

# Big Data, Big Opportunity

## Survey results yield checklist for turning big data into bankable results.

MARKETPULSE WHITE PAPER MARKETPULSE WHITE PAPER MARKETPULSE WHITE PAPER MARKETPULSE WHITE PAPER

### Executive summary

Industry stats repeatedly point to the simple fact that big data isn't just here to stay—its prominence is growing. In fact, big data is redefining how organizations operate and make business decisions. While a diverse group of stakeholders are driving big data's utilization—from IT and execs to marketing and finance—it is still up to IT to make it work. And, as with any initiative, IT needs to make choices around which tools to use both short and long term. This paper provides a checklist for IT to use when managing the ever-growing workloads in a big data environment—and its fit within the enterprise.

### Evolving Environment

**B**ig data is growing at an unprecedented pace. According to a recent survey by IDG Research Services, the amount of data managed is expected to increase by 43 percent on average during the next year. And, as big data finds its way into various aspects of the business, it dramatically changes how organizations go about operations and make decisions. It's capable of impacting product production, service offerings, as well as tactics utilized to interact with customers.

Big data is so impactful that we are seeing new companies that are built entirely around its capability. Birmingham, Ala.-based ChipRewards, whose focus is on reducing the cost of healthcare through healthy lifestyles, is a prime example. By leveraging relationships with insurance compa-

nies, healthcare providers, and individual members, the behavioral scientists at ChipRewards are cost-effectively collecting detailed information around dietary choices, exercise patterns, etc., which enables the firm to develop individualized profiles and programs to incentivize healthy lifestyles. Rather than making generalized assumptions using sample data, these big data powered profiles enable ChipRewards to provide individualized help with modifying behaviors—help that would otherwise be impossible on a large scale.

Big data is also providing cost efficiencies in a variety of IT areas, such as archiving, data extract and transformation, and analytics. For example, keeping up with a government mandate to maintain seven years or more of data often meant dealing with costly storage. And, if there was ever a need to retrieve data, it wasn't a trivial task be-

SPONSORED BY



cause of inefficient search options. Well-designed big data environments eliminate the issue by enabling cheap storage, fast batch processing, and detailed analysis—all leading to low-cost, highly efficient archive and retrieval.

### Driving Force

While IT is building out the capabilities, and in many instances taking ownership for the actual systems, the survey confirms that executive leadership, marketing, business analysts, and finance are playing an instrumental role in driving organizations to embrace big data solutions. And, when asked how organizations are leveraging or planning to leverage big data, respondents cite enabling business intelligence and analytics (76 percent), business strategy and direction (55 percent), as well as data discovery and exploration (44 percent) as primary objectives.

“The importance of partnering with lines of business when building a big data environment is essential, especially with so many different users and uses,” says Robin Reddick, solutions marketing manager at Houston-based BMC Software, a leading provider of IT automation solutions. “Developing high-value big data analytics requires these teams to work together. There are a lot of options when it comes to collecting and analyzing data, so IT and application development need to work closely with the line of business or these projects become nothing more than IT experiments.”

For business leadership, big data is proving instrumental in guiding strategic investment decisions, such as whether or not to expand into

new geographies, specific market segments, or product line extensions. For marketing, adopting big data means an improved ability to profile individual customers and effectively understand and better satisfy their purchasing behaviors and habits. More than 50 percent of the respondents state that big data analytics is essential for competitive differentiation.

“Big data analytics allow a business to understand an individual’s behavior and purchasing decisions—maybe even better than the individual knows them—by objectively understanding and influencing behavior patterns. Being able to identify and understand these trends is very powerful,” says Reddick. “It’s dramatically different from the past, when sample data was all a business unit could leverage. Big data provides a complete picture, which completely changes the meaning of customer relationships.”

### Understanding Obstacles

Of course, as IT leaders work to build big data applications capable of providing the business with the desired results, they face an array of challenges. Specifically, respondents cite the lack of user-friendly tools (44 percent) and inability to control the resources used for workload processing (38 percent) as two of the top obstacles.

