

# The Forrester Wave™: Machine Learning Data Catalogs, Q2 2018

The 12 Providers That Matter Most And How They Stack Up

by Michele Goetz

June 21, 2018

## Why Read This Report

In our 29-criteria evaluation of machine learning data catalogs (MLDCs) providers, we identified the 12 most significant ones — Alation, Cambridge Semantics, Cloudera, Collibra, Hortonworks, IBM, Infogix, Informatica, Oracle, Reltio, Unifi Software, and Waterline Data — and researched, analyzed, and scored them. This report shows how each provider measures up and helps enterprise architecture (EA) professionals make the right choice.

## Key Takeaways

### **IBM, Reltio, Unifi Software, Alation, And Collibra Lead The Pack**

Forrester's research uncovered a market in which IBM, Reltio, Unifi Software, Alation, and Collibra are Leaders; Informatica, Oracle, Waterline Data, Infogix, Cambridge Semantics, and Cloudera are Strong Performers; and Hortonworks is a Contender.

### **Data Pros Are Looking For Data**

#### **Understanding That Everyone Can Access**

The MLDC market is growing because firms want to scale data to the masses through self-service. However, back-end data management technology can't support tribal knowledge, provide a good user experience (UX) for data consumers, and scale across a highly federated data ecosystem. MLDCs solve this and scale elastically by leveraging their machine learning (ML) capabilities.

### **Machine Learning, Collaboration, And Activation Are Key Differentiators**

Combining ML with collaboration and activation scales out data understanding and speeds up use. Thus, MLDCs are demonstrating ROI in many cases within four weeks. Additionally, ML provides certain insights from data as part of its analytic process that previously required an analyst to see facts, trends, and causal relationships.

# The Forrester Wave™: Machine Learning Data Catalogs, Q2 2018

## The 12 Providers That Matter Most And How They Stack Up



by [Michele Goetz](#)  
with [Gene Leganza](#), Elizabeth Hoberman, and Kara Hartig  
June 21, 2018

### Table Of Contents

- 2 MLDCs Are The Stepping Stone For The Intelligent Business
- 4 Machine Learning Data Catalogs Evaluation Overview
  - Evaluated Vendors And Inclusion Criteria
- 6 Vendor Profiles
  - Leaders
  - Strong Performers
  - Contenders
- 14 Supplemental Material

### Related Research Documents

- [Data Engineers Have Become More Important Than Data Scientists](#)
- [Data Stewards Are Set Up To Fail](#)
- [Now Tech: Machine Learning Data Catalogs, Q1 2018](#)
- [The Top 10 Technology Trends To Watch: 2018 To 2020](#)



**Share reports with colleagues.**  
[Enhance your membership with Research Share.](#)

**The Forrester Wave™: Machine Learning Data Catalogs, Q2 2018**

The 12 Providers That Matter Most And How They Stack Up

## MLDCs Are The Stepping Stone For The Intelligent Business

The four V's of big data (i.e., volume, variety, velocity, and veracity) may be a cliché. But firms are still struggling under the weight of their data: 36% to 38% of global data and analytics decision makers reported that their structured, semistructured, and unstructured data each totaled 1,000 TB or more in 2017, up from only 10% to 14% in 2016 (see Figure 1). And the growth of data is outpacing organizations' ability to get value from it. The two biggest challenges our respondents reported in using systems of insight were 1) merging existing business processes to source data to analyze it and implement insights and 2) sourcing, gathering, managing, and governing the data as it grows.

For EA professionals, relying on people and manual processes to provision, manage, and govern data simply does not scale. Enterprises are waking up to this fact and turning to data catalogs to democratize access to data, enable tribal data knowledge to curate information, apply data policies, and activate all data for business value quickly. Data catalog investment links to:

- › **The number of data lakes.** According to global data and analytics decision makers at organizations investing in, implementing, or expanding their data catalogs, they have more than seven data lakes across their enterprise.<sup>1</sup>
- › **Prioritization of insights.** In addition, 51% of our survey respondents at organizations that are expanding/upgrading their data catalog implementations said that leveraging big data and analytics in decision making was a critical priority.<sup>2</sup>
- › **Competitive advantage of AI.** Those expanding their data catalog implementations are more likely to mention the use of AI for product testing and innovation (new value-based offerings) rather than traditional customer experience and operational business scenarios.<sup>3</sup>

