February 2, 2012

The Forrester Wave™: Enterprise Hadoop Solutions, Q1 2012

by James G. Kobielus
for Application Development & Delivery Professionals

February 2, 2012

The Forrester Wave™: Enterprise Hadoop Solutions, Q1 2012

Amazon Web Services, IBM, EMC Greenplum, MapR, Cloudera, And Hortonworks Lead This Emerging Market, With Seven Others Serving Key Niches Close Behind

by James G. Kobielus
with Stephen Powers, Brian Hopkins, Boris Evelson, and Shannon Coyne

EXECUTIVE SUMMARY

In Forrester's 15-criteria evaluation of enterprise Hadoop solution providers, we found that in the Leaders category, Amazon Web Services led the pack due to its proven, feature-rich Elastic MapReduce subscription service; IBM and EMC Greenplum offer Hadoop solutions within strong EDW portfolios; MapR and Cloudera impress with best-of-breed enterprise-grade distributions; and Hortonworks offers an impressive Hadoop professional services portfolio. Strong Performer Pentaho provides an impressive Hadoop data integration tool. Of the Contenders, DataStax provides a Hadoop platform for real-time, distributed, transactional deployments; Datameer has a user-friendly Hadoop/MapReduce modeling tool; Platform Computing and Zettaset offer best-of-breed Hadoop cluster management tools; and Outerthought has optimized its Hadoop platform for high-volume search and indexing. HStreaming is a Risky Bet with a solution that is strong in real-time Hadoop.

TABLE OF CONTENTS

NOTES & RESOURCES

Forrester conducted evaluations in August and September 2011 and interviewed 13 solution provider companies, including Amazon Web Services, Cloudera, Datameer, DataStax, EMC Greenplum, Hortonworks, HStreaming, IBM, MapR, Outerthought, Pentaho, Platform Computing, and Zettaset.

Related Research Documents

"Enterprise Hadoop: The Emerging Core Of Big Data"
October 20, 2011

"Case Study: Yahoo Pioneers Hadoop In Operational Big Data Programs"
October 19, 2011

"Enterprise Hadoop Best Practices: Concrete Guidelines From Early Adopters In Online Services"
October 11, 2011

HADOOP: THE OPEN SOURCE HEART OF BIG DATA

Big data is a core theme in the evolution of the enterprise data warehousing (EDW) and advanced analytics markets. A growing number of EDW vendors support such key big data features as shared-nothing massively parallel processing (MPP), petabyte scaling, and in-database analytics.[1]

Most enterprises have built big data initiatives on the tried-and-true approach of EDWs that support MPP. However, the cost, proprietary nature, inflexibility, and scalability issues of some MPP EDWs have spawned the development of an emerging open source, cloud-oriented approach known as Hadoop. Originating in the mid-2000s via technologies from Yahoo, Google, and other Web 2.0 pioneers, Hadoop is now central to the big data strategies of enterprises, service providers, and other organizations.

The Apache Hadoop community has spawned many promising startups and has resulted in new products and features from established vendors of MPP EDW platforms, data management tools, and business analytics. The Hadoop and EDW markets are rapidly converging as EDW vendors add Hadoop technologies to their solution portfolios and Hadoop tool vendors build tighter alliances with EDW providers. Providers of traditional EDW solutions see great advantages in adopting Hadoop for handling complex content, advanced analytics, and massively parallel in-database processing in the cloud.

New And Established Vendors Alike Are Getting Into The Hadoop Game

Forrester regards Hadoop as the nucleus of the next-generation EDW in the cloud. Hadoop implements the core features that are at the heart of most modern EDWs: cloud-facing architectures, MPP, in-database analytics, mixed workload management, and a hybrid storage layer. Essentially, application development and business process professionals should regard today's Hadoop market as the reinvention of the EDW for the new age of cloud-centric business models that require rapid execution of advanced, embedded analytics against big data.

