bmc

Solution White Paper

Get Big Data into Production at the Speed of Business

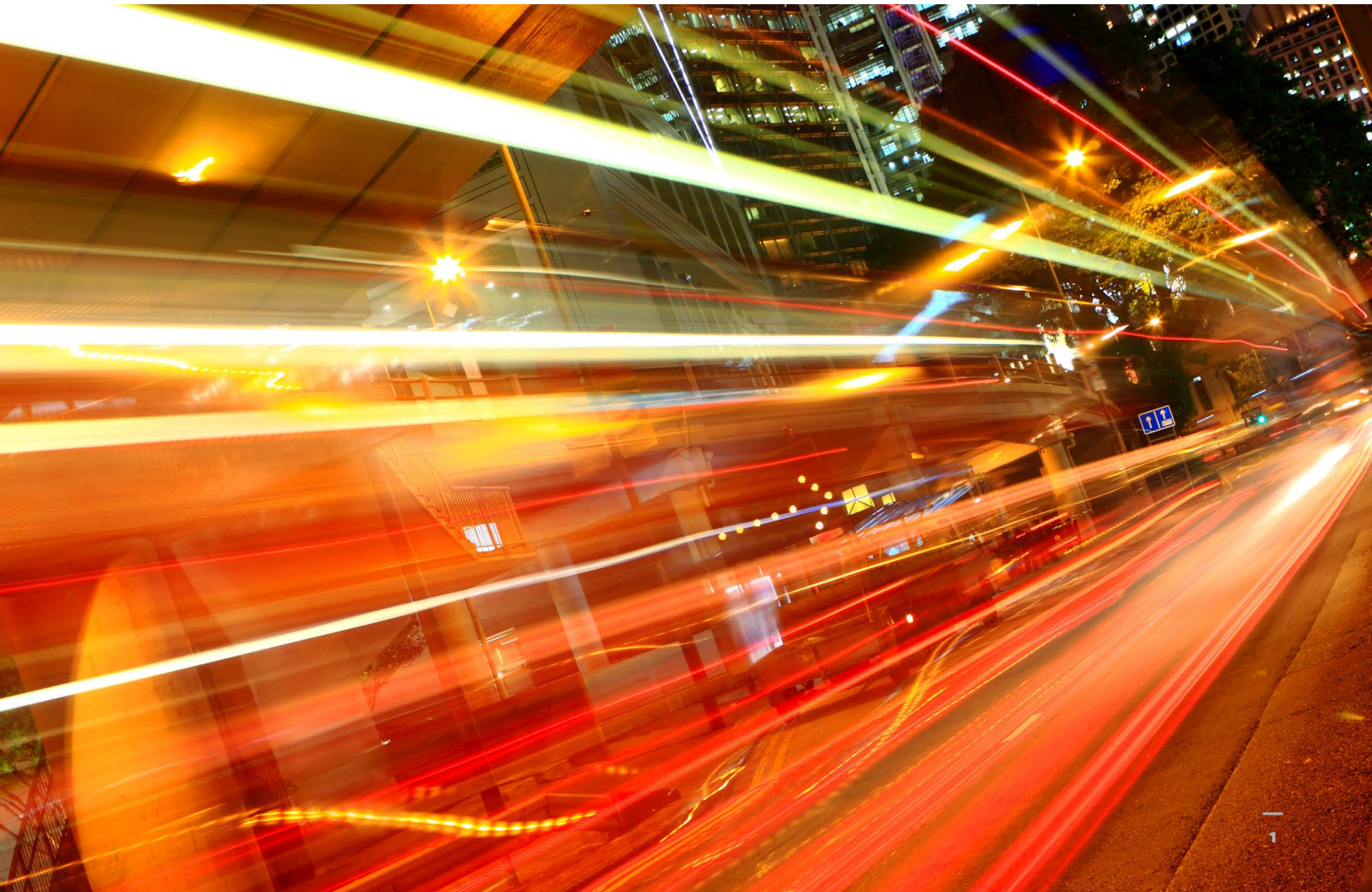
Deliver innovative digital services faster to drive growth and competitive advantage

Table of Contents

Executive Summary

Businesses rely on IT to deliver the digital services that drive growth and innovation. CIOs also face a mandate from executive business leaders to leverage new trends for competitive advantage. In both cases, the challenge is for IT to adopt new technologies as quickly as possible, in a way that helps accelerate time to market for new services. When it comes to big data, this remains a struggle for many organizations. In spite of enthusiastic predictions and clear potential, relatively few Hadoop projects have actually reached production, in part due to the complexity of the ecosystem. **Automation can help IT overcome this complexity**, making it possible to deliver on the promise of big data and create the digital services the business needs, when it needs them.

