A Forrester Consulting
Thought Leadership Paper
Commissioned By BMC

October 2015

Hadoop Workflow Automation Improves Efficiency And Increases Competitive Advantage

FORRESTER®

Table Of Contents

ABOUT FORRESTER CONSULTING

Forrester Consulting provides independent and objective research-based consulting to help leaders succeed in their organizations. Ranging in scope from a short strategy session to custom projects, Forrester's Consulting services connect you directly with research analysts who apply expert insight to your specific business challenges. For more information, visit forrester.com/consulting.

FORRESTER®

Executive Summary

Hadoop is about gaining new business insights from large data sets — insights that were previously impossible to gain due to prohibitive infrastructure costs and lack of scalable solutions. However, organizations are realizing that Hadoop deployments are often not straightforward. They require planning, advanced skills, a scalable hardware platform, architectural design, and knowledge of the Hadoop ecosystem to succeed. Today, organizations are spending months to set up and deploy Hadoop clusters, slowing decision-making and limiting competitive advantage.

Hadoop workflow automation is an emerging framework that enables advanced automation to support faster deployment and management of Hadoop clusters. Through improved Hadoop deployments, enterprises see several key benefits, including:

- Faster time-to-value.
- Improved IT productivity.
- Fewer human errors.
- Reduced IT costs.
- Standardization of Hadoop clusters.

In August 2015, BMC commissioned Forrester Consulting to conduct a study on how customers are using Hadoop in their organizations. Then to further explore this trend, Forrester developed a hypothesis that a successful Hadoop initiative cannot be conducted in isolation, but instead requires an integrated effort that grows and extends throughout the enterprise. And given that batch processing is a significant portion of the Hadoop workflow, workflow automation and job scheduling should be considered at the beginning of a Hadoop project rather than at a later stage.

"Without Hadoop workflow automation, we were behind on several big data projects that were keeping us in the dark on valuable insights and analytics. Today, Hadoop workflow automation is part of our big data strategy — period."
— IT manager, large financial services institution, North America

KEY FINDINGS

Forrester's study yielded several key findings:

- **Most enterprises deploy Hadoop on-premises, with hybrid deployments gaining popularity.** Based on the Forrester-commissioned survey, most enterprises cited support for on-premises Hadoop deployments, with cloud strategy on the road map in the coming years.
- **The majority of enterprises take several months and tens of professionals to deploy Hadoop.** Today, organizations are building many Hadoop clusters for their business requirements. Based on our study, the majority of enterprises take between one and six months to deploy Hadoop. More than a third of the enterprises needed more than 10 personnel to build and manage Hadoop.
- **Integration of disparate data, technical skill sets and security are the top challenges while deploying Hadoop.** Most organizations claim they struggle with some core data management functions such as integration, security, and governance when implementing mission-critical Hadoop clusters.
- **The majority of enterprises write scripts to automate workloads on Hadoop.** Scripts still rule when it comes to Hadoop implementation. However, that's changing as organizations look to deploy more Hadoop clusters, using automation to streamline their operations.
- **Hadoop workflow automation is gaining ground.** A majority of companies have less than half of their Hadoop use cases deployed through workflow automation but plan to deploy many more in the future.
- **Enterprises are realizing benefits with Hadoop workflow automation.** Faster processing, fewer manual errors, and better utilization of system resources are key benefits of using workflow automation in Hadoop.
- **Few enterprises are replacing their data warehouses with Hadoop.** Although Hadoop has evolved over the years, it still has limitations around data quality, governance, security, and integration. Most organizations supplement Hadoop with their data warehouse strategy.

