growing. When analytics becomes part of a business process, insights are also more likely to be acted upon, which drives value. As explained elsewhere in this report (Figure 4), being able to take action on analysis is the top driver for real-time analytics.

Our data also suggests that real-time analytics can drive measurable value. To examine this, we separated those respondents actually using real-time technology (either in a proof-of-concept or in a production deployment) into two groups: those who measured a top-line or bottom-line impact (22 respondents) versus those who did not (68 respondents). Although the size of the former group is too small for statistical testing, we qualitatively compared the two groups across a number of dimensions, including demographics, the variety of data used, infrastructure, kinds of real-time applications, and level of satisfaction. The data indicates that those who measured value were more likely to be completely satisfied in three areas—software and tools, infrastructure, and skills—versus those who did not measure value. In addition, those who measured value also appear to use more disparate data types and data that lends itself to real-time analytics, such as log data or event data. This suggests that these organizations are relatively mature in terms of infrastructure and capabilities. They may also have some standardized processes in place to gain and measure value because they are thinking analytically. This can take time, but the benefits can be enormous.

Status of Real-Time Analytics. We asked respondents several questions about real-time analytics and advanced analytics in order to understand the current use of high-frequency analytics or advanced analytics against real-time data. As illustrated in Figure 17 (later in this report), there are many kinds of analytic applications that could be enabled with real-time functionality and that have growth potential. After standard reports, operational BI and reporting take the lead with 43% of companies using this currently. However, 34% of respondents stated they are already doing some sort of operational analytics, with the same percentage planning to undertake operational analytics in the next three years. This is a strong growth potential and speaks to the value of operationalizing analytics.

The majority of respondents do not support advanced analytics with real-time or near-real-time data.

In a separate question, we asked which statement best describes the status of real-time or near-real-time advanced analytics in a respondent's organization. As illustrated in Figure 12, the majority of respondents (59%) don't currently support advanced analytics with real-time or near-real-time data. In fact, only 18% of respondents were using analytics against data with a frequency of seconds or more often. If they use advanced analytics, it appears to be in the traditional way of analyzing data offline and using the insight manually for decision making. However, a small percentage (11%) are embedding advanced analytics into systems to operationalize the analytics; another group (14%) is performing some sort of in-database scoring, which involves using models. Embedding analytics and in-database scoring are both forms of operationalizing analytics.

In your organization, which BEST describes the status of real-time or near-real-time where more advanced analytics (e.g., predictive, social media, text mining, other classification or regression) are involved?

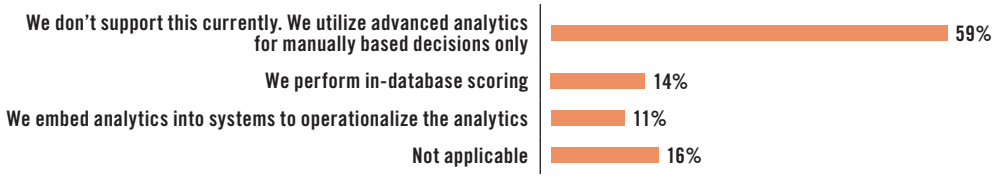


Figure 12. *Based on 239 respondents.*

Operationalizing Analytics

Operationalizing—making something part of a business process—is an important use case for real-time analytics. Of course, as we saw previously, this operationalization can happen in different real-time intervals, with daily updates being the most common interval for monitoring/alerting operations. (See Figure 17.) The use cases for operationalizing analytics are wide and varied.

When analytics are operationalized, they are more likely to be acted upon.

Operations monitoring/alerting. In operations monitoring, the goal is to gain visibility into business operations. This is sometimes called operational intelligence or continuous monitoring. Examples include monitoring manufacturing lines, an electric grid, or a telecommunications network. Often, this begins as alerting when there is a problem. This might require that rules be utilized. Data is captured and, if it exceeds a certain threshold based on a calculation (such as exceeding the average of previous observations), an alert is generated. As previously mentioned, 34% of respondents (in Figure 17) claim to be using operational analytics. Although it's still relatively early, some organizations are utilizing more advanced analytics, such as predictive analytics, to determine the probability of when a failure might occur. For example, predictive analytics might be used to determine network failure based on historical observations of when failures previously occurred. This is a move away from query-and-respond to actually analyzing the data, in motion, at or close to the time it was created, together with data that might be stored. Currently, 19% of respondents (in Figure 17) are using some type of predictive analytics applications with real-time data. Interestingly, this small group of respondents seemed more likely to be utilizing continuous data streams and performing some kind of operational analytics on the data than the respondent pool as a whole, although the exact kind of application was not clear from our survey data.

CRM and customer service. A number of respondents stated that customer service can greatly benefit from real-time analytics. CRM and call-center applications ranked highest on the list when we asked, "What kinds of operational applications and business processes are integrated with your organization's solutions for real-time data, BI, and analytics today?" In fact, 36% of respondents (not charted) are already using some real-time BI or analytics for customer service and call center applications. A good example of how this might be done involves in-database scoring, which was cited by 14% of the respondents as a use case for more advanced analytics. Here, a predictive model might be built for determining the probability that a telecommunications customer might be at risk for discontinuing service. When a customer calls a call center agent, their data is put through the model and the probability of churn is calculated. If the probability is high, the agent is given suggestions about how to try to retain the caller. In this case, the data is being processed in near real time to determine churn probability, although it is not being continuously mined.