This paper explores the challenges that slow many Hadoop projects and how they can be solved through automation at the setup, development, and production stages.



THE CHALLENGE FOR IT: DRIVE BUSINESS VALUE AND INNOVATION

CEOs today are intensely focused on growth, and technology plays a key role in their strategies to expand their market reach and capture new opportunities. Corporate leaders now recognize the value of making their business more digital. Instead of viewing IT solely in terms of services to support business functions, they want to introduce digital services that operate at the very core of their business, from financial management tools for banking customers to the delivery of entertainment content for consumers.

The mandate for CIOs is clear—deliver high-quality digital services that put the latest innovations to work for the business. In many cases, CEOs and CFOs are actually pushing IT toward specific trends such as big data, not just asking for a particular business outcome, a clear indication of the transformative value they perceive in the new technology. Unfortunately, this direction may also reflect a lack of confidence in IT, as if the CIO would otherwise be unable to deliver this value effectively. Surveys have shown that many CEOs consider much of their IT budget to be wasted or poorly spent. While any IT leader understands the many factors that can undermine an initiative, many beyond their control, such explanations do little to mollify executives. IT must demonstrate more value to prove its ability to fulfill the more strategic role now expected of it.

Above all, IT must transform its ability to bring digital services to market at the speed of business. In fast-moving markets, competitive advantage can come down to the ability to deliver a digital service sooner rather than later. When a key technology is plagued with challenges that slow this delivery, IT has no way to meet these expectations. Instead of continuing to struggle in the face of hopeless business-mandated timelines, IT must cut to the heart of the matter and remove the hurdles that stand in the way of crucial initiatives.

SPEED THE PATH TO PRODUCTION

Hadoop has become a major force in the enterprise—at least in theory. A Forrester survey[1] last year found that 82 percent of the companies performing proofs-of-concept for big data planned on using Hadoop or were already using it. However, only 16 percent of these companies were actually using Hadoop in production, showing a gap between their hopes for the technology and their ability to fulfill them in a timely manner.

The enthusiasm for Hadoop is easy to understand. Companies can harness data in new ways to drive business value, from reducing risk and fraud, to speeding the development of new products and services, to enabling one-to-one marketing across channels. But the difficulties Hadoop presents are equally evident—even beyond its unique development challenges. While sometimes discussed as if it were a magic formula for creating business value, Hadoop is nothing more nor less than a data store embedded within the enterprise environment. Below its layer lie data sources such as email, CRM and ERP systems, logs, and so on, which feed data into it. Above it are the applications into which it feeds data in turn, either directly or through an analytics layer. Surrounding databases and processes enhance its value. The complexity of this ecosystem is one of the main big data issues IT faces.

To deliver data-powered digital services, IT must adopt, manage, and exploit Hadoop effectively while meeting enterprise requirements for self-service, compliance, forecasting, SLAs, and more. This poses challenges in three areas: setting up the environment, including all the components required for big data projects; developing applications and code using this new big data layer; and running applications and analytics in production, where complexity can be especially daunting.

ENVIRONMENT CHALLENGES

Simply building out the Hadoop environment can mean long hours and sleepless nights for IT. The team must engineer the Hadoop cluster, install it, figure out the relevant dependencies, and configure the supporting networking infrastructure and software. As an open source technology, Hadoop also involves frequent changes, with patches to be applied and new versions to update. IT must also develop the tools to keep the environment measurable—including defining what it means for the system to be functionally up or down, based on the metrics that matter to the business, and monitoring these factors effectively. The only way to manage this challenge—especially when dealing with large Hadoop clusters—is through automation.

To avoid introducing additional complexity or risk, this automation should be enabled by provisioning Hadoop under a mature solution that already works with the rest of your environment. By the same token, Hadoop must be treated as a full citizen of the IT ecosystem, not a separate island with its own separate team and stack. It should run as part of the enterprise, within the same context of security, compliance, patch management, remediation, and other functions as any other enterprise system.