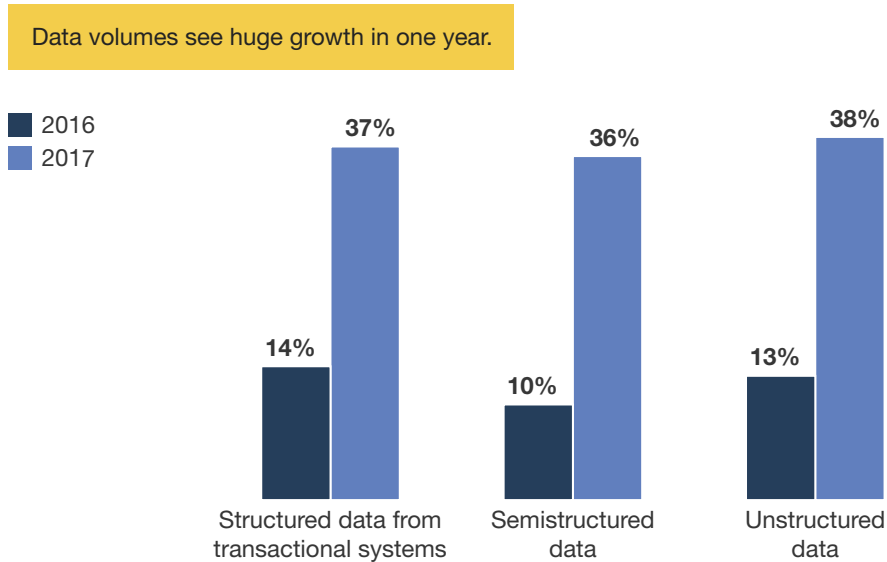
**The Forrester Wave™: Machine Learning Data Catalogs, Q2 2018**

The 12 Providers That Matter Most And How They Stack Up

**FIGURE 1** Data Catalogs Solve Big Data Challenges And Guide Business Insights Strategy

**“Using your best estimate, about how much data is currently stored within your company?”**

(% with 1,000 TB+)



Base: 1,246 (2017) and 1,244 (2016) global data and analytics technology decision makers at enterprises (1,000+ employees)

Source: Forrester Analytics Global Business Technographics® Data And Analytics Surveys, 2016 and 2017

**“What [could be/are] the biggest challenges for your firm in using systems of insight?”**

Organizations are struggling to keep up.

**1 40%**  
merging existing business processes to source data to analyze it and implement insights

**2 39%**  
sourcing, gathering, managing, and governing the data as it grows

Base: 1,733 global data and analytics decision makers at enterprises (1,000+ employees) that are interested in, planning to, implementing, or expanding their use of systems of insight

Source: Forrester Analytics Global Business Technographics Data And Analytics Survey, 2017

**The Forrester Wave™: Machine Learning Data Catalogs, Q2 2018**

The 12 Providers That Matter Most And How They Stack Up

## Machine Learning Data Catalogs Evaluation Overview

To assess the state of the MLDC market and see how the vendors stack up against each other, Forrester evaluated the strengths and weaknesses of top MLDC vendors. After examining past research, user need assessments, and vendor and expert interviews, we developed a comprehensive set of evaluation criteria. We evaluated vendors against 29 criteria, which we grouped into three high-level buckets:

- › **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 include profiling and analytics, stewardship and governance, cataloging, collaboration, data activation and integration, ML, and security.
- › **Strategy.** Placement on the horizontal axis indicates the strength of the vendors' strategies. We evaluated planned enhancements, ML teams, market alignment, product strategy, and research and development.
- › **Market presence.** Represented by the size of the bubbles on the Forrester Wave graphic, our market presence scores reflect each vendor's regional coverage and growth.

### Evaluated Vendors And Inclusion Criteria

Forrester included 12 vendors in the assessment: Alation, Cambridge Semantics, Cloudera, Collibra, Hortonworks, IBM, Infogix, Informatica, Oracle, Reltio, Unifi Software, and Waterline Data. Each of these vendors has (see Figure 2):

- › **An MLDC product that has been in the market for more than four months.** The evaluated version was generally available by April 12, 2018.
- › **Support for data engineer use cases.** The catalog is positioned as a solution, not a feature.
- › **A product renewal rate of more than 85%.** The rate was more than 85% during the past two years.
- › **Demonstrated machine learning and innovations in the category.** The vendor offers machine learning in at least two of three analytics cataloging capabilities in the "Now Tech: Machine Learning Data Catalogs, Q1 2018" Forrester report.<sup>4</sup>
- › **More than 15 customers and a customer base in multiple regions.** The vendor has these customer minimums.
- › **Consistent appearances in Forrester inquiries.** These inquiries include questions from clients about purchases and competitive situations.

