



# How to Choose the Best Cloud Provider

## *Data Science and Machine Learning Platforms*

---

Data, IT and E-com teams can use to source data, build models, and operationalize machine learning. It will help C Levels make the right choice for their business from a crowded field in a maturing DSML platform market that continues to show rapid product development.

**Dr.Keskin**

**August 2021**

### Market Definition/Description

Machine learning (DSML) platform is a core product and supporting portfolio of coherently integrated products, components, libraries, and frameworks (including proprietary, partner-sourced and open-source). Its primary users are data science professionals, including expert data scientists, citizen data scientists, data engineers, application developers and machine learning (ML) specialists. The core product and supporting portfolio:

- Are sufficiently well-integrated to provide a consistent “look and feel.”
- Create a user experience in which all components are reasonably interoperable in support of an analytics pipeline.

The DSML platform:



- Offers a mixture of basic and advanced functionality essential for building DSML solutions (primarily predictive and prescriptive models).
- Supports the incorporation of these solutions into business processes, surrounding infrastructure, products, and applications.
- Supports the sustainable consumption of insights derived from the platform and offers functionality to quantify and track the value of data science projects.
- Supports variously skilled data science professionals (“data scientist” is an inconsistently applied job title and professional distinction) — a DSML platform’s user base is often made up of professionals with diverse technical and business backgrounds.
- Supports multiple tasks across the data science life cycle, including:
  - Problem and business context understanding
  - Data ingestion
  - Data preparation
  - Data exploration
  - Feature engineering
  - Model creation and training
  - Model testing
  - Deployment
  - Monitoring
  - Maintenance
  - Data and model governance
  - Explainable artificial intelligence (XAI)
  - Business value tracking
  - Collaboration



# BEST CLOUD PROVIDER

Figure 1: Best Cloud Provider for Data Science and Machine Learning Platforms





## Vendor Strengths and Cautions

### Alibaba Cloud

Alibaba Cloud is a Niche Player in this BEST CLOUD PROVIDER LIST. It provides two software products that together make up its core DSML platform: Platform for AI (PAI) Studio and Data Science Workshop.

In addition to China, Alibaba has a strong customer base in the South and Southeast Asian markets, but it does not have many clients elsewhere. Its current platform focuses on applications for the retail, internet and data service sectors.

Alibaba Cloud's product and roadmap are well-suited to expert data scientists and data engineers in sectors such as internet technology, data services, retail and government. Alibaba Cloud emphasizes support for augmentation of certain tasks in the DSML workflow, but its platform lacks functionality and ease of use for citizen data scientists, which slows its adoption by less mature organizations.

### Strengths

- Strong community built in China: Alibaba Cloud showcases its community's strength with the Tianchi platform, a Kaggle-like platform for collaboration, competition and knowledge sharing. The platform is widely adopted within the Chinese market.
- Advanced use-case modeling: Alibaba provides strong solutions for advanced use cases such as image labeling, image recognition and segmentation, and recommendation engines, which can be useful to expert data scientists.
- Seamless integration that creates coherence: Alibaba provides a coherent platform that integrates well with its other offerings for data preparation, exploration, ML, augmentation and delivery. It offers drag-and-drop interactive modeling features across its platforms, which can be used by expert data scientists to support the ML pipeline.

### Cautions

- Geographic strategy: Although Alibaba Cloud has offices and service locations in many countries, the clients it serves are mostly in Asia/Pacific. Prospective customers should ensure they are satisfied with the vendor's presence and support in their region.



- Product vision: Given the current pace of development by other vendors, Alibaba Cloud will have to be swift and agile, as this market is likely to remain highly competitive. Some key themes of, and items on, its product roadmap are already available as standard features from many other vendors.
- Narrow usage and lack of citizen data scientist support: The current PAI Studio and Data Science Workshop offerings offer limited ML and advanced analytics capabilities, such as agent-based modeling, discrete-event modeling, Monte Carlo simulation, support for generative adversarial networks and self-supervised learning. Currently, the platform is suitable for advanced users but may not be a good choice for citizen data scientists or business analysts.

## **Altair**

Altair is a Niche Player in this Best Cloud Provider List. It offers a suite of products called Altair Knowledge Works, and the core product considered in this Best Cloud Provider evaluation is Altair Knowledge Studio. The Knowledge Works suite also includes Knowledge Studio for Apache Spark, Knowledge Hub, Panopticon and Monarch.