Real-Time Embedded Analytics and High-Frequency Analytics

The call center example above is one type of embedded analytics. When analytics is embedded, it is actually integrated into business applications. In general, this opens the analytics to more users. There are different variations and degrees of embedding analytics, from BI self-service dashboards to real-time fraud detection, and from human interaction to applications that are more programmatic, where the events might be happening too fast for humans to respond. We asked respondents a specific question about what they would expect in terms of response time for embedded analytics for automated analysis. (See Figure 13.) Results varied, although the largest percentage response appeared to be in the minutes-to-seconds time frame. Different respondents had different ideas when thinking about *automated* real-time analysis. This tracks with the notion that the meaning of real time varies.

For any analytics that are operationalized or embedded in a business process for automated analysis, what is the expected response time?

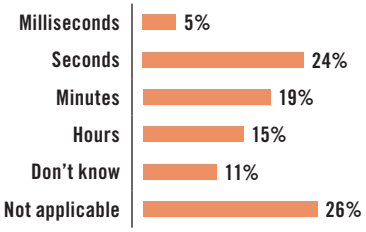


Figure 13. *Based on 239 respondents.*

We then asked respondents about the status of extremely high-frequency, real-time analytics in their organizations, where we defined extremely high frequency as seconds or shorter. (See Figure 14.) Although 39% had no plans for this, 18% were starting to work on it or already had something going, and 33% were thinking about it.

Some respondents won't disclose details about real-time analytics because it's a competitive advantage.

What is the status of extremely high-frequency real-time analytics (i.e., seconds or shorter) in your organization?

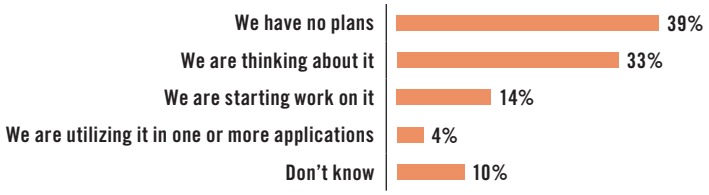


Figure 14. *Based on 239 respondents.*

Although it is still early days for programmatic, real-time analysis, we did ask those who were undertaking extremely high-frequency, real-time analytics if they could specify what they are doing. A few of the responses are shown below. They illustrate that real time is not new in certain industries as well as that some organizations are keeping their cards close to the vest.

- "Loan approval"—Architect, insurance, U.S.
  This kind of application represents analytics that is embedded in an application and might require complex analysis, and require a quick response, but may not be acting on a continuous stream of data—although there may be many requests for loan approvals coming through a Web-based system.
- "Operational analytics for smart grid electric utility"—IT professional, utilities, U.S.
  As stated earlier, operational analytics is gaining steam, and this often involves embedding higher-frequency analytics. A number of utilities companies were analyzing the grid. Likewise, a few manufacturing companies were analyzing their production lines in close to real time. There were also comments from healthcare organizations that were analyzing patient data at higher frequencies.
- "Not at liberty to say"—Data scientist, utilities, U.S.
  Interestingly, a number of respondents indicated they could not say what they were doing with high-frequency analytics. Perhaps they feel that it provides competitive advantage.

Stream Computing and Mining

When many people think about real-time analytics, they are often thinking about stream mining or CEP. Some people confuse the two; others believe that stream computing is the next evolution of CEP. Let's examine both.

Stream mining is poised for growth.

Stream mining is nascent. Stream computing typically deals with a continuous stream of high-volume and high-velocity data such as analyzing disparate data (e.g., structured and unstructured data) collected in a hospital's ICU. Typically, a streaming platform with in-memory capability for analysis is utilized to support stream mining. In our survey, only 7% of respondents are currently using stream mining. However, that number is set to triple (with an additional 24%) in the next three years if respondents keep to their plans. This growth is due to a number of factors, including the increase in machine-generated data, such as data from sensors that can provide a valuable source of information for certain industries, as well as the fact that the infrastructure exists to handle large streams. Industries such as telecommunications, utilities, financial services, and others are interested in stream mining for predictive maintenance or to manage risk. In fact, the percentage of respondents who were using stream mining was higher in these industries than in the overall group of respondents.

Complex event processing is also poised for growth. CEP, a technology that has been around for at least a decade, is primarily a rules-based engine that processes event flows that may be streams. Traditionally, it dealt with a complex series of events that needed to be correlated with a specific process, such as correlating a purchase with a loyalty card at the point of sale, which then might drive an action such as a discount. The technology is evolving to include automatic updates to the rules engine, with analytics sometimes built in. Forty-two percent of respondents stated they would be using this technology in the next three years.

Real-Time Analytics in the Public Cloud

Real-time data collection is a top reason for using the public cloud for data, BI, and analytics.