Consistent with this trend, many EDW vendors, such as EMC Greenplum, IBM, Microsoft, and Oracle, are evolving their offerings to support Hadoop. Both incumbent EDW providers and startups now provide enterprise-grade distributions of Apache Hadoop that incorporate many of the core Hadoop subprojects with various proprietary revisions, extensions, and tools to add functionality, performance, high availability, security, and manageability (see Figure 1). Application development pros must consider the new approach to big data: the Apache Hadoop open source codebase and the commercial offerings that leverage and extend these technologies to help enterprises address critical business challenges that demand extreme scalability in EDW, advanced analytics, business intelligence (BI), online transaction processing (OLTP), and data integration.

Figure 1 Principal Hadoop Subprojects By Functional Layer

Functional layers	Hadoop subprojects
Hadoop modeling and development	MapReduce, Pig, Mahout
Hadoop storage and data management	HDFS, HBase, Cassandra
Hadoop data warehousing, summarization, and query	Hive, Sqoop
Hadoop data collection, aggregation, and analysis	Chukwa, Flume
Hadoop metadata, table, and schema management	HCatalog
Hadoop cluster management, job scheduling, and workflow	Zookeeper, Oozie, Ambari
Hadoop data serialization	Avro

60755   Source: Forrester Research, Inc.

Caution: The Hadoop Market Is Full Of Immature Offerings

Big data practitioners face considerable risk, uncertainty, and confusion, as the market for enterprise-grade Hadoop solutions is currently still immature. Established EDW and advanced analytics platforms vary in how well they support Hadoop in terms of scalability, functionality, flexibility, performance, and the price of commercial offerings. Some commercial EDW vendors, such as EMC Greenplum and IBM, have begun to integrate Apache Hadoop to enable cloud-based deployment of increasingly complex analytics against unstructured and real-time data. EDW vendors have begun to offer, develop, and preannounce big data solutions that incorporate Apache Hadoop technologies.

ENTERPRISE HADOOP SOLUTION EVALUATION OVERVIEW

To assess the state of the enterprise Hadoop solution market and see how the vendors stack up against each other, Forrester evaluated the strengths and weaknesses of top enterprise Hadoop solution providers.

Evaluation Criteria Focus On Present And Future Solutions

After examining past research and user needs assessments and conducting vendor and expert interviews, Forrester developed a comprehensive set of evaluation criteria. We evaluated vendors against 15 criteria, which we grouped into three high-level buckets:

- **Current offering.** To assess the breadth and depth of each vendor's enterprise Hadoop offering, we evaluated each Hadoop solution's architectural and operational functionality, including supported functionality, subproject integration, modeling, storage, acceleration and optimization, real-time and low latency, cluster management, packaging, distributed EDW file store connectors, and business applications.

- **Strategy.** We reviewed each vendor's strategy to assess how it plans to evolve its enterprise Hadoop solution to meet emerging customer demands. We also evaluated each vendor's strategic direction and professional services capabilities.

- **Market presence.** To establish each enterprise Hadoop solution's presence in this emerging market, we assessed each solution provider's Hadoop adoption, revenues, and partnerships.

Evaluated Vendors Must Meet Functional, Architectural, And Market Presence Criteria

Forrester included 13 enterprise Hadoop solution providers in this assessment: Amazon Web Services, Cloudera, Datameer, DataStax, EMC Greenplum, Hortonworks, HStreaming, IBM, MapR, Outerthought, Pentaho, Platform Computing (which announced during the research phase of the Wave that IBM is acquiring it), and Zettaset. Each of these solution providers has (see Figure 2):[2]

- **Core Hadoop functionality.** The chief criterion for inclusion in the Forrester Wave on the emerging market of enterprise Hadoop solutions is whether a vendor offers one or more generally available solutions — software, appliance, and/or cloud or software-as-a-service (SaaS) — that incorporate a MapReduce-enabled Hadoop distribution and/or MapReduce-enabled data integration layer.[3]

- **MapReduce as a mandatory subproject.** MapReduce is the only subproject universally implemented in Hadoop initiatives; thus, it is the only mandatory Hadoop subproject that a vendor must support to be considered for inclusion in the Wave. Forrester defines "MapReduce-enabled Hadoop distribution" as software that at the very least includes code from MapReduce, plus any or all of the other open source Apache Hadoop subprojects.[4]

- **At least two in-production customers.** The Hadoop solution provider must also provide at least two reference customers that have deployed the solution(s) in a production environment, which may include uses in testing and development, R&D, or operations.

