



Predictions 2015: Hadoop Will Become A Cornerstone Of Your Business Technology Agenda

Hadoop Has Already Disrupted The Economics Of Data. What's Next?

by [Mike Gaultieri](#)

with [Holger Kisker, Ph.D.](#), [Sharyn Leaver](#), [James Staten](#), [Brian Hopkins](#), David Murphy, and Sophia Christakis

WHY READ THIS BRIEF

Hadoop is the rising star of the business technology agenda for a simple reason — it disrupts the economics of data, analytics, and someday soon, all enterprise applications; it is secretly becoming an application platform too. Application development and delivery (AD&D) professionals should be aware of and take action on these eight predictions, including the disruptive power of “Hadooponomics,” Hadoop’s current killer app, the closing data management gap, and the emergence of brand new distros.

HADOOP ADOPTION IS NOT AN OPTION

Forrester believes that Hadoop is a must-have for large enterprises, forming the cornerstone of any flexible future data platform needed in the age of the customer.¹ But, we also believe that Hadoop is becoming more than just a data platform. Given its economics, performance and flexibility Hadoop will become an essential piece of every company’s business technology (BT) agenda. Application development and delivery professionals will build smart applications on Hadoop that not only use analytics, but also become an integral part of many of their applications. Hadoop is hot. We know that. You know that. Now on to our Hadoop predictions for 2015:

1. **Hadooponomics makes enterprise adoption mandatory.** The jury is in. Hadoop has been found *not guilty* of being a hyped-up open source platform. Hadoop has proven real value in any number of use cases including data lakes, traditional and advanced analytics, ETL-less ETL, active-archive, transactional data, and more.² All these use cases are powered by what Forrester calls “Hadooponomics” — its ability to linearly scale both data storage and data processing and leverage pay-per-use public cloudonomics.³ Many enterprises are dabbling in Hadoop to see what it can do. Many already have mission-critical capabilities running on Hadoop, including (to mention a few): Wal-Mart, Fidelity Investments, Sears, Verizon, USAA, Cardinal Health, Wells Fargo, Proctor & Gamble, Cablevision, Nasdaq, AutoTrader, Netflix, and Yelp.

What it means: The remaining minority of dazed and confused CIOs will make Hadoop a priority for 2015. AD&D professionals should be ready and waiting with a compelling use case to get started. Need an idea? See the next prediction.

2. **SQL becomes Hadoop’s killer app for 2015.** SQL is the primary lingua franca for structured enterprise data and is used by application developers to read and write data to databases. It is often used by business intelligence professionals to explore data. And, it is certainly supported by all major business intelligence and analytics tools. That’s why SQL in and on Hadoop is one of the most



Headquarters

Forrester Research, Inc., 60 Acorn Park Drive, Cambridge, MA, 02140 USA
Tel: +1 617.613.6000 • Fax: +1 617.613.5000 • www.forrester.com

prolific use cases in the Hadoop ecosystem. Fast and ANSI-compliant SQL on Hadoop creates immediate opportunities for Hadoop to become a useful data platform for enterprises. AD&D professionals already know SQL and many existing technologies can already interface to it. Certainly Hadoop has many more use cases than SQL, but many enterprises will start with it because it is the low-hanging fruit.

What it means: SQL on Hadoop creates an instant, easy-to-use and justify use case for enterprises. Create a data sandbox in Hadoop using structured data that is siloed and not readily accessible from other sources such as a data warehouse. Let your technology management professionals use SQL to analyze the data.

- 3. Enterprise software vendors close Hadoop's data management and governance gaps.** Hadoop is a general-purpose platform. That means it can store and process any kind of data. That's the good news. The bad news is that all parties agree that Hadoop-based data management and governance solutions have a ways to go before they provide the functionality sophisticated enterprises expect from their app platforms. We believe enterprise software vendors will beat the open source community to the punch because they have a broader perspective than just Hadoop. We are already seeing mature vendors such as SAS Institute, Informatica, Teradata, BMC, Protegrity, Talend, IBM, Oracle, Microsoft, and many, many others offer data analytics, management, and governance tools that run natively in Hadoop, but also work with other enterprise platforms, which provides them an advantage.

What it means: Customers will have to endure fits and starts when it comes to choosing data management and governance. Take a minimum viable product (MVP) approach to implementing data management and governance until the tools mature and leaders emerge.