Altair's operations are geographically diversified, and the vendor maintains strong offerings for service-centric industries (particularly banking and financial services). It also offers various simulation and high-performance computing solutions that appeal to customers in the automotive, aerospace and manufacturing sectors and to other asset-based organizations.

Knowledge Studio's capabilities for automated ML (AutoML), XAI and enhanced open-source integration have strengthened, but it is still catching up with other vendors' products in terms of native capabilities for delivery and deployment and model management.

### **Strengths**

- Ease of use: Altair Knowledge Studio offers an intuitive, easy-to-use interface for both coders and noncoders. Additionally, by exposing and allowing editing of the underlying open-source code, it enables expert data scientists and data engineers to augment the platform's standard functionality.
- Tools for building strong data pipelines: Altair Knowledge Studio offers strong capabilities for augmented data preparation. Integration with advanced data preparation



tools (Monarch and Knowledge Hub) enables semistructured data to be easily extracted and included in ML modeling. Knowledge Hub also offers strong data governance and metadata management capabilities.

- Operations and customer experience: Altair customers report high satisfaction with the vendor's operations, including in the areas of deployment, service and support. Altair has a team of product design experts who have a keen understanding of customer needs and processes for simulation-based design activities.

### Cautions

- Gaps in current offering: Altair has added new features in areas such as augmented DSML, MLOps and XAI. However, other advanced analytics and delivery capabilities are weaknesses for this vendor. Altair also needs to strengthen its decision modeling and composite artificial intelligence (AI) capabilities.
- Limited resonance across industries: Although Altair has a strong focus on both service- and asset-centric industries, its product marketing needs improvement to resonate with a wider set of client needs. Existing and prospective customers need to work with the vendor to understand its full suite of products, which are applicable to a range of use cases.
- Comparatively slow growth: Adoption of Altair's core Knowledge Works products has been slow, compared with competitors' offerings. Several competitors are sustaining extremely strong growth and offering market-leading products.

### Alteryx

Alteryx is a Challenger in this Best Cloud Provider. It has repositioned its offering by introducing Analytics Process Automation (APA) technology to provide building blocks for automating the analytics process and integrating with applications and robotic process automation (RPA). The platform includes Alteryx Designer, Alteryx Intelligence Suite, Alteryx Server, Alteryx Connect and Alteryx Promote. Alteryx Analytics Hub provides an environment for workflow automation and scheduling, collaboration, multitenancy and data connection management.

Alteryx's operations are geographically diversified, and this vendor has clients in most domains and industries. Top verticals include manufacturing, financial services, consumer packaged goods, retail, healthcare and government.



Alteryx's broad revamping is a work in progress. The newly introduced Alteryx Analytics Hub provides a centralized approach to orchestrating workflow and collaboration when managing analytics and data connection environments.

### Strengths

- Ease of use for diverse personas: A collaborative user experience leveraging code-free and expert modes contributes to ease of use by all personas. Alteryx also provides line-of-business (LOB) and industry solution templates and jump-start kits to accelerate onboarding and use.
- Go-to-market strategy: With APA, Alteryx emphasizes the creation of analytic content and progression from insight to action. Strong channel and independent software vendor partnerships and a verticalized go-to-market strategy, including Alteryx-developed and joint partner-developed solutions, create momentum and increase visibility.
- Customer experience and operational support: Alteryx has consistently delivered excellent functionality and support, judging by feedback from customers. Customers generally respond very positively when asked about their overall experience with Alteryx.

### Cautions

- Changing product portfolio: With the introduction of APA, Alteryx is making many changes to its portfolio. Customers should seek clarification and verify that the evolving APA framework is a good fit for their DSML strategy and users.
- Perceived high cost: Pricing is commonly identified as a concern by Alteryx customers. They report good value for money, but often also evaluate less costly alternatives as their data science initiatives develop.
- Innovation: Although Alteryx has delivered some good innovation with RPA integration, augmentation and a multipersona approach, other vendors are leading the way in terms of cutting-edge ML and key areas such as streaming, the Internet of Things (IoT) and XAI.

