

Magic Quadrant for Data Science Platforms

Published: 14 February 2017 **ID:** G00301536

Analyst(s): Alexander Linden, Peter Krensky, Jim Hare, Carlie J. Idoine, Svetlana Sicular, Shubhangi Vashisth

Data science platforms are engines for creating machine-learning solutions. Innovation in this market focuses on cloud, Apache Spark, automation, collaboration and artificial-intelligence capabilities. We evaluate 16 vendors to help you make the best choice for your organization.

Market Definition/Description

This Magic Quadrant evaluates vendors of data science platforms. These are products that organizations use to build machine-learning solutions themselves, as opposed to outsourcing their creation or buying ready-made solutions (see "Machine-Learning and Data Science Solutions: Build, Buy or Outsource?").

There are countless tasks for which organizations prefer this approach, especially when "good enough" packaged applications, APIs and SaaS solutions do not yet exist. Examples are numerous. They include demand prediction, failure prediction, determination of customers' propensity to buy or churn, and fraud detection.

Gartner previously called these platforms "advanced analytics platforms" (as in the preceding "Magic Quadrant for Advanced Analytics Platforms"). Recently, however, the term "advanced analytics" has fallen somewhat out of favor as many vendors have added "data science" to their marketing narratives. This is one reason why we now call this category "data science platforms," but it is not the main reason. Our chief reason is that it is commonly "data scientists" who use these platforms.

We define a data science platform as:

A cohesive software application that offers a mixture of basic building blocks essential for creating all kinds of data science solution, and for incorporating those solutions into business processes, surrounding infrastructure and products.

"Cohesive" means that the application's basic building blocks are well-integrated into a single platform, that they provide a consistent "look and feel," and that the modules are reasonably interoperable in support of an analytics pipeline. An application that is not cohesive — that mostly

uses or bundles various packages and libraries — is not considered a data science platform, according to our definition.

Readers of this Magic Quadrant should also understand that:

- Adoption of open-source platforms is an important market dynamic that is helping increase awareness and adoption of data science. However, Gartner's research methodology prevents evaluation of pure open-source platforms (such as Python and R) in a Magic Quadrant, as there are no vendors behind them that offer commercially licensable products.
- We invited a heterogeneous mix of data science platform vendors to subject themselves to evaluation for potential inclusion in this Magic Quadrant, given that data scientists have different UI and tool preferences. Some data scientists prefer to code their models in Python using scikit-learn; others favor Scala or Apache Spark; some prefer running data models in spreadsheets; and yet others are more comfortable building data science models by creating visual pipelines via a point-and-click UI. Diversity of tools is an important characteristic of this market.

Target audience: There are many different types of data scientist. In practice, any tool suggested for such a versatile group of people must be a compromise. This Magic Quadrant is aimed at:

- Line-of-business (LOB) data science teams, which typically solve between roughly three and 20 business problems over the course of three to five years. They have their LOB's executive as sponsor. They are formed in the context of LOB-led initiatives in areas such as marketing, risk management and CRM. Recently, some have adopted open-source tools (such as Python, R and Spark).
- Corporate data science teams, which typically solve between 10 and 100 business problems over the course of three to five years. They have strong and broad executive sponsorship and can take cross-functional perspectives from a position of enterprisewide visibility. They typically have more funding to access IT services.
- Adjacent, maverick data scientists. Of the tools covered in this Magic Quadrant, they should consider only those with an exceedingly small footprint and low cost of ownership. These data scientists strongly favor open-source tools, such as H2O, Python, R and Spark.

Other roles, such as application developer and citizen data scientist, may also find this Magic Quadrant of interest, but it does not fully cover their specific needs. Citizen data scientists may also consult "Magic Quadrant for Business Intelligence and Analytics Platforms." Application developers are only just starting to get more involved with data science, and many are trying the more rudimentary tools provided by cloud providers such as Amazon and Google.

Changes to Methodology

For this Magic Quadrant, we have completely revamped the inclusion criteria to give us more flexibility to include the vendors — both established and emerging — that are most relevant and representative in terms of execution and vision. This revamp has resulted in the inclusion of more innovative — and typically smaller — vendors. In our judgment, however, even the lowest-scoring inclusions are still among the top 16 vendors in a software market that supports over 100 vendors and that is becoming more heated and crowded every year.

The most significant change from the methodology used for 2016's "Magic Quadrant for Advanced Analytics Platforms" is in the process we used to identify the vendors eligible for inclusion.

We designed a stack-ranking process that examined in detail how well vendors' products support the following use-case scenarios:

- **Production refinement:** This is the scenario on which the majority of data science teams spend most of their time. Their organization has implemented several data science solutions for the business, which these teams need to keep refining.
- **Business exploration:** This is the classic scenario of "exploring the unknown." It requires many data preparation, exploration and visualization capabilities for both existing and new data sources.
- **Advanced prototyping:** This scenario covers the kinds of project in which data science solutions, and especially novel machine-learning solutions, are used to significantly improve on traditional approaches. Traditional approaches can involve human judgment, exact solutions, heuristics and data mining. Projects typically involve some or all of the following:
 - Many more data sources
 - Novel analytic approaches, such as deep neural nets, ensembles and natural-language processing (see also "Innovation Insight for Deep Learning").
 - Significantly greater computing infrastructure requirements
 - Specialized skills

We used the following 15 critical capabilities when scoring vendors' data science platforms across the three use-case scenarios:

1. **Data access:** How well does the platform support access and integrate data from various sources (both on-premises and in the cloud) and of different types (for example, textual, transactional, streamed, linked, image, audio, time series and location data)?
2. **Data preparation:** Does the product have a significant array of coding or noncoding features, such as for data transformation and filtering, to prepare data for modeling?
3. **Data exploration and visualization:** Does the product allow for a range of exploratory steps, including interactive visualization?
4. **Automation:** Does the product facilitate automation of feature generation and hyperparameter tuning?
5. **User interface:** Does the product have a coherent "look and feel," and does it provide an intuitive UI, ideally with support for a visual pipelining component or visual composition framework?
6. **Machine learning:** How broad are the machine-learning approaches that are either shipped with (prepackaged) or easily accessible from the product? Does the offering also include

support for modern machine-learning approaches like ensemble techniques (boosting, bagging and random forests) and deep learning?

7. **Other advanced analytics:** How are other methods of analysis (involving statistics, optimization, simulation, and text and image analytics) integrated into the development environment?
8. **Flexibility, extensibility and openness:** How can open-source libraries be integrated into the platform? How can users create their own functions? How does the platform work with notebooks?
9. **Performance and scalability:** How can desktop, server and cloud deployments be controlled? How are multicore and multinode configurations used?
10. **Delivery:** How well does the platform support the ability to create APIs or containers (such as code, Predictive Model Markup Language [PMML] and packaged apps) that can be used for faster deployment in business scenarios?
11. **Platform and project management:** What management capabilities does the platform provide (such as for security, compute resource management, governance, reuse and version management of projects, auditing lineage and reproducibility)?
12. **Model management:** What capabilities does the platform provide to monitor and recalibrate hundreds or thousands of models? This includes model-testing capabilities, such as K-fold cross-validation, training, validation and test splits, AUC, ROC, loss matrices, and testing models side-by-side (for example, champion/challenger [A/B] testing).
13. **Precanned solutions:** Does the platform offer "precanned" solutions (for example, for cross-selling, social network analysis, fraud detection, recommender systems, propensity to buy, failure prediction and anomaly detection) that can be integrated and imported via libraries, marketplaces and galleries?
14. **Collaboration:** How do users with different skills work together on the same workflows and projects? How can projects be archived, commented on and reused?
15. **Coherence:** How intuitive, consistent and integrated is the platform to support an entire data analytics pipeline? The platform itself must provide metadata and integration capabilities for the preceding 14 capabilities and provide a seamless end-to-end experience, to make data scientists more productive across the whole data and analytics pipeline. This metacapability includes ensuring data input/output formats are standardized wherever possible, so that components have a consistent "look and feel" and terminology is unified across the platform.

In addition, to be included in this Magic Quadrant, vendors had to satisfy revenue threshold requirements and identify reference customers demonstrating significant cross-industry and cross-geographic traction (see the Inclusion and Exclusion Criteria section below).

This process enabled us to identify 16 vendors eligible for inclusion in this Magic Quadrant.

Figure 1. Magic Quadrant for Data Science Platforms



Source: Gartner (February 2017)

Magic Quadrant

Vendor Strengths and Cautions

Alpine Data

Alpine Data is based in San Francisco, California, U.S. It offers a citizen data science platform, Chorus, with a focus on enabling collaboration between business analysts and front-line operational users in building and running analytic workflows. Chorus does not require users to move their data from where it resides, be that within a traditional database or Apache Hadoop. This is important for industries in which organizations are unable to move their data between servers for security reasons.

