

The Forrester Wave™: Big Data Hadoop Distributions, Q1 2016

Five Top Vendors Have Significantly Improved Their Offerings

by Mike Gualtieri and Noel Yuhanna

January 19, 2016

Why Read This Report

Hadoop is data's darling for a reason — it thoroughly disrupts the economics of data, analytics, and data-driven applications. Enterprise adoption is mandatory for firms that wish to double-down on advanced analytics and create insights-driven applications to help them succeed in the age of the customer. Application development and delivery (AD&D) pros can start here with Forrester's 35-criteria evaluation of five Hadoop distributions from vendors: Cloudera, Hortonworks, IBM, MapR Technologies, and Pivotal Software.

Key Takeaways

Five Hadoop Solutions Vie For Enterprise Adoption

Among the Hadoop distribution vendors Forrester evaluated, we found four Leaders and one Strong Performer. For cloud-only Hadoop solutions, look for Forrester's upcoming Hadoop Cloud Solutions Wave.

Hadoop Has Both Open Source And Commercial Friends

Hadoop is not one simple thing. It is more of a rallying point for distribution data storage and processing. For example, every Hadoop distribution we evaluated includes Apache Spark.

Your Choice Is Tough, Because Competition Is So Vibrant

Choosing the best Hadoop distribution for your organization is not easy. Each vendor's solution has nuanced sweet spots. Use Forrester's downloadable Excel tool to deeply analyze the criteria and scores for each.

The Forrester Wave™: Big Data Hadoop Distributions, Q1 2016

Five Top Vendors Have Significantly Improved Their Offerings

by [Mike Gualtieri](#) and [Noel Yuhanna](#)

with [Holger Kisker, Ph.D.](#), [Rowan Curran](#), [Brandon Purcell](#), [Sophia Christakis](#), [Shreyas Warriar](#), and [Matthew Izzi](#)

January 19, 2016

Table Of Contents

- 2 Hadoop Becomes The New Core Of The Analytical Enterprise**
- 3 Big Data Hadoop Distributions Evaluation Overview**
 - Evaluation Criteria: Current Offering, Strategy, And Market Presence
 - Our Hadoop Distribution Evaluation Assessed The Capabilities Of Five Vendor Offers
- 5 Solid Choices All Around Make For A Tough Decision**
- 7 Vendor Profiles**
 - Leaders Differentiate On Their Approach To Innovation
 - Strong Performer Can Still Be The Right Choice For Some Enterprises

- 9 Supplemental Material**

Notes & Resources

Forrester conducted product evaluations in November 2015 and interviewed five vendors: Cloudera, Hortonworks, IBM, MapR Technologies, and Pivotal Software.

Related Research Documents

- [Apache Spark Is Powerful And Promising](#)
- [The Forrester Wave™: Big Data Predictive Analytics Solutions, Q2 2015](#)
- [SQL-For-Hadoop: 14 Capable Solutions Reviewed](#)

The Forrester Wave™: Big Data Hadoop Distributions, Q1 2016

Five Top Vendors Have Significantly Improved Their Offerings

Hadoop Becomes The New Core Of The Analytical Enterprise

The proliferation of data generated by enterprise applications, consumer web/mobile apps, and the Internet of Things (IoT) has afforded an unprecedented opportunity for businesses to know more about their customers, competitors, and business. However, the unfortunate truth is that the potential of most data lies dormant. On average, between 60% and 73% of all data within an enterprise goes unused for business intelligence (BI) and analytics (see Figure 1).¹ That's unacceptable in an age where deeper, actionable insights, especially about customers, are a competitive necessity.

Most enterprises see the opportunity of using more data, and they're not standing idly by as their data gathers dust. AD&D pros are adding Hadoop, en masse, to their mix of BI and data warehousing technologies to:

- › **Economize the analysis of more data.** Data warehouses have long been the gold standard for accumulating, organizing, and analyzing large volumes of structured enterprise data. But gold is what they cost too. Hadoop can be a more economical solution for both structured and unstructured data because it is based on open source, it processes large data sets more efficiently than data warehouses, and it can handle both SQL and general-purpose workloads.² Arguments can ensue about the price/performance of data warehouses versus Hadoop, but most enterprises cite cost effectiveness as a key reason they turn to Hadoop.
- › **Create lakes of big data from multiple tributaries.** Analyzing data from a large portfolio of dozens (or even hundreds) of internal business applications and external sources is a challenge, but it's essential to finding the most valuable actionable insights. To overcome this challenge, many enterprises create a "data lake," which is akin to a data warehouse except that it is built using Hadoop instead of traditional data warehouse technology. For most firms, a data lake is less formal than a data warehouse; they often use it in an ad-hoc fashion to discover new insights from both structured and unstructured data that get elevated to the data warehouse after the insights are discovered and formalized. However, many firms aspire to have data lakes built in Hadoop to completely replace their data warehouses.
- › **Perform advanced analytics.** Traditional BI is mostly about producing reports and populating dashboards with historical business performance analytics. That's enormously valuable, but to get the most out of their data, enterprises must also use advanced analytics, such as machine learning to build predictive models, text analytics to analyze unstructured data, and streaming analytics to capture real-time, perishable insights. Hadoop is a general-purpose data processing platform and has an increasing number of both open source and commercial partners. One of the best is Apache Spark.³ All Hadoop distributions come with Spark, and AD&D pros can use it to load data from the Hadoop distributed file system (HDFS) into memory for general-purpose processing or advanced analytics such as machine learning. Many commercial analytics vendors also have ported their tools and technologies to work with both Hadoop and Spark running on YARN, Hadoop's resource manager.

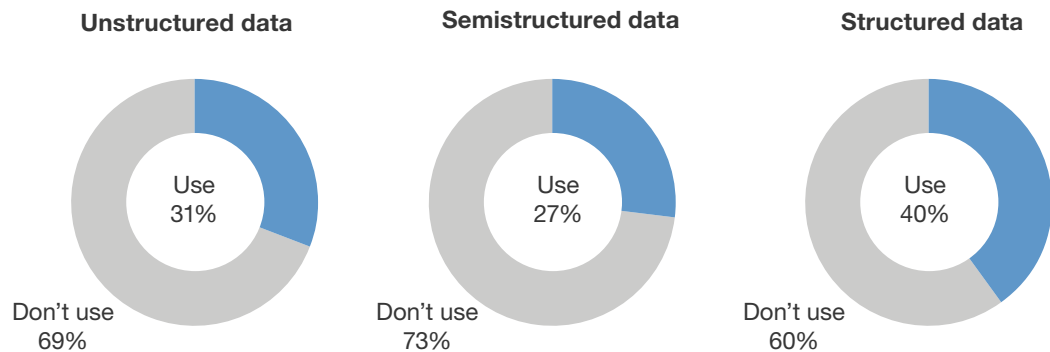
The Forrester Wave™: Big Data Hadoop Distributions, Q1 2016

Five Top Vendors Have Significantly Improved Their Offerings

- › **Provide self-service analytics to all who need it.** A past criticism of Hadoop was that only a Java programmer could use it to analyze data by creating custom MapReduce jobs or writing Pig scripts. Fortunately, business intelligence professionals, data scientists, and nontechnical businesspeople can also use Hadoop through a bevy of commercial ETL, BI, and advanced analytics tools too long to list here.⁴ Many of these tools have been enabled by SQL-for-Hadoop solutions that provide a standardized interface for data stored in Hadoop.⁵

FIGURE 1 Most Enterprise Big Data Goes Unused For Analytics

“Please estimate what percentage of the total size/volume of data within your company your company is currently using for business intelligence (BI).”



Base: 1,805 global technology decision-makers who know how much BI data their firm uses

Note: The percentages shown are estimates based on reported ranges; the values are not exact.
Source: Forrester's Global Business Technographics® Data And Analytics Survey, 2015

Big Data Hadoop Distributions Evaluation Overview

To assess the state of the market and see how the vendors stack up against each other, Forrester evaluated the strengths and weaknesses of five top commercial Hadoop distribution vendors: Cloudera, Hortonworks, IBM, MapR Technologies, and Pivotal Software.