There are several other vendors that are in the Hadoop market or have announced plans to enter the market but were not included in this Wave because they did not meet the above criteria.[5]

Figure 2 Evaluated Vendors: Product Information And Selection Criteria

Vendor	Product evaluated	Date evaluated
Amazon Web Services (AWS)	Amazon Elastic MapReduce	Q3 2011
Cloudera	Cloudera Distribution for Hadoop v. 3.x	Q3 2011
Datameer	Datameer Analytics Solution	Q3 2011
DataStax	DataStax Brisk	Q3 2011
EMC Greenplum	EMC Greenplum HD Enterprise Edition, EMC Greenplum HD Community Edition, EMC Greenplum HD Module	Q3 2011
Hortonworks	Hortonworks	Q3 2011
HStreaming	HStreaming Enterprise	Q3 2011
IBM	IBM InfoSphere BigInsights V1.2, Netezza Analytics	Q3 2011
MapR	MapR M3, MapR M5	Q3 2011
Outerthought	Lily	Q3 2011
Pentaho	Pentaho Data Integration 4.2	Q3 2011
Platform Computing	Platform MapReduce	Q3 2011
Zettaset	Zettaset Data Platform	Q3 2011

Vendor selection criteria

Vendors must offer one or more generally available solutions (software, appliance, and/or cloud/SaaS) incorporating a MapReduce-enabled Hadoop distribution and/or a MapReduce-enabled data integration layer as of August 2, 2011.

Vendors must support MapReduce as a mandatory Hadoop subproject.

Vendors must have at least two reference customers that have deployed its Hadoop solution(s) in a production environment, which may include uses in testing and development, R&D, and/or operations.

Source: Forrester Research, Inc.

THE RESULTS: ENTERPRISE HADOOP DISTRIBUTIONS DOMINATE THE MARKET

The evaluation uncovered a market in which (see Figure 3):

- **Amazon Web Services, IBM, EMC Greenplum, Cloudera, and Hortonworks are Leaders.** All of the Leaders have a strong Hadoop presence. Amazon leads the pack due to its proven, feature-rich Elastic MapReduce subscription service. IBM and EMC Greenplum offer Hadoop solutions within strong EDW portfolios. Cloudera and MapR impress with best-of-breed enterprise-grade distributions. And Hortonworks is building an impressive Hadoop professional services portfolio.

- **Pentaho is a Strong Performer with an impressive Hadoop data integration tool.** Among data integration vendors that have added Hadoop functionality to their products over the past year, it has the richest functionality and the most extensive integration with open source Apache Hadoop and with the Amazon, Cloudera, EMC Greenplum, MapR, and Hortonworks distributions of Hadoop.

- **DataStax, Datameer, Platform Computing, Zettaset, Outerthought, and HStreaming are Contenders.** DataStax provides a Hadoop platform for real-time, distributed, transactional deployments. Datameer offers a user-friendly Hadoop/MapReduce modeling tool. Platform Computing and Zettaset offer best-of-breed Hadoop cluster management tools. Outerthought has optimized its Hadoop platform for high-volume search and indexing.

- **HStreaming is a Risky Bet.** HStreaming is strong in real-time Hadoop and supports complex event processing (CEP). However, it lacks several key solution components — including a Hadoop modeling tool, an appliance or cloud/SaaS version, and business applications — and has a small professional services team.