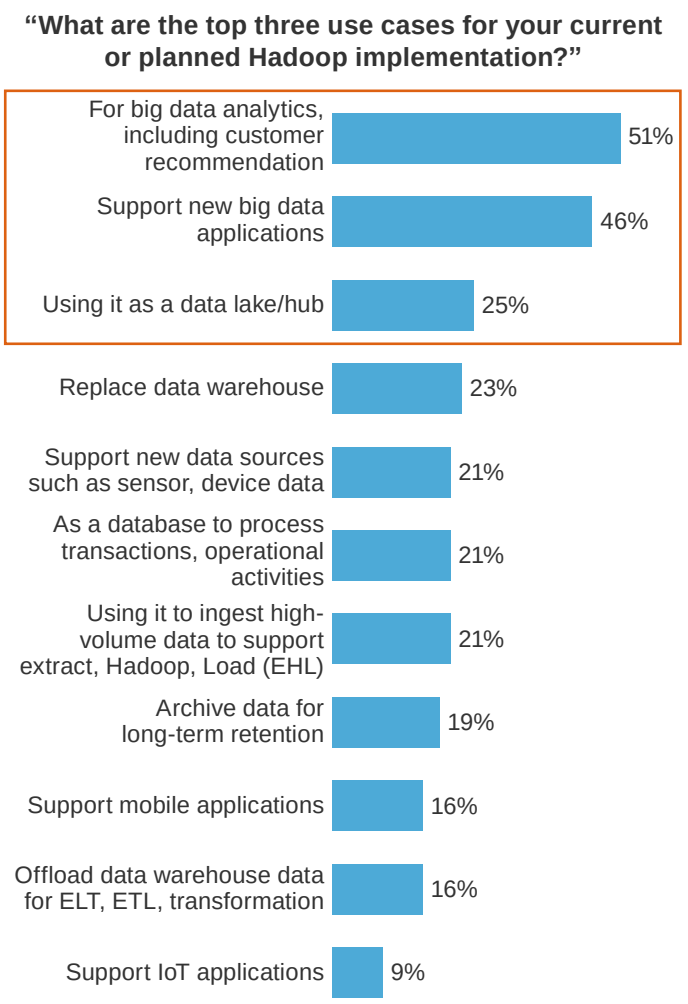
Hadoop Deployments Are Growing Rapidly, Supporting Many Use Cases

Hadoop is unstoppable as its open source roots grow widely and deeply into enterprise data management architectures. It's transforming how companies store, process, analyze, and share data of any size and structure. Hadoop is a must-have data platform for large enterprises, forming the cornerstone of any flexible future data management architecture. Our study shows that:

- **Hadoop deployments are currently on-premises, but hybrid deployments will grow.** Our research showed that currently, 43% of Hadoop deployments in large enterprises (1,000-plus employees) are on-premises. Looking toward the future, more than half (54%) of those with on-premises deployments are considering deploying Hadoop through hybrid methods, while 35% will consider Hadoop deployments in the cloud.
- **Hadoop use cases are dominated by big data analytics.** According to our research, 51% are using or considering Hadoop for big data analytics use cases including customer recommendations, and 46% are supporting new big data applications. A quarter reported that they are using Hadoop as a data lake/hub. Only 23% are replacing an enterprise warehouse, and 19% are archiving data for long-term retention (see Figure 1).
- **More than half (56%) of on-premises Hadoop clusters are deployed in one to six months.** Enterprises are deploying Hadoop clusters at a higher-than-expected rate, with 28% deploying on-premises in less than five weeks, 31% deploying in one to three months, and the remaining 34% deploying in three months or more. This indicates some improvements in automation deployment (see Figure 2).
- **Building out Hadoop projects is resource intensive.** When implementing, 37% of companies with 1,000 or more employees needed 10 or more full-time employees (FTEs), and another 22% need six to nine people.
- **Size matters.** Nearly half (46%) of enterprise Hadoop production clusters are 1 petabyte (PB) or larger, and 43% have 500 or more nodes in their largest Hadoop cluster. The remaining deployments are small to mid-size and are likely growing as more use cases are added.

