FORRESTER®

The Forrester Wave™: Cloud Hadoop/Spark Platforms, Q1 2019

The 11 Providers That Matter Most And How They Stack Up

by Noel Yuhanna and Mike Gualtieri
February 13, 2019

Why Read This Report

Cloud Hadoop/Spark (HARK) platforms accelerate insights by automating the storage, processing, and accessing of big data. In our 25-criterion evaluation of HARK providers, we identified the 11 most significant ones — Amazon Web Services (AWS), Cloudera, Google, Hortonworks, Huawei, MapR, Microsoft, Oracle, Qubole, Rackspace, and SAP — and researched, analyzed, and scored them. This report shows how each provider measures up and helps enterprise architecture (EA) professionals select the right one for their needs.

Note: Cloudera and Hortonworks completed their planned merger on January 3, 2019, and will continue as Cloudera. This Forrester Wave reflects our evaluation of each company's independent HARK platforms prior to the completion of the merger.

Key Takeaways

AWS, Hortonworks, Microsoft, And Cloudera Lead The Pack
Forrester's research uncovered a market in which AWS, Hortonworks, Microsoft, and Cloudera are Leaders; Google, MapR, SAP, Oracle, and Huawei are Strong Performers; Qubole is a Contender; and Rackspace is a Challenger.

Performance, Data Management, And Security Are Key Differentiators
The Leaders we identified support a broader set of use cases, enhanced automation, and good scalability and performance. The Strong Performers have turned up the heat on the incumbents. Contenders offer more data management features and lower cost, and Challengers are ramping up their core cloud HARK functionality.

The Forrester Wave™: Cloud Hadoop/Spark Platforms, Q1 2019
The 11 Providers That Matter Most And How They Stack Up

by Noel Yuhanna and Mike Gualtieri
with Gene Leganza and Elizabeth Hoberman
February 13, 2019

Table Of Contents

Related Research Documents

The Forrester Wave™: Big Data Fabric, Q2 2018

The Forrester Wave™: Cloud Data Warehouse, Q4 2018

Predictions 2019: Business Insights

Share reports with colleagues.
Enhance your membership with Research Share.

FORRESTER®

Forrester Research, Inc., 60 Acorn Park Drive, Cambridge, MA 02140 USA
+1 617-613-6000 | Fax: +1 617-613-5000 | forrester.com

Cloud HARK Platforms Are For Serious Business

Modern enterprises are insights driven. They have to serve customers, scale, adapt, and compete — all with great efficiency. The veracity, timeliness, and scope of their insights will determine enterprise winners and losers. Getting data is not the problem — enterprises have gobs of it, generated from dozens, hundreds, and sometimes thousands of applications and systems. The perennial challenge faced by business insights pros is to turn the data-to-insights wheel faster. That "wheel" includes data acquisition, data security, data transformation, data governance, data exploration, analytics, and machine learning that is massively scalable, performant, and fault tolerant.

It's a lot to ask, and the need is great. That's why enterprises adopt and implement cloud HARK platforms for serious business. Forrester defines enterprise cloud HARK platforms as:

> *Distributed computing software and services that are rooted in open source Apache Hadoop and Apache Spark to ingest, store, process, manage, and analyze data to generate insights that drive business outcomes.*

Cloud HARK Platforms Vendors Have Taken Big Leaps Forward

Hadoop is more than 10 years old. It has gone from a batch-oriented storage, processing, and highly programmatic data platform to become a cloud-enabled, highly interconnected analytical platform that can support a wide range of use cases and stakeholders. It particularly shines when running as a cloud service because the workloads tend to be compute-heavy and elastic, and the insights generated are often integrated as essential elements of business processes, applications, and customer experiences. Enterprises seeking cloud HARK platforms should look for vendors that:

› **Add differentiated enterprise value on top of open source.** HARK platforms are based on Apache Hadoop, Apache Spark, and dozens of other open source projects that work in concert to support broader data and analytics use cases. HARK vendors must add value to succeed in the enterprise market. Pricing, packaging, support services, applications, and proprietary features differentiate the vendors. This Forrester Wave evaluation teases out those differences in 25 criteria. Enterprise buyers should take heed of these differences and carefully map them to their specific requirements, now and for the future.