This evaluation of the enterprise Hadoop solutions market is intended to be a starting point only. We encourage readers to view detailed product evaluations and adapt the criteria weightings to fit their individual needs through the Forrester Wave Excel-based vendor comparison tool.

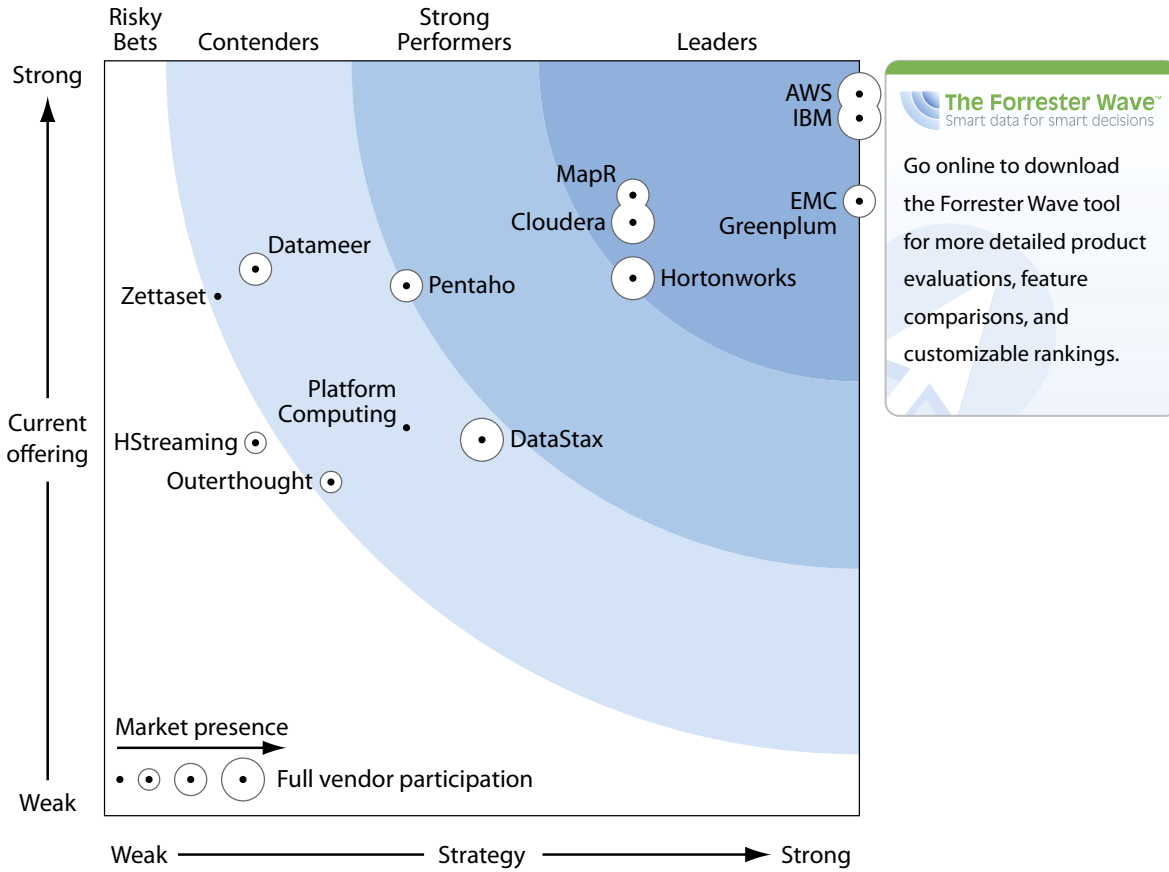**Figure 3** Forrester Wave™: Enterprise Hadoop Solutions, Q1 '12

Source: Forrester Research, Inc.

Figure 3 Forrester Wave™: Enterprise Hadoop Solutions, Q1 '12 (Cont.)