**The Forrester Wave™: Machine Learning Data Catalogs, Q2 2018**

The 12 Providers That Matter Most And How They Stack Up

**FIGURE 2** Evaluated Vendors: Product Information And Inclusion Criteria

Vendor	Product evaluated	Product version evaluated
Alation	Alation Data Catalog	5.0
Cambridge Semantics	Anzo Smart Data Lake	4.0
Cloudera	Cloudera Enterprise	5.14
Collibra	Data Governance Center	5.3
Hortonworks	Hortonworks Data Platform (HDP) Hortonworks DataPlane Service (DPS) Hortonworks Data Steward Studio (DSS)	2.6.4 1.0 1.0
IBM	IBM Watson Knowledge Catalog	Continuous release
Infogix	Infogix Data3Sixty	2.1.1
Informatica	Enterprise Data Catalog	As of April 12, 2018
Oracle	Oracle Metadata Management Oracle Data Relationship Management Suite Oracle Enterprise Data Management Cloud	12.2.1.2 11.1.2.x 18.03
Reltio	Reltio Cloud	2018.1
Unifi Software	Unifi Data Platform	2.6
Waterline Data	Waterline Data Catalog	4.0

**The Forrester Wave™: Machine Learning Data Catalogs, Q2 2018**

The 12 Providers That Matter Most And How They Stack Up

**FIGURE 2** Evaluated Vendors: Product Information And Inclusion Criteria (Cont.)**Vendor inclusion criteria**

**An MLDC product that has been in the market for more than four months.** The evaluated version was generally available by April 12, 2018.

**Support for data engineer use cases.** The catalog is positioned as a solution, not a feature.

**A product renewal rate of more than 85%.** The rate was more than 85% during the past two years.

**Demonstrated machine learning and innovations in the category.** The vendor offers machine learning in at least two of three analytics cataloging capabilities in the “Now Tech: Machine Learning Data Catalogs, Q1 2018” Forrester report.

**More than 15 customers and a customer base in multiple regions.** The vendor has these customer minimums.

**Consistent appearances in Forrester inquiries.** These inquiries include questions from clients about purchases and competitive situations.