As last year, Alpine is in the Visionaries quadrant. It has limited visibility in this market, despite its continued growth in enterprise data science environments. Its vision-related scores suffered due to a lack of enterprise app connectivity, limited native analytic operators and the absence of plans to add "smart" capabilities that would help increase productivity.

Strengths

- **Collaboration capabilities.** These remain a core strength of Alpine's offering, and a key reason why customers choose it. They are increasingly important for enabling business users and analytics teams to work together more closely and efficiently, especially in organizations struggling to hire expert data scientists.
- **Support for big data analytics.** Reference customers scored Alpine highly for its scalability and wide support for data types, including unstructured data. A differentiating characteristic is its ability to analyze big data by running analytic workflows natively within existing Hadoop (and similar) platforms, on-premises or in the cloud. The platform supports a broad spectrum of open-source big data technologies (Hadoop, Spark, MADlib and MLlib).
- **Enhanced model portability.** Alpine has added support for portable format for analytics (PFA), which joins PMML support and enables some of the most commonly used models to be exported and run in its PFA scoring engine in platform-as-a-service environments, such as those of Cloud Foundry and Amazon Web Services (AWS). Alpine has also announced a partnership and integration with Trifacta, a data preparation provider, to make it easier to prepare and cleanse data with Chorus.

Cautions

- **Weak mind share and market visibility.** Due to its small size, Alpine is struggling to gain significant market visibility, which accounts for the drop in its Ability to Execute score. Of the vendors surveyed, it submitted the fewest reference customers, and 20% of those expressed concern about the small size of the community of users with whom they could network and share knowledge. One-fifth of its reference customers also identified a need for better vendor support.

- **Limited native analytic operators.** The choice of algorithms available natively in Alpine Chorus should be sufficient for most citizen data scientist teams. However, 40% of Alpine's reference customers expressed a desire for more analytic operators in this product, rather than having to rely on open-source capabilities or build their own custom operators.
- **Lack of enterprise app connectors.** Alpine offers few built-in integrations to connect directly to popular enterprise applications, such as those of Salesforce. Custom integration requires more technical expertise and maintenance. A work-around is to copy and store enterprise app data in one of the supported data platforms (such as Apache Hadoop, Greenplum Database, Oracle Database, Microsoft SQL Server and PostgreSQL).
- **Need for more analytics automation.** Alpine Chorus is lacking in several categories: visualization and exploration features, guided analytics/model search automation, and native support for more advanced analytics (such as affinity/graph, streaming and geospatial). With its focus on citizen data scientists, Alpine has an opportunity to innovate and carve out an important niche, but it must focus its product roadmap on adding smart capabilities.

Alteryx

Alteryx is based in Irvine, California, U.S. It offers a data science platform geared toward citizen data scientists. The platform's self-service data preparation capabilities and advanced analytics enable business users to blend data from internal and external sources and then to analyze it using predictive, prescriptive tools, using the same UI in a single workflow. Integration with open-source R enables expert data scientists to extend functionality by creating and running custom R scripts. Alteryx also offers a cloud-based analytics gallery for collaboration, sharing and version control of workflows.

This year, Alteryx is at the bottom right of the Challengers quadrant. Its move out of the Visionaries quadrant is due partly to solid customer growth, which has resulted in a higher score for Ability to Execute. It is also due to a lack of automated modeling features, a lack of Python or Scala programming support, and limited visual exploration and industry offerings, which together have reduced its Completeness of Vision score.

Strengths

- **Expansion into prescriptive analytics.** Alteryx has added simulation and optimization capabilities that enable users to create models spanning predictive and prescriptive analytics. Users can now simulate alternatives, predict outcomes, and discover strategic, tactical and operational efficiencies with optimization analysis.
- **Customer traction.** Alteryx has strong traction in the market with its "land and expand" strategy for moving customers from self-service data preparation to predictive analytics. It does normalization and fuzzy matching very well, which makes it easier for data scientists to prepare data. It offers spatial analytics capabilities, and it is likely to start promoting these as interest in location intelligence increases.

- **Customer satisfaction.** Ease of learning and use is the main reason for choosing Alteryx, this being identified by over half its reference customers. Other important reasons were its ability to support a wide variety of data types, and high flexibility and control for citizen data scientists in a code-free environment. Reference customers placed Alteryx almost at the top for overall customer satisfaction and delivery of business value.

Cautions

- **Lack of automated modeling and key language support.** Although it offers functions to help automate data preparation, Alteryx does not provide capabilities to automate feature engineering or model building. Its product does not currently support Python or Scala, which have become the programming languages of choice for expert data scientists. However, automated feature modeling and support for Python and Scala are on its product roadmap.
- **Visual data exploration.** Although Alteryx does offer some interactive visualization capabilities, 16% of its reference customers stated that they would like better reporting, data exploration and visualization natively. More robust capabilities are available through Alteryx's visual data discovery vendor partners, which include Tableau, Qlik and Microsoft (Power BI).
- **Scalability and performance concerns.** Reference customers want greater scalability and performance, and improved online documentation and support. Challenges reported by almost 30% of its reference customers included difficulty working with extremely large datasets, inability to run complex algorithms, and lack of documentation with examples for power users. Ten percent of its reference customers also noted that the high license cost limits broader use of Alteryx's platform.

Angoss

Angoss is based in Toronto, Ontario, Canada. It provides a suite of visual-based advanced data mining and predictive analytics tools for expert data scientists and citizen data scientists. It also offers prescriptive analysis and optimization capabilities that can be used to solve complex decision-making problems. In addition, its platform supports custom programming in R, Python, SQL and SAS through a built-in editor. Angoss focuses on specific use cases by offering packaged models for financial services, risk, marketing and sales.

Angoss is again in the Challengers quadrant. Although its overall scores from reference customers for product satisfaction were good, its scores for vision were lower, primarily due to performance and scalability challenges for emerging use cases using very large datasets (which are, however, partially addressed by its new Hadoop/Spark integration).

Strengths

- **Wide spectrum of analytics.** Breadth of use across multiple data science audiences is a core strength. Angoss is known for its decision and strategy trees. As well as supporting a variety of machine-learning techniques, its suite includes prescriptive analysis and optimization capabilities. Customer survey results show that the platform is most often used for business exploration.

- **Intuitive software.** Angoss offers easy-to-use software that is well-suited to citizen data scientists — it uses a wizard-driven approach to model building that saves time, compared with coding. Other key reasons why customers chose Angoss were its ability to build models with exceptional accuracy and its low total cost of ownership (TCO). Reference customers rated highly the software's "look and feel," automated model development and flexibility.
- **Packaged industry solutions.** Angoss offers packaged models and SaaS-based applications for several key industries, including financial services and insurance. These solutions cater to organizations that either have no data science capability or want to use Angoss' domain expertise.

Cautions

- **Market traction and growth.** Angoss has not yet been able to turn its 30 years' experience of data mining into growth matching that of other vendors in this market. Reference customers want better account management and a larger user community for knowledge sharing and networking. In response, Angoss is increasing its sales presence in several key territories in the U.S., U.K. and Canada, and strengthening its partnerships in the data science ecosystem.
- **Difficulty handling large datasets.** Several surveyed customers of Angoss reported challenges with performance on large datasets (over 200GB, for example) for certain use cases. To help address this concern, Angoss recently announced support for Hadoop as a storage repository, with Spark as the execution engine, so that advanced analytics can be run on very large datasets.
- **Data preparation and UI.** Although overall product satisfaction scores were good, several reference customers stated that they would like Angoss to provide more intuitive data preparation capabilities, along with an updated UI.

Dataiku

Dataiku is headquartered in New York City, U.S., and has a main office in Paris, France. It has chosen to take a very ambitious path with its data science platform, Data Science Studio (DSS).

Dataiku is a new entrant to the Magic Quadrant. Its placement as a Visionary is due to the innovative nature of the DSS, especially its openness and ability to cater to different skill levels, which enables better collaboration. Dataiku's Ability to Execute suffers from limited user adoption and deficiencies in its data access and exploration capabilities.

Strengths

- **Bring your own algorithm.** Dataiku DSS is a wrapper/shell environment that enables machine-learning algorithms and other important capabilities to be plugged in from the open-source community. It emphasizes openness and flexibility, and can run high compute loads in the Spark runtime environment. Users are offered a selection of native machine-learning algorithms, and a choice of leading machine-learning engines (such as H2O.ai Sparkling Water, Hewlett

Packard Enterprise Vertica and MLlib). They can also choose different language wrappers for Python, R, Scala and Spark.