	Forrester's Weighting	AWS	Cloudera	Datameer	DataStax	EMC Greenplum	Hortonworks	HStreaming	IBM	MapR	Outerthought	Pentaho	Platform Computing	Zettaset
CURRENT OFFERING	50%	4.78	3.99	3.92	2.39	4.07	3.56	2.47	4.62	4.11	2.21	3.61	2.57	3.44
Functionality	35%	5.00	5.00	5.00	3.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	3.00	5.00
Subproject integration	35%	5.00	5.00	4.00	3.00	4.00	4.00	1.00	5.00	5.00	1.00	4.00	3.00	4.00
Modeling	2%	5.00	0.00	5.00	0.00	5.00	5.00	0.00	1.00	5.00	1.00	5.00	0.00	0.00
Storage	2%	5.00	3.00	5.00	4.00	4.00	2.00	5.00	3.00	5.00	0.00	4.00	3.00	1.00
Acceleration and optimization	2%	3.00	1.00	5.00	3.00	3.00	3.00	5.00	3.00	3.00	0.00	5.00	5.00	0.00
Real-time/low-latency data management	2%	5.00	3.00	0.00	5.00	3.00	3.00	5.00	1.00	5.00	1.00	3.00	3.00	1.00
Cluster management	5%	5.00	3.00	3.00	1.00	1.00	3.00	1.00	3.00	3.00	1.00	0.00	5.00	5.00
Packaging	2%	1.00	5.00	1.00	0.00	1.00	0.00	1.00	3.00	5.00	1.00	1.00	0.00	0.00
Distributed EDW file store connector	5%	3.00	0.00	0.00	0.00	5.00	0.00	0.00	5.00	0.00	0.00	0.00	0.00	0.00
Business applications	10%	5.00	1.00	3.00	0.00	3.00	0.00	0.00	5.00	0.00	0.00	1.00	0.00	0.00
STRATEGY	50%	5.00	3.50	1.00	2.50	5.00	3.50	1.00	5.00	3.50	1.50	2.00	2.00	0.75
Strategic direction	75%	5.00	3.00	1.00	3.00	5.00	3.00	1.00	5.00	3.00	1.00	1.00	1.00	1.00
Professional services	25%	5.00	5.00	1.00	1.00	5.00	5.00	1.00	5.00	5.00	3.00	5.00	5.00	0.00
MARKET PRESENCE	0%	5.00	5.00	3.00	4.80	3.20	4.80	2.60	5.00	3.20	2.50	3.20	1.80	1.20
Solution adoption	80%	5.00	5.00	3.00	5.00	3.00	5.00	3.00	5.00	3.00	3.00	3.00	1.00	1.00
Solution revenues	10%	5.00	5.00	3.00	5.00	5.00	5.00	1.00	5.00	5.00	1.00	5.00	5.00	3.00
Solution partners	10%	5.00	5.00	3.00	3.00	3.00	3.00	1.00	5.00	3.00	0.00	3.00	5.00	1.00

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

Source: Forrester Research, Inc.

VENDOR PROFILES

Leaders

- **Amazon is the most prominent Hadoop cloud service provider.** Amazon Web Services (AWS) is the principal cloud/SaaS service provider in the emerging Hadoop market; its subscription-based Elastic MapReduce (EMR) service has already achieved considerable adoption in both large and medium-size enterprises. For a wide range of users, AWS EMR is the principal on-ramp to an enterprise-grade Hadoop platform that requires no investment in internally deployed hardware or software (after a pure Apache Hadoop open source distribution, that is). AWS does not offer a Hadoop hardware appliance and has only unidirectional integration with third-party EDWs. Nevertheless, AWS has a wide range of partners in Hadoop data access/query, modeling and development, data integration, cluster management, and business applications.