## Vendor Profiles

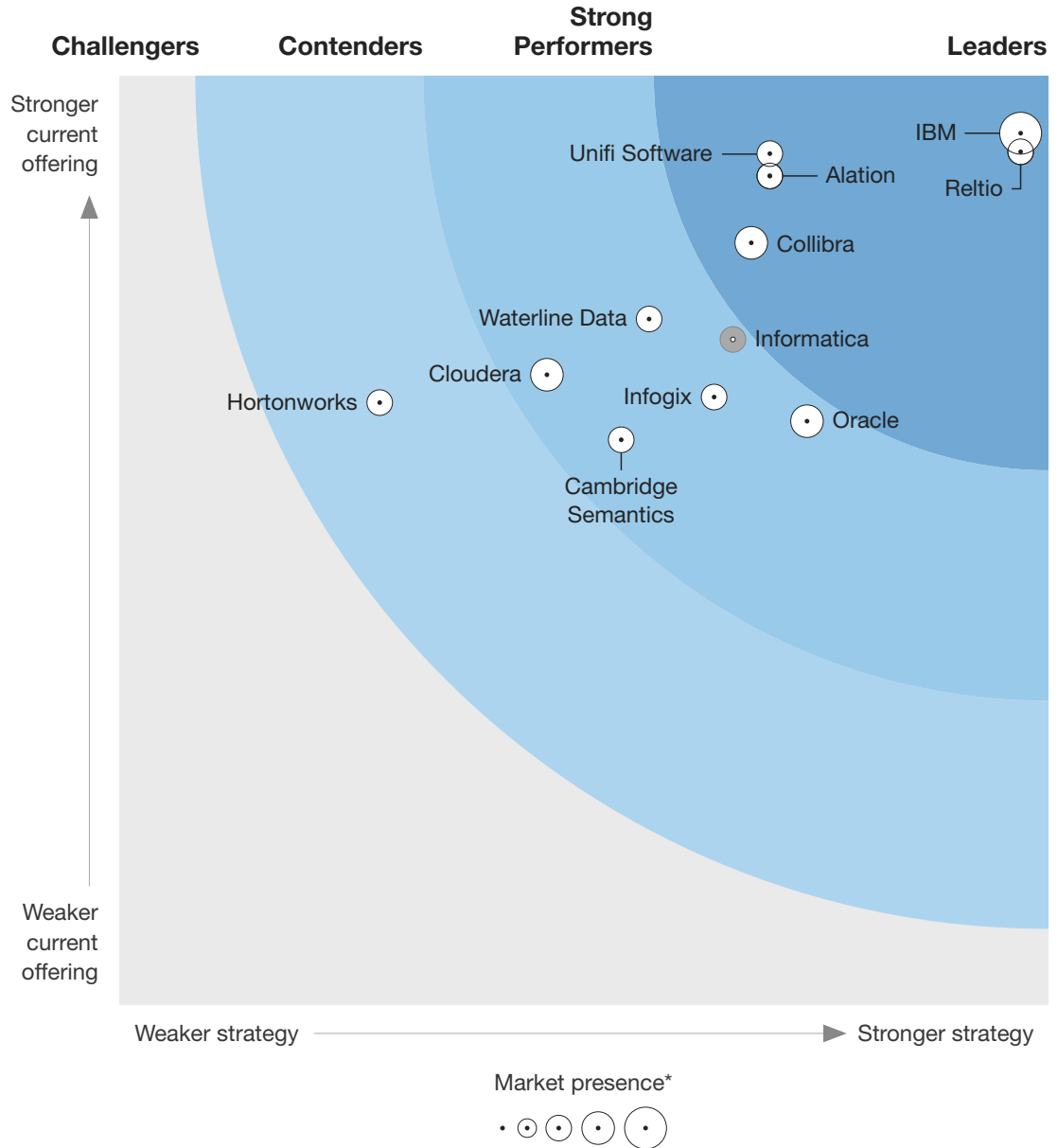
This evaluation of the MLDC market is intended to be a starting point only. We encourage clients to view detailed product evaluations and adapt criteria weightings to fit their individual needs through the Forrester Wave Excel-based vendor comparison tool (see Figure 3 and see Figure 4). Click the link at the beginning of this report on Forrester.com to download the tool.

**The Forrester Wave™: Machine Learning Data Catalogs, Q2 2018**

The 12 Providers That Matter Most And How They Stack Up

**FIGURE 3** Forrester Wave™: Machine Learning Data Catalogs, Q2 2018

**THE FORRESTER WAVE™**  
**Machine Learning Data Catalogs**  
 Q2 2018



\*A gray bubble indicates a nonparticipating vendor.



**The Forrester Wave™: Machine Learning Data Catalogs, Q2 2018**

The 12 Providers That Matter Most And How They Stack Up

**FIGURE 4** Forrester Wave™: Machine Learning Data Catalogs Scorecard, Q2 2018

	Forrester's weighting	Alation	Cambridge Semantics	Cloudera	Collibra	Hortonworks	IBM
<b>Current offering</b>	50%	4.46	3.04	3.39	4.10	3.24	4.69
Profiling and analytics	15%	4.40	3.20	3.00	3.20	2.60	5.00
Cataloging	20%	4.60	3.80	3.90	3.80	3.30	4.50
Stewardship and governance	5%	4.00	3.50	2.50	5.00	3.00	5.00
Collaboration	20%	5.00	3.00	3.50	4.50	3.00	4.50
Data activation and integration	10%	3.40	3.50	2.70	3.70	2.70	3.90
Machine learning (ML)	25%	4.34	1.68	3.66	4.34	3.66	5.00
Security	5%	5.00	5.00	3.00	5.00	5.00	5.00
<b>Strategy</b>	50%	3.50	2.70	2.30	3.40	1.40	4.85
Planned enhancements	30%	3.00	3.00	1.00	5.00	1.00	5.00
ML team	20%	3.00	1.00	5.00	1.00	1.00	5.00
Market alignment	15%	3.00	5.00	3.00	3.00	1.00	4.00
Product strategy	25%	5.00	3.00	1.00	3.00	1.00	5.00
Research and development	10%	3.00	1.00	3.00	5.00	5.00	5.00
<b>Market presence</b>	0%	3.00	3.00	4.00	4.00	3.00	5.00
Regional coverage	50%	1.00	1.00	5.00	5.00	3.00	5.00
Growth	50%	5.00	5.00	3.00	3.00	3.00	5.00