› **Press the advantages of cloud infrastructure.** The best cloud HARK platforms are cloud-savvy, meaning they are designed to make the best of cloud infrastructure and related data services. Cloud HARK platforms must provision and deprovision cloud infrastructure resources, integrate with administration consoles, offer pay-as-you-go billing systems, and provide other cloud services, such as data storage, query services, analytical tools, business intelligence (BI) tools, and applications services, with minimal effort. Enterprise buyers should look to what extent a shortlisted solution truly takes advantage of the native cloud infrastructure versus abstracting away potential performance-boosting and cost-efficiency features.

› **Integrate with large ecosystems for analytical tools and technologies.** HARK platforms are the foundation on which great insights-driven organizations are built. That foundation is solid because it is highly scalable, is secure, includes enterprises tools, and has become the vessel of injection for open source insights innovation. That foundation has also attracted a huge following of data and analytics vendors, such as those providing data governance, ETL, BI, data science solutions, and many thousands of others.[1] In acquiring a cloud HARK platform, enterprise buyers should also look at the partnerships the vendor supports to take advantage of their tools and services.

Evaluation Summary

The Forrester Wave evaluation highlights Leaders, Strong Performers, Contenders, and Challengers. It's an assessment of the top vendors in the market and does not represent the entire vendor landscape. You'll find more information about this market in our overview report of HARK providers.[2]

We intend this evaluation to be a starting point only and encourage clients to view product evaluations and adapt criteria weightings using the Excel-based vendor comparison tool (see Figure 1 and see Figure 2). Click the link at the beginning of this report on Forrester.com to download the tool.