### **Amazon Web Services -AWS**

Amazon Web Services (AWS) is a Visionary in this Best Cloud Provider. Its vision is for data science teams to use the entire breadth of the AWS portfolio and ML stack, with Amazon SageMaker at its core. Many of the supporting AWS components and services were considered



in evaluating AWS's offering. These included the SageMaker Studio IDE (which includes Autopilot, Notebooks, Model Monitor, Experiments and Debugger), Amazon EMR (including S3), AWS Glue, Amazon SageMaker Neo, Amazon SageMaker Ground Truth, Amazon SageMaker Clarify, Amazon SageMaker Data Wrangler, Amazon SageMaker Pipelines, AWS CloudWatch, AWS CloudTrail and others.

AWS is geographically diversified, and its client base spans many industries and business functions.

Amazon SageMaker continues to demonstrate formidable market traction, with a powerful ecosystem and considerable resources behind it.

### Strengths

- Breadth and depth of cloud platform: Users can directly leverage AWS's prepackaged AI services (such as Amazon Lex, Polly and Transcribe). SageMaker is also natively integrated with AWS's many cloud data and analytics tools. Additionally, SageMaker provides extensive support for a broad range of popular and niche open-source software (OSS) libraries and frameworks.
- Performance, scalability and granularity of control: Amazon SageMaker and its supporting portfolio offer best-in-class performance and scalability. The platform supports a significant selection of hardware options optimized for various ML and deep learning frameworks, and features a pay-as-you-go pricing model with no minimum fees or upfront commitment, thus encouraging experimentation.
- Data labeling and human-in-the-loop capabilities: Amazon SageMaker Ground Truth supports labeling of training data, and Amazon's Augmented AI (Amazon A2I) helps build optimal workflows for human review of deployed models. AWS connects customers with third-party marketplace vendors and the Amazon Mechanical Turk (MTurk) workforce for human labeling of data.

### Cautions

- Evolving citizen data science appeal: AWS has made its platform more accessible, mainly through Autopilot, Data Wrangler, Pipelines and continued development of the SageMaker Studio IDE. Still, the platform is more popular among coders — it is not as intuitive for nontechnical users, compared with leading tools for citizen data scientists.



- **Rapid pace of development needed to match competitors’ functionality:** AWS’s flurry of new components and services is filling important gaps in its platform. However, these new capabilities are neither as proven nor as strong as other vendors’ capabilities for data preparation, user interfaces, collaboration and coherence.
- **Maturing on-premises, hybrid and multicloud support:** The majority of Amazon SageMaker customers operate in purely cloud environments. Some capabilities within the AWS portfolio change or become more complicated in hybrid, multicloud or on-premises environments. Multicloud support is evolving, however, and today most customers manage data, models and ML workloads within AWS.

## **Anaconda**

Anaconda is a Niche Player in this Best Cloud Provider. It offers Anaconda Enterprise, a data science development environment based on the interactive notebook concept that supports use of open-source Python and R-based packages. (This evaluation excludes the Anaconda Individual Edition, formerly known as Anaconda Distribution Version.)

Anaconda is geographically diversified. The majority of its users are in the financial services sector, but it is also used in sectors such as energy and utilities, healthcare, manufacturing and retail.

Anaconda has made noteworthy innovations in the areas of model governance and scalability. It has partnerships with vendors such as Google, IBM and Microsoft to drive DSML innovation with the use of open-source technologies.

### **Strengths**

- **Trusted and flexible platform:** Anaconda offers a popular and trusted platform within the coding community, one with options for both beginners and experts. The GUI is intuitive, gives access to all R and Python libraries, and offers users the flexibility to work on several IDEs of choice, including Jupyter and RStudio.
- **Optimization of open-source technologies:** To optimize open-source technologies and support scalability, Anaconda provides upscaling options using GPUs, managed within the Anaconda environment. Users can also use Apache Hadoop, Apache Hadoop YARN and Kubernetes clusters, on-premises or in the cloud.



- Culture of collaboration and accompanying features: The Anaconda community supports the Python open-source contributions, thus fostering a culture of code integrity and integration with other open-source data science projects. Anaconda Cloud, for data scientists, provides ways to collaborate, share deployments and exchange code libraries. It also enables developers to explore and accelerate model development and deployment.