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

**The Forrester Wave™: Machine Learning Data Catalogs, Q2 2018**

The 12 Providers That Matter Most And How They Stack Up

**FIGURE 4** Forrester Wave™: Machine Learning Data Catalogs Scorecard, Q2 2018 (Cont.)

	Forrester's weighting	Infogix	Informatica*	Oracle	Reltío	Unifi Software	Waterline Data
<b>Current offering</b>	50%	3.27	3.58	3.14	4.59	4.58	3.69
Profiling and analytics	15%	2.60	3.20	3.40	4.20	5.00	3.00
Cataloging	20%	2.70	3.30	3.30	4.10	3.90	3.50
Stewardship and governance	5%	3.70	3.10	3.70	5.00	5.00	4.50
Collaboration	20%	4.50	3.00	2.50	5.00	5.00	4.50
Data activation and integration	10%	3.50	3.50	3.80	3.90	4.60	3.40
Machine learning (ML)	25%	3.00	4.34	3.00	5.00	4.34	3.68
Security	5%	3.00	5.00	3.00	5.00	5.00	3.00
<b>Strategy</b>	50%	3.20	3.30	3.70	4.85	3.50	2.85
Planned enhancements	30%	3.00	3.00	3.00	5.00	5.00	3.00
ML team	20%	3.00	1.00	3.00	5.00	1.00	3.00
Market alignment	15%	3.00	3.00	3.00	4.00	3.00	2.00
Product strategy	25%	3.00	5.00	5.00	5.00	5.00	3.00
Research and development	10%	5.00	5.00	5.00	5.00	1.00	3.00
<b>Market presence</b>	0%	3.00	3.00	4.00	3.00	3.00	3.00
Regional coverage	50%	5.00	5.00	5.00	1.00	1.00	3.00
Growth	50%	1.00	1.00	3.00	5.00	5.00	3.00

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

\*Informatica declined to participate in or provide information for our research. Scores are based on Forrester estimates.

**Leaders**

- › **IBM reimagines data.** Launched in 2018, the Watson Knowledge Catalog disguises its ML and cataloging power behind a simple role-based workspace. Revisiting its traditional data management and governance approach to enablement, IBM designed its MLDC from the ground up around role intent and behavior, with ML at the core and the ability to tap into Watson APIs. The UI lets roles work the way they want to and not reorient their data sourcing, stewardship, or

**The Forrester Wave™: Machine Learning Data Catalogs, Q2 2018**

## The 12 Providers That Matter Most And How They Stack Up

administrative processes to match another role's workspace. Data search builds confidence in data by allowing social sharing, ratings and communication, usage metrics, drill-down to sample views, and ML for classification and tagging.

Customers should be aware that full lineage analysis won't be available until the summer of 2018 (shortly after this report release), that the enterprise version only just came to market, and that customer deployments are still ramping up. For enterprises ready to scale their data strategies, IBM provides the environment that makes this possible.

- › **Reltio challenges data assumptions and then innovates data.** The only master data management (MDM) vendor in this evaluation, Reltio continues to show that labels don't always tell the full story. By building an MDM capability on ML, graph, big data, metadata, and services, Reltio was a data catalog all along. Data engineers and stewards will be comfortable in the environment and can continue to take advantage of the self-data linking and curation, workflows, rich profiling, and actionable charts. Social capabilities behaviorally inform ML capabilities. Orienting to Reltio as a catalog takes getting used to and is thus a potential barrier to consider.

Customers indicate that the environment is fast to deploy, allowing users to see data in one place. Customers also see benefit in a convergence of MDM, metadata management, and reference data management into a single capability rather than three. Enterprises that are ready to simplify their data management stack and get closer to the architecture principles that the internet giants use will find Reltio to be a good starting point.