FIGURE 1

Big Data Is Prevalent Use Case For Hadoop Deployments



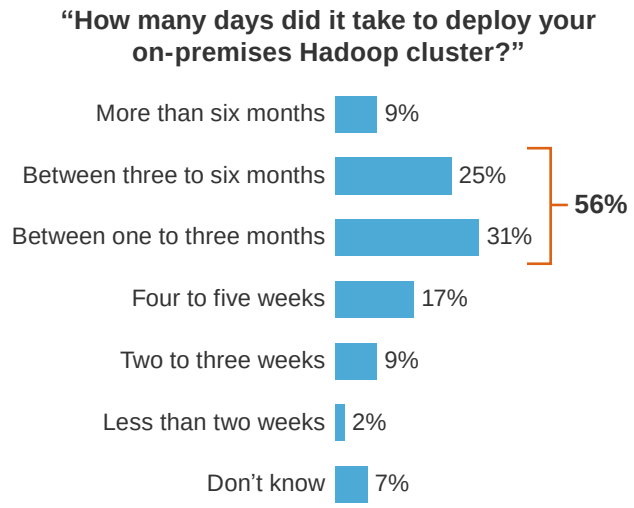
Base: 150 IT and business managers involved in Hadoop projects in large enterprises in the US

(percentages are summary of top three selections)

Source: A commissioned study conducted by Forrester Consulting on behalf of BMC, September 2015

FIGURE 2

Fifty-Six Percent Of On-Premises Hadoop Clusters Are Deployed In One To Sixth Months



Base: 150 IT and business managers involved in Hadoop projects in large enterprises in the US

Source: A commissioned study conducted by Forrester Consulting on behalf of BMC, September 2015

Implementing Hadoop Hasn't Been Straightforward

Hadoop offers great value to any organization big or small. However, there are many challenges to overcome to succeed with big data projects. These include:

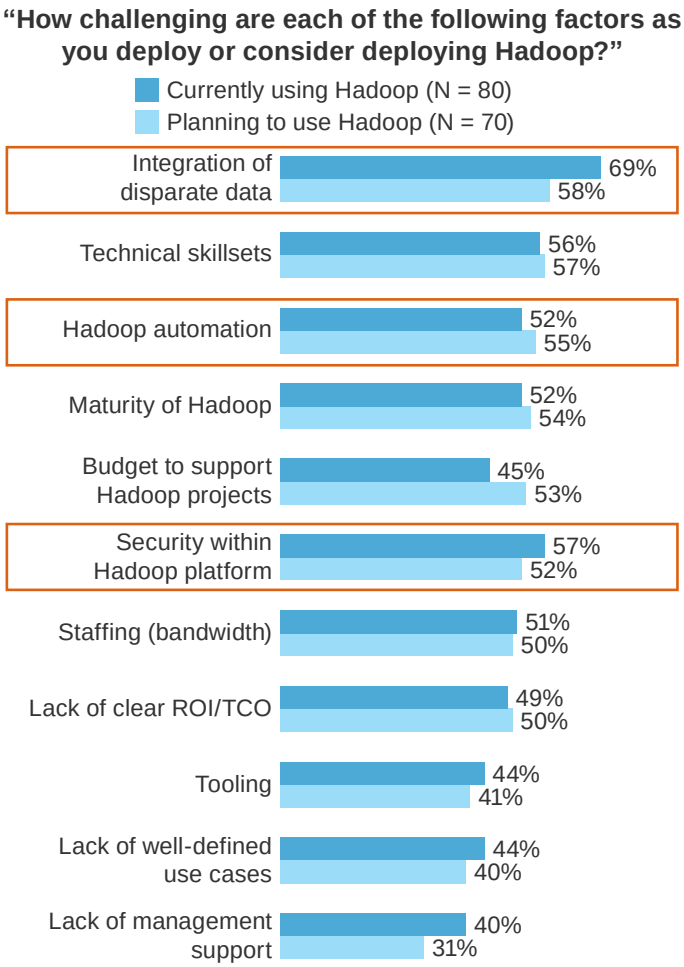
› **Data integration and security challenges.** Sixty-nine percent of large enterprises already deploying Hadoop have data integration challenges, while 57% have security challenges (see Figure 3). More than half of those already deploying are also challenged with Hadoop automation. Those planning to deploy anticipate similar challenges, but over half are concerned about having budget to support Hadoop projects.

› **Enterprises underestimating the employee resources needed to manage Hadoop clusters.** According to our research, 37% of enterprises need 10 or more FTEs to manage their Hadoop platform on a daily basis. And larger Hadoop deployments over 1 PB require even more resources, with 31% of companies needing 10 to 24 FTEs and 12% requiring 25 or more employees (see Figure 4).