### Cautions

- Focus on technical audience: Anaconda targets a technical audience that prefers to code in R or Python languages for data science. The platform lacks features that enable citizen data scientists to take advantage of it.
- Lack of some critical model operationalization capabilities: Anaconda's platform lacks model management capabilities such as dependency management, explainability and bias detection, as well as model inventory features. Anaconda does, however, provide some model-monitoring and governance features, such as scheduling of deployments, user information and resource consumption for ML models (via the scheduler UI).
- Stability: Anaconda users highlight compatibility and runtime issues with the platform. Nonexpert users often find it challenging to keep their projects coherent when new platform or package updates are released.

### Cloudera

Cloudera is a Niche Player in this Best Cloud Provider. It has a core ML product, Cloudera Machine Learning (CML), supported by Cloudera Data Engineering (CDE) and Cloudera Data Visualization (CDV). These products are interconnected and delivered as services on top of the Cloudera Data Platform (CDP). CML has replaced and extended Cloudera's previous on-premises DSML platform, Cloudera Data Science Workbench (CDSW), to provide hybrid and multicloud capabilities.

Cloudera is geographically diversified, and its client base spans many industries and various business functions.

Cloudera's heritage as a big data company is reflected in its ML offering being part of the CDP. The vendor's vision focuses on unifying ML workflows across data warehousing, data engineering, DSML and operationalization.

### Strengths



- Native use of Spark on Kubernetes: With CDE and CML, Cloudera aims to overcome the overhead associated with managing Spark clusters and dependencies by maintaining containerized, repeatable workflows that can be scaled on demand. CML enables data science teams to use a variety of ML runtimes without prescribing underlying frameworks.
- Processing of complex data workloads: CDP is designed for creating and managing high-volume data integration and preparation processes across hybrid and multicloud environments. CML and CDE are part of CDP, and thus provide control over data processing infrastructure and ML execution environments from a single platform.
- Metadata management for DataOps and MLOps: The central framework that enables the building of scalable and repeatable DSML pipelines is Cloudera's Shared Data Experience (SDX), based on Apache Atlas, which stores metadata on each step of execution. An MLOps SDK enables programmatic interaction with SDX.

### Cautions

- Code-first focus: The majority of DSML tasks undertaken in CML require coding and use of open-source libraries in Python, R, Scala and similar languages with no visual workflow interface. There is little augmentation in the platform to help citizen data scientists build their own models.
- Coherence of product offerings: CDP is the platform on which CML and CDE are offered. CDP can also include Cloudera Data Hub (CDH), Cloudera Data Warehouse, Cloudera Operational Database and Cloudera DataFlow. These services may be used to migrate on-premises deployments to the cloud. Even with centralized access to these components from CDP, the learning curve may be steep, even for experts.
- Domain-specific solutions: The prototypes provided by Cloudera's Fast Forward Labs in the form of Applied ML Prototypes (AMPs) are still small in number. The goal of taking cutting-edge ML from research and applying it to enterprise environments in a packaged form has great potential. However, organizations with limited in-house expertise will have to rely on Cloudera's professional services.

### Databricks

Databricks is a Leader in this Best Cloud Provider. Its Unified Data Platform, available in multiple clouds and with an emphasis on scalability, spans data science, ML, analytics and data engineering.



Databricks is geographically diversified, and its client base spans many industries and various business functions.

The company is evolving beyond its perception as merely the leader of the Apache Spark community, as is reflected by the renaming of its Spark + AI Summits as Data + AI Summits. Databricks keeps contributing to the open-source community — for example, by leading the Delta Lake and MLflow projects. It has also extended its offering with the acquisition of Redash, which enables users to query and visualize data more easily, using SQL.

### Strengths

- **Multicloud performance at scale:** Databricks enables its customers to experiment and train their models fast and then to scale them quickly. It offers automanaged and scalable CPU and GPU clusters on multiple cloud platforms, preconfigured with the most popular ML frameworks, with built-in optimizations. MLflow offers flexibility to deploy models to different cloud environments.
- **Empowerment of mature data scientists:** Databricks' notebook-centric vision and optimization of OSS appeals to expert data scientists who demand high performance and early access to the latest innovative ML technology. This appeal is enhanced by an extensive collection of training materials, other documentation and access to a large community of knowledgeable users.
- **Execution and expansion:** Databricks has sustained strong revenue growth, catalyzed by its successful partnerships with Microsoft (Azure), AWS and hundreds of other organizations across the world. The company has a well-executed vertical sales strategy, with strong commitment to customer value creation.