- › **Unifi brings insight where none existed before.** Data users are immediately struck by the Unifi environment's simplicity: a single search field, designed around the same concept as commercial search engines, so that user intent is center stage for immediate insights. Search is through natural language where users can not only find data they need but also ask questions of the data such as: "What was my 2018 revenue?" or, "What was the revenue trend from the past three years?" Most data preparation vendors see catalogs as a function; Unifi treats it as a core differentiator.

Customers can't stop talking about the ease of deployment, the ability of nontechnical and executive stakeholders to understand and get value from the tool, and the ability to govern data more holistically, compared with other stewardship and data prep environments. While the solution is powerful and prominent investors back Unifi, the data science workbench is still small, and investment is still primarily going to product improvements. But that shouldn't deter enterprises from putting Unifi on a shortlist because it shows how a consolidated offering for data prep, cataloging, and self-service can democratize data and speed time-to-insight.

- › **Alation started the MLDC trend.** In 2012, Alation wanted to re-envision metadata management and governance and be first to market with an ML data catalog. Today, it provides deep data introspection with its behavioral I/O analysis of data use and queries. Its strategy is to stay true to data cataloging and leverage point solution and platform partnerships, such as with Trifacta and Paxata for anomaly detection, to further extend its footprint. But in a market that is rapidly moving toward platform incumbents, there will be pressure on Alation to grow beyond its point solution footprint.

**The Forrester Wave™: Machine Learning Data Catalogs, Q2 2018**

## The 12 Providers That Matter Most And How They Stack Up

Customer satisfaction is currently high, demonstrating fast time-to-value. One reference customer said, “The product has been a huge time saver, cutting query writing by two-thirds and helping analysts easily find the data they need,” while another indicated that the product is easy to use by nontechnical employees. This will help Alation support data governance efforts for areas such as the EU General Data Protection Regulation (GDPR) and personally identifiable information (PII) detection. Clients considering Alation will find a strong MLDC to help understand and make sense of the vast sources that exist on-premises, in the cloud, and across their legacy and modern systems.

- › **Collibra steps further into data management.** Known for its data governance prowess, Collibra now goes beyond semantic metadata management, business glossary functionality, and stewardship cockpits. Adding support for systems and logical metadata into its core catalog, Collibra better manages data models, schemas, classification, tagging, and certification. Adding a data shopping environment makes Collibra a hub for data democratization and activation. Collibra’s innovative alliance with the University of Brussels taps into data management and data science expertise. Forrester believes that Collibra still has work to do to differentiate itself from other strong MLDC competitors and provide a more extensive capability in the enterprise data ecosystem.

Customers provide strong reviews on a business user and analyst friendly environment to search, understand, and gather data. Collaboration is at the core of the product through social mechanisms, workflows, and usage analytics. Enterprises seeking to turn on self-service data sourcing but still maintain oversight will find that Collibra fits the bill.

**Strong Performers**

- › **Informatica revives existing data investments.** Informatica declined to participate in or provide information for our research. Scores are based on Forrester estimates. Informatica recently entered the MLDC market with its Enterprise Data Catalog (EDC). Going beyond its metadata management and business glossary capabilities, Informatica has now evolved its linked data and graph-based prototypes into an environment that is intelligent, democratized, and user friendly. EDC blends the search-like experience for data with metadata and glossary capabilities that data stewards and engineers have come to expect from their information management solutions.

Clients we speak with would like to see more governance and stewardship support rather than having to invest in another Informatica tool such as Axon Data Governance. The real potential of EDC is the ability to integrate and take advantage of the wider Intelligent Data Platform. While it’s still early to assess EDC in customer deployments, firms that want to take advantage of ML capabilities, more data consumer friendly environments, and wide connectivity options for interoperability and extend their cloud strategies for the logical data system will be well served with Informatica’s EDC.

- › **Oracle powers up data management.** Oracle is a quiet giant when it comes to data management. Don’t be fooled by the legacy name of its data catalog: Oracle Enterprise Metadata Management (OEMM). One of the most compelling aspects of OEMM is the recognition that metadata and models

**The Forrester Wave™: Machine Learning Data Catalogs, Q2 2018**

## The 12 Providers That Matter Most And How They Stack Up

exist beyond data sources and live within extract, transform, and load (ETL) and open source environments (e.g., Kafka). The ability to bring these data models into the catalog gives a level of data visibility that most modern and traditional tools have lacked. Currently, Oracle uses public and open source models. Moving forward, it can customize and extend models through its acquisition of DataScience.com. The environment locks data down (data stays encrypted) so that users can understand data elements but can't view sample data. Data shopping is not yet available.

