

Solution Spotlight

Why Hadoop Needs a Unified Integration Platform

Date: September 2016 **Author:** Nik Rouda, Senior Analyst

Big Data Success Is More than Volume and Velocity

When creating an environment for big data and analytics, it is too easy to focus on building only the right data infrastructure. Yet big data initiatives aren't successful simply because of Apache Hadoop's clever distributed analytics architecture—they are successful if they enable the development of applications that provide your business with new insights and capabilities.

ESG research (right) shows which attributes are most important to respondent organizations when considering big data solutions, closely matching traditional enterprise operational values.¹ Hadoop-based analytics applications must bring these qualities.

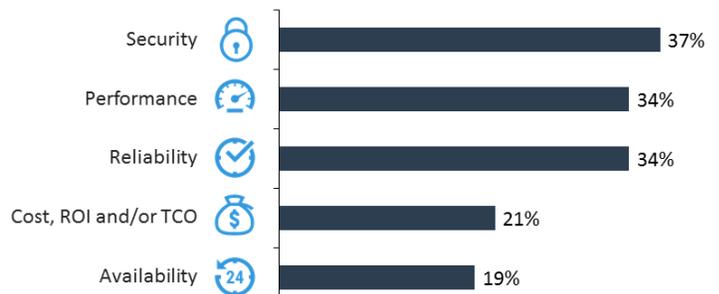
Although many vendors are building point products to complement a Hadoop data infrastructure, most offerings are narrow in scope. The result of too many disparate tools is more complexity, rather than high-quality applications that deliver against enterprise requirements.

ESG research found 45% of organizations report that big data analytics initiatives originate with the IT applications team. Similarly, 52% of respondents stated that applications groups provide the manpower to implement and manage such projects, and 82% say it's crucial or important for them to be engaged for a big data project to be successful. So if applications teams are central to success, where are their tools? Traditional databases had middleware platforms, but unfortunately these necessary capabilities end up being mostly self-built for Hadoop at great effort. The lack of a holistic big data application and integration platform is made worse by skills gaps, as 36% of businesses report a problematic shortage of skills for business intelligence and analytics, and 28% don't have enough skilled application developers.²

The Big Data Market Needs a Unified Integration Platform

Then how can you build secure, fast, reliable, cost-effective, highly available, well integrated, and agile applications? There is clearly a need for a merging of the application management and data integration tools to simplify the building and

Which of the following attributes are most important to your organization when considering technology solutions in the area of big data and analytics? (Percent of respondents, N=475, three responses accepted)



¹ Source: ESG Research Report, [Enterprise Big Data, Business Intelligence, and Analytics Trends: Redux](#), July 2016. All ESG research references and charts in this solution spotlight have been taken from this research report unless otherwise noted.

² Source: ESG Research Report, [2016 IT Spending Intentions Survey](#), February 2016.

running of Hadoop-based applications, and to handle common tasks like data ingestion, integration, extract/transform/load (ETL), and analytics modeling. And good data governance means you need to be able to easily see and control how data is flowing, and control access by role without making it frustrating. This combined functionality would facilitate self-service big data and data lake environments for data scientists, engineers, and business analysts.

To get more value from efforts and investments in big data, you need more of your organization to be able to leverage basic analytics without requiring data scientists or architects to enable each instance. There should be automatic integration of Hadoop into your existing ecosystem of BI tools. Compared to a “do-it-yourself” approach, a comprehensive integration platform would accelerate development, simplify access and access control, and prove a less brittle solution. This would improve your time to value and time to market, offer better governance, and remove barriers to innovation.

How Cask Data Application Platform (CDAP) Helps the Business Build Better Data Applications

Cask’s CDAP brings together a unified, open source integration platform for data applications and data lakes. The solution is Hadoop-native, lowering your risk, and giving you a reference design template from which to begin. Even better, the platform is readily extensible, so you can add plug-ins as needed, and future proof your applications against the continuous changes in the big data market. Maybe you developed for MapReduce yesterday, but are more focused on Spark today—who knows what the hottest new project will be tomorrow? Similarly, your applications will run uniformly on-premises or in public clouds like Amazon Web Services, Microsoft Azure, or Google Cloud Platform. Cask also supports heterogeneous Hadoop distributions including Cloudera, Hortonworks, and MapR. CDAP therefore brings you much more flexibility to keep innovating without having to keep rebuilding for different production environments.

At the same time, CDAP lets you get started without reinventing your existing best practices of application building. You can bring new applications into production with the same governance and security principles that have already been vetted and approved for your organization, reducing both effort and risk.

This robust set of capabilities lets you focus on creating and refining big data applications for your prioritized use cases, such as price optimization, customer 360 profiling, data warehouse offload, or perhaps the Internet of Things (IoT). You can build an application just once, but adapt it many times with reusable components, which improves developer productivity.

The Bigger Truth

The goal is to build business value with big data applications, and spend more time on code for the logic of each application instead of operationalizing it for different distributed environments. The benefits of this will directly apply to IT operations, DevOps, developers, executives, and architects. Cask has tilted the playing field, earning a massive unfair advantage over proprietary point products for data integration and ingest. Don’t get locked in. Set yourself up for hybrid app portability and provide the same quality and consistency even as you move to new environments. The Cask integration layer gives you much needed consistency across different environments and access to the latest technology, taking away the pain of managing big data applications.

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of The Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at 508.482.0188.



Enterprise Strategy Group is an IT analyst, research, validation, and strategy firm that provides market intelligence and actionable insight to the global IT community.

© 2016 by The Enterprise Strategy Group, Inc. All Rights Reserved.