### Cautions

- **Citizen data science support:** Databricks still targets a mainly technical audience of data engineers and data scientists with a coding background. Its platform offers collaboration support and recently also gained new SQL analytics capabilities aimed at data analysts. However, the platform is not well-suited to citizen data scientists and other low/no-code users.
- **Governance and responsible AI:** Databricks offers support for General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA) compliance, and has embedded open-source techniques for bias mitigation and explainability. However, the



company's vision and messaging should pay more explicit attention to the responsible, ethical and trustworthy use of DSML, including the need for governance, risk management and compliance with emerging laws, policies and guidelines.

- Growing competition from cloud partners: Databricks has a cloud-first strategy, offering its customers a choice of a growing number of cloud platforms for scalable ML. However, the company faces growing competition from its key cloud partners (and others), which all have their own DSML offerings and visions that overlap with its.

## **Dataiku**

Dataiku is a Leader in this Best Cloud Provider. Its core product is Data Science Studio (DSS), which provides one platform for all DSML tasks, with a focus on multidisciplinary data science teams, collaboration and ease of use.

Dataiku is geographically diversified, and its client base spans many industries and various business functions.

The company announced a Series D funding round of \$100 million in August 2020. It has also formed partnerships with global system integrators and vendors including Tableau, Snowflake and UiPath. It has a strong roadmap and vision in the areas of responsible AI, collaboration and business applications, one that points to continued growth and innovation.

## **Strengths**

- Understanding of citizen data scientists: Dataiku has added augmented functionality to every stage of the DSML cycle. Citizen data scientists are well-supported with everything from detailed information on data quality and profiling to guided control over AutoML and explainability features. Users who want to build models in a no-code manner have a wealth of tools at their disposal.
- Focus on business value: Dataiku understands the need for performance metrics that go beyond model accuracy and provides the ability to create custom business metrics optimized to deliver a particular business benefit and to monitor concept drift. A newly formed professional services team that focuses on business value is also a testament to Dataiku's vision.
- Increasing market traction: Dataiku remains on an impressive growth trajectory. The company continues to expand its ecosystem of partners in order to build targeted industry



and function-specific analytic solutions. This expansion includes an increase in its OEM and managed service provider (MSP) programs.

## Cautions

- Heavy use of extensions and plugins: Deep learning support within DSS is primarily code-driven using Keras and TensorFlow, while deep learning support for visual users is achieved via an extension to DSS. Extensions are also needed for time-series processing, connecting to certain data sources, signal processing and mining. Navigating and installing these features increases platform management overhead and the complexity of containerized deployments.
- Emerging vision for unifying XOps: Dataiku is working on better capabilities for full model management that can be used by multiple operations teams (those for data, ML, models and platforms). The personas the platform is built for will need to expand, as will associated functionality for deployment in complex heterogeneous environments.
- Pricing model for smaller teams: DSS is available in several versions that have increasing levels of functionality. The prices of versions that do not offer full enterprise capabilities for model scalability and deployment are higher than those of other vendors' offerings that do have these capabilities.

## DataRobot

DataRobot is a Visionary in this Best Cloud Provider. The DataRobot Enterprise AI Platform consists of Paxata Data Preparation, Automated Machine Learning, Automated Time Series, MLOps and AI applications. Its augmented approach enables both citizen and expert data scientists to productively use data science.

DataRobot's operations are geographically diversified. The vendor has a strong presence in the banking, insurance, other financial services, manufacturing, retail, life sciences and healthcare sectors.

DataRobot delivers trusted AI through its Humble AI initiative, which enables prediction quality management. The vendor prioritizes measuring business value with its Use Case Value Tracker, a centralized hub for managing ROI. A Series F funding round in November 2020 and additional investments in December raised \$320 million.