We also asked respondents about using the cloud for real-time data, BI, and analytics. Interestingly, more than half of respondents had no plans for using the public cloud for any of their real-time needs. (See Figure 15.) Only 35% of respondents were using the public cloud today or planning to use it in the future.

Real-time data collection is a top reason to use the public cloud. Fifty-three percent (not charted) of the respondents who were either using or planning to use the public cloud for real-time data, BI, or analytics were using it for data collection. Given the high volume of disparate data types associated with real-time data, the elasticity of the cloud for data collection and storage makes sense.

BI and analytics in the public cloud is also being utilized for real-time data. Interestingly, 44% of those respondents who were either using or planning to use the public cloud for data, BI, or analytics were actually using it for BI; 42% use or plan to use it for analytics. Of course, the actual sample size for this group of respondents is small (81 respondents), but it speaks to the potential importance of the cloud for real-time BI and analytics. Depending on the kind of real-time BI and analytics organizations are performing, the data warehouse may not be the most efficient way to process continuous data. Some organizations are analyzing their real-time data in the cloud and then bringing relevant data back on-premises to the data warehouse.

The cloud seems to drive value. When we examined those companies that had deployed real time and measured its value versus those who had deployed real time and couldn't measure value, the data suggested that those who measured value were more likely to use the cloud for data collection or other activities. This same group seemed more advanced analytically, indicating that as companies build up their analytics practice and gather increasing amounts of disparate data, they are more likely to utilize the cloud because they appreciate its extensibility and flexibility.

What is the status of extremely high-frequency real-time analytics (i.e., seconds or shorter) in your organization?

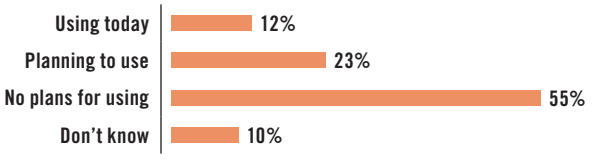


Figure 15. *Based on 232 respondents.*

EXPERT COMMENT DECISION MANAGEMENT CAN PROVIDE STRUCTURE FOR AUTOMATING DECISIONS IN REAL TIME.

"Simply dumping more data on people will not help them succeed," said James Taylor, CEO of Decision Management Solutions. "To exploit new sources of data, we have to move to increasingly advanced analytics. Yet it is not enough to develop these advanced analytics; we must also *act* on them. We have to move beyond the idea that simply extracting our operational data and conducting analysis is enough. We have to close the loop to operationalize these analytics. We need to stop thinking about a cottage industry of a few experts developing models and think about industrializing advanced analytics. This means both broadening the community of those able to work with advanced analytics *and* making sure that the advanced analytics we develop are widely consumed and used.

"The key to real-time, analytic decision making is a focus on decisions—those points in time when you and your organization must make choices. To do this requires a three-step process. The first step is decision discovery, where you identify the decisions that are most important to your operational success. Next come decision services, where you design and build decision management systems to automate these decisions. The final step is decision analysis, where an organization creates a 'closed loop' between operations and analytics to measure results and drive improvement."

Future Trends in Real-Time Data, BI, and Analytics

In this report so far, we've defined real-time operations in the context of data management, BI, and analytics, and we examined many of its business use cases and technical best practices. Let's now quantify usage to see which real-time practices and technologies successful users are employing today, as well as which they anticipate employing in a few years. From this information, we can quantify trends and project future directions for real-time functionality and data types. We can also deduce priorities that can guide users in planning their future efforts in real time relative to data warehousing, business intelligence, and advanced analytics.

Data Types Handled in Real Time, Today and Tomorrow

Structured data is the most common real-time data type today.

Data types handled in real time today. Numerous TDWI surveys have shown that structured data (which includes relational data) is by far the most common class of data types handled for BI and analytic purposes, as well as many operational and transactional ones. It's no surprise that structured data bubbled to the top of Figure 16. Other data types and sources commonly handled in real time today include application logs (33%), event data (26%), semi-structured data (26%), and hierarchical and raw data (24% each).

Social media, Web logs, and unstructured data will get more real-time handling soon.

Data types to be handled in real time within three years. Looking ahead, a number of data types are poised for greater real-time usage. Some are in limited use today but will experience aggressive adoption within three years, namely social media data (38%), Web logs and clickstreams (34%), and unstructured data (34%). Others are handled in real time today and will become even more so, namely event (36%), semi-structured (33%), structured (31%), and hierarchical (30%) data.

Data types rarely handled in real time. Many data types are selective, seeing use in some organizations or industries but not others. As discussed earlier, real-time operations are likewise selective, in that a small subset of data is time-sensitive enough to merit real-time treatment. Put the two together, and some data types are not applicable (N/A) to real-time treatment in a majority of organizations, as with scientific (85%), raw (60%), unstructured (59%), spatial (56%), and machine (55%) data.

In your organization, which of the following data types are captured and processed in real time or near real time for BI or analytic purposes today? Which will be processed in real time or near real time within three years? Which are not applicable (N/A) to you?