Evaluation Criteria: Current Offering, Strategy, And Market Presence

After examining past research, user requirements, and vendor interviews, we developed a comprehensive set of 35 evaluation criteria, which we grouped into three high-level buckets:

The Forrester Wave™: Big Data Hadoop Distributions, Q1 2016

Five Top Vendors Have Significantly Improved Their Offerings

- › **Current offering.** We evaluated each product's configuration options, scalability, performance, high availability, disaster recovery, administration, security, core data capabilities, development tools, and other features to establish the capabilities of the vendor's current offering. All evaluated products must have been publicly available by October 1, 2015.
- › **Strategy.** We reviewed each vendor's strategy to assess its ability to compete and grow in the commercial Hadoop distribution market. Key criteria include Forrester's confidence in the vendor's ability to execute on its stated strategy and support current and future customers. We also reviewed each vendor's product road map to assess how this will affect the vendor's competitive position compared with the other vendors in this evaluation.
- › **Market presence.** To determine each vendor's market presence, we evaluated overall Hadoop revenue, number of paying customers, global distribution of paying customers, market awareness of the vendor's product, and partnerships with other technology and services firms.

Our Hadoop Distribution Evaluation Assessed The Capabilities Of Five Vendor Offers

Each of the five vendors (Cloudera, Hortonworks, IBM, MapR Technologies, and Pivotal Software) included in this evaluation has (see Figure 2):

- › **Core Apache Hadoop functionality.** All five vendors provide a Hadoop distribution that is based on the Apache Hadoop and related open source projects. Distribution features and add-on technologies must be aimed at the requirements of the world's largest enterprises, organizations, and government agencies.
- › **General-purpose technology.** The products are general-purpose Hadoop distributions products that aren't embedded or functionally focused within domain-specific applications.
- › **Software solution focus.** The vendors provide a software solution that organizations can install on their own on-premises, private cloud, and/or public cloud infrastructure. Cloud-only solutions such as Altiscale, Amazon Web Services' Elastic MapReduce, and Microsoft Azure HDInsight are not included because they will be covered in our upcoming Forrester Wave evaluation of Hadoop cloud solutions.
- › **A significant number of customers using the platform.** Each vendor must have at least 10 paying customers using its product in production (most have over 500 paying customers). Each vendor also provided at least two customer references who were willing to be interviewed by Forrester about their experience using the product.

The Forrester Wave™: Big Data Hadoop Distributions, Q1 2016

Five Top Vendors Have Significantly Improved Their Offerings

FIGURE 2 Evaluated Vendors: Product Information And Selection Criteria

Vendor	Product evaluated	Product version evaluated
Cloudera	Cloudera Enterprise	5.50
Hortonworks	Hortonworks Data Platform	2.30
IBM	IBM BigInsights for Apache Hadoop	4.10
MapR Technologies	The MapR Distribution including Apache	5.00
Pivotal Software	HadoopPivotal HD	3.x

Vendor selection criteria

Core Apache Hadoop functionality. The vendors included in this evaluation provide a Hadoop distribution that is based on Apache Hadoop and related open source projects. Distribution features and add-on technologies must be aimed at the requirements of the world's largest enterprises, organizations, and government agencies.

General-purpose technology. The products included in this evaluation are general-purpose Hadoop distributions products that aren't embedded or functionally focused within domain-specific applications.

Software solution focus. The vendors included in this evaluation provide a software solution that organizations can install on their own on-premises, private cloud, and/or public cloud infrastructure. Cloud-only solutions such as Altiscale, Amazon Web Services' Elastic MapReduce, and Microsoft Azure HDInsight are not included because they will be covered in our upcoming Forrester Wave™ evaluation of Hadoop cloud solutions.

A significant number of customers using the platform. Each vendor must have at least 10 paying customers using their product in production (most have over 500 paying customers). Each vendor must also provide at least two customer references who are willing to be interviewed by Forrester about their experience using the product.

Solid Choices All Around Make For A Tough Decision

Forrester's evaluation of big data Hadoop distributions uncovered a market with four Leaders and one Strong Performer (see Figure 3):

- › **Cloudera, MapR Technologies, IBM, and Hortonworks are Leaders.** Enterprise Hadoop is a market that is not even 10 years old, but Forrester estimates that 100% of all large enterprises will adopt it (Hadoop and related technologies such as Spark) for big data analytics within the next two years. The stakes are exceedingly high for the pure-play distribution vendors Cloudera, Hortonworks, and MapR Technologies, which have all of their eggs in the Hadoop basket. Currently, there is no absolute winner in the market; each of the vendors focuses on key features such as security, scale, integration, governance, and performance critical for enterprise adoption.