- **Bridging of different skills.** Dataiku DSS is differentiated by having different GUI endpoints that allow users of differing skill levels to collaborate — from dashboard-type GUIs for less skilled users to a visual pipelining tool for more advanced ones. Expert data scientists who can work on the code line can benefit from the integration of notebooks that allow different machine-learning engines to be integrated and to work in concert.
- **Ease of use.** Reference customers praised DSS for being easy to learn and use. They also appreciated its support for open-source capabilities and the overall speed of model development. Some customers selected Dataiku for its intuitive data preparation capabilities. Of the vendors in this Magic Quadrant, Dataiku had the lowest percentage of customers reporting platform-related challenges.

Cautions

- **Adoption and partnerships.** Dataiku is still a young company, with a nascent user community and a lack of significant partnerships with major system integrators.
- **Deficiencies upstream in the data pipeline.** As is typical of products early in their life cycle, DSS focuses on the midstream machine-learning phase. Dataiku's reference customers noted deficiencies upstream in the data pipeline, with below-average data access, exploration and visualization capabilities. A significant number of customers expressed concerns about pricing and ROI.
- **Growing pains.** Dataiku DSS may require advanced internal support, as some reference customers identified "rough edges" (concerning, for example, the installation process and the number of releases). Also, history shows that young companies such as Dataiku may become victims of their own success. Recent injections of venture capital suggest fast growth to come, but technical support might suffer.

Domino Data Lab

Domino Data Lab is headquartered in San Francisco, California, U.S. It offers the Domino Data Science Platform, with a focus on openness, collaboration and reproducibility of models. Founded in 2013, Domino has gained significant attention and momentum in only a few years.

A new entrant to the Magic Quadrant, Domino is a Visionary due to its support for a wide range of open-source technologies and offer of freedom of choice to data scientists. Its innovation scores are among the highest of any vendor in this Magic Quadrant. To raise its Ability to Execute score, Domino will need to improve its data access and data preparation capabilities, and devote resources to strengthening its UI and functionality for the business exploration use case. Vision alone will not be enough to compete with the traditional powerhouse vendors at the enterprise level.

Strengths

- **Support for open-source capabilities.** The overwhelming majority of Domino's reference customers chose it for its native support for open-source capabilities. Data scientists can use the platform to build and deploy models using the language and techniques of their choice.
- **Customer support and satisfaction.** Reference customers identified Domino's customer support and collaboration features as particular strengths. Domino's customer satisfaction scores for collaboration features were among the highest of any vendor. Customers recognize Domino's speed of model development and experimentation.
- **Recognition of trends and growing momentum.** Domino offers support for many data science technologies, such as notebooks, Python, R and Spark. The company doubled its number of customers in 2016, and will attract more prospective customers as interest in, and adoption of, these technologies continues to grow.
- **Sales relationship.** Domino received the highest overall rating from customers for its sales relationship with them.

Cautions

- **High technical bar for use.** Domino's code-based approach requires the user to have significant technical knowledge. Multiple reference customers cited ease of use of the platform as a weakness, and its data exploration and visualization capabilities are limited. The platform is not a good fit for citizen data scientists.
- **Delivery of project management and workflow capabilities.** Although Domino trumpets its platform's support for workflow reproducibility, and offers functionality to audit and recreate past projects, reference customers' overall scores for satisfaction with its project management were middling.
- **Data access and data preparation capabilities.** Domino overestimates its platform's strengths in certain areas, particularly data access and data preparation, relative to more established solutions.
- **Lack of "precanned" solutions.** Of the vendors in this Magic Quadrant, Domino received the lowest scores from reference customers for satisfaction with its precanned solutions. This is because users have to rely primarily on open-source packages for common data science use cases.

FICO

FICO is based in San Jose, California, U.S. Its Decision Management Suite (DMS) offers analytic tools, including Model Builder, Optimization Modeler and Decision Modeler.

FICO is again positioned as a Niche Player. It is an especially strong choice for organizations in the financial services sector and for those that depend on scorecard modeling. FICO's placement stems largely from its need to catch up with, and innovate at the level of, many other vendors, which

results in a low overall score for Completeness of Vision. Its Ability to Execute score is dragged down by low market traction in other vertical areas and limited machine-learning capabilities. FICO needs to deliver on its many product roadmap promises in the near future.

Strengths

- **Presence and experience in financial services.** FICO's significant presence in the financial services industry and particular strength for financial services use cases (particularly relating to risk and fraud) are the main reasons why customers choose it.
- **Functionality in key areas for decision management.** FICO's scorecard approach is best-in-class and ideal for highly regulated industries that need transparency. Its product offers strong functionality for tasks such as facilitation of decision flows, for complex rule logic and for custom rules. FICO emphasizes governance and model management.
- **Project management and model management.** FICO's DMS received high customer satisfaction scores for platform/project management and model management, as well as for speed of model development and product quality. Customers find FICO's pricing to be predictable and controllable, but a few expressed concerns about high costs.

Cautions

- **Critical capability scores.** FICO received below-average scores (customer and analyst evaluations combined) for a number of critical capabilities. Reference customers raised concerns regarding data access, preparation, and exploration and visualization.
- **Open-source tool support.** Beyond R integration, the current version of FICO's Model Builder does not adequately address the growing demand for flexibility and extensibility with open-source tools and libraries (such as H2O, Spark and TensorFlow). Some reference customers are asking for Python support and cohesion with other products.
- **Algorithm selection.** Reference customers mentioned few challenges regarding FICO, though some reported a lack of access to required or desired algorithms and some wanted broader machine-learning capabilities. Several popular machine-learning algorithms are not offered natively (for example, support vector machines and gradient-boosted machines).
- **Innovation around key trends.** FICO has not acted sufficiently on many market trends (such as demand for product cohesion and deep-learning capabilities) — it has not innovated to the same degree as many other vendors in this Magic Quadrant. To increase its visibility in the greater machine-learning market, FICO needs to catch up to its competitors in the near future and better engage the data science community.

H2O.ai

H2O.ai is based in Mountain View, California, U.S. It offers an open-source data science platform, H2O, with a focus on fast execution of cutting-edge machine-learning capabilities. This evaluation focuses on the following products and versions, which were generally available at the end of July

2016: H2O Flow, Steam and Sparkling Water. The company has recently launched deep-learning capabilities in the form of Deep Water.

A new entrant to the Magic Quadrant, H2O.ai is a Visionary because of its solid range of highly scalable machine-learning implementations. It is one of the machine-learning partners most frequently mentioned by companies such as IBM and Intel. However, its execution is lacking, especially as regards its commercial strategy and its lower-ranking features upstream in the data pipeline.

Strengths

- **Market awareness.** Although a small company, H2O.ai has gained considerable mind share and visibility, due to its strategy of offering most of its software on an open-source basis. The H2O community is active and growing, and the company has significant partners, such as IBM and Intel, which are integrating H2O components into their upcoming data science platforms.
- **Customer satisfaction.** Overall customer satisfaction with H2O.ai is very high and likely rooted in the perception that organizations do not need to pay for H2O products. Reference customers identified the "ability to build models with exceptional accuracy" as a key strength. With its Java export and one-click REST API publication capabilities, H2O enables fast deployment of data science models into production environments. It is especially suited to Internet of Things (IoT) edge and device scenarios.
- **Flexibility and scalability.** Customers choose H2O.ai for its flexibility, and it received the highest score from reference customers for analytics support of any vendor in this Magic Quadrant. Other top reasons for choosing it include its support for open-source technologies, speed of model development, ability to build models that are exceptionally accurate, and ease of learning for data scientists.

Cautions

- **Data access and preparation.** The H2O platform currently lacks many of the data access and preparation features available in other data science platforms. Out-of-the-box connectors to enterprise applications such as those of Salesforce and NetSuite are absent. Users report that the platform's native data preparation is tedious, so they use other tools before importing data.
- **High technical bar for use.** The documentation is fairly limited and, without programming skills, prospective users are likely to struggle to learn the H2O stack of products.
- **Visualization and data exploration.** H2O.ai's platform has inferior data exploration and visualization capabilities. Reference customers scored its model management and collaboration aspects below the average.
- **Sales execution.** With just a few dozen paying clients, H2O.ai must focus on how to turn its phenomenal platform adoption into revenue.

IBM

IBM is based in Armonk, New York, U.S. As a large enterprise software vendor, IBM offers a wide array of analytics solutions. For this Magic Quadrant, we evaluated SPSS Modeler and, to a lesser extent, SPSS Statistics.