[1] *Is Your big data Solution Production-Ready?*, Forrester Research, January 2014.

DEVELOPMENT CHALLENGES

Hadoop poses challenges for developers due to its highly iterative nature. This is fundamentally different from the iterations involved in the Agile development of a Java application, for example, in which developers aim for a clear requirement and can quickly test whether it has been met. A big data application is typically more fuzzy, such as a predictive algorithm for customer preference, which must be fine-tuned repeatedly to improve its accuracy in more relative terms. Moreover, its code must be run against data as similar as possible to production data in terms of both scale and quality, each of which can influence the results achieved. Each time through, development comes to a halt while developers wait for a result.

Like other big data technologies, Hadoop also involves extensive use of batch workloads. Data from diverse sources must often be vetted for quality, prepared, and updated before a Hadoop job can commence. Again, this is an entirely different prospect than writing a Java application for WebLogic, where data is secondary; in this case, data is the primary focus for developers—and its scale is enormous.

Automation can help address problems related to development and testing, but it must be applied intelligently, with full sensitivity to business issues, such as which workloads should take priority over others. Intelligent automation is also critical for leveraging production data in development, which would be far too risky using brute-force automation. Again, Hadoop should leverage the same automation engine as the rest of the environment, as part of its full integration into the ecosystem, to eliminate the need for IT to move from tool to tool as part of a single motion. A mature solution for workload automation will help IT build big data services more efficiently and innovate more quickly for the business.

The problem with scripting

When developing applications in Hadoop, programmers often write the code itself first, and then write scripts to run it. This poses a number of issues. First, scripting is far from cheap, introducing another piece of code to write, test, debug, manage, and maintain—always with the risk of human error, as with any new software. These risks compound existing systems risk.

Scripts also lack security and authentication, and fail to understand rules or change management, requiring additional layers to enable these functions—further adding cost and risk. Consider the experience of a developer writing a script for a MapReduce job, who must also deal with surrounding infrastructure that must be written around the actual value they're trying to produce.

A good workload automation product can help you shorten development time and get apps into production more quickly. By replacing scripting with the standard functions provided by a workload automation system, you can also deliver an application that is simpler, more reliable, and easier to debug once in production.

PRODUCTION CHALLENGES

Even when a big data application has reached production, challenges remain. The ecosystem involved in reaching an end result is tremendously complex, requiring diverse components to operate in concert, such as systems for ERP, ETL, data integration, analytics, and file transfers, as well as IT services such as email, helpdesk, monitoring, and change management. It's not enough to go into an individual module to verify its output; IT must also make sure that the application as a whole is delivering what the business requires. This comes back to the notion of treating Hadoop as a full citizen of the environment, operating within a larger context to drive value.

As part of this context, IT must have a way to ensure that the application is secure and compliant, as well as maintaining some control over data flow. At the same time that the application is in production, non-production systems might need to access its data. This needs to be done in a managed and controlled manner to ensure that SLAs are met.

Maintaining this level of visibility and control is impossible without automation. A Hadoop system might have thousands of nodes—IT can't hope to perform checks and react manually. Automated compliance checks must be complemented with automated remediation. A Hadoop application must be able to report on itself, including performance measurement from the infrastructure level to the application and user experience levels, as well as utilization. Is the application running at capacity, will it require additional servers, or are its supporting components underutilized? Workload automation is especially valuable in this regard, helping IT ensure that workloads will be executed not just on time, but at the right time based on business requirements.

CONCLUSION

To prove its ability to drive real business value, growth, and competitive advantage, IT must be able to deliver the innovative digital services the business needs, when it needs them. Given the complexity of the Hadoop environment, this can only be achieved through intelligent automation applied throughout the initiative, from setup to development to production. **By bringing Hadoop into the existing enterprise environment as a full citizen, IT can leverage proven, familiar tools while gaining complete visibility and control over the application and its supporting technologies.** Security, compliance, change management, and other key functions can be applied through the same environment, rather than requiring separate technologies and the cost and complexity they add. For IT, Hadoop and big data become as rational, reliable, and manageable as any other enterprise technology—while delivering transformative value for the business.

FOR MORE INFORMATION

To learn more about Hadoop and workload automation, please visit **bmc.com/hadoop**

460787