The Forrester Wave™: Big Data Hadoop Distributions, Q1 2016

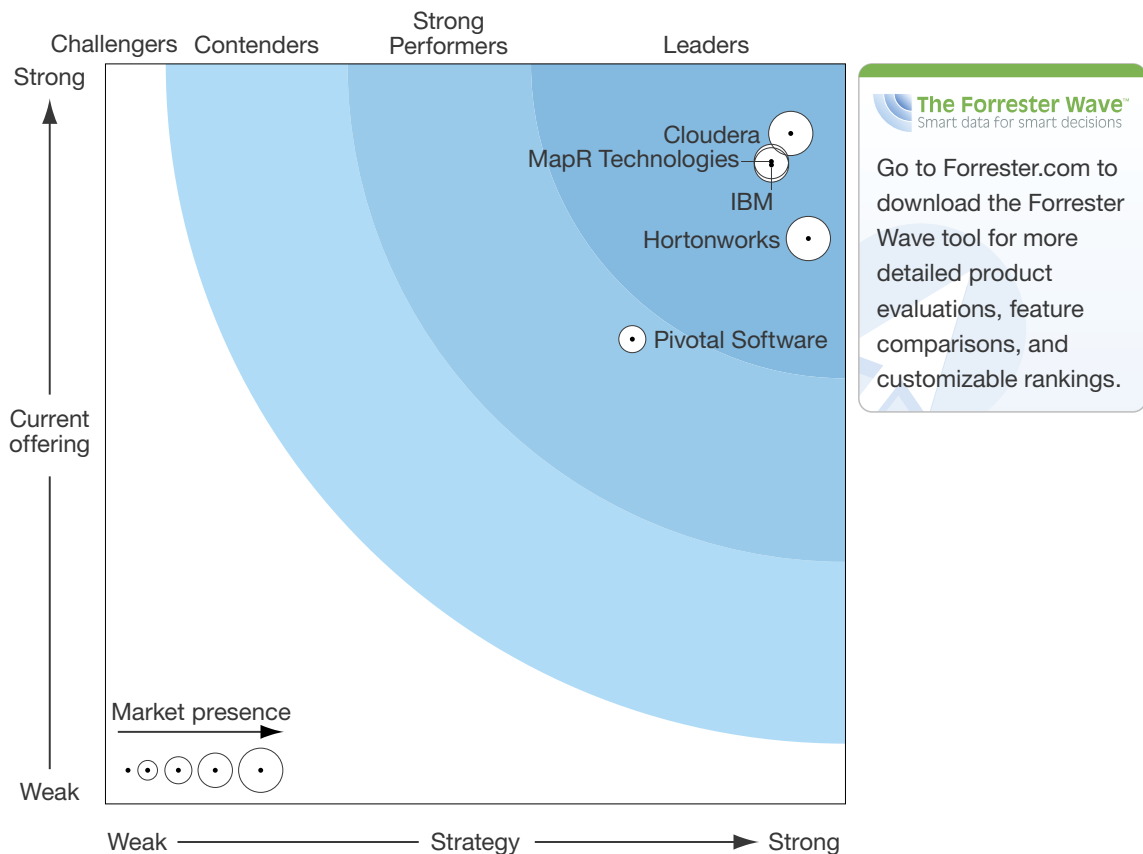
Five Top Vendors Have Significantly Improved Their Offerings

However, each pure-play vendor has a sweet spot strong enough to vigorously compete in the market (read the vendor profiles, below). IBM has the market strength, engineering prowess, and portfolio of analytics products to compete against the hot Hadoop startups. Choosing a Hadoop distribution will be difficult for most AD&D pros who carefully consider each of these Leaders. Forrester doesn't think there is a wrong choice among the Leaders in this evaluation. This is still a neck-and-neck market.

- › **Pivotal Software is a Strong Performer.** A Strong Performer among a Forrester Wave dominated by Leaders can still be a strong choice for an enterprise, especially if you value Pivotal Software's HAWQ SQL-for-Hadoop engine and MADlib machine learning library. Pivotal is an ODPI member, and thus some components of the Hadoop distribution will be equivalent to that of Leaders Hortonworks and IBM.