IBM is again a Leader. Its position on the Ability to Execute axis has dropped slightly, as its attention is split between SPSS Modeler and its new data science platform, IBM Data Science Experience (DSx). DSx is not evaluated in this Magic Quadrant, but it does strengthen IBM's position on the Completeness of Vision axis.

Strengths

- **Customer base and continued innovation.** IBM has a vast customer base and remains committed to modernizing and extending its data science and machine-learning capabilities. DSx is likely to be one of the most attractive platforms in the future — modern, open, flexible and suitable for a range of users, from expert data scientists to business people.
- **Commitment to open-source technologies.** IBM has been focusing on support for open-source technologies, which is a top requirement for data scientists. It is strongly committed to Spark and other open-source technologies, and has contributed to open-source SystemML technology from IBM Research, as well as over 170 extensions that allow SPSS users to access external APIs and Algorithmia machine-learning models. IBM also partners with numerous open-source ecosystem providers (such as Databricks, H2O.ai and Continuum Analytics).
- **Support for a broad range of data types.** Customers choose IBM SPSS because of its ability to support a broad range of data types, including unstructured data, and its solid product quality. SPSS supports all leading Hadoop distributions, NoSQL DBMSs and a variety of relational databases (such as IBM DB2 on mainframes and Amazon Redshift). It can work with very large datasets and multiple streams.
- **Model management and governance.** Surveyed IBM customers rated SPSS's model management highly, with praise for its breadth of models, accuracy and transparency in workflows, model deployment, monitoring for degradation and automatic retuning. SPSS provides excellent features for analytics governance: versioning, metadata and audit capabilities.

Cautions

- **Confusion about product offering.** IBM's branding and breadth of analytics offerings across a wide range of roles remain confusing in terms of go-to-market approach and product roadmap. SPSS has interoperability problems with adjacent platforms (such as IBM Watson, Watson Analytics and Cognos Analytics). Customers are often confused by mismatches between marketing messages and actual, purchasable products.
- **Dual focus.** To many new users, IBM SPSS Modeler and Statistics seem outdated and overpriced. IBM's roadmap focuses on modernizing SPSS, and IBM DSx could address these

concerns in the long run. IBM intends to eventually converge SPSS and DSx, which would be a substantial change, albeit for the better.

- **Customer support and bureaucracy.** Reference customers expressed dissatisfaction with IBM's support and bureaucracy; they reported difficulties finding the right liaisons and technical help, despite high maintenance fees. Some customers expressed concerns about purchasing products from IBM, as the company reportedly often tries to bring its consulting organization, IBM Global Business Services, into data science projects.

KNIME

KNIME ("Konstanz Information Miner") is based in Zurich, Switzerland. Its open-source KNIME Analytics Platform is a fully functional and scalable platform for advanced and expert data scientists. Its commercial offering provides extended value-added proprietary tools, service layers and support. KNIME's platform is used in a number of industries. Its strongest presence is in manufacturing and life sciences, but it also has a strong presence in the financial services, education, government and retail sectors.

KNIME remains a Leader and a popular choice for data science needs. Its strong Ability to Execute is attributable to solid interactions with customers via its sales team, its responsiveness and its community support. Its position for Completeness of Vision, however, has worsened due to weaker marketing, sales innovation and product innovation than is shown by the other Leaders.

Strengths

- **Low barrier to entry and TCO.** The KNIME Analytics Platform presents a low barrier to entry and offers a low TCO as an open-source solution that is flexible and extensible by design. Its low TCO and ease of use are consistently identified as reasons for choosing it. Survey respondents rated KNIME highest of all participating vendors for flexibility, openness and extensibility.
- **Data access and transformation.** KNIME provides strong data access and data preparation capabilities, including for combining and blending data, verifying data quality, transforming and aggregating values, binning, smoothing, dataset partitioning, and feature generating and selecting.
- **Sales and vendor relationship.** Survey respondents commended KNIME for a positive sales relationship, as well as for its product enhancements and ability to include requested features in subsequent releases.
- **Active community and partnerships.** KNIME has strong partnerships with independent software vendors and system integrators, including some (such as Teradata) that embed parts of the KNIME platform into their own solutions. It also has a very strong and active user community and an enthusiastic user base, as demonstrated by its many pleased survey respondents.

Cautions

- **Model management.** Model management is challenging in KNIME's platform, which diminishes its appeal for use by the larger data science teams. Many survey respondents indicated issues with managing and documenting large workflows and with managing use by large work teams.
- **Scalability.** KNIME's platform can be difficult to scale and share across enterprise deployments. Several survey respondents indicated that their main challenge with KNIME is slow performance or an inability to build and deploy models in the needed time frame. Other issues with performance and scalability could reflect a need for additional components beyond the KNIME Analytics Platform.
- **Exploring and visualizing data.** KNIME's data exploration and visualization capabilities are limited, as it lacks the ability to easily create interactive dashboards, or to natively conduct robust conjoint and survey analysis. Its improved, but still unpolished, UI is a potential barrier to extending the use of its platform to business analysts and citizen data scientists.
- **Sales approach.** KNIME has a low-touch sales approach focused on community support. Domain experts support customers within specific industries or cross-industry areas. Paid support for Asia, Latin America, Africa, India and Australia is delivered either by KNIME remotely or through a certified network of local partners.

MathWorks

MathWorks is a privately held corporation headquartered in Natick, Massachusetts, U.S. Its two major products are Matlab and Simulink, but only Matlab met the inclusion criteria for this Magic Quadrant.

A new entrant to the Magic Quadrant, MathWorks is a Challenger. Its Ability to Execute is driven by its prominent presence in data-science-related domains, but it still has to demonstrate its vision by producing a significantly easier-to-use platform that can address the concerns of corporate data scientists, especially for customer-facing use cases like marketing, sales and CRM.

Strengths

- **Advanced prototyping for engineering-savvy data scientists.** Popular with engineers, MathWorks' Matlab platform builds on one of the four major quantitative software programming languages, MATLAB (the other three being Python, R and SAS). MATLAB is one of the simplest and most concise languages for anything involving matrix operations, and it works well for anything represented as a numeric feature matrix.
- **Rich set of toolboxes.** MathWorks' Matlab platform is optimized for solving engineering and scientific problems. It provides engineers and data scientists with a comprehensive set of toolboxes (such as for statistics, machine learning, optimization, intelligent control and signal/image processing). The Matlab platform is strong in areas such as advanced prototyping for engineering, manufacturing and quantitative finance purposes.

- **Market presence.** MathWorks has long-established relationships with many organizations, including universities. New software engineers will find its platform simple to use and can test run it easily. Pricing for versions of Matlab and for toolbox extensions is highly transparent, and most functionality can be purchased and downloaded directly from the company's website.

Cautions

- **Customer-facing use cases.** MathWorks Matlab is not the best choice for many common data science use cases in marketing, sales and CRM (such as recommender systems, cross-selling and churn prediction), as it lacks some of the necessary functionality (such as champion/challenger). Lacking visual pipelining functionality, it is less user-friendly than some other data science platform options. However, there are point-and-click apps that guide the user through basic data science workflows.
- **Python and R not yet "first-class citizens."** Although MathWorks Matlab is one of the major engineering platforms worldwide, with a very large and active community of users, it is not the top choice for many data scientists and statisticians. They often prefer to work with more open environments, using Java, Python, R and Scala.
- **Difficulty of use for many data scientists.** As Matlab is a very code-centric environment, it is not suitable for many citizen data scientists or business analysts. MathWorks' current offering cannot bridge the skills gap that affects many data science initiatives.
- **Data science and big data capabilities.** Reference customers noted poor capabilities for machine learning with larger datasets. Many also found the operationalization of models difficult.

Microsoft

Microsoft is based in Redmond, Washington, U.S. It has bundled its data science tools into its cloud-based Microsoft Cortana Intelligence Suite and Microsoft R Server. This evaluation concentrates on the Azure Machine Learning platform, part of the Cortana Intelligence Suite, which includes many additional components, such as Azure Data Factory, Azure Stream Analytics and Power BI.

During the past three years, Microsoft has undertaken a remarkable revamp in the context of machine learning. It entered the market with a very limited product offering and remains a Visionary for its market-leading data science cloud solution. The omission of a comparable on-premises offering continues to pose challenges for customers and significantly reduces Microsoft's Ability to Execute.

Strengths

- **Cloud flexibility and scalability.** Microsoft Azure Machine Learning is a comprehensive cloud offering — one of the primary reasons customers choose it. The platform enables easy integration with cloud data sources, as well as with on-premises sources via its gateway. This

model lends itself to flexibility and scalability. Weekly updates to the platform give Microsoft the ability to enhance and extend functionality rapidly.