One customer described how Oracle's data catalog helps with new processes, and it "can turn things around the same business day, which is a huge win." When paired with Oracle's data preparation tool, the environment becomes suggestive. Enterprises will find that Oracle's data catalog is a solid investment to manage data they have within and beyond Oracle systems.

- › **Waterline Data keeps back the big data swamp monster.** As firms began adopting data lakes, Waterline recognized the need to navigate across the ecosystem that fed the lake and automatically inventory and tag raw data with familiar business terms. Waterline's data catalog provides deep profiling of the data and incorporates tribal knowledge that connects system, logical, and semantic insights about data, its lineage, models, and fitness for purpose. Deployment of Waterline Data's MLDC takes longer than average at three to four months. Partnerships are extensive and strong including some seemingly strange bedfellows with other catalog vendors. This allows Waterline customers to synchronize metadata with preexisting or embedded catalogs that may have overlapping functionality such as quality scores, security and access controls, and workflow/task management.

Customers appreciate the ability to support self-service and data democratization to citizen analysts, with one reference indicating that Waterline Data supported up to 800 users within the organization. Enterprises in need of a capability to de-swamp their data lakes, find and secure sensitive data, and prepare their data migrations to new platforms will find a good solution in Waterline Data.

- › **Infogix moves from data auditor to data activator.** Infogix began with a robust tool to audit data against governance policies, but with the acquisition of Lavastorm, Infogix with Data3Sixty is now a complete stewardship, quality, and cataloging solution. The environment is intuitive and business-oriented for data stewards and data management teams to understand the conditions of the data and create data flows. There is still room for improvement to help data consumers find data through a marketplace experience, and expansion of ML for data models could go further than data mapping to broader schema inference. But ML is an engineering priority, and Infogix has a dedicated data science workbench.

Reference customers indicate high levels of satisfaction with the product. One customer described choosing Infogix over a large data management vendor because there was no need to "shoot a mosquito with a shotgun." Another indicated that Infogix understood the difference between governing data and enabling data. Where data governance and stewardship are bottlenecks to democratizing data, Infogix provides the oversight for firms shifting toward scale and self-service.

**The Forrester Wave™: Machine Learning Data Catalogs, Q2 2018**

The 12 Providers That Matter Most And How They Stack Up

- › **Cambridge Semantics brings context to data for insight.** Cambridge Semantics' cohesive platform of semantic tools, text analytics, cataloging, and insight capabilities on a big data foundation helps enterprises across multiple industries overcome challenges of interpreting and standardizing complex data. ML is embedded in the platform to alleviate data science efforts of customers. However, Cambridge Semantics can improve its ML results by augmenting its engineers with data scientists. Customers overall speak highly about the product capabilities but look forward to refinements. Shifting investments to post-sale customer success and sales and marketing will optimize the road map for product refinement while growing the business. Implementation and time-to-value vary.

One customer found it quite easy to implement and highlighted the ontology builder that allows small and medium-size enterprises to inform semantics, data management, and insight. However, another customer indicated some challenges due to the product's early maturity. Enterprises seeking to gain insight from structured and semistructured data will find Cambridge Semantics to be one of the only integrated platforms in the market that supports holistic information management and analytics.

- › **Cloudera makes sense of the data within data lakes.** Cloudera introduced Navigator to capture, catalog, and update files in its ML and analytics platform. Cloudera offers advanced cataloging with sophisticated ML capabilities to understand, classify, and catalog data ingested into the data lake. The environment provides the right foundation to catalog and search data at scale, but there is an assumption that the data consumer will have expertise in structured query language (SQL) and database environments. However, Cloudera has one of the largest data science workbenches of all vendors in this evaluation and a product road map oriented to extending ML for metadata capture and data management. So Cloudera has the potential to take on standalone MLDC vendors for data science and analytic scenarios in their environment.

Customers appreciate Cloudera's cataloging capabilities but indicate there is room for more fine-grained management to support real-time user management on the cluster. Enterprises that see Cloudera as their strategic big data partner will see the data cataloging capabilities as a good foundation to keep their data lake from turning into a swamp.