› **Limited deployment of use cases through workflow automation.** According to the survey, 73% of enterprise organizations have less than half of Hadoop use cases deployed through workflow automation. Survey

FIGURE 3

Challenges With Hadoop Deployments Include Disparate Data, Security, And Automation



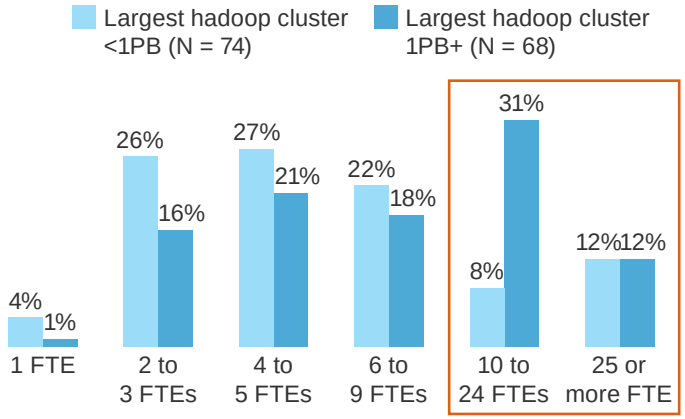
Base: 150 IT and business managers involved in Hadoop projects in large enterprises in the US

(percentages represent "very challenging" rating)

Source: A commissioned study conducted by Forrester Consulting on behalf of BMC, September 2015

FIGURE 4
Larger Deployments Require More Resources

"How many IT staff are required to manage your Hadoop platform on a daily basis?"



Base: 150 IT and business managers involved in Hadoop projects in large enterprises in the US

(managing includes loading, unloading, transforming, securing, backup, administration, and upgrading tasks)

("don't know" responses not shown)

Source: A commissioned study conducted by Forrester Consulting on behalf of BMC, September 2015

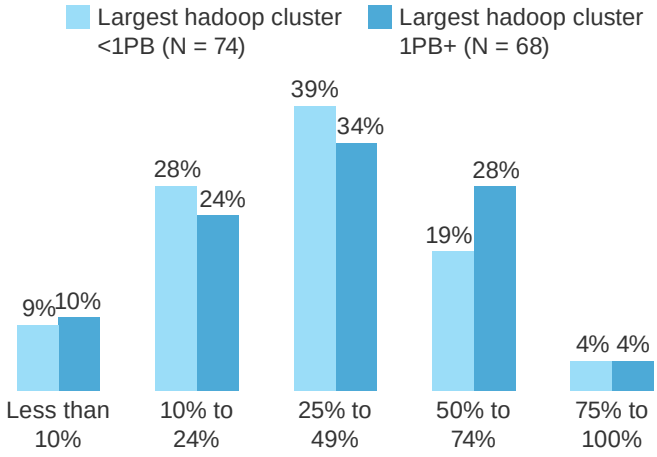
respondents agreed that larger deployments require more automation. Enterprises with Hadoop clusters of 1 PB or more are embracing deployment through workflow automation, with nearly a third (32%) deploying half or more of Hadoop use cases this way, compared with only 23% of companies with Hadoop clusters under 1 PB (see Figure 5).

› **Using scripts to automate Hadoop workloads.** Those companies currently deploying Hadoop have a mix of methods for automating workloads, including writing scripts and using various Apache modules. Forty-nine percent of companies planning to deploy Hadoop anticipate writing scripts (see Figure 6).

› **Growing use of third-party automation tools.** Survey respondents are using various forms of Apache open source projects, including Apache Hue, Azkaban, and Luigi to collaborate and manage Hadoop workflows across the enterprise. Although only 6% of large enterprises reported using third-party commercial tools, for automation, Forrester anticipates this adoption will grow, especially as organizations look at expanding their use cases.

FIGURE 5
More Automation In Large Deployments

"What percent of your Hadoop use cases are currently deployed through some level of workflow automation?"