- 4. The Hadoop skills shortage disappears.** Hadoop is not that hard to understand. It is a file system, albeit distributed, and it is a computing platform, albeit distributed. The APIs are Java. What's the big deal? Digging in to a new open source platform and learning the APIs is nothing new to enterprise Java application developers. They have to do it continuously to keep up and keep their résumés fresh. Hadoop is no different. The shortage of Hadoop skills will quickly disappear as enterprises turn to their existing application development teams to implement projects such as filling data lakes and developing MapReduce jobs using Java. Let's not forget about the operations skills needed to maintain a Hadoop cluster. Same thing. DevOps continuously deal with a complex heterogeneous environment. For those familiar with business intelligence, SQL on Hadoop will open the door to familiar access to their data. And while experienced technology management professionals take the bull by the horns, the open source community and commercial vendors are building better tools to make using Hadoop easier to use.

What it means: CIOs won't have to hire high-priced Hadoop consultants to get projects done. Hadoop projects will get done faster because the enterprise's very own application developers and operations professionals know the data, the integration points, the applications, and the business challenges. Additional skills for more complicated applications, such as predictive analytics running inside Hadoop, can be built when needed over time.

- 5. Enterprises will let thousands of Hadoop clusters bloom in the cloud.** Hadoop is both a data storage and data processing system. That means storage, compute, and network resources are required to run a Hadoop cluster. But, clusters don't always need all of these resources at once. If there is no job running in a Hadoop cluster, then the compute and network on the nodes are largely sitting idle. That's a waste. In the cloud, those resources can be managed more elastically. Data can sit idle in a file system and then compute and network resources can be added to the cluster when jobs need to run. Enterprises that implement static on-premises clusters may suffer the cost of inefficiency if they run jobs sporadically or even on a regular basis but only for a fraction of the day. Hadoop cloud services offer enterprises a way to manage the resources much more efficiently.⁴ They can spin up a thousand nodes to perform a particular gnarly job for a few hours and then knock back the compute and network resources to next to nothing when not needed. Cloud economics also minimizes that cost of deploying and maintaining Hadoop.

What it means: Early adopters of Hadoop will increasingly use Hadoop in the cloud to optimize the cost of Hadoop clusters and to meet demand for ad hoc analytics on Hadoop. This is Hadooponomics at its best. This won't, however, slow down on-premises adoption.

- 6. Hadoop won't be just for analytics anymore.** Hadoop is as much a computing platform that can become the foundational component of enterprise applications. With better resource management features provided by YARN, database options such as HBase, and in-memory overlay Apache Spark, Hadoop becomes an application platform.⁵ It is designed to run code with data and there is no reason why that code cannot be application specific. We also will see more database options and middleware running directly on Hadoop. HBase is already the de facto database specifically designed to run on Hadoop. Advertising technology firm Rocket Fuel uses HBase running on Hadoop to instantly access online profiles stored in HBase to determine the best ad to display on a web page or mobile device. That puts Hadoop squarely in the mode of being part of an application.

What it means: Meet Hadoop, the multi-function, enterprise application platform. As Hadoop starts to collapse the application stack, some vendors will get crushed. Vendors that offer application middleware, databases, search engines, integration tools, and other software infrastructure must move to develop versions that will run natively inside Hadoop.

7. **Yet another new Hadoop distribution emerges.** The large enterprise vendors such as HP, Oracle, SAP, Software AG, and Tibco will create their own distribution, as IBM and Pivotal already have. We will also see startups emerge that focus on a distribution that is specifically focused on public and private cloud to join the market for Hadoop in the cloud and disintermediate the pure-plays. New distribution will also emerge specifically designed to be embedded in even the smallest of devices. With all the data generated toward airplanes and automobiles, we can image a distribution designed to run in these systems to collect and process data on the fly — so to speak. The market is primed for a bold, new startup to take on the entire Hadoop world by declaring that Apache Spark, not Hadoop, is the future of computing. Apache Spark runs on Hadoop and is deliverable in many of the current Hadoop distributions. But, Apache Spark doesn't *have* to run on Hadoop. It could support other file systems or data sources.

What it means: This is an emerging market with no locked-down leader yet. More competition will drive more innovation, but AD&D leaders should not wait until the dust settles. Pick a distro and implement a use case that justifies the investment in time and money.