## Contenders

- › **Hortonworks knows your data.** The Data Steward Studio helps organizations understand data within the Hortonworks ecosystem through extensive metadata capture about data, data models, and schemas from source systems at the file, table, and column level. Classifications and ongoing changes are automated in ingestion processes from connected source systems. Hortonworks' capabilities span from stewardship of data policies and administrative capabilities to manage the clusters to maintenance of connectivity and data science efforts through Zepplin integration.

The inability to view data to understand the insights about assets, policies, and lineage may challenge customers. Reference customers also indicate that the product takes a bit more time

**The Forrester Wave™: Machine Learning Data Catalogs, Q2 2018**

The 12 Providers That Matter Most And How They Stack Up

to install and keep running and stable and requires a higher-level skill set due to its open source nature. But customers were willing to make this tradeoff for open source. Customers with open source strategies will find Hortonworks provides sufficient cataloging capabilities to meet their data engineering and data science needs.

## 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

The online version of Figure 3 is an Excel-based vendor comparison tool that provides detailed product evaluations and customizable rankings. Click the link at the beginning of this report on Forrester.com to download the tool.

**The Forrester Wave™: Machine Learning Data Catalogs, Q2 2018**

The 12 Providers That Matter Most And How They Stack Up

**Data Sources Used In This Forrester Wave**

Forrester used a combination of three data sources to assess the strengths and weaknesses of each solution. We evaluated the vendors participating in this Forrester Wave, in part, using materials that they provided to us by June 11, 2018.

- › **Vendor surveys.** Forrester surveyed vendors on their capabilities as they relate to the evaluation criteria.
- › **Product briefings and demos.** We asked vendors to conduct product briefings and demonstrations of their products' functionality. We used findings from these 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 at least two of each participating vendor's current customers.

**The Forrester Wave Methodology**

We conduct primary research to develop a list of vendors that meet our criteria for evaluation in this market. From that initial pool of vendors, we 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. Vendors marked as incomplete participants met our defined inclusion criteria but declined to participate or contributed only partially to the 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 evaluation — and then score the vendors based on a clearly defined scale. We intend these default weightings to serve only as a starting point and 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, please visit [The Forrester Wave™ Methodology Guide](#) on our website.

**Integrity Policy**

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



**The Forrester Wave™: Machine Learning Data Catalogs, Q2 2018**

The 12 Providers That Matter Most And How They Stack Up

## Survey Methodology

The Forrester Analytics Global Business Technographics® Data And Analytics Survey, 2017, was fielded between February and April 2017. This online survey included 3,378 respondents in Australia, Brazil, Canada, China, France, Germany, India, New Zealand, the UK, and the US from companies with 100 or more employees.

Forrester Analytics Global Business Technographics Data And Analytics Survey, 2016, was fielded in March 2016. This online survey included 3,343 respondents in Australia, Brazil, Canada, China, France, Germany, India, New Zealand, the UK, and the US from companies with 100 or more employees.

Forrester Analytics Business Technographics ensures that the final survey population contains only those with significant involvement in the planning, funding, and purchasing of business and technology products and services. Research Now fielded these surveys on behalf of Forrester. Survey respondent incentives include points redeemable for gift certificates.

Please note that the brand questions included in these surveys should not be used to measure market share. The purpose of Forrester Analytics Business Technographics brand questions is to show usage of a brand by a specific target audience at one point in time.

## Endnotes

<sup>1</sup> Source: Forrester Analytics Global Business Technographics Data And Analytics Survey, 2017.

<sup>2</sup> Source: Forrester Analytics Global Business Technographics Data And Analytics Survey, 2017.

<sup>3</sup> We asked global data and analytics decision makers whose firms are planning to use or are currently using AI technologies which use cases/application scenarios their firms use those AI technologies for: Those expanding or upgrading implementation of their data catalogs are more likely to use AI to test new products (40%) and innovate product design and development (35%) than to create and deliver a better customer experience (29%) or to improve efficiencies in business operations (29%). Source: Forrester Analytics Global Business Technographics Data And Analytics Survey, 2017.

<sup>4</sup> See the Forrester report "[Now Tech: Machine Learning Data Catalogs, Q1 2018.](#)"

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](mailto:clientsupport@forrester.com). We offer quantity discounts and special pricing for academic and nonprofit institutions.