## Strengths



- Sales strategy and execution: DataRobot’s simplified pricing structure and self-service trial drives consistent addition of new customers and growth within its existing client base. In addition, a relationship with Boston Consulting Group (BCG) enables joint sales motion and co-development of industry-specific applications.
- High-touch customer service: DataRobot’s AI Success Plan provides a proven structured delivery approach focused on providing business value. The Customer-Facing Data Science and AI Success teams, repeatable playbooks and user training accelerate delivery of a pipeline of prioritized use cases and help customers achieve value quickly.
- Acquisitions and key developments to fill gaps and link to business value: The acquisition of Paxata, a company focused on data preparation, is the most recent example of DataRobot’s use of acquisitions to fill functionality gaps in its platform. Paxata and all previous acquisitions were quickly incorporated. AI applications and the Use Case Value Tracker help track the value achieved by models.

#### Cautions

- Increasing complexity of product portfolio: DataRobot’s offering now consists of multiple components, including those for data preparation, model development and MLOps. Understanding and using the various components, the capabilities included within each, and their interoperability becomes increasingly complex as the portfolio grows.
- Resource-heavy onboarding: DataRobot’s AI Success Plan and Customer-Facing Data Science and AI Success teams provide an approach that some believe is difficult to scale. In addition, this approach must specifically address and demonstrate the ability to lead clients to self-sufficiency quickly. DataRobot’s new self-service program offers customers a simpler way to get started.
- Capability gaps: Although it is strengthening its capabilities in areas such as model management and data access, DataRobot still has capability gaps in areas such as other advanced analytics, including decision modeling and precanned solutions.

#### Domino

Domino is a Niche Player in this Best Cloud Provider. Its core product is the Domino Data Science Platform, which is supported by Domino Model Monitor to provide end-to-end DSML capabilities in the cloud or on-premises.



Domino's operations are primarily in North America and EMEA. The vendor has a significant presence in the banking, financial services, manufacturing and life sciences sectors, but its platform is used in most industries.

The introduction of Domino Model Monitor in 2020 shows this vendor's clear commitment to enterprise MLOps. Domino's market positioning is deliberate and R&D for its platform will remain focused on large, code-first data science teams.

### Strengths

- Support for large, expert data science teams: Domino's focus on large, code-first data science teams gives it a deep understanding of emerging enterprise needs and challenges. The platform is well liked by chief data and analytics officers — roles that many organizations will add or expand in the near future to set up systems that can orchestrate and govern an entire firm. Collaboration capabilities remain among the best.
- MLOps capabilities and maturity of vision: Domino released Domino Model Monitor in 2020 to supplement the MLOps capabilities of its core platform. The product supports enterprise MLOps, where models are monitored across different teams, deployment infrastructures and languages. The platform's highly auditable workflows enable rapid and targeted responses to deteriorating model health.
- On-premises, hybrid and multicloud support for modern ML deployment: Domino Data Science Platform is the rare code-focused platform that offers a first-class experience both on-premises and in the cloud. Domino is among the best options for supporting complex hybrid and multicloud model development and deployment. Domino provides support for all major cloud providers' versions of Kubernetes, and the vision for the platform's future is cloud-agnostic.

### Cautions

- Support for small and immature data science teams: An organization that does not plan to expand past 20 data scientists on its staff should not consider Domino. Typical Domino deployments involve more than 25 data scientists and ML engineers. The platform is designed to support a highly interactive community of data scientists, not just a few loosely connected individuals or people who depend on augmented approaches to ML.
- Low visibility: Judging from client inquiry levels and Applant.com searches, interest in Domino is consistently low in this market. Additionally, requests for Domino experience



appear in only a small number of data science job postings, compared with competing vendors.

- Increasing flexibility and openness of most vendors: Domino helped transform the definition of DSML platforms into what it is today, with vendors providing curation and optimization of open-source technologies in addition to proprietary functionality. However, a highly flexible, OSS-fueled collaboration hub is now a common platform vision. Domino still receives a top score for flexibility and openness, but differentiation in this area is narrowing.

## Google

Google is a Visionary in this Best Cloud Provider. It offers the Google Cloud AI Platform as its core DSML platform. The platform has an expanded suite of components that includes Cloud Data Fusion, Cloud AutoML, BigQuery ML, AI Platform Notebooks and TensorFlow. Google will launch its unified AI Platform in the first quarter of 2021 (after the cut-off date for evaluation in this Best Cloud Provider). Key features and services that will be released with this new platform include AutoML tables, XAI, AI platform pipelines and other MLOps services.