· **IBM has the deepest Hadoop platform and application portfolio.** IBM, an established EDW vendor, has its own Hadoop distribution; an extensive professional services force working on Hadoop projects; extensive R&D programs developing Hadoop technologies; connections to Hadoop from its EDW products; a considerable number of Hadoop solutions and services customers; and software, appliance, and cloud offerings. IBM lacks a full Hadoop appliance, doesn't provide a visual MapReduce modeling tool, and hasn't fully integrated its CEP technologies into its Hadoop portfolio. Nevertheless, it is developing a full-featured Hadoop appliance product, has strong Hadoop modeling partnerships, and has recently tightened its CEP integration with InfoSphere Streams.

· **EMC Greenplum is the first mover in Hadoop appliances.** EMC Greenplum the first EDW vendor to provide a full-featured enterprise-grade Hadoop appliance and roll out an appliance family that integrates its Hadoop, EDW, and data integration in a single rack. It provides its own open source Hadoop distribution software, integrates EMC's strong storage product portfolio in its appliances, and has an extensive professional services force of EMC technical consultants and data scientists with Hadoop expertise. EMC Greenplum has no Hadoop-powered business applications or Hadoop data integration tools of its own and offers no cloud/SaaS Hadoop service for production applications — but it does have extensive Hadoop partnerships with complementary solution vendors.

· **MapR has a strong OEM business for its Hadoop distribution.** MapR has become one of the principal Hadoop distribution vendors, as attested by the wide range of solution providers that partner with the company and integrate its distribution into their solutions. MapR's flagship offerings incorporate one of the most feature-rich distributions of Apache Hadoop, plus the vendor's own proprietary software. MapR does not natively embed HDFS, lacks an EDW product of its own, does not provide Hadoop-based business applications, and lacks an internal Hadoop professional services and consulting team. Nevertheless, MapR supports the HDFS API, integrates with a broad range of EDWs, has strong Hadoop modeling tool partnerships, and has OEM partners that distribute the software under their own label and provide software and appliance form factors.

· **Cloudera is the Hadoop pure play with the greatest adoption.** Cloudera is a well-established startup among pure-play vendors in the emerging Hadoop solution market. Its core open source product, Cloudera's Distribution including Apache Hadoop (CDH), has strong uptake among many early adopters and is popular with cloud/SaaS providers that have built Hadoop-based services. Cloudera has software, cloud/SaaS, and appliance offerings, and a substantial and growing professional services force focused on Hadoop training and consulting. Cloudera offers no EDW of its own, provides no Hadoop modeling tools, and does not offer real-time/low-latency data integration. Nevertheless, Cloudera has strong partnerships with other technology vendors in most of the areas in which its own portfolio lacks an offering.

- **Hortonworks provides professional services to the Hadoop ecosystem.** Yahoo and Benchmark Capital created Hortonworks as a joint venture in June 2011. In addition to further developing the open source Apache Hadoop distribution, Hortonworks also provides Hadoop professional services to both vendors and users. It is the technology leader and ecosystem builder for the entire Hadoop industry and has recently released its Hortonworks Data Platform, which incorporates purely open-source Apache Hadoop software. However, Hortonworks provides no Hadoop modeling or development tools, does not work with other Hadoop database options such as Cassandra, offers no Hadoop business applications or library of MapReduce models, has no connectors to third-party EDWs or other non-Hadoop big data platforms, and has Hadoop training and professional services offerings that are still embryonic.

Strong Performers

- **Pentaho executes Hadoop MapReduce models and Pig scripts in its data integration product.** Pentaho, an established open source data analytics solution vendor, calls Hadoop-based extract, transform, and load (ETL) jobs from its Pentaho Data Integration (PDI) 4.2 and Pentaho Kettle products, can execute Pig scripts, and can run its data integration engine inside third-party Hadoop clusters. It has the richest functionality and most extensive integration with open source Apache Hadoop among those data integration vendors that have added Hadoop functionality to their products over the past year. It also provides certified integration with distributions from Amazon, Cloudera, and EMC Greenplum. Pentaho has no EDW of its own, no appliance or cloud/SaaS Hadoop offerings, no Hadoop cluster management tools, and no Hadoop business applications. But Pentaho has OEMed PDI 4.2 for embedding into numerous third-party business applications.