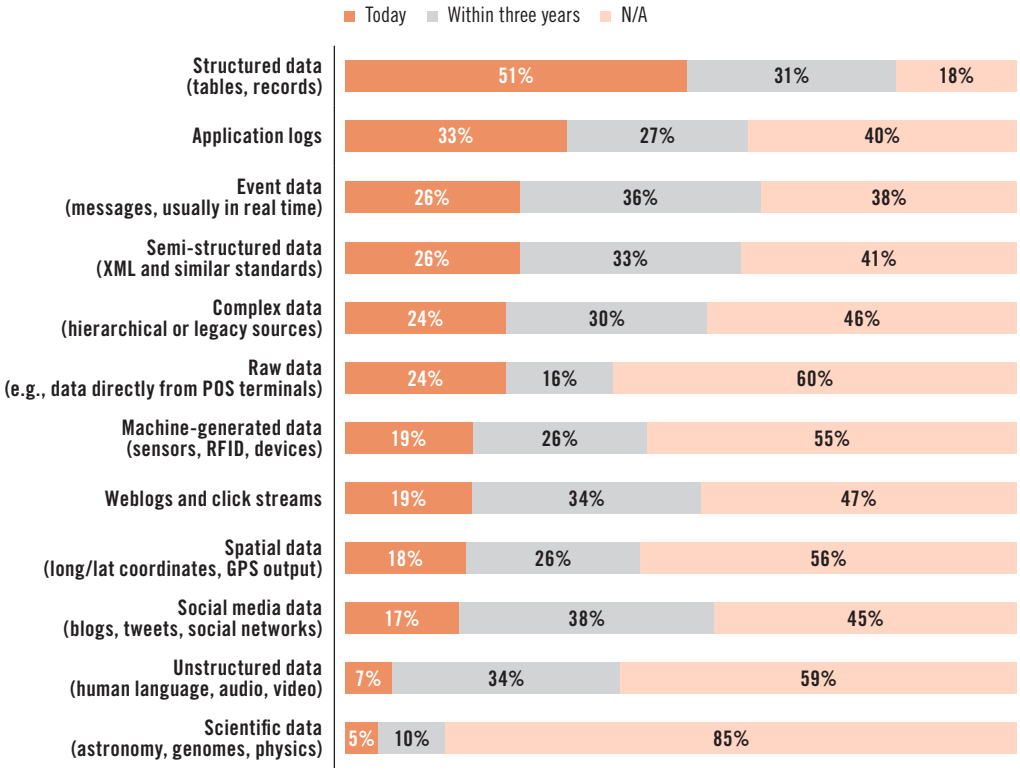


Figure 16. *Based on 282 respondents. Sorted by "Today."*

Real-Time BI and Analytics, Today and Tomorrow

The survey question charted in Figure 17 asked, "What kinds of analytic applications and BI practices are enabled with real-time functionality today? Which will be integrated within three years? For which do you have no plans?" Based on the responses, we can make a number of observations about what's in use today and how usage will increase in the near future, and identify areas where usage may be limited.

Today

- The good news is that users today have already enabled real-time functionality for some of the most common BI and analytics practices and tools, including standard reports (46% using today), operational BI (43%), BI portals (39%), OLAP (37%), management dashboards (36%), operational analytics (34%), and data visualization (31%).
- The slightly bad news is that some of the most often discussed practices for BI and analytics still aren't widely operating in real time today, as is the case with complex event processing (only 13% using today), geospatial analytics (13%), inline analytics (12%), social media (11%), text analytics (10%), and stream mining (7%).

In three years

- The greatest increases for real-time functionality will be for practices that are established but not all that common, such as real-time data warehousing (an additional 47% will use it in three years), predictive analytics (45%), customer intelligence and analysis (42%), data visualization (37%), and social media analytics (37%).
- Some well-established practices will get more real-time treatment in the next three years, such as management dashboards (an additional 35%), operational analytics (34%), active data warehousing (34%), business activity monitoring (33%), BI portals (32%), and operational BI (31%).
- Expect moderate gains in real-time treatment with some technologies and practices that are still emerging, namely text analytics (32%), Hadoop (31%), complex event processing (29%), geospatial analytics (27%), and stream mining (24%).
- Data warehousing stands out as a practice that will move aggressively into real time. In this report's survey results, real-time data warehousing is in use by 15% of respondents today, while another 47% say they'll adopt it in three years. With a delta of 32 percentage points, that's the greatest change revealed in this survey. Similar practices are also poised for greater adoption, as seen in dynamic data warehousing (15% today, with an additional 32% in three years) and active data warehousing (31% and 34%, respectively).

No plans

- Some real-time data, BI, and analytic practices are for specific types and sources of data or for specific types of businesses. Other practices work just fine in latent time frames without any pretense of real-time operation. These situations help explain why large percentages of respondents currently have no plans to apply real-time functionality to text analytics (58% have no plans), complex event processing (58%), geospatial analytics (60%), recommendation engines (65%), inline analytics (67%), and stream mining (69%).