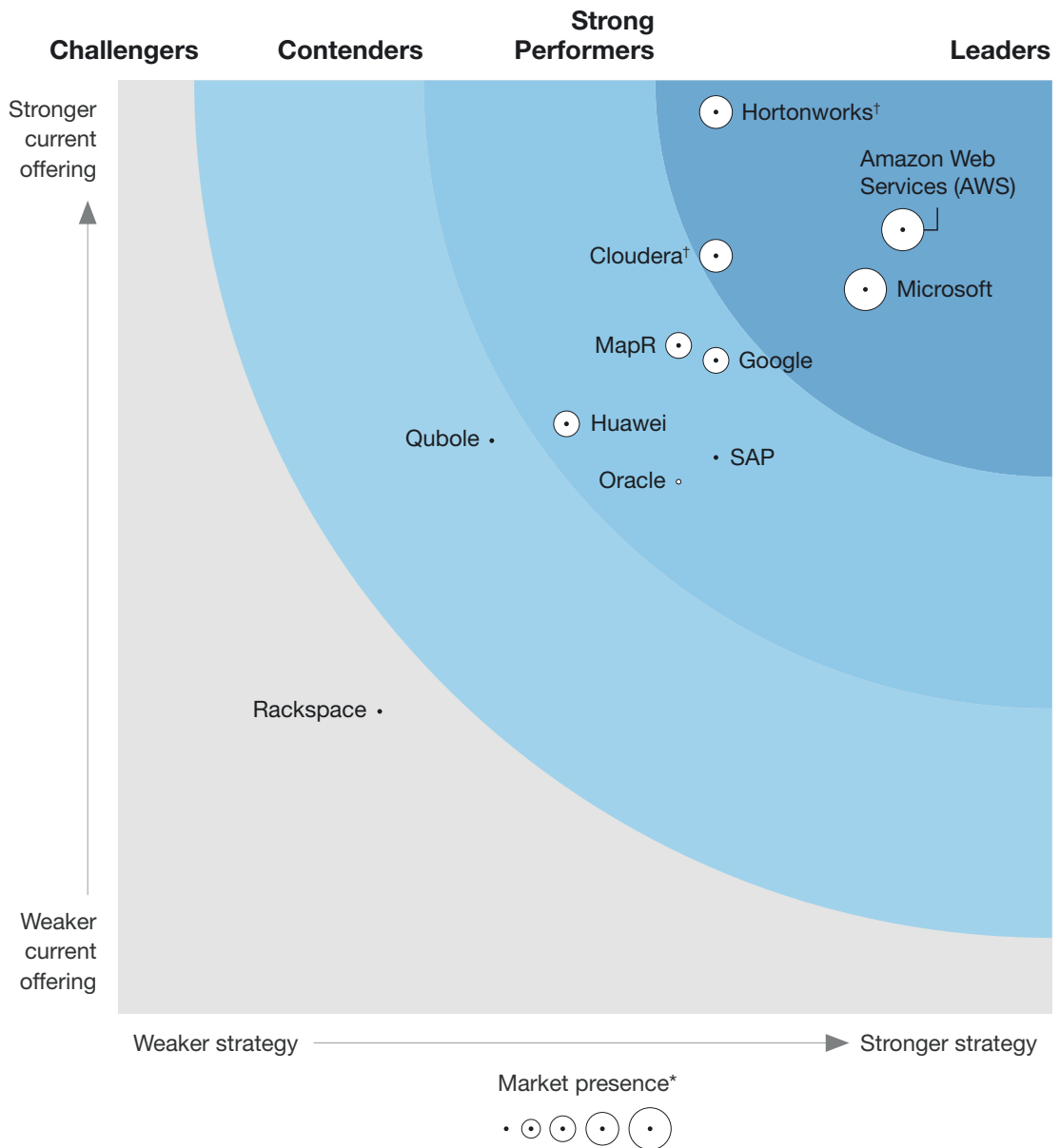**FIGURE 1** Forrester Wave™: Cloud Hadoop/Spark Platforms, Q1 2019

THE FORRESTER WAVE™

Cloud Hadoop/Spark Platforms

Q1 2019

*A gray marker indicates incomplete vendor participation.

†Cloudera and Hortonworks announced the completion of their merger on January 3, 2019. This graphic does not reflect the positioning of the combined entity.

FIGURE 2 Forrester Wave™: Cloud Hadoop/Spark Platforms Scorecard, Q1 2019

	Forrester's weighting	Amazon Web Services (AWS)	Cloudera	Google	Hortonworks	Huawei	MapR	Microsoft	Oracle*	Qubole	Rackspace	SAP
Current offering	50%	4.20	4.06	3.50	4.83	3.16	3.58	3.88	2.85	3.07	1.62	2.98
Packaging	15%	3.00	4.00	3.00	5.00	4.00	5.00	3.00	5.00	4.00	2.00	3.00
Architecture	20%	4.60	2.90	3.60	4.60	4.10	4.10	3.50	2.40	2.60	1.40	2.90
Administration	15%	5.00	5.00	5.00	5.00	3.00	3.60	4.40	3.00	3.80	1.60	3.00
Security	15%	3.60	5.00	3.80	5.00	3.00	3.60	3.80	3.00	3.80	1.60	3.00
Data management	20%	5.00	3.60	2.80	5.00	2.40	3.00	4.20	2.40	2.40	1.60	3.60
Applications	15%	3.60	4.40	3.00	4.40	2.40	2.20	4.40	1.60	2.20	1.60	2.20
Strategy	50%	4.20	3.20	3.20	3.20	2.40	3.00	4.00	3.00	2.00	1.40	3.20
Ability to execute	40%	5.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	1.00	1.00	3.00
Solution road map	40%	3.00	3.00	3.00	3.00	1.00	3.00	5.00	3.00	3.00	1.00	3.00
Pricing and acquisition	10%	5.00	3.00	5.00	3.00	3.00	3.00	3.00	3.00	1.00	3.00	3.00
Partners	10%	5.00	5.00	3.00	5.00	5.00	3.00	5.00	3.00	3.00	3.00	5.00
Market presence	0%	5.00	3.60	2.40	3.60	2.40	2.20	4.40	1.00	1.00	1.00	1.00
Customer adoption	40%	5.00	3.00	3.00	3.00	3.00	1.00	5.00	1.00	1.00	1.00	1.00
Evaluated product revenue	30%	5.00	3.00	3.00	3.00	3.00	1.00	3.00	1.00	1.00	1.00	1.00
Market awareness	30%	5.00	5.00	1.00	5.00	1.00	5.00	5.00	1.00	1.00	1.00	1.00