- **Innovation in image and speech recognition and deep learning.** Microsoft is delivering significant breakthroughs in image and speech recognition, and is developing one of the leading open-source deep-learning toolkits (Cognitive Toolkit). The Cortana Intelligence Gallery and Cognitive Services are further evidence of Microsoft's strong vision.
- **Machine-learning algorithms.** Of the capabilities they rated, Microsoft's reference customers scored Azure Machine Learning highest for its machine-learning capabilities. The platform supports customization of some algorithms, providing the ability to "open the black box" and examine the calculations. Other capabilities that scored highly were collaboration and automation.
- **Roadmap and breadth of offering.** Microsoft's product roadmap, broad offering, continual innovation, frequent upgrades and internal expertise attract a wide range of customers to Azure Machine Learning.

Cautions

- **No on-premises Azure Machine Learning offering.** Microsoft Azure Machine Learning's inability to run on-premises is a severe limitation that significantly reduces its Ability to Execute score. Microsoft does have on-premises machine learning capabilities with SQL Server R Services and R Server for Linux, Hadoop and Teradata, but these capabilities are not as strong as those in the Azure Machine Learning offering.
- **Product immaturity.** Despite its growing popularity, the Microsoft Cortana Intelligence Suite is new and immature. There are some challenging aspects to Azure Machine Learning: usability, absence of some desired algorithms, a lack of integration of various components, product instability and poor vendor support. Over time, updates to the entire suite might bring unexpected changes.
- **Delivery options.** Delivery options are limited to REST API and R. There is no ability to export code, as is offered by other data science vendors.
- **Lacking capabilities.** One-third of Microsoft's reference customers indicated that they would like better supporting documentation. In addition, 13% stated that they would like more robust data preparation capabilities, improvements to platform project management and more prepackaged industry solutions.

Quest

Quest is headquartered in Aliso Viejo, California, U.S. As a result of the sale of Dell Software to Francisco Partners and Elliott Management, completed in November 2016, Quest now sells the Statistica Analytics Platform.

Quest is a Challenger in this year's Magic Quadrant, whereas Dell was a Leader in last year's. This shift is largely due to the second change in ownership for Statistica in three years and to the lack of some product improvements central to native cloud and some Spark capabilities (which are,

however, on the product roadmap). The platform's large customer base and all-around strength on-premises across the production refinement, business exploration and advanced prototyping use cases merit a position as a strong Challenger.

Strengths

- **Diversity of use cases and well-rounded on-premises capabilities.** Statistica has strong use-case diversity across many industries and can fulfill a wide range of data science needs. The platform is flexible and extensible, with integration for R, Python, algorithm marketplaces and H2O. Statistica's ability to handle various types of data, such as image and text data, should be emphasized. Its new capabilities for handling and visualizing graph data are especially noteworthy.
- **Model sharing and deployment.** Statistica has best-in-class deployment options for IoT and edge analytics. These are strengthened by capabilities for creating C, C#, Java and SQL code, which allow for easy deployment on edge devices. Quest also offers distribution capabilities to run analytic models in-database for a wide range of open-source and commercial databases via its native distributed analytics architecture (NDAA).
- **Reusable workflow templates.** Organizations can create reusable data preparation and analytic workflow templates. This makes it easier for expert data scientists to build analytics models and workflows once and to distribute to citizen data scientists, who can reuse the workflow templates repeatedly.
- **Production refinement.** Overall, Statistica received one of the highest scores from reference customers for production refinement of models, encompassing capabilities such as scalability, model management and delivery. Reference customers chose Statistica for its TCO and praised its outstanding customer service and technical support.

Cautions

- **Getting started with the platform.** Reference customers reported difficulties learning and using the platform, and with initial deployment or migration to newer versions. Quest emphasizes the features it offers citizen data scientists (such as reusable workflows and guided analytics), but it still lacks the model factory capabilities and superb data exploration and visualization capabilities that would appeal to a broader base of users.
- **Performance issues and calls for product enhancements.** Some Quest reference customers cited stability and speed as weaknesses of the platform, alluding to occasional crashes and long compute times. Many reference customers were unsatisfied with the responses to their requests for product enhancements.
- **Cloud and Spark capabilities.** Quest's current cloud capability (which still has Dell branding) is implemented on the Microsoft Azure Marketplace and lacks cloud-native capabilities, such as elastic scale-out. Its Spark capabilities are also lagging behind: although model scoring can be done via NDAA, workflows for training cannot be pushed seamlessly into the distributed Spark runtime.

- **Uncertainty created by the recent acquisition.** The sale of Statistica represents more changes in top-level management for the Statistica product group, which has a decades-long history of providing solid customer satisfaction. In the short term, Gartner does not believe the acquisition will have an immediate material effect. However, to alleviate long-term concerns, Quest needs to focus on retaining key Statistica employees.

RapidMiner

RapidMiner is based in Boston, Massachusetts, U.S. It offers its RapidMiner GUI-based data science platform for the full spectrum of data scientists, including beginners. It also offers access to the core open-source code for expert data scientists who prefer to program. RapidMiner's data science studio is available both as a free edition and as a commercial edition, which offers additional functionality for working on larger datasets and connecting to more data sources.

RapidMiner is again a Leader, owing to its market presence, the volume of client inquiries that Gartner receives about it, its user community, and its well-rounded product that addresses most data science use cases well.

Strengths

- **Platform breadth.** RapidMiner products are used in different industries for a wide range of use cases. Reference customers praised many facets of the platform — its large selection of algorithms, flexible modeling capabilities, data source integration and consequent data preparation. The platform's strength lies not just in particular areas, but also in its all-around consistency.
- **Ease of use for the entire spectrum of data scientists.** Reference customers chose RapidMiner's platform for its ease of learning and use, speed of model development and support for open-source capabilities. Expert data scientists get access to advanced analytics and their choice of tools and programming language. Citizen data scientists can use guided model development and many precanned solutions.
- **User community.** Over the years, RapidMiner has focused on fostering a large and vibrant community of users, which was an early differentiator. This community has augmented RapidMiner's customer support. A partnership with Experfy has continued this effort by offering clients access to a marketplace of validated data science experts.

Cautions

- **Data usage cap.** Although RapidMiner offers a free edition, with an open-source core, its real-world use by organizations is limited by its data usage cap. Regarding the commercial edition, many RapidMiner reference customers felt that its data usage pricing model made TCO unpredictable, and some questioned the value they received.
- **Issues with documentation.** Many of RapidMiner's reference customers reported issues with its documentation, which they identified as being of poor quality, needing simplification and having insufficient examples of how to use specific operators.

- **Global presence.** RapidMiner has just one U.S. office location and three European office locations. This could limit its worldwide growth and ability to provide high-quality enterprise support worldwide. The company's technical support was criticized by some reference customers. RapidMiner does not offer professional services directly, but relies on channel partners to help organizations build and deploy models.

SAP

SAP is based in Walldorf, Germany. It has rebranded its data science platform, which is now called SAP BusinessObjects Predictive Analytics (BOPA). In addition, SAP has a broad array of other analytics offerings, including the SAP BusinessObjects Business Intelligence platform and SAP BusinessObjects Lumira for self-service data discovery, which are not included in this evaluation.

A fall in SAP's Ability to Execute score has caused it to drop slightly, from near the bottom of the Challengers quadrant to near the top of the Niche Players quadrant. SAP is lagging behind in terms of Spark integration, open-source support, Python and notebook integration, and cloud deployment.

Strengths

- **SAP Hana integration.** SAP BOPA's unique strength is its ability to integrate with SAP's Hana platform, an in-memory DBMS. This opens up significant functionality in terms of support for different data types, R integration, spatial analytics, time series analysis and text analytics.
- **Support for citizen data scientists.** SAP BOPA has an "automated analysis" mode that enables models to be built in a wizard-driven and fairly automated fashion. Customers choose BOPA for its speed and ease of use in enabling both citizen data scientists and expert data scientists to build and deploy models.
- **Deployment.** SAP BOPA's deployment and delivery options are in the top quartile for combined analyst and customer ratings. SAP has added new components, such as SAP BusinessObjects Predictive Factory and SAP Analytics Extensions Directory, which clients said were useful but still immature.
- **Broad footprint.** SAP supports diverse use cases. It is making progress on a much simpler tiered pricing model, which should address most clients' concerns about costs and pricing. It is making a significant commitment to machine learning and data science, as highlighted by its many forward-looking partnerships (such as one with Nvidia).

Cautions

- **Mind share in the data science community.** SAP has yet to gain much mind share and traction outside its traditional community. "Alignment with existing infrastructure investments" was among the most frequently cited reasons why customers selected SAP BOPA. This helps to explain why SAP was among the least-considered alternatives in reference customers' selection processes.