What kinds of analytic applications and BI practices are enabled with real-time functionality today? Which will be integrated within three years? For which do you have no plans?

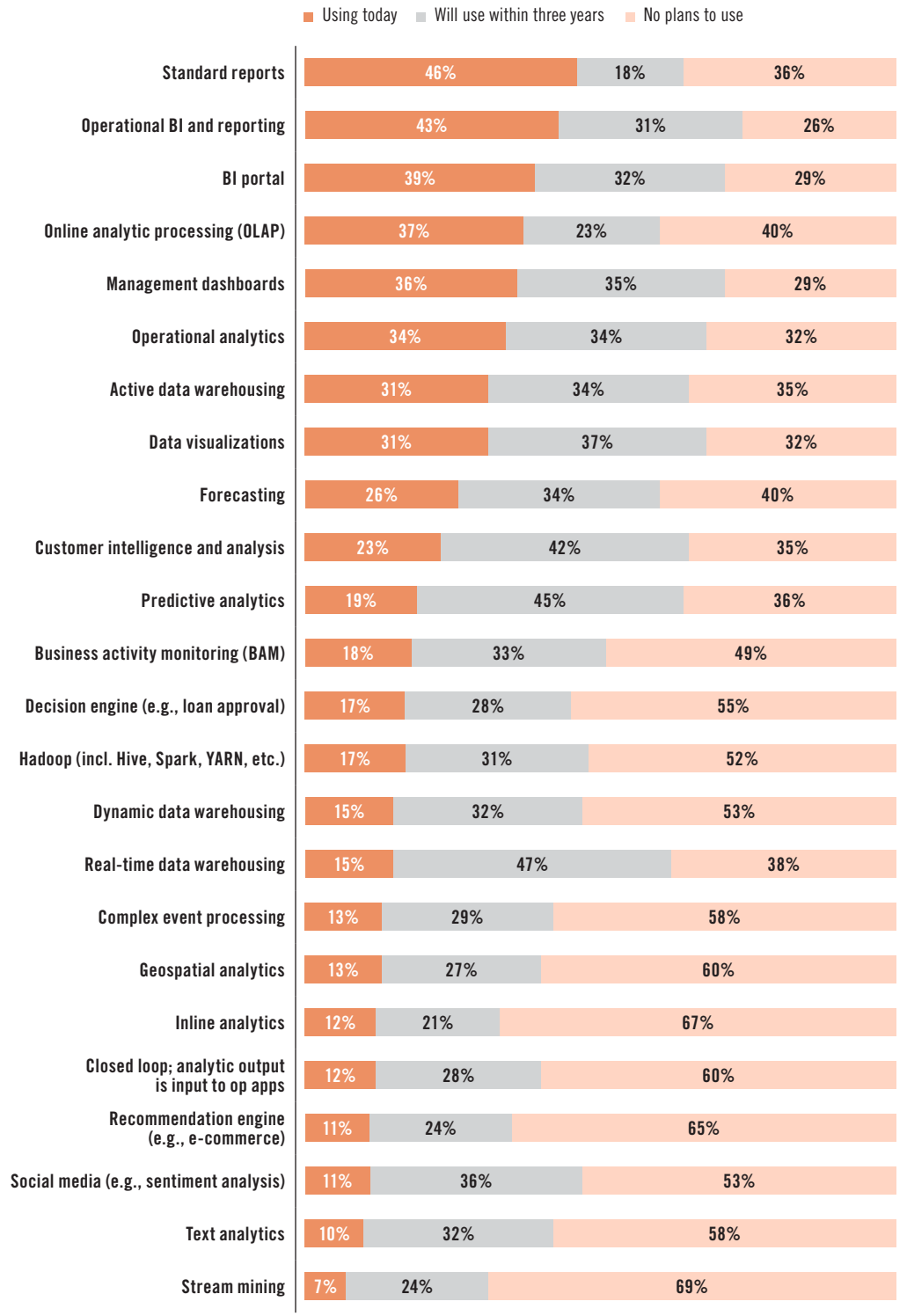


Figure 17. *Based on 243 respondents. Sorted by "Today."*

USER STORY OUT ON THE CUTTING EDGE, HYBRID NEAR-REAL-TIME AND REAL-TIME SOLUTIONS ARE IN PLACE TODAY.

A number of user organizations apply multiple system types in a hybrid solution that leverages the near-real-time analytic power of one system, which is fed into the real-time analytic functionality of another. In many configurations, there is a specialized back-end analytic database or analytic tool that operates on data of various latencies to recalculate analytic models, probabilities, segments, clusters, and so on. The output from the first system becomes input (typically as one of many streams) to a real-time engine for CEP or streaming analytics, which can assess a situation and make a decision automatically as events occur. Here are some real-world examples:

- A large government agency calculates complex network intrusion models in batch and near real time, then feeds them into real-time intrusion detection systems.
- A leading annuity firm regularly recalculates complex fraud analytics and passes the intelligence to a true real-time fraud detection system.
- One of the largest Internet firms recalculates 10,000 customer segments and feeds the output into a real-time advertising optimization engine, which in turn serves up more than 300 million ads per day.