Contenders

- **DataStax embeds Cassandra for real-time Hadoop applications.** DataStax is a promising Hadoop startup with a differentiated focus on Apache Cassandra as a real-time distributed database for analytic and transactional applications. DataStax provided DataStax Brisk as its current Hadoop offering for evaluation in this Wave but released DataStax Enterprise as its next-generation offering during the Wave research phase. DataStax's primary differentiators include Cassandra, its cluster management capabilities, and its extensive Hadoop customer base. But DataStax lacks an EDW product of its own, Hadoop modeling tools, appliance or cloud/SaaS offerings, Hadoop business applications, and HBase support. Nevertheless, DataStax has a solid core of partners of complementary offerings in modeling, professional services, and other areas.

- **Datameer provides a user-friendly Hadoop modeling tool.** Datameer provides Hadoop modeling, data integration, and cluster management tools. Its Datameer Analytics Solution (DAS) ships with Apache Hadoop but supports and runs distributions from Amazon, Cloudera, DataStax, EMC Greenplum, IBM, and MapR. Datameer's chief differentiator is DAS' Hadoop modeling tool, which provides a browser-based spreadsheet user interface for self-service modeling, visualization, and data analytics tasks directly on data within Hadoop. However, Datameer lacks an EDW

product, a Hadoop appliance offering, and support for real-time data integration and processing and Hadoop business applications. It also has a small professional services team. Nevertheless, DAS ships with connectors to pull or push data from/to several third-party EDWs.

- **Platform Computing brings proven cluster management tools to Hadoop.** Platform Computing is an established vendor of cluster management tools for many high-performance computing platforms, including big data environments like Hadoop. Its Platform MapReduce product provides a feature-rich integrated cluster management across diverse third-party distributions. Platform Computing lacks a Hadoop distribution or an EDW of its own, Hadoop modeling tools or business applications, Hadoop appliance or cloud/SaaS offerings, and a wide range of Hadoop technology and solution partners. But Platform Computing's future parent, IBM, has offerings in many (but not all) of the Hadoop functional areas where this niche vendor is absent.

- **Zettaset specializes in Hadoop cluster management tools.** Zettaset, a startup cluster management tool vendor, focuses entirely on the Hadoop market, specifically on working tightly with all Hadoop distributions. Its Zettaset Data Platform incorporates the open source Apache Hadoop distribution as well as tools for ETL and data import and export. Zettaset's product also works with any distribution that incorporates the core Apache. But Zettaset has few customers or partners, provides no professional services, and lacks an EDW solution, graphing or charting tools, appliance or cloud packaging, Hadoop modeling tools, acceleration or low-latency tools, and business applications. However, Zettaset has an aggressive partnering program that will enable it to address a much wider range of Hadoop appliance and cloud opportunities.

- **Outerthought focuses on Hadoop search applications.** Outerthought, a promising Hadoop startup, offers a hybrid HBase/HDFS solution optimized for rich indexing and search applications. Outerthought incorporates cluster management tools that allow users to set up the product in both cloud (i.e., Amazon EC2) and on-premises setups and can integrate with existing managed Hadoop/HBase clusters based on the Cloudera CDH distribution. Outerthought has no EDW solution of its own; lacks support for Pig, Hive, Flume, Chukwa, and Mahout; has no appliance or cloud packaging; provides no Hadoop modeling, acceleration, or low-latency tools; and has no Hadoop business applications or solution partners.

Risky Bets

- **HStreaming provides complex event processing middleware for Hadoop.** HStreaming provides a CEP solution for real-time low-latency Hadoop applications. HStreaming Enterprise extends MapReduce/Pig's batch processing model with an additional continuous stream-processing mode where applications run continuously. HStreaming Enterprise is distribution-agnostic, comes in software and cloud editions, and includes Hadoop data integration, access/query, and dashboard visualization and exploration of Hadoop data. However, HStreaming lacks an EDW of its own, a modeling tool, an appliance version, and Hadoop business applications. It also has a small professional services team. Nevertheless, HStreaming is establishing partnerships to deliver its streaming Hadoop features to a larger market.