- **Customer satisfaction.** Similar to last year, reference customers scored SAP in the lowest decile for overall customer satisfaction. For most critical capabilities, SAP (BOPA) received the lowest scores of the vendors in this Magic Quadrant.
- **Hana-first strategy.** SAP has been relying too long on Hana as its premier high-performance computing runtime. Its reliance on Hana for advanced features deters small or cutting-edge teams, who fear Hana's footprint — although this problem should be alleviated by the new SAP Hana, express edition. Existing integrations of Hana and R are too slow for low-latency applications, where very fast data transfer between BOPA, Hana and the R runtime is required.
- **Spark, open-source support and other innovation.** SAP lags behind most of the Leaders and Challengers in terms of Spark integration, open-source support, and Python and notebook integration. SAP seems to be behind other large vendors in providing cutting-edge machine-learning components via APIs.

SAS

SAS is based in Cary, North Carolina, U.S. It provides a vast array of software products for analytics and data science. This evaluation covers SAS Enterprise Miner (EM) and the line of SAS products with names starting with the word "Visual," such as Visual Statistics and Visual Data Mining and Machine Learning, which we refer to collectively as the Visual Analytics suite (VAS).

SAS's focus is now interactive modeling in its VAS, but it continues to support its traditional programmatic approach (for Base SAS). SAS's delivery of these capabilities has enabled it to retain a strong position in the Leaders quadrant. However, confusion about its multiple products and concerns about licensing costs continue to impact its Ability to Execute.

Strengths

- **Presence and mind share.** SAS is the top vendor in the data science market, in terms of total revenue and number of paying clients. SAS VAS is the most frequently cited toolset by users of Gartner's client enquiry service. It was also considered by some of the surveyed reference customers who subsequently chose a different vendor's platform.
- **Appeal to different users.** SAS EM provides strong data access and preparation capabilities, as well as excellent data visualization and exploration capabilities. It is used mainly by expert data scientists. By contrast, VAS is used more by business analysts and citizen data scientists. But VAS's new Visual Data Mining component, partially released in 2016, will soon increase its appeal to specialist data scientists.
- **Infrastructure and breadth of capability.** SAS is usually chosen for its alignment with existing infrastructure. Reference customers also identified product quality as a top reason for selecting SAS's platform. Other key reasons were the ability to support a wide variety of data sources, EM's breadth of algorithms and the speed of model development.
- **Presence and ecosystem.** SAS has a very strong customer base and a large ecosystem of users and partners. It has widespread geographical presence and market penetration across all

major industries. SAS Viya, a revamped cloud-ready architecture, and SAS Analytics as a Service are positive signs of a focus on cloud and multiple deployment options.

Cautions

- **Management difficulty.** Although they provide a good breadth of functional capability, the SAS products considered in this Magic Quadrant are difficult to manage. Over half of SAS's reference customers pointed to difficulties with the initial deployment or version migration. Several indicated instability and bugs. Many encountered difficulties in learning and using SAS EM.
- **Production refinement capabilities.** Although many reference customers are using SAS's products specifically for production refinement, SAS EM and SAS VAS both received low scores for platform and project management, as well as delivery. SAS Viya is intended partly to address SAS EM's need for an updated, easier-to-use UI.
- **Pricing.** SAS's pricing remains a concern. Open-source data science platforms are often used along with SAS's products as a way to control costs, especially for new projects.
- **Confusion about what to use and when.** Multiple SAS products with similar functionality continue to cause confusion, and the need to have multiple products for complete functionality adds to the complexity. SAS EM and VAS are largely not interoperable at present, although interoperability is planned. SAS's current focus seems to be on VAS, which could distract it from continuing to invest in the more established EM.

Teradata

Teradata is headquartered in Dayton, Ohio, U.S. It offers a data science platform called Aster Analytics, which has three layers: analytic engines, prebuilt analytic functions, and the Aster AppCenter for analysis and connectivity to external business intelligence (BI) tools. Aster Analytics can be shipped as software only, as an appliance, or as a service in the cloud on AWS or a Teradata-managed cloud. Configuration is available on the platform's own massively parallel processing shared-nothing database or directly on Hadoop. The most popular use of Aster Analytics is for customer analytics in its various forms.

Teradata is a new entrant to the Magic Quadrant. It is a Niche Vendor, largely due to its low level of adoption and lack of broad usability and applicability. However, it excels in situations where Aster Analytics fits into an organization's existing infrastructure and there are significant requirements for high performance.

Strengths

- **Performance and scalability.** Aster Analytics shines when it comes to performance and scalability, with Teradata focusing on enabling rapid data discovery and fast iterations. As most data scientists lack the time and skills for sophisticated subsampling, Aster Analytics runs very quickly on full datasets, although it still has prebuilt sampling capabilities. Reference customers praised Aster Analytics' performance and scalability, as well as Teradata's support.

- **Multifaceted SQL analytics.** Aster Analytics is a good choice for customers who prefer SQL or who want to simplify analytics on Hadoop. The analytics engines include Aster SQL, SQL-MapReduce and SQL-Graph, and its second layer contains over 150 prebuilt in-database analytics functions that can be called directly in SQL. As an SQL frontend, Teradata QueryGrid overlays complex programming that can access many other data sources and is optimized to access Hadoop. Users run Spark functions and scripts as Aster's query operators.
- **Prebuilt analytics solutions.** Aster Analytics offers a portfolio of analytics and business solution bundles in the Aster AppCenter, its application development framework. It includes prebuilt apps tailored to specific industries, and an analytics library that includes time series, graph, text, behavior and marketing analytics, among others.

Cautions

- **Market awareness.** Although Teradata demonstrates a strong understanding of the data science platform market, its execution is limited by its target market focus. Aster Analytics is most often used inside Teradata's own ecosystem — many customers select it to complement their Teradata data warehouse or Hadoop in the Teradata Unified Data Architecture. The strength of Teradata's installed base does, however, give its product good long-term viability.
- **Use alongside other data science platforms.** Most of Teradata's reference customers use Aster Analytics in addition to other data science platforms — predominantly those of SAS or IBM. Aster Analytics received low overall scores for advanced prototyping, and is not suitable for expert data scientists seeking cutting-edge techniques. It compensates for this by connecting to external functions that improve its decision management and visualization deficiencies and provide more cutting-edge predictive analytics.
- **Agility and automation.** Reference customers criticized Aster Analytics' lack of agility or automation. App implementation often involves a Teradata service engagement. Aster Analytics does, however, offer capabilities to develop custom analytics in a variety of languages and in Aster AppCenter, which can mitigate these issues. In addition, the Aster Analytics community provides tips and workarounds to improve automation.

Vendors Added and Dropped

We review and adjust our inclusion criteria for Magic Quadrants as markets change. As a result of these adjustments, the mix of vendors in any Magic Quadrant may change over time. A vendor's appearance in a Magic Quadrant one year and not the next does not necessarily indicate that we have changed our opinion of that vendor. It may be a reflection of a change in the market and, therefore, changed evaluation criteria, or of a change of focus by that vendor.

Added

- Dataiku
- Domino Data Labs
- H2O.ai

- MathWorks
- Teradata

Note also that Quest takes the place of Dell.

Dropped

- Accenture
- Lavastorm
- Megaputer
- Predixion Software
- Prognoz

Inclusion and Exclusion Criteria

As noted earlier, the inclusion and exclusion criteria have changed significantly from those of last year's Magic Quadrant.

To qualify for inclusion in this Magic Quadrant, each vendor had to pass the following assessment "gates":

Gate 1

For each vendor, we gathered information about the following:

- **Perpetual-license-model revenue:** Software license, maintenance and upgrade revenue, excluding hardware and professional service revenue, for the calendar year 2015.
- **SaaS subscription model revenue:** Annual contract value (ACV) at year-end 2015, excluding any professional services included in annual contracts. For multiyear contracts, only the contract value for the first 12 months was used.
- **Customer adoption:** Number of active paying client organizations using the vendor's data science platform (excluding trials).

To progress to the next assessment gate, vendors had to have generated revenue from data science platform software licenses and technical support of:

- At least \$5 million in 2015 (or the closest reporting year) in combined-revenue ACV, *or*
- At least \$1 million in 2015 (or closest reporting year) in combined-revenue ACV, *and either*
 - At least 150% revenue growth from 2014 to 2015, *or*
 - At least 200 paying end-user organizations

Only vendors that passed this initial requirement progressed to Gate 2.

Gate 2

Vendors were evaluated on the basis of the reference customers they identified. In contrast to last year, we required vendors to show significant cross-industry and cross-regional traction.