A Sample of Relevant Vendor Platforms and Tools

Because the firms that sponsored this report are all good examples of vendors that offer real-time functionality of various kinds for data, BI, and analytics, let's take a brief look at the product portfolio of each. The sponsors form a representative sample of the vendor community, yet their offerings illustrate different approaches.[8]

Actian

The latest version of the Actian Analytics Platform—the Hadoop SQL Edition—provides far deeper, faster, and more complete support for SQL than has been seen on Hadoop to date. First-generation SQL support for Hadoop typically involves a subset of SQL functionality, shored up by programming in MapReduce. This is so onerous that mature data management organizations are moving data out of Hadoop and into relational DBMSs. Such big data movement is time consuming, taxing for IT systems, and anathema to real-time and iterative analytic methods. The new generation makes a "SQL on Hadoop" approach a reality, with users conducting advanced SQL-based analytics directly on data in HDFS. The new platform offering also includes a visual data flow framework for data blending, enrichment, and data science. 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 many real-time and low-latency use cases, including fraud detection, trading surveillance, network intrusion detection, digital ad optimization, campaign optimization, and so on.

Cloudera

Cloudera is a leading provider of Apache Hadoop-based software, services, and training. Organizations can subscribe to Cloudera Enterprise, comprising CDH (Cloudera's distribution of Apache Hadoop), Cloudera Support, and the Cloudera Manager to simplify and reduce the cost of Hadoop configuration, rollout, upgrades, and administration. Cloudera Enterprise now includes commercial support for open-source Spark, thanks to collaboration with Databricks, a leading developer of Spark. For analysts and data scientists who rely on iterative algorithms (e.g., clustering, classification), Spark is 10–100 times faster than MapReduce, delivering faster time to insight on more data. With Spark, Cloudera's enterprise data hub enables powerful end-to-end analytic workflows comprising batch data processing, interactive query, deep data mining, and real-time applications. All that stems from a single platform with common shared data, metadata, security, and management. Furthermore, Cloudera's Spark Streaming extends Spark with an API for working with real-time data streams. Related to real time, Cloudera also provides Cloudera Enterprise Real-Time Query (RTQ), powered by Cloudera Impala—the first low-latency SQL query engine that operates directly with data in HDFS and HBase.

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

Datawatch

Datawatch provides visual data discovery software based on its 2013 acquisition of Panopticon Software. The company's Datawatch Desktop and Datawatch Server enable users to interactively explore data through a range of visualizations. Rather than provide only snapshots, Datawatch's visual discovery software can consume streaming data from CEP engines, message brokers, and more. Sources may include social media, energy grids, healthcare data, EDI streams, PDF files, and so on. This capability enables users to build sophisticated visualizations that combine streaming and historical data. 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 KPIs in real time.

HP Vertica

HP Vertica provides solutions to big data challenges. The HP Vertica Analytics Platform was purpose-built for advanced analytics against big data. It consists of a massively parallel database with columnar support, plus an extensible analytics framework optimized for the real-time analysis of data. It is known for high performance with very complex analytic queries against multi-terabyte data sets.

Vertica's analytics engine allows you to query data that is stored in Hadoop or in Vertica's optimized cluster storage. Although SQL is the primary query and analysis language, Vertica also supports Java, R, and C. Furthermore, the HP Vertica Flex Zone feature enables users to define and apply schema during query and analysis, thereby avoiding the need to preprocess data or deploy Hadoop or NoSQL platforms for schema-free data.

HP Vertica is part of HP's HAVEn platform, which integrates multiple products and services into a comprehensive big data platform that provides end-to-end information management for a wide range of structured and unstructured data domains. To simplify and accelerate the deployment of an analytic solution, HP offers the HP ConvergedSystem 300 for Vertica—a prebuilt and pretested turnkey appliance.

IBM

IBM InfoSphere Streams is an analytic platform that enables real-time analytics, including machine learning, cognitive, predictive, and prescriptive analyses. Applications quickly ingest, analyze, and correlate information as it arrives from thousands of streaming data sources. Published case studies show that IBM can handle extreme throughput in real time, up to millions of events or messages per second. InfoSphere Streams v3.2.1 (the latest release) delivers enhancements, including improvements to social and machine data accelerators, extended R support, and real-time scoring across InfoSphere Streams and InfoSphere BigInsights. For Hadoop environments, there's now deeper support for Kerberos authentication, HBase, and GPFS integration. New adaptors support the MQ Telemetry Transport (MQTT) protocol (for accessing the Internet of things), IBM Operational Decision Management (ODM), and HP Vertica.

The new Remote Streams Studio for Windows and Linux is a big step forward beyond the earlier Eclipse-based tool. Developers can now work natively on a Windows desktop to develop Streams applications. End users will appreciate the simple approach to visualizing streaming data in charts, graphs, and tables.