8. **Hadoop gets added to Linux and Windows operating systems — for free.** Hadoop software runs on clusters of a few to thousands of nodes. There no reason why Microsoft, Red Hat, VMware, and other operating system vendors couldn't include Hadoop and make it a configurable option within their operating systems. If a technology management professional wanted to add a node to join a Hadoop cluster they could simply "turn it on" through configuration. We already have seen this phenomenon occur with other technologies. In the mid-1990s Netscape Communication Corporation was selling web servers at a hefty price; now nearly all Windows and Linux-based distributions come with an HTTP server built in. The same thing could happen with Hadoop. A courageous or desperate vendor will make it happen now.

What it means: This would be a game-changer because it would mean that every node could be configured as a Hadoop node. This would disrupt the existing model used by Hadoop distribution vendors to charge \$2,000 to \$3,000 per node per year for their current distributions.

SUPPLEMENTAL MATERIAL

Acknowledgements

Forrester would like to thank the many technology professionals that we have interviewed for this brief (and about Hadoop in general) for providing input into this research. Your perspective on the market, appreciation of the complexity and nuances of enterprises application and data management architectures, and experience implementing real enterprise applications has been invaluable in conducting this research and all other research in big data and application development and delivery.

Companies whose technology professionals were interviewed specifically for this research include:

Companies Interviewed For This Report

Altiscale	MapR Technologies
Amazon Web Services	Microsoft
AutoTrader	Oracle
Cloudera	Pivotal Software
Hortonworks	Teradata
IBM	Verizon

ENDNOTES

- ¹ In Forrester's 32-criteria evaluation of big data Hadoop solutions, we evaluated nine solutions from Amazon Web Services (AWS), Cloudera, Hortonworks, IBM, Intel, MapR Technologies, Microsoft, Pivotal Software, and Teradata. This report details our findings about how well each solution fulfills the criteria, shows where the vendors stand in relation to each other, and helps technology management professionals select the right Hadoop solution for their business needs. For more information, see the February 27, 2014, "[The Forrester Wave™: Big Data Hadoop Solutions, Q1 2014](#)" report.
- ² The proliferation of customer-facing data-intensive systems in almost every modern enterprise has catalyzed the rapid deployment of big data environments, commonly with Hadoop as the underlying processing environment. This report helps I&O professionals understand the basics of Hadoop infrastructure and includes guidelines for system configuration, rough data sizing, and suggestions on how to plan for the inevitable growth of the Hadoop big data environment. Read our latest reports on Big Data use cases or for more information, see the April 9, 2014, "[Building The Foundation For Customer Insight: Hadoop Infrastructure Architecture](#)" report.
- ³ Cloudonomics is an IT financial model that is based on two fundamental differences in how IT resources are provided: elastic scale and resources priced on a purely pay-per-use basis. This concept is fully fleshed out in Joe Weinman's book, Cloudonomics. Source: Joe Weinman, Cloudonomics, John Wiley & Sons, 2012 (<http://www.cloudonomics.com/>).
- ⁴ What are your cloud options for building predictive applications, doing real-time analytics and driving new business insights? It starts with understanding the landscape of choices, your business objectives and what skills you bring to the question. An upcoming report on big data options in the cloud will help AD&D professionals map the right solution to the right need.
- ⁵ YARN stands for "Yet Another Resource Negotiator." It is a component of Hadoop that manages cluster resources and scheduling applications.

Forrester Research (Nasdaq: FORR) is a global research and advisory firm serving professionals in 13 key roles across three distinct client segments. Our clients face progressively complex business and technology decisions every day. To help them understand, strategize, and act upon opportunities brought by change, Forrester provides proprietary research, consumer and business data, custom consulting, events and online communities, and peer-to-peer executive programs. We guide leaders in business technology, marketing and strategy, and the technology industry through independent fact-based insight, ensuring their business success today and tomorrow.

© 2014 Forrester Research, Inc. All rights reserved. Forrester, Forrester Wave, RoleView, Technographics, TechRankings, and Total Economic Impact are trademarks of Forrester Research, Inc. All other trademarks are the property of their respective owners. Reproduction or sharing of this content in any form without prior written permission is strictly prohibited. To purchase reprints of this document, please email clientsupport@forrester.com. For additional reproduction and usage information, see Forrester's Citation Policy located at www.forrester.com. Information is based on best available resources. Opinions reflect judgment at the time and are subject to change.

117705