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

Vendor Offerings

Forrester included 11 vendors in this assessment: AWS, Cloudera, Google, Hortonworks, Huawei, MapR, Microsoft, Oracle, Qubole, Rackspace, and SAP (see Figure 3).

FIGURE 3 Evaluated Vendors And Product Information

Vendor	Product evaluated	Product version evaluated
Amazon Web Services (AWS)	Amazon EMR	
Cloudera	Cloudera Enterprise (Altus)	6.0
Google	Cloud Dataproc	
Hortonworks	Hortonworks Data Platform (HDP) Hortonworks DataPlane	HDP 3.0.1 Data Life Cycle Manager 1.2 Data Analytics Studio 1.0 Data Steward Studio 1.2
Huawei	FusionInsight	2.8
MapR	MapR Data Platform	6.1
Microsoft	Azure HDInsight Azure Databricks	3.6
Oracle	Oracle Big Data Cloud Services	
Qubole	Qubole Data Service	R53
Rackspace	Managed Big Data	
SAP	SAP Cloud Platform Big Data Services (SCP BDS) SAP Data Hub	5.0 2.3

Vendor Profiles

Our analysis uncovered the following strengths and weaknesses of individual vendors.

Leaders

› **AWS offers comprehensive cloud HARK services for most use cases.** AWS offers versions of Hadoop, Spark, and Presto that can work off data stored in Amazon S3. With Amazon EMR, organizations can leverage multiple data stores, including Amazon S3 using the EMR File System

(EMRFS), the Hadoop Distributed File System on local storage or Amazon EBS, Amazon RDS, Amazon DynamoDB, and Amazon Kinesis. Among evaluated vendors, AWS offers one of the broader ranges of partners for data integration, modeling and development, and analytics. Customers like its ability to separate compute and storage for on-demand scaling, automation of HARK deployments, faster time-to-value for new insights, and lower cost. However, some are concerned that it does not have comprehensive security features or data integration, and it lags in multiregion capabilities.

› **Hortonworks focuses on global multicloud and hybrid cloud HARK platforms.** Hortonworks' strategy has been to drive all innovation through the open source community and create an ecosystem of partners and build services that accelerate Hadoop and Spark adoption. Today, organizations are supporting new and emerging business use cases with Hortonworks, such as the IoT, customer 360, advanced analytics, and real-time insights.[3] The vendor provides a cost-effective, nimble, and scalable architecture to implement HARK, whether on-premises, multicloud, or hybrid cloud. Customers like its flexible open source platform, multicloud support, data ingestion capabilities, performance and scale, and broad ecosystem of partners and tooling. However, some claim it lags in support of comprehensive data integration and requires a lot of training and understanding for complex and large deployments. Cloudera and Hortonworks recently completed their merger.[4]