This evaluation of the big data Hadoop distribution market is intended to be a starting point only. We encourage clients to view detailed product evaluations and adapt criteria weightings to fit their individual needs through the Forrester Wave Excel-based vendor comparison tool. Clients can also schedule an inquiry with the authors to have a conversation about the market and specific vendor products.

FIGURE 3 Forrester Wave™: Big Data Hadoop Distributions, Q1 '16



The Forrester Wave™: Big Data Hadoop Distributions, Q1 2016

Five Top Vendors Have Significantly Improved Their Offerings

FIGURE 3 Forrester Wave™: Big Data Hadoop Distributions, Q1 '16 (Cont.)

	Forrester's Weighting	Cloudera	Hortonworks	IBM	MapR Technologies	Pivotal Software
CURRENT OFFERING	50%	4.53	3.82	4.32	4.34	3.14
Solution configuration	5%	5.00	5.00	5.00	5.00	4.00
Architecture	20%	4.20	3.40	4.00	4.80	2.40
Administration	15%	5.00	4.75	3.75	4.25	3.75
Security	10%	5.00	3.00	4.32	4.34	3.00
Data	15%	4.25	3.50	3.50	4.75	3.00
Data governance	10%	5.00	3.00	5.00	3.00	3.00
Workload flexibility	10%	3.00	3.00	5.00	5.00	3.00
Development	10%	5.00	5.00	5.00	3.00	3.00
Platform integrations	5%	5.00	5.00	5.00	5.00	5.00
STRATEGY	50%	4.63	4.75	4.50	4.50	3.56
Acquisition and pricing	25%	4.50	5.00	3.00	5.00	2.25
Solution road map	25%	5.00	5.00	5.00	4.00	3.00
Ability to execute	25%	5.00	5.00	5.00	5.00	5.00
Implementation support	25%	4.00	4.00	5.00	4.00	4.00
MARKET PRESENCE	0%	4.56	4.45	3.33	3.78	2.21
Evaluated product revenue	33%	4.00	4.00	3.00	3.00	2.00
Customer base	33%	4.67	4.34	4.00	4.67	3.00
Partnerships	34%	5.00	5.00	3.00	3.66	1.66

All scores are based on a scale of 0 (weak) to 5 (strong).

Vendor Profiles

Leaders Differentiate On Their Approach To Innovation

- › **Cloudera thinks big and acts fast.** Cloudera's scope and pace of innovation is astounding, but not surprising for the first commercial Hadoop startup, founded in 2008. Cloudera started the SQL-for-Hadoop craze with Impala. It offered the first visual cluster management tool and continues to put significant effort into key features such as security, high availability, governance, and administration. The vendor is anything but shy when it comes to making strategic acquisitions and partnerships to fill enterprise gaps in security, data management, and analytics. Cloudera's customers value the commercial add-on tools like Cloudera Manager, Cloudera Navigator, and Impala, as well as the company's overall vision for an enterprise big data platform.

The Forrester Wave™: Big Data Hadoop Distributions, Q1 2016

Five Top Vendors Have Significantly Improved Their Offerings

- › **MapR Technologies innovates to deliver extreme performance and reliability at scale.** From day one, MapR Technologies' strategy has been to engineer a distribution that would allow Hadoop to reach its full performance and scale potential with minimal effort. Enter MapR Technologies' linchpin — the MapR filesystem, which implements the HDFS API, is fully read/write, and can store trillions of files (versus the complex configuration for HDFS that requires separated namespaces). MapR has also done more than any other distribution vendor under the covers of Hadoop to deliver a reliable and efficient distribution for large-cluster implementations. Its customers typically have or are planning large, mission-critical Hadoop clusters and want to use MapR-DB and MapR Streams (which implement the HBase and Kafka APIs, respectively).
- › **IBM differentiates BigInsights with end-to-end advanced analytics.** IBM integrates a wealth of key data management components and analytics assets into the open source core of its Hadoop distribution. Enterprises using IBM's data management stack will find BigInsights a natural extension to their existing data platform. The company has also launched an ambitious open source project, Apache SystemML, for machine learning on Apache Spark from its newly minted Spark Technology Center. IBM's customers value the maturity and depth of its Hadoop extensions, such as BigSQL, which is one of the fastest and most SQL-compliant of all the SQL-for-Hadoop engines. In addition, BigQuality, BigIntegrate, and IBM InfoSphere Big Match provide a mature and feature-rich set of tools that run natively with YARN to handle the toughest Hadoop use cases.
- › **Hortonworks doubles-down on inclusive, broad community innovation.** Hortonworks is a rock when it comes to its promise to offer a 100% open source distribution. All of the technology built into HDP is an Apache open source project. Hortonworks will acquire companies to fill enterprise gaps and immediately contributes the code to an Apache project for the good of the community. For example, Hortonworks acquired XA Secure, a company with a commercially licensed security solution, and contributed the code to Apache as Apache Ranger. Hortonworks is also an important member of the Open Data Platform initiative (ODPi) formed earlier this year with IBM, Pivotal Software, and 12 other technology vendors, because the group has adopted Hortonworks-initiated projects such as Apache Ambari.⁶ Customers value Hortonworks' approach to open source innovation.