Cross-industry reference customers

Each vendor had to identify reference customers for its main data science platform products in production environments. For each product, we required at least 12 unique reference customers, which had to be using predictive analytics solutions in production environments and come from at least four of the following industries:

- Banking, insurance and other financial services
- Education and government
- Healthcare
- Logistics and transportation
- Manufacturing and life sciences
- Mining, oil and gas, agriculture
- Retail
- Telecommunications
- Utilities
- Other

To avoid possible bias, we did not accept more than 40 reference customers per product.

Cross-region reference customers

For each vendor, the identified reference customers had to include at least two from each of the following areas:

- North America
- European Union
- Rest of the world

Only vendors that passed Gate 2 progressed to Gate 3.

Gate 3

We used a scoring system to measure how well vendors' product(s) addressed the 15 critical capabilities (listed in the Market Definition/Description section above).

Product capabilities were scored as follows:

- 0 = rudimentary capability or capability not supported by the data science platform
- 1 = capability partially supported by the data science platform
- 2 = capability fully supported by the data science platform

As there were 15 critical capabilities, a product could achieve a maximum score of 30 points.

Only products that scored at least 20 points were considered for assessment in this Magic Quadrant. Also, because the number of vendors that can appear in a Magic Quadrant is limited, only products with critical capability scores among the top 16 continued to the detailed evaluation phase.

(If Gate 3 had resulted in a tie between three or fewer products, we would have included those products in the assessment, thereby increasing the maximum number of vendors in the Magic Quadrant to 18. If more than three products had received the same score, we would have used internet search, Gartner search and inquiry data to determine which vendor's products had the greater market traction. In no case would more than 18 vendors be included in the Magic Quadrant.)

This year, 17 products from 16 vendors qualified for assessment. Only SAS had more than one qualifying product: SAS Enterprise Miner and the SAS Visual Analytics suite.

Honorable Mentions

Vendors that did not qualify for inclusion in this Magic Quadrant, but which clients may also wish to consider, include:

- Amazon
- BigMLDassault Systèmes
- Databricks
- DataRobot
- Google
- Infosys
- Oracle
- Salford Systems
- Yhat

Evaluation Criteria

Ability to Execute

The Ability to Execute criteria have the same weightings as last year (see Table 1).

Product or Service: A use-case-weighted average of the scores in "Critical Capabilities for Data Science Platforms" (forthcoming).

Overall Viability: An evaluation of the viability of best-of-breed vendors and of the importance of this product line to other, larger vendors.

Sales Execution/Pricing: An evaluation based on customer feedback about the vendor's sales process and pricing.

Market Responsiveness/Record: An evaluation based on the size of the vendor's active customer base and sales traction since last year.

Marketing Execution: An evaluation based on how well the vendor's product has achieved market awareness and the market's understanding of its value proposition.

Customer Experience: An evaluation of feedback from users about their overall satisfaction with the vendor and its product, and the product's integration.

Operations: An evaluation based on customer feedback about the vendor's product development process and customer support capabilities.

Table 1. Ability to Execute Evaluation Criteria

Evaluation Criteria	Weighting
Product or Service	High
Overall Viability	Medium
Sales Execution/Pricing	Low
Market Responsiveness/Record	Medium
Marketing Execution	Medium
Customer Experience	High
Operations	Medium

Source: Gartner (February 2017)

Completeness of Vision

The Completeness of Vision criteria have the same weightings as last year (see Table 2).

Market Understanding: An evaluation of the vendor's market understanding and how prepared it is to track the market's evolution.

Marketing Strategy: An evaluation of the vendor's marketing strategy, which contributes to its overall vision for this market.

Sales Strategy: An evaluation of the vendor's sales strategy, which contributes to its overall vision for this market.

Offering (Product) Strategy: An evaluation of the vendor's product strategy, which contributes to its overall vision for this market and includes the product roadmap.

Innovation: An evaluation of the vendor's support for emerging areas of innovation, such as alternative data types, Python and Spark integration, analytic marketplaces, collaboration and edge deployment.

Table 2. Completeness of Vision Evaluation Criteria

Evaluation Criteria	Weighting
Market Understanding	Medium
Marketing Strategy	Low
Sales Strategy	Low
Offering (Product) Strategy	High
Business Model	Not Rated
Vertical/Industry Strategy	Not Rated
Innovation	High
Geographic Strategy	Not Rated

Source: Gartner (February 2017)

Quadrant Descriptions

Leaders

Leaders have a strong presence and significant mind share in the market. Resources skilled in their tools are readily available. Leaders demonstrate strength in depth and breadth across a full model

development and implementation process. While maintaining a broad and long-established customer base, Leaders are also nimble in responding to rapidly changing market conditions, driven by the overwhelming interest in data science across all industries and domains.

Leaders are in the strongest position to influence the market's growth and direction. They address all industries, geographies, data domains and use cases. This gives them the advantage of a clear understanding and strategy for the data science market, with which they can become disrupters themselves and develop thought-leading and differentiating ideas.

Leaders are suitable vendors for most organizations to evaluate. They should not be the only vendors evaluated, but at least two are likely to be included in the typical shortlist of five to eight vendors.

Challengers

Challengers have an established presence, credibility, viability and robust product capabilities. They may not, however, demonstrate thought leadership and innovation to the same degree as Leaders.

There are two main types of Challenger:

1. Long-established data science vendors that succeed because of their stability, predictability and long-term customer relationships. They need to revitalize their vision to stay abreast of market developments and become more broadly influential. If they simply continue doing what they have been doing, their growth and market presence may decrease.
2. Vendors well-established in adjacent markets that are entering the data science market with solutions that can reasonably be considered by most customers. As these vendors prove they can influence this market, they may eventually become Leaders. These vendors must avoid the temptation to introduce new capabilities quickly but superficially.

Challengers are well-placed to succeed in this market. However, their vision may be hampered by a lack of market understanding, excessive focus on short-term gains, and strategy- and product-related inertia. Equally, their marketing efforts, geographic presence and visibility may be lacking, in comparison with Leaders.

Visionaries

Visionaries are typically smaller vendors or newer entrants which embody trends that are shaping, or will shape, the market. There may, however, be concerns about these vendors' ability to keep executing effectively and to scale as they grow. They might also be hampered by a lack of awareness of them in the market, and therefore by insufficient momentum.

Visionaries have a strong vision and a roadmap for achieving it. They are innovative in their approach to their platform offerings and provide strong functionality for the capabilities they address. Typically, however, there are gaps in the breadth and completeness of their capabilities.

Clients should consider Visionaries because these vendors might:

- Represent an opportunity to skip a generation of technology
- Provide some compelling capability that offers a competitive advantage as a complement to, or substitute for, existing solutions
- Be more easily influenced with regard to their product roadmap

But Visionaries also pose increased risk. They might be acquired, or they might struggle to gain momentum, develop a presence or increase their market share.

As Visionaries mature and prove their Ability to Execute over time, they may eventually become Leaders.

Niche Players

Niche Players demonstrate strength in a particular industry, or pair well with a specific technology stack.

Some Niche Players demonstrate a degree of vision, which suggests they might become Visionaries, but they are struggling to make this vision compelling. They may also be struggling to develop a track record of continual innovation.

Other Niche Players have the opportunity to become Challengers if they continue to develop their products with a view to improving their overall execution.

Context

The data and analytical needs of organizations continue to evolve rapidly. Many organizations are extending the breadth of their analytical capability to include data science.

Similarly, the data science platform market is in a state of flux. In addition to the traditional market leaders, emerging vendors provide strong capabilities and solutions.

Even Leaders are not immune to potential market disruption, now that the hype about machine learning and artificial intelligence is at its peak and new entrants with substantial resources, such as Amazon, Baidu and Google, are targeting the data science market. Open-source software, which permeates this market, adds to the potential for disruption: For example, Spark, which became a top-level Apache project in 2014, is now a trend-setter, a foundation for most data science advances. The Leaders may also encounter disruption from the proliferation of startups, 2016 having brought huge growth in venture capital investment in this market.

Users of data science platforms should keep their finger on the pulse of this changing market and their changing internal needs. They should evaluate their current platform's ability to deliver additional, extended capabilities and address new needs. They should only consider purchasing a new platform to replace or augment their existing capability if there is a specific business need for this.

Organizations new to the field of data science should look beyond the Leaders to see if there is a specialized solution that meets their specific needs (see "Machine-Learning and Data Science Solutions: Build, Buy or Outsource?").

Market Overview

Data science, analytics, machine learning and artificial intelligence are the engines of the future. They represent a springboard for a growing number of businesses and a generation of software engineers and algorithmic business designers.