› **Microsoft's HARK offering has evolved to support most use cases.** Microsoft's strong presence in the database, data warehouse, cloud, spreadsheet, collaboration, BI, OLAP, and development tools markets is helping customers build and support big data stacks for customers.[5] Azure HDInsight uses the HDP Hadoop and Spark distribution, designed for the Microsoft Azure cloud for scale, performance, and security. Enterprise architects can use C#, Java, and .NET to create, configure, submit, and monitor Hadoop and Spark jobs, in addition to a user interface to accelerate deployments. PolyBase allows SQL Server customers to execute queries against data stored in Hadoop. Microsoft also offers Azure Databricks, a Spark-based analytics platform. Customers like its support for multiple use cases, integration with the SQL stack, and ease of use. However, some report that it requires a lot of training and understanding for large and complex deployments, and some criticize its performance, data integration, and security issues as well as the high cost of running the solution.

› **Cloudera's innovation is helping it gain momentum.** Cloudera's approach to innovation has been core to meeting customer demands and differentiating its solution. Cloudera leverages Apache Hadoop, Spark, Impala, HBase, Kudu, and Kafka to support low-latency, high-concurrency interactions that deliver real-time insights from big data. Enterprises have successfully deployed a wide range of use cases with Cloudera, including fraud detection, real-time analytics, IoT analytics, and customer 360. Customers like Cloudera's HARK platform performance, technical support, road map, continuous innovation, and partner ecosystem. However, some claim the platform's many components require highly skilled resources to deploy and operate. Cloudera and Hortonworks recently completed their merger.[6]

Strong Performers

› **Google's HARK offering is starting to gain momentum across verticals.** Cloud Dataproc, Google's managed Hadoop and Spark service, integrates with other Google Cloud Platform services such as BigQuery and BigTable, giving customers a more comprehensive platform for data processing and machine learning. Although Cloud Dataproc's level of integration with BigQuery and Cloud ML Engine can be improved, customers like its ease of use for managing large and complex HARK workloads, data pipeline support, global high availability across regions and tools, and support for broad analytical use cases. However, some claim that it lacks generally available integrated autoscaling with YARN and that its enterprise technical support is not on par with leading HARK vendors. They also voice concerns about the solution's data integration and graph processing.

› **MapR offers a low-cost HARK platform to support multiple use cases.** MapR Technologies continues to add unique innovations to its Hadoop and Spark distribution. MapR delivers a converged data platform that enables customers to leverage big data by combining real-time analytics with operational applications to support actionable insights. MapR has full platform integration with containers and Kubernetes, as well as good support for hybrid and multicloud environments. Customers like the platform's ease of use for data engineers, support for broad use cases, faster time-to-value, and lower cost. However, some customers claim the platform lags in data integration capabilities, graph processing, data preparation, and master data management.

› **SAP's HARK platform focuses on real-time use cases across various data sets.** SAP Cloud Platform Big Data Services (formerly Altiscale) is a fully managed HARK service to support big data workloads. It handles all of the provisioning, scaling, security, availability, and tuning to support many workloads, including real-time analytics, customer 360, IoT analytics, and integrated analytics. In addition, SAP customers can use SAP Data Hub for data management and SAP HANA for in-memory processing of non-Hadoop data sets. Customers like SAP's data platform to support a broader ecosystem with Spark, SAP HANA, Relational, and Hadoop sources; its technical support; and its broad set of real-time use cases. However, some claim it hasn't yet fully extended data security for HARK.

› **Oracle's Big Data Cloud has some capabilities, but adoption is lagging.** Oracle Big Data Cloud Services (BDCS) is composed of Apache Hadoop, Apache Spark, Connectors, Oracle Data Integrator, Oracle Big Data Spatial and Graph, and Oracle R Distribution. It focuses on automation, ease of use through single-command patching, upgrading, and comprehensive dashboards. In addition, Oracle's Big Data SQL delivers federation with Hadoop and non-Hadoop sources to support a broader set of analytical use cases. Customers like its on-demand provisioning, data integration, and end-to-end platform support for hybrid cloud with common tooling, integration, security, and delivery. However, customers claim it lags in comprehensive big data support, data security, ease of use, and management of Hadoop and Spark deployments. Oracle declined to participate in or provide information for our research. Scores are based on Forrester estimates.