Tableau Software

Tableau Software's focus on providing highly visual, easy-to-use tools enables business users to avoid the delays to insight that are often caused by BI's complexity. Business-side subject matter experts can use Tableau Desktop, Server, and Online products (cloud) to realize value from data without having to perform all the setup and formatting required by standard BI tools. Tableau visualizations allow users to be inventive and take BI out of its usual numeric context so that BI practices can be applied to many other types of data. Tableau customers have been able to expand BI use beyond expert power users and data analysts so that managers, line staff, and other employees can access live data sets and make decisions within the context of their responsibilities. Tableau can employ in-memory computing, and offers connectors for working with live data in in-memory databases such as SAP HANA. Tableau tools can also access machine data flows collected by Splunk, enabling users to build visualizations and analytics for interacting with real-time data from sensors, networks, devices, and other machine applications.

Treasure Data

Treasure Data provides a cloud-based, managed service for real-time data collection, storage, and analysis, often called software-as-a-service (SaaS) for big data. A monthly subscription gives you instant big data capabilities without a major investment in infrastructure or specialized skills. The service includes unique technology for real-time data collection, and it is optimized for high-volume log and event data types, including clickstream, Web logs, application logs, mobile data, sensor data, and more. Treasure customers process millions to billions of records a day with its massively scalable service.

The service is set up for you in hours or days, managed by Treasure staff, and accessed via a simple Web console so users can focus on analyzing data for business value, not managing infrastructure. Use SQL to query data, and keep it all in the cloud or export it any time for free. Treasure easily works with other local or cloud databases and BI tools, so you can quickly add big data to your existing environment while maintaining existing processes and investments.

WebAction

Founded in 2012 by data industry veterans and emerging from stealth in 2014, WebAction, Inc. provides the end-to-end Real-time App Platform that enables the next generation of data-driven Apps for the agile enterprise. The platform acquires data from any structured or unstructured source, processes the data using an in-memory grid, and delivers the data streams via alerts, visualizations, and persistence to existing data stores. Running on commodity hardware and using standards-based integration, the WebAction platform lives in harmony with existing infrastructure. Data-driven Apps continuously query and monitor your streams of data to filter, enrich with history and context, and correlate events across streams to create highly actionable big data records moments after data is generated. Apps are easy to build, modify, and deploy in days using a SQL-like language. WebAction offers a suite of data-driven App solutions for data center monitoring, quality of service management, cloud application control, security event processing, consumer analytics, and risk and fraud management.

USER STORY SOMETIMES A PLATFORM CHANGE AND/OR REDESIGN IS REQUIRED FOR REAL-TIME OPERATIONS.

"Our vice president of business reporting convinced everyone that the business needs real-time data for BI and analytics," said a data modeling specialist and data architect at a global insurance company, "so we redesigned and re-platformed our data warehouse to get ready for real time. The new architecture includes high-performance features that we consider close enough to real-time performance to suffice, namely changed data capture, columnar, and in-memory software functions. We also acquired really fast appliance hardware. Using those, we're now providing near-real-time reporting and some analytics for the call center, actuarials, and fraud detection. Annuities and social media are next."

Top 10 Priorities for Real-Time Data, BI, and Analytics

In closing, let's summarize the report with the top 10 priorities for real-time operation and a few comments about why each is important. Think of the priorities as requirements, rules, or recommendations that can guide organizations into success with real-time technologies and practices.

Embrace real time for the business improvements it fosters.

1. **Enable new fast-paced business practices.** Business empowerment is what real-time data, BI, and analytics are really about. They enable managers to make tactical and operational decisions at a level of temporal granularity that's unprecedented. In turn, that allows managers to step into the 21st century with modern practices such as operational BI and analytics, ensuring SLA compliance, monitoring commodity investments, hourly performance management, and fraud prevention (not just detection), plus cutting-edge practices such as stream mining, social media monitoring, and automated responses executed by software.
2. **Don't expect the new stuff to replace the old stuff.** Interactive dashboards and analyses based on real-time data will complement—but not replace—the traditional products of BI, such as static reports and offline analytics based on historical data.

Real-time operation never involves all data, and it never replaces latent processes and their data.

3. **Do not assume that all the data needs to be in real time, all the time.** BI directors and data warehouse managers need to guide users to focus requests for real-time data on metrics or updates that truly have a business reason for that level of timeliness. As we have seen, operational BI applications need the most frequent updates; organizations should focus technology and practice investment on reducing latency for operational BI first. However, as more organizations deploy self-service visual data discovery and exploratory analytics, organizations should expect more requests for more frequent data updates. Organizations should examine both cloud and on-premises technology options for refreshing data more frequently for these newer BI applications.
4. **Correlate real-time data with data from other sources and latencies.** Analytic correlation across multiple real-time data sources and latent enterprise data sets is the epitome of modern operational intelligence and situation awareness. Real-time data in a vacuum rarely tells the whole story. Latent data from a data warehouse or similar database provides a historical and seasonal context for more accurate appraisals of what's happening now.

Identify non-risky but valuable starting points.