IT leaders also face a number of tool choices, including open source and commercial distributions of Hadoop and related tools on the market today. “IT leaders need to weigh the pros and cons of using open source vs. a commercial distribution and

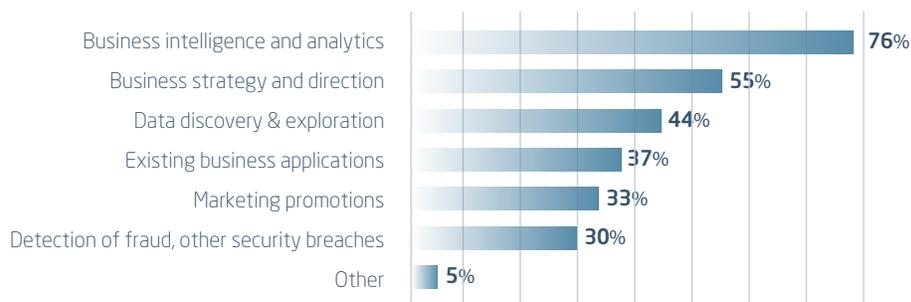


### PARTNERING

with lines of business when building a big data environment is essential, especially with so many different users and uses.”

**ROBIN REDDICK**  
Solutions Marketing Manager,  
BMC Software

## TYPES OF WORK OR PROCESSING FOR WHICH THE BIG DATA ENVIRONMENT IS OR WILL BE LEVERAGED



BASE - 108 QUALIFIED RESPONDENTS

toolset,” Reddick says. “Many times the choice is a matter of policy or practice within a company. As a rule, large enterprises are choosing supported distributions and commercial tools to ensure they have the support they need for successful operations.”

In addition, more than half (53 percent) of survey respondents report that batch-processing delays sometimes impact service levels for their big data applications. Among those who have already implemented a big data environment, more than three-quarters (77 percent) report at least an occasional impact on service levels as a result of batch-processing delays.

The majority of the data is batch processed at this time, explains Reddick. “Organizations own job scheduling and workload automation tools and know the requirements for successful batch processing. Those requirements are no different for big data applications. All of the same scheduling, monitoring, recovery, and compliance issues apply,” she says. “This is an area where businesses should look at using a commercial tool, as they are much more mature and functionally rich than the open source job scheduling tools.”

### Forward Motion

One of the biggest motivations in embracing big data isn’t simply to provide a means to manage data; rather, it represents a significant route to achieving a competitive advantage. According to survey respondents, 54 percent see big data analytics as essential to competitive differentiation. After all, being able to serve the customer exactly as desired, when desired, is extremely powerful. However, to make this a reality, businesses need to understand the key components of a purpose-driven big data platform.

### Checklist for Success

**Seek Integration.** According to survey results, 67 percent of respondents consider tool integration with big data applications to be highly important. And, a majority of respondents envision big data applications being used for both batch and online analytics within three years.

With big data applications usually starting as independent projects, new infrastructure tools

are often considered. However, IT leaders need to take stock of the infrastructure tools already used within the enterprise, since many are suitable for big data and often support Hadoop frameworks. “As Hadoop environments mature and big data applications evolve, business and IT leaders will want to connect the big data processing capability to other enterprise applications and infrastructure,” says Reddick. “As you build out your big data applications and Hadoop environments you need to have a long-range view in mind. This starts with extending the use of your enterprise infrastructure tools, where possible, to your big data environment.”

According to Reddick, there is an assumption that businesses have to build from the ground up with a new set of tools. “In many instances, the tools IT already owns may support Hadoop application and system needs, and are far more mature, robust, and stable than new Hadoop tools,” she says. “Find out what tools you already have that work with your Hadoop environment, such as workload automation, application release management, and server provisioning. If you are unsure, talk to your preferred vendors to see if they have or are extending their enterprise infrastructure tools to work in the Hadoop environment.”

**Embrace Single-View Solutions.** Sixty-one percent of survey respondents identify a single view for monitoring workflows across the enterprise as a highly important capability. Having a single view of an entire workflow—not just the processing in a Hadoop environment—is directly linked to successfully meeting service levels. Without it, detecting slowdowns and failures in workflow processing becomes a difficult if not impossible task.