› **Huawei's low-cost HARK platform is starting to expand beyond China.** As an OpenStack-based offering, Huawei Enterprise Cloud simplifies HARK development and deployment for various use cases, including customer 360, IoT, and line-of-business-driven analytics. The platform leverages open community software that's optimized to deliver high performance and scale. Although most of Huawei's HARK deployments are in China, its aggressive global marketing and sales campaign is boosting adoption in Europe, Africa, and the US. Customers like Huawei's low-cost platform, faster time-to-value, and support for hybrid cloud deployment. However, some report that its data security isn't comprehensive, the platform lacks comprehensive tools and partner support for broader analytical use cases, and it lags in automation capabilities.

Contender

› **Qubole is a cloud-agnostic, viable HARK platform.** Qubole, launched in 2013, offers HARK cloud services that run on AWS, Google Cloud Platform, Oracle, and Microsoft Azure. Qubole offers the ability to automate, simplify, and integrate data from sources and deliver advanced analytics. It elastically scales Hadoop resources up or down based on workload needs. Qubole focuses on simplifying the provisioning, management, scale, and security of big data analytics workloads, using data stored in AWS, Google Cloud Storage, Oracle Cloud, or Microsoft Azure infrastructure. Customers like its ability to do ad hoc analysis, real-time analytics, technical support, and faster time-to-value. However, some customers claim that its data security is lagging, and they cite the high cost of running the solution and that it often requires a lot of training and understanding for complex deployments.

Challenger

› **Rackspace's HARK services are viable for some use cases.** Rackspace offers enterprises a broad range of cloud services, including fully managed Hadoop and Spark as-a-service. It offers flexible multitenant or fully dedicated and isolated clusters. Rackspace provides support for Hadoop, Spark, and other Apache tools to support various stacks. Although Rackspace still has a way to go before becoming a threat to the larger cloud HARK vendors, its commitment to expanding key partnerships and increasing its focus on innovation will likely interest customers. Customers like its flexible infrastructure platform, ease of use for smaller to midsize deployments, high availability and disaster recovery offering, and lower cost. However, some claim that it lacks a strong big data offering to support large and complex deployments, that its data security isn't strong, and that it has limited automation capabilities.

Evaluation Overview

We evaluated vendors against 25 criteria, which we grouped into three high-level categories:

> › **Current offering.** Each vendor's position on the vertical axis of the Forrester Wave graphic indicates the strength of its current offering. Key criteria for these solutions are scalability, performance, administration, provisioning, auditing, data pipeline, data access, data governance, developer tools, and deployment options.

> › **Strategy.** Placement on the horizontal axis indicates the strength of the vendors' strategies. We evaluated ability to execute, solution road map, pricing and acquisition, and partners.

> › **Market presence.** Represented by the size of the markers on the graphic, our market presence scores reflect each vendor's customer adoption, product revenue, and market awareness.

Vendor Inclusion Criteria

Forrester included 11 vendors in the assessment: AWS, Cloudera, Google, Hortonworks, Huawei, MapR, Microsoft, Oracle, Qubole, Rackspace, and SAP. Each of these vendors has:

> › **A comprehensive cloud HARK offering.** Evaluated vendors provide a cloud HARK platform/service based on the Apache HARK open source and include value-added features and add-on tools appropriate for use by enterprises, such as security, data management, and cloud HARK management tools. The vendors also need to provide high availability, disaster recovery, provisioning, and administration services.

> › **An independent cloud HARK platform/service.** We included only cloud HARK platform/service that are not technologically embedded into any particular application, BI, predictive analytics, ETL, or middleware stacks. The solution should be supported in a standalone platform for use on-premises or in the cloud, or both.