Strong Performer Can Still Be The Right Choice For Some Enterprises

- › **Pivotal Software launches an open source big data stack.** Pivotal Software made a big strategic move when it decided to contribute many of its key data platforms to the open source community, including Greenplum data warehouse, GemFire in-memory data grid as Apache Geode, HAWQ SQL-for-Hadoop as Apache HAWQ, and MADlib machine learning as Apache MADlib. The vendor's strategy is to make Pivotal HD Hadoop distribution a component of a larger big data platform that leverages its other database, data management, and application development assets. Pivotal Software is an ODPi member along with Hortonworks and IBM. Its customers are often also Greenplum users and value the compatibility, performance, and SQL compliance of the HAWQ SQL-for-Hadoop engine.

The Forrester Wave™: Big Data Hadoop Distributions, Q1 2016

Five Top Vendors Have Significantly Improved Their Offerings

Engage With An Analyst

Gain greater confidence in your decisions by working with Forrester thought leaders to apply our research to your specific business and technology initiatives.

Analyst Inquiry

Ask a question related to our research; a Forrester analyst will help you put it into practice and take the next step. Schedule a 30-minute phone session with the analyst or opt for a response via email.

Learn more about inquiry, including tips for getting the most out of your discussion.

Analyst Advisory

Put research into practice with in-depth analysis of your specific business and technology challenges. Engagements include custom advisory calls, strategy days, workshops, speeches, and webinars.

Learn about interactive advisory sessions and how we can support your initiatives.

Supplemental Material

Survey Methodology

Forrester's Global Business Technographics® Data And Analytics Survey, 2015, was fielded from January through March 2015 of 3,005 business and technology decision-makers located in Australia, Brazil, Canada, China, France, Germany, India, New Zealand, the UK, and the US from companies with 100 or more employees.

Forrester's Business Technographics provides demand-side insight into the priorities, investments, and customer journeys of business and technology decision-makers and the workforce across the globe. Forrester collects data insights from qualified respondents in 10 countries spanning the Americas, Europe, and Asia. Business Technographics uses only superior data sources and advanced data-cleaning techniques to ensure the highest data quality.

Online Resource

The online version of Figure 3 is an Excel-based vendor comparison tool that provides detailed product evaluations and customizable rankings.

The Forrester Wave™: Big Data Hadoop Distributions, Q1 2016

Five Top Vendors Have Significantly Improved Their Offerings

Data Sources Used In This Forrester Wave

Forrester used a combination of three data sources to assess the strengths and weaknesses of each solution:

- › **Vendor surveys.** Forrester surveyed vendors on their capabilities as they relate to the evaluation criteria. Once we analyzed the completed vendor surveys, we conducted vendor calls where necessary to gather details of vendor qualifications.
- › **Product briefings and demos.** We asked vendors to conduct briefings and demonstrations of their product's functionality. We used findings from these product briefings and demos to validate details of each vendor's product capabilities.
- › **Customer reference calls.** To validate product and vendor qualifications, Forrester also conducted reference calls with two of each vendor's current customers.

The Forrester Wave Methodology

We conduct primary research to develop a list of vendors that meet our criteria to be evaluated in this market. From that initial pool of vendors, we then narrow our final list. We choose these vendors based on: 1) product fit; 2) customer success; and 3) Forrester client demand. We eliminate vendors that have limited customer references and products that don't fit the scope of our evaluation.