Revenue growth in the data science platform segment was almost double that of the overall BI and analytics software market in 2015, with a 20.8% rise bringing its total to \$2.4 billion in constant currency (see "Market Snapshot: BI and Analytics Software, Worldwide, 2016"). Gartner estimates that, by 2020, predictive and prescriptive analytics will attract 40% of enterprises' new investment in BI and analytics technologies.

Key drivers of the data science market's strong growth include:

- **Increased digitalization:** The imperative for organizations to transform into digital businesses is forcing them to move beyond descriptive and diagnostic analytics into more complex use cases that require predictive and prescriptive analytics.
- **Access to far more information:** With access to an explosion of information from an increasingly wide range of sources (applications, sensors, log files, smartphones and more), data scientists have stockpiles of data with which to fuel more advanced algorithms. In addition, the cost of storing data continues to decrease rapidly, making it cost-effective to hold on to more historical data.
- **Increased computing power:** The availability of advanced system architectures, in-memory computing and storage, as well as more powerful and efficient chipsets in a highly scalable, cloud-based architecture, has removed many of the inhibitors of infrastructure implementation by organizations. As a result, machine-learning solutions are now vastly more powerful and affordable.
- **Wide availability and affordability of machine-learning tools:** Once, machine-learning tools were prohibitively expensive and complex, but now they are relatively inexpensive — or even free. Data scientists and researchers can now choose from a growing number of open-source and commercial options.

The market for data science platforms is one of the most vibrant and collaborative technology markets. It embraces open-source technologies. Organizations often start by using open-source offerings to build up their knowledge and explore possibilities. They then often buy commercial versions to tackle broader use cases, team collaboration, and deploy and manage models in production environments.

Open-source technology is also fundamental to this market in several other aspects:

- All the vendors in this Magic Quadrant — indeed, in the market as a whole — have moved to include data in the open-source Hadoop ecosystem, which is now considered first-class. As such, it is equal in status to the proprietary stores that are predominantly used for traditional data warehouses, which are rapidly becoming part of the logical data warehouse architecture that combines the strengths of traditional repository warehouses with alternative data management and access (see "Magic Quadrant for Data Warehouse and Data Management Solutions for Analytics").
- Spark is becoming a de facto data science foundation for the vendors in this Magic Quadrant, as well as for other participants in this market. Analytics leaders should understand how Spark could benefit their data science projects, regardless of their choice of data science platform vendor. This understanding will help them migrate between platforms, if necessary, while preserving their core advanced analytics systems. Alternatively, they could start with Spark and only later select a data science platform vendor, as many organizations already do.
- Open-source languages — Python, R and Scala — dominate this market. Almost all data science platform vendors support Python and R, and many of the vendors in this Magic Quadrant also support Scala.

Many data science platform vendors are innovative. Nevertheless, organizations still face key challenges in terms of a lack of qualified staff with data science skills, long cycle times for building and deploying data science models, and difficult collaboration between business teams and data science teams. Many data science vendors are addressing these challenges by adding capabilities to their platforms that make it easier for citizen data scientists to build and deploy models visually, without coding, along with tools to make it easier for data scientists and business users to collaborate. They are also adding more capabilities to automate data preparation, exploration, model building, model running, and even the generation of presentation-quality results.

With so many data science platform options available, and with open-source options creating pricing pressure, it is inevitable that this market will consolidate during the next two years. Some of the smaller vendors will probably survive by focusing on specific market niches (industries or groups of users). The larger vendors will have to offer differentiated functionality to justify any premium over open-source solutions.

Gartner Recommended Reading

Some documents may not be available as part of your current Gartner subscription.

"How Markets and Vendors Are Evaluated in Gartner Magic Quadrants"

"Machine-Learning and Data Science Solutions: Build, Buy or Outsource?"

"Toolkit: RFP for Data Science Platforms"

"Survey Analysis: How Customers Rate Their Advanced Analytics Platforms, 2016"

"Staffing Data Science Teams"

"How Data Science Teams Leverage Advanced Analytics"

"Innovation Insight for Deep Learning"

"Hype Cycle for Data Science, 2016"

Evidence

Information used in this Magic Quadrant came from many sources:

- Gartner analysts undertake thousands of interactions with hundreds of clients in the data science sector. We also interact with many vendors. To a large degree, this report reflects what we have learnt from such interactions.

We also:

- Conducted a series of briefings and demonstrations with the selected vendors.
- Studied the instruction manuals and further documentation provided by the vendors.
- Analyzed survey responses completed by over 600 organizations, regarding their experience of working with the selected vendors and their products.
- Selectively interviewed reference clients about their experiences with those products.

Evaluation Criteria Definitions

Ability to Execute

Product/Service: Core goods and services offered by the vendor for the defined market. This includes current product/service capabilities, quality, feature sets, skills and so on, whether offered natively or through OEM agreements/partnerships as defined in the market definition and detailed in the subcriteria.

Overall Viability: Viability includes an assessment of the overall organization's financial health, the financial and practical success of the business unit, and the likelihood that the individual business unit will continue investing in the product, will continue offering the product and will advance the state of the art within the organization's portfolio of products.

Sales Execution/Pricing: The vendor's capabilities in all presales activities and the structure that supports them. This includes deal management, pricing and negotiation, presales support, and the overall effectiveness of the sales channel.

Market Responsiveness/Record: Ability to respond, change direction, be flexible and achieve competitive success as opportunities develop, competitors act, customer

needs evolve and market dynamics change. This criterion also considers the vendor's history of responsiveness.

Marketing Execution: The clarity, quality, creativity and efficacy of programs designed to deliver the organization's message to influence the market, promote the brand and business, increase awareness of the products, and establish a positive identification with the product/brand and organization in the minds of buyers. This "mind share" can be driven by a combination of publicity, promotional initiatives, thought leadership, word of mouth and sales activities.

Customer Experience: Relationships, products and services/programs that enable clients to be successful with the products evaluated. Specifically, this includes the ways customers receive technical support or account support. This can also include ancillary tools, customer support programs (and the quality thereof), availability of user groups, service-level agreements and so on.

Operations: The ability of the organization to meet its goals and commitments. Factors include the quality of the organizational structure, including skills, experiences, programs, systems and other vehicles that enable the organization to operate effectively and efficiently on an ongoing basis.

Completeness of Vision

Market Understanding: Ability of the vendor to understand buyers' wants and needs and to translate those into products and services. Vendors that show the highest degree of vision listen to and understand buyers' wants and needs, and can shape or enhance those with their added vision.

Marketing Strategy: A clear, differentiated set of messages consistently communicated throughout the organization and externalized through the website, advertising, customer programs and positioning statements.

Sales Strategy: The strategy for selling products that uses the appropriate network of direct and indirect sales, marketing, service, and communication affiliates that extend the scope and depth of market reach, skills, expertise, technologies, services and the customer base.

Offering (Product) Strategy: The vendor's approach to product development and delivery that emphasizes differentiation, functionality, methodology and feature sets as they map to current and future requirements.

Business Model: The soundness and logic of the vendor's underlying business proposition.

Vertical/Industry Strategy: The vendor's strategy to direct resources, skills and offerings to meet the specific needs of individual market segments, including vertical markets.

Innovation: Direct, related, complementary and synergistic layouts of resources, expertise or capital for investment, consolidation, defensive or pre-emptive purposes.

Geographic Strategy: The vendor's strategy to direct resources, skills and offerings to meet the specific needs of geographies outside the "home" or native geography, either directly or through partners, channels and subsidiaries as appropriate for that geography and market.

GARTNER HEADQUARTERS**Corporate Headquarters**

56 Top Gallant Road
Stamford, CT 06902-7700
USA
+1 203 964 0096

Regional Headquarters

AUSTRALIA
BRAZIL
JAPAN
UNITED KINGDOM

For a complete list of worldwide locations,
visit <http://www.gartner.com/technology/about.jsp>

© 2017 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. or its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. If you are authorized to access this publication, your use of it is subject to the [Usage Guidelines for Gartner Services](#) posted on gartner.com. The information contained in this publication has been obtained from sources believed to be reliable. Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information and shall have no liability for errors, omissions or inadequacies in such information. This publication consists of the opinions of Gartner's research organization and should not be construed as statements of fact. The opinions expressed herein are subject to change without notice. Although Gartner research may include a discussion of related legal issues, Gartner does not provide legal advice or services and its research should not be construed or used as such. Gartner is a public company, and its shareholders may include firms and funds that have financial interests in entities covered in Gartner research. Gartner's Board of Directors may include senior managers of these firms or funds. Gartner research is produced independently by its research organization without input or influence from these firms, funds or their managers. For further information on the independence and integrity of Gartner research, see "[Guiding Principles on Independence and Objectivity](#)."