5. **Start with a proof of value with measurable outcomes.** Some of our respondents said their management thought that real time sounded like "voodoo." In other words, they didn't trust what they didn't understand and couldn't see the value through real results. Often, the best way to get started is to build a proof of concept or proof of value that is tied to a value measure. It is important to make these measures reasonable and achievable in the context of the problem you are trying to solve.
6. **As a safe starter project, accelerate successful latent processes into near real time.** TDWI has seen many organizations begin their real-time journey by identifying operational reports, dashboards, and analyses that already work well for management. While still refreshing these reports and their data overnight, frequency is increased to also include two intraday refreshes. Do the same with more reports and business processes. It's a risk-free yet value-adding starting point for real-time data, BI, and analytics.

Design real-time analytics with action in mind.

7. **Think about operationalizing analytics.** Analytics—even against real-time data—is not valuable if you don't take action on analyses. Making analytics part of a business process that generates real-time data can be very important because it enables your organization to take action on the results. This is how your analytics becomes even more useful.

8. **Think about the skills you need.** The more advanced the technology and analytics, the more important the right skills are. Organizations that can't measure value typically do not have the right skills in place. Whether they include the development skills to bring real-time access to the data, the analytic skills to develop a model that is operationalized, or the skills to understand in-memory technologies, skills are a key factor for success.

As you embrace real time, expect to adjust business processes, user expectations, and application functionality.

9. **Examine application business rules to ensure they are ready for real-time data flows.** Business rules written into ERP and other business applications and transaction processing systems may not have been written with the expectation of real-time data flows. They may need to be rewritten to synchronize rules in applications and processes with real-time data requirements for BI and analytics.

10. **Evaluate technology platforms and expertise for availability and reliability.** Organizations need to determine how real-time or near-real-time data flows will affect the entire stack for provisioning data. Is this entire stack ready to deliver on the promise of real time, all the time? Most likely, organizations will need to set user expectations for when they can expect to see fast, continuous data updating and what they should expect in terms of the quality of that data. Some organizations arrange to deliver raw, real-time data during morning hours but let users know that this data has not yet been scrubbed for the highest quality. Later in the day, users will receive fully cleansed and prepared data, often in reports.

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 with 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. Founded in 2008, Cloudera was the first and still is the leading provider and supporter of Hadoop for the enterprise. Cloudera also offers software for business-critical data challenges including storage, access, management, analysis, security, and search. Cloudera works with over 1,000 hardware, software, and services partners to meet customers' big data goals.

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 and unstructured 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, including 99 of the *Fortune* 100.

HP Vertica

www.vertica.com

Have your enterprise data warehouse and "business-as-usual" practices stalled your big data initiative? With the HP Vertica Analytics Platform, you don't need to make any compromises and there are no longer any limits. Purpose-built from the very first line of code for big data analytics, HP Vertica provides you with the blazing-fast speed (queries run 50–1,000x faster), petabyte scale (store 10–30x more data per server), and openness and simplicity (use any BI/ETL tools, Hadoop, etc.)—all at a lower TCO. See why Facebook, Cerner, Comcast, and others modernized their enterprise data warehouse with HP Vertica.
—No Limits. No Compromises—

IBM

www.ibm.com

IBM is unique in having developed an enterprise-class big data and analytics platform—Watson Foundations—that allows you to address the full spectrum of big data business challenges. Information management is key to that platform helping organizations discover fresh insights, operate in a timely fashion, and establish trust to act with confidence. IBM InfoSphere BigInsights brings the power of Apache Hadoop to the enterprise with application accelerators, analytics, visualization, development tools, performance, and security features. IBM InfoSphere Streams efficiently delivers real-time analytic processing on constantly changing data in motion and enables descriptive and predictive analytics to support real-time decisions.

Tableau Software

www.tableausoftware.com

Tableau Software helps people see and understand data. Anyone can analyze, visualize, and share information quickly. More than 19,000 customer accounts get rapid results with Tableau in the office and on the go. And tens of thousands of people use Tableau Public to share data in their blogs and websites. See how Tableau can help you by downloading the free trial at www.tableausoftware.com/trial.

Treasure Data, Inc.

www.treasuredata.com

Treasure Data provides a cloud-based, managed service for real-time data collection, storage, and analysis, often called software-as-a-service (SaaS) for big data. A monthly subscription gives you instant big data capabilities without a major investment in infrastructure or specialized skills. The service includes unique technology for real-time data collection, and it is optimized for high-volume log and event data types, including clickstream, Web logs, application logs, mobile data, sensor data, and more. Treasure customers process millions to billions of records a day with our massively scalable service. Learn more or try the service for free at www.treasuredata.com.

WebAction, Inc.

www.webaction.com

Founded in 2012 by data industry veterans, WebAction delivers the leading real-time app platform enabling the next generation of data-driven apps for the sensing enterprise. Data-driven apps continuously query structured and unstructured data streams, filtering, enriching with history and context, and correlating events across streams, moments after data is generated. Apps are easy to build, modify, and deploy in days. Scaling on commodity hardware and using standards-based integration, WebAction enables existing infrastructure. Data-driven app solutions include: data center management, quality of service management, cloud application control, security event processing, risk and fraud management, and consumer analytics.

TDWI RESEARCH

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