SUPPLEMENTAL MATERIAL

Online Resource

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

Data Sources Used In This Forrester Wave

Forrester used a combination of two 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.

· **Customer reference survey responses.** To validate product and vendor qualifications, Forrester also conducted elicited online survey feedback from 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.

ENDNOTES

[1] Over the next several years, cloud and software-as-a-service (SaaS) EDWs will gain greater adoption as a complement or outright replacement for appliance- and software-based EDWs. A growing number of EDW vendors now offer cloud/SaaS offerings in diversified portfolios that usually include one or more appliances. See the February 10, 2011, "The Forrester Wave™: Enterprise Data Warehousing Platforms, Q1 2011" report.

[2] Forrester regards any company that functions as a committer (i.e., contributing new, revised, and/or patch code to one or more Apache Hadoop subprojects) as being a vendor for possible inclusion in the Wave. Integration of an Apache Hadoop open source distribution with vendor-developed proprietary software does not disqualify the vendor's total Hadoop offering — open source plus commercial — from being evaluated in the Wave.

[3] Hadoop, as implemented by vendors and by users, at heart refers to various approaches for executing advanced analytic models defined under the Apache MapReduce specification across one or more nodes in a MPP architecture. Other features or components of a vendor's enterprise Hadoop solution portfolio may be incorporated into evaluation criteria but are not factored into the Wave inclusion criteria.

[4] Based on the list of Apache Hadoop open source subprojects on the Apache Hadoop website (http://hadoop.apache.org/).

Forrester defines "MapReduce-enabled data integration layer" as any offering that executes MapReduce models in addition to or in conjunction with its core function of performing ETL jobs on the bilateral movement of data between Hadoop clusters and/or between Hadoop clusters and other databases and repositories. Any data integration solution that lacks a MapReduce execution engine is not considered a MapReduce-enabled data integration layer.

[5] Forrester excluded several vendors from the Wave, due to not meeting all inclusion criteria, not having a generally available product as of August 2, 2011, or not having the minimum number of reference customers. These vendors include Composite Software, Concurrent, Dell, Endeca, Hadapt, HP/Vertica, Informatica, NetApp, Microsoft, Oracle, Pervasive, Platfora, Quest Software, RainStor, SAP/Sybase, SGI, and StackIQ. Karmasphere and Teradata/Aster Data were both invited to participate in this Wave but declined to do so.

FORRESTER®

Making Leaders Successful Every Day

Headquarters

Forrester Research, Inc.
60 Acorn Park Drive
Cambridge, MA 02140 USA
Tel: +1 617.613.6000
Fax: +1 617.613.5000
Email: forrester@forrester.com
Nasdaq symbol: FORR
www.forrester.com

Research and Sales Offices

Forrester has research centers and sales offices in more than 27 cities internationally, including Amsterdam, Netherlands; Beijing, China; Cambridge, Mass.; Dallas, Texas; Dubai, United Arab Emirates; Frankfurt, Germany; London, UK; New Delhi, India; San Francisco, Calif.; Sydney, Australia; Tel Aviv, Israel; and Toronto, Canada.

For the location of the Forrester office nearest you, please visit: www.forrester.com/locations.

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.

Forrester Research, Inc. (Nasdaq: FORR) is an independent research company that provides pragmatic and forward-thinking advice to global leaders in business and technology. Forrester works with professionals in 19 key roles at major companies providing proprietary research, customer insight, consulting, events, and peer-to-peer executive programs. For more than 28 years, Forrester has been making IT, marketing, and technology industry leaders successful every day. For more information, visit www.forrester.com.

FORRESTER®