After examining past research, user need assessments, and vendor and expert interviews, we develop the initial evaluation criteria. To evaluate the vendors and their products against our set of criteria, we gather details of product qualifications through a combination of lab evaluations, questionnaires, demos, and/or discussions with client references. We send evaluations to the vendors for their review, and we adjust the evaluations to provide the most accurate view of vendor offerings and strategies.

We set default weightings to reflect our analysis of the needs of large user companies — and/or other scenarios as outlined in the Forrester Wave document — and then score the vendors based on a clearly defined scale. These default weightings are intended only as a starting point, and we encourage readers to adapt the weightings to fit their individual needs through the Excel-based tool. The final scores generate the graphical depiction of the market based on current offering, strategy, and market presence. Forrester intends to update vendor evaluations regularly as product capabilities and vendor strategies evolve. For more information on the methodology that every Forrester Wave follows, go to <http://www.forrester.com/marketing/policies/forrester-wave-methodology.html>.

Integrity Policy

All of Forrester's research, including Forrester Wave evaluations, is conducted according to our Integrity Policy. For more information, go to <http://www.forrester.com/marketing/policies/integrity-policy.html>.

The Forrester Wave™: Big Data Hadoop Distributions, Q1 2016

Five Top Vendors Have Significantly Improved Their Offerings

Endnotes

- ¹ These are estimates based on ranges and are not exact. Source: Forrester's Global Business Technographics Data And Analytics Survey, 2015.
- ² There's a new dynamic duo in big data town: SQL and Hadoop. If you think this an unlikely combination, think twice. Enterprises have gobs of structured and semi-structured data generated by all sorts of transactional applications, and SQL is still the best option for querying it. Much of that data is increasingly finding its way into Hadoop clusters for analytics because of its versatility and the economical, linear scalability of both data storage and compute. In this report, Forrester has identified and reviewed open source and commercial SQL engines for Hadoop to help AD&D professionals learn about the maturity and sweet spot for each and choose the best for their enterprise's needs. See the "[SQL-For-Hadoop: 14 Capable Solutions Reviewed](#)" Forrester report.
- ³ You'll often hear Hadoop and Spark mentioned in the same breath. That's because, although they are independent platforms in their own right, they have an evolving, symbiotic relationship. AD&D pros must understand the key differences and synergies between this next-generation cluster-computing power couple to make informed decisions about their big data strategy and investments. See the "[Apache Spark Is Powerful And Promising](#)" Forrester report.
- ⁴ ETL stands for extract, transform, and load.
- ⁵ Forrester has identified and reviewed open source and commercial SQL engines for Hadoop to help AD&D pros learn about the maturity and sweet spot for each and choose the best for their enterprise's needs. See the "[SQL-For-Hadoop: 14 Capable Solutions Reviewed](#)" Forrester report.

Forrester examined more than a dozen Hadoop SQL solutions organized by three distinct approaches. Enterprise architects should read this report to understand the architectures of these solutions and what to look for as they plan to meet business needs with big data platforms and support emerging systems of insight. See the "[SQL-For-Hadoop: Critical Technology For Customer-Obsessed Firms](#)" Forrester report.
- ⁶ This brief analyzes the goals of the Open Data Platform initiative, the challenges and opportunities it will bring to the market, and the consequences that AD&D professionals will need to consider when planning and executing their company's Hadoop strategy. See the "[Brief: Can Hadoop's Enterprise Loose Ends Be Tied By The Open Data Platform Initiative?](#)" Forrester report.

We work with business and technology leaders to develop customer-obsessed strategies that drive growth.

PRODUCTS AND SERVICES

- › Core research and tools
- › Data and analytics
- › Peer collaboration
- › Analyst engagement
- › Consulting
- › Events

Forrester's research and insights are tailored to your role and critical business initiatives.

ROLES WE SERVE

Marketing & Strategy Professionals

CMO
B2B Marketing
B2C Marketing
Customer Experience
Customer Insights
eBusiness & Channel Strategy

Technology Management Professionals

CIO
› Application Development & Delivery
Enterprise Architecture
Infrastructure & Operations
Security & Risk
Sourcing & Vendor Management

Technology Industry Professionals

Analyst Relations

CLIENT SUPPORT

For information on hard-copy or electronic reprints, please contact Client Support at +1 866-367-7378, +1 617-613-5730, or clientsupport@forrester.com. We offer quantity discounts and special pricing for academic and nonprofit institutions.