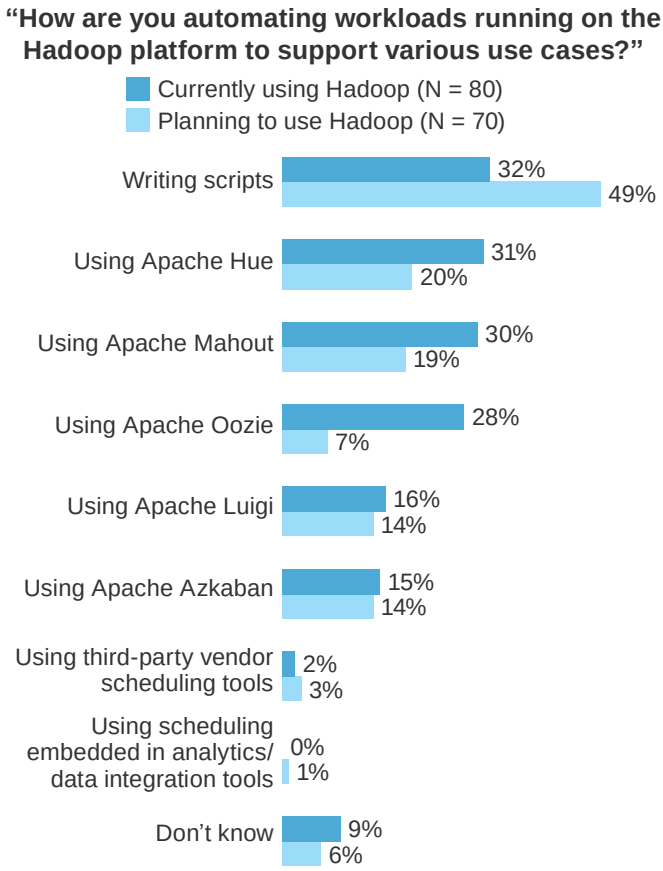
Base: 150 IT and business managers involved in Hadoop projects in large enterprises in the US

(percentages may not total 100 because of rounding)

Source: A commissioned study conducted by Forrester Consulting on behalf of BMC, September 2015

FIGURE 6

Enterprises Using Scripts And Apache To Automate Hadoop Workloads



Base: 150 IT and business managers involved in Hadoop projects in large enterprises in the US

Source: A commissioned study conducted by Forrester Consulting on behalf of BMC, September 2015

Workflow Automation Helps Enable Faster Time-To-Value For Hadoop

Recent technology advances, lower hardware costs, and innovation have helped mitigate Hadoop deployment issues to deliver an economical Hadoop platform that all enterprises can leverage. Workflow automation adoption will increase in the coming years as more use cases are deployed (given the fact that automation is required to achieve faster time-to-value). Forrester's in-depth surveys with IT and business managers yielded several important observations:

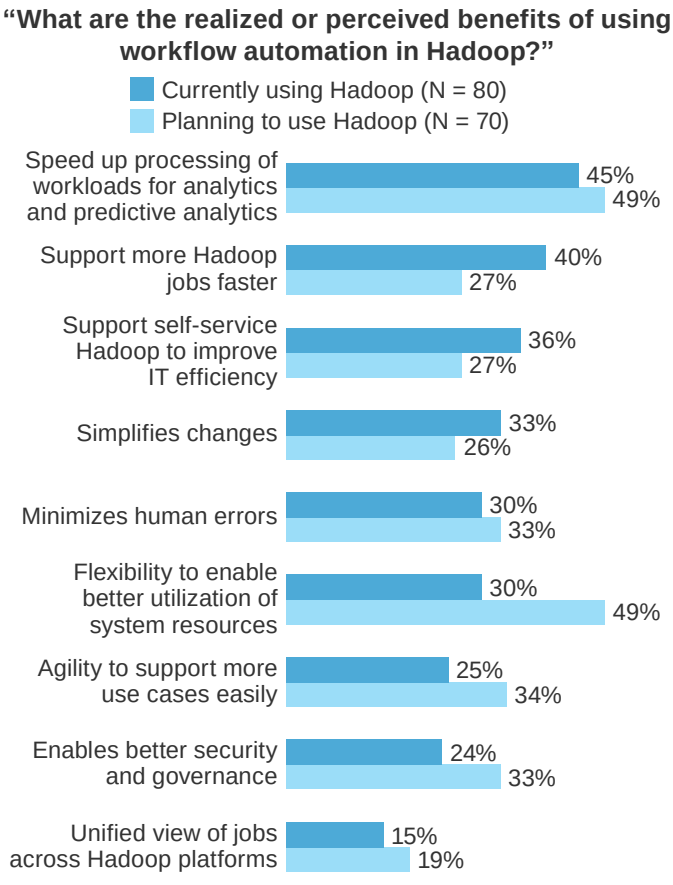
› **Workflow automation benefits are clear.** For those enterprise organizations currently using workflow automation in Hadoop, the benefits include:

- Faster big data processing (45%).
- Improved efficiency (36%).
- Utilization of system resources (30%).
- Fewer human errors (30%).

Nearly half of enterprises planning to deploy Hadoop anticipate increased flexibility and utilization of system resources and faster processing of workloads (see Figure 7).

FIGURE 7

Benefits Of Workload Automation In Hadoop

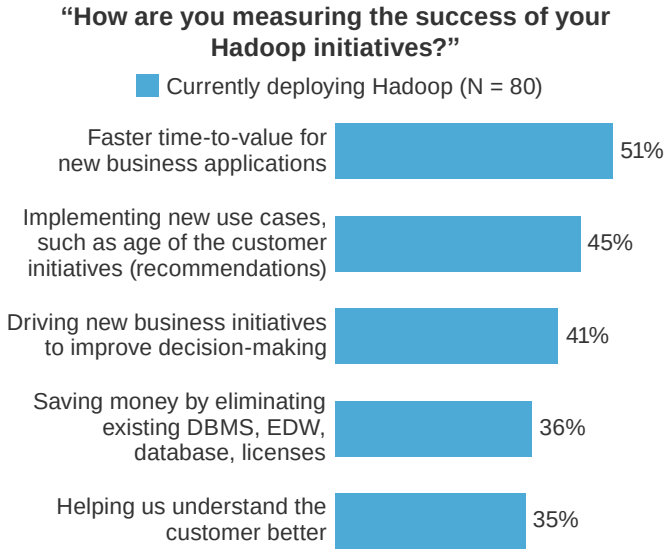


Base: 150 IT and business managers involved in Hadoop projects in large enterprises in the US

Source: A commissioned study conducted by Forrester Consulting on behalf of BMC, September 2015

› **Success of Hadoop initiatives is measured by time-to-value for new business applications.** Over half of large companies currently deploying Hadoop measure the success of these initiatives based on a faster time-to-value, while 45% measure by new use case implementation. More than a third (36%) measure success by cost savings found through eliminating databases and database management systems (see Figure 8).

FIGURE 8
Measuring Success Of Hadoop Initiatives



Base: 150 IT and business managers involved in Hadoop projects in large enterprises in the US

Source: A commissioned study conducted by Forrester Consulting on behalf of BMC, September 2015

Key Recommendations

Enterprises are spending more time writing scripts for Hadoop automation than on other critical big data management functions such as security and integration. Enterprises should seriously consider Hadoop workflow automation to streamline their deployment, support faster time-to-value for new business use cases, and make IT more productive by minimizing big data complexity. Without automation, you run the risk of falling behind in your big data strategy, delaying timely insights that can affect business growth and innovation. IT and business managers should:

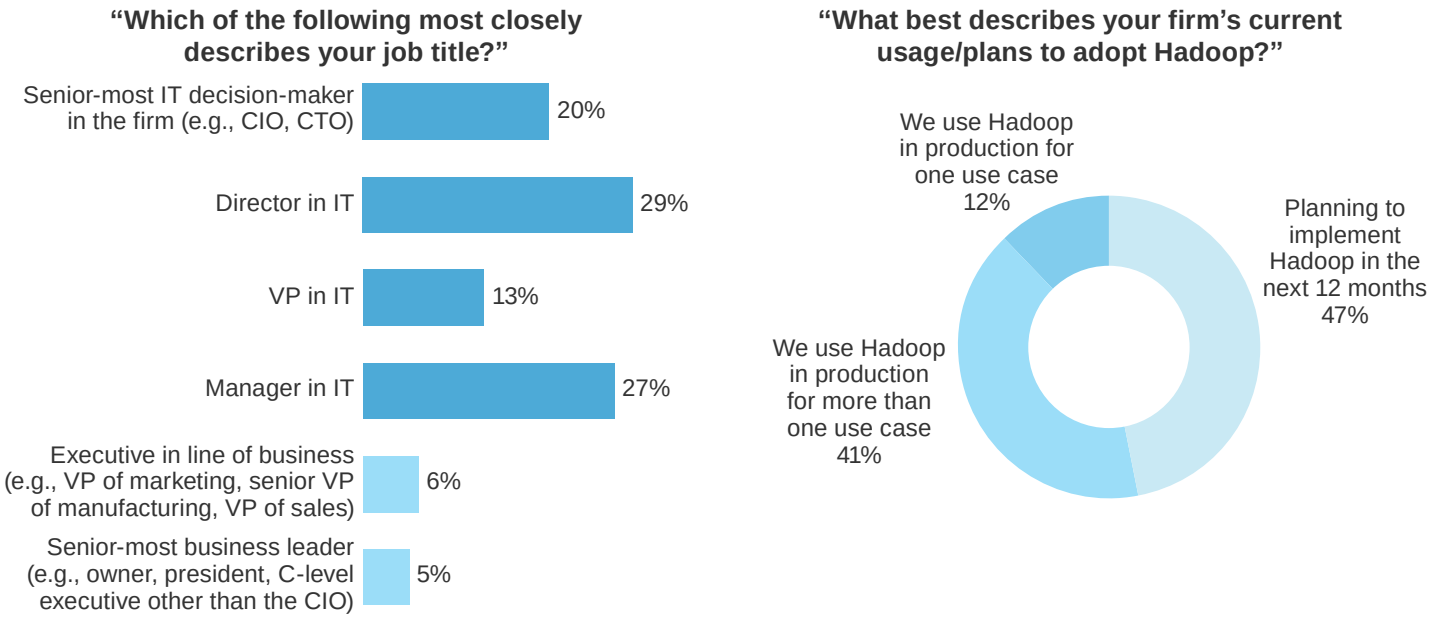
- **Use Hadoop to support multiple use cases.** Hadoop has evolved over the years to support many different use cases, including fraud detection, customer recommendations, predictive analytics, Internet-of-Things (IoT) analytics, customer analytics, and other insights. Enterprises should start with use cases that are clearly established and ones where the benefit can be achieved quickly.
- **Look at workflow automation to help accelerate deployment.** Although you can run Hadoop without workflow automation, it'll take you more time and effort to streamline your deployment. Without automation, most enterprises spend months of effort in setting up, installing, configuring, and managing Hadoop clusters.
- **Focus on integration, security, and skills to minimize failures.** Remember that there are still some gaps with Hadoop, especially around security, governance, integration, and quality. Look for tools and best practices to address these gaps, focusing on Hadoop clusters that are mission-critical and contain sensitive data.
- **Make cloud part of your Hadoop strategy.** Forrester expects a hybrid world will dominate big data deployments by 2018, with organizations having mission-critical Hadoop clusters on both on-premises and in the public cloud.
- **Look at third-party tools to help with automation of Hadoop workflow.** Organizations that have deployed Hadoop workflow automation have not looked back. Many claim to have saved more than 30% of time and effort with Hadoop workflow automation, while others have stated 65%. Look at third-party tools that can help with automating the implementation of Hadoop clusters, focusing on those that can improve productivity, minimize human errors, and support enterprise capabilities.

Appendix A: Methodology

In this study, Forrester conducted an online survey of 150 US large enterprises with 1,000-plus employees to evaluate Hadoop use cases and the adoption of workflow automation in Hadoop deployments. Survey participants included decision-makers in IT and business management. The study was completed in August and September 2015.

Appendix B: Demographics

FIGURE 9
Job Title And Hadoop Adoption



Base: 150 IT and business managers involved in Hadoop projects in large enterprises in the US

Source: A commissioned study conducted by Forrester Consulting on behalf of BMC, September 2015