“A single view is especially important as organizations seek advances in performance monitoring and data loading and unloading. The Hadoop environments today are predominately batch processing, and a big part of the workflow is moving data into and out of the Hadoop file system [HDFS],” says Reddick. “Successful data loads and unloads, as well as file transfers, are often pre- and post-processing steps for big data application processing. It’s important to monitor the work-

## Taking Control of Big Data

As a leading workload automation solution, BMC Control-M allows IT to easily manage all of its scheduling and monitoring tasks from one new streamlined, intuitive interface. It is designed for automation of various functions in the enterprise, including traditional IBM mainframe OS/ MVS JCL, batch files, shell scripts, as well as routine functions such as invoking database stored procedures, invoking and using Web services, and handling file transfers inside and outside the organization.

flow from beginning to end, not just the individual jobs. A single view makes it possible to proactively address potential problems, reprioritize work as needed, and better satisfy service delivery.”

**Empower with Self-Service.** New applications are continually entering the market for big data Hadoop environments. That means a lot of application workflow testing. Self-service for application developers and business users is identified as highly important by 58 percent of respondents.

“The days of relying on IT support staff to keep us informed of the state and status of business processing and services are waning. That is a very inefficient and costly way to run a business. With the fast pace of new development in these new Hadoop environments, it makes sense to empower the application developers with tools that offer self-service capabilities,” says Reddick.

Application developers and business users want to know that the workflows driving their business activities are being executed on time—or know when they are not. Users today expect to be informed about their work-related tools and activities—in real time.

**Use the Cloud.** As organizations build their big data environments, cloud computing is surfacing as a tool for early proof-of-concept (POC) validation and testing, as well as testing and training. Accord-

ing to survey results, 44 percent of respondents consider cloud enablement and integration for workload monitoring to be highly important.

“Cloud resources are not yet heavily used in big data production environments,” said Reddick. “Most big data environments are predominately hardware dependent thanks to the low-cost appliances available. But there is still a place for cloud computing, including POCs, tools testing, and application testing. And it is also being used for peak period resource needs—just as it is for more traditional application processing. Organizations are factoring the use of cloud into the future growth of their Hadoop environments.”

The data growth transpiring within big data environments is unprecedented; making resource usage a potential problem. Nearly half of the respondents indicated data volumes are expected to increase 50 percent or more in the next 12 months in their Hadoop file system. With this rapid rise, IT organizations need to consider cloud-enabled processing to supplement growth.

**Bottom line:** “When companies have the right perspective, work as a team, and embrace tools that can guide the organization well into the future, the potential around big data is astronomical,” says Reddick. “As people see the possibilities, they will continue to figure out creative new ways to leverage the data IT makes available.”

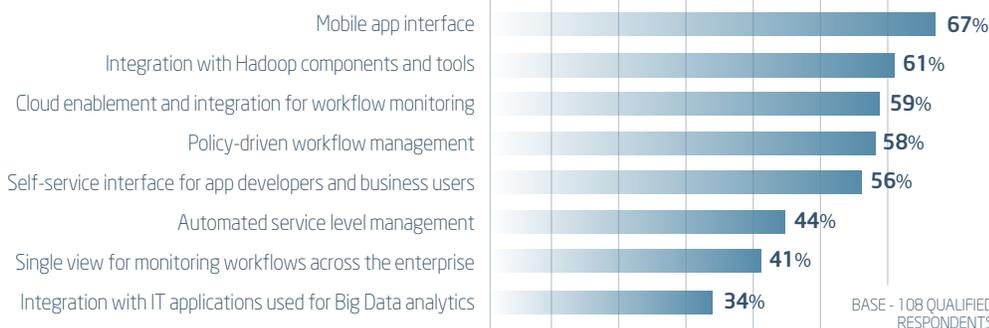


**THE DAYS OF** relying on IT support staff to keep us informed of the state and status of business processing and services are waning.”

**ROBIN REDDICK**  
Solutions Marketing Manager,  
BMC Software

## KEY CAPABILITIES IN BIG DATA ENVIRONMENTS WITH REGARDS TO WORKLOAD AUTOMATION-BATCH PROCESSING

Respondents who ranked each capability either critical or very important



Learn more about managing Hadoop batch processing at [www.bmc.com/hadoop](http://www.bmc.com/hadoop).