> › **A referenceable install base.** Vendors must have at least 10 paying, enterprise customers using the cloud HARK platform that span more than one major geographical region.

> › **Client inquiries and/or technologies that put the vendor on Forrester's radar.** Forrester clients often discuss the vendor and its products through inquiries and interviews; alternatively, the vendor may, in Forrester's judgment, warrant inclusion or exclusion in this evaluation because of technology trends, market presence, the financial condition of the vendor, or lack of client interest.

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

To help you put research into practice, connect with an analyst to discuss your questions in a 30-minute phone session — or opt for a response via email.

Learn more.

Analyst Advisory

Translate research into action by working with an analyst on a specific engagement in the form of custom strategy sessions, workshops, or speeches.

Learn more.

Webinar

Join our online sessions on the latest research affecting your business. Each call includes analyst Q&A and slides and is available on-demand.

Learn more.

Forrester's research apps for iOS and Android.
Stay ahead of your competition no matter where you are.

Supplemental Material

Online Resource

We publish all our Forrester Wave scores and weightings in an Excel file that provides detailed product evaluations and customizable rankings; download this tool by clicking the link at the beginning of this report on Forrester.com. We intend these scores and default weightings to serve only as a starting point and encourage readers to adapt the weightings to fit their individual needs.

The Forrester Wave Methodology

A Forrester Wave is a guide for buyers considering their purchasing options in a technology marketplace. To offer an equitable process for all participants, Forrester follows The Forrester Wave™ Methodology Guide to evaluate participating vendors.

In our review, we conduct primary research to develop a list of vendors to consider for the evaluation. From that initial pool of vendors, we narrow our final list based on the inclusion criteria. We then gather details of product and strategy through a detailed questionnaire, demos/briefings, and customer reference surveys/interviews. We use those inputs, along with the analyst's experience and expertise in the marketplace, to score vendors, using a relative rating system that compares each vendor against the others in the evaluation.

We include the Forrester Wave publishing date (quarter and year) clearly in the title of each Forrester Wave report. We evaluated the vendors participating in this Forrester Wave using materials they provided to us by October 31, 2018, and did not allow additional information after that point. We encourage readers to evaluate how the market and vendor offerings change over time.

In accordance with The Forrester Wave™ Vendor Review Policy, Forrester asks vendors to review our findings prior to publishing to check for accuracy. Vendors marked as nonparticipating vendors in the Forrester Wave graphic met our defined inclusion criteria but declined to participate in or contributed only partially to the evaluation. We score these vendors in accordance with The Forrester Wave™ And The Forrester New Wave™ Nonparticipating And Incomplete Participation Vendor Policy and publish their positioning along with those of the participating vendors.

Integrity Policy

We conduct all our research, including Forrester Wave evaluations, in accordance with the Integrity Policy posted on our website.

Endnotes

[1] ETL: extract, transform, load.

[2] See the Forrester report "Now Tech: Hadoop/Spark Platforms, Q3 2018."

[3] IoT: internet of things.

[4] Source: "Cloudera and Hortonworks Complete Planned Merger," Cloudera press release, January 3, 2019 (https://www.cloudera.com/about/news-and-blogs/press-releases/2019-01-03-cloudera-and-hortonworks-complete-planned-merger.html).

[5] OLAP: online analytical processing.

[6] Source: "Cloudera and Hortonworks Complete Planned Merger," Cloudera press release, January 3, 2019 (https://www.cloudera.com/about/news-and-blogs/press-releases/2019-01-03-cloudera-and-hortonworks-complete-planned-merger.html).

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	**Technology Management Professionals**	**Technology Industry Professionals**
CMO	CIO	Analyst Relations
B2B Marketing	Application Development & Delivery	
B2C Marketing		
Customer Experience	› Enterprise Architecture	
Customer Insights	Infrastructure & Operations	
eBusiness & Channel Strategy	Security & Risk	
	Sourcing & Vendor Management	

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.