Google is geographically diversified and its client base spans many industries and various business functions.

Google's Completeness of Vision is boosted by thought leadership in ML research and responsible AI, as well as by the roadmap for its unified AI Platform. The coherence of, and learning curve for, Google's platform are key aspects to monitor in the coming year.

## Strengths

- Responsible AI vision and capabilities: Google has taken a clear thought leadership position in the area of AI explainability and responsibility. Google shares and productizes its learnings on these subjects through responsible AI practices, fairness best practices, technical references and other materials.
- Research contributions and impact: Google's leadership in AI research includes the prominent work of Google Research, Google Brain and DeepMind, as well as ongoing significant contributions to scholarship, open-source projects and communities — TensorFlow, Kubernetes/Kubeflow and Kaggle stand out.
- Consolidation, cohesion and simplification: Google has made a significant effort to reorganize and redesign not just its DSML platform, but also the way it releases software.



The unified AI Platform will seek to address past issues of coherence, interoperability and ease of use. Google has also introduced simplified New Product Introduction (NPI) stages to provide more predictability and transparency about launch timelines.

### Cautions

- Transition of portfolio: Google is developing capabilities for data science professionals at a rapid pace. This means a period of transition and learning for the market in general and adopters of its unified AI Platform in particular. Google's new product release standards and timelines will be put to the test in 2021.
- Steepness of learning curve: Although Google has made improvements in terms of accessibility and augmentation, its platform presents a steep learning curve and requires technical expertise. Supplementary tools for citizen data scientists and developers new to ML may be necessary.
- Maturing on-premises, hybrid and multicloud support: The majority of Cloud AI Platform customers operate in purely cloud environments. Some capabilities of the Cloud AI Platform change and may become more complicated in hybrid, multicloud or on-premises environments. Multicloud support is evolving, and today most customers manage data, models and ML workloads within Google Cloud. New services like BigQuery Omni for viewing data across clouds are indicative of Google's next steps in the multicloud field.

### H2O.ai

H2O.ai is a Visionary in this Best Cloud Provider. H2O Driverless AI is this vendor's commercial product, for which there are additional modules such as MLOps and AutoDoc. H2O.ai also offers open-source products with optional enterprise support, such as the H2O 3 platform and AutoML for ML, Sparkling Water for Spark integration, and Wave for app development. H2O Driverless AI can be extended and customized with open-source or custom-made "recipes."

H2O.ai is geographically diversified. About one-third of its customers can be found in the financial services sector. Other industries are represented more or less equally among the company's client base.



H2O.ai's roadmap and innovation earned it the highest overall score for Completeness of Vision. H2O.ai is a thought leader in the automation and augmentation of DSML, including time-series analysis.

### Strengths

- Vision for value creation: In addition to its vision for democratizing AI through automation and augmentation, H2O.ai has extended its offering with Wave, an open-source product for building AI apps. Wave appeals to the corporate developer community and integrates with H2O AI Hybrid Cloud, Driverless AI and other components. This reflects a strong vision for streamlining value creation with AI, and this vision is further highlighted by H2O.ai's contributions to AI for Good and investments in responsible AI capabilities.
- Extensive augmentation (automation): H2O Driverless AI eases the adoption of DSML by offering augmentation in multiple areas: in addition to augmented feature engineering, model selection and parameter tuning, the company stands out for its sophisticated automation of time-series modeling. In the past year, H2O.ai has invested significantly in augmentation and automation for innovative natural language and image processing.
- Rich XAI: H2O.ai offers multiple explainability capabilities throughout the ML life cycle, not just for modeling but also for feature engineering. Supported methods include K-LIME, LIME-SUP, Shapley, decision tree surrogates, causal graphs, NLP explainability and more.

### Cautions

- Lack of certain data access and preparation capabilities: H2O.ai has room for improvement in terms of data access and aspects of data preparation. These include data refresh, data lineage, access governance, metadata management and data catalogs.
- OEM partner strategy: H2O Driverless AI's capabilities for augmentation and automation depend on OEM partnerships with other DSML vendors. If those vendors' platforms outperform H2O's in terms of capabilities for data preparation, for example, then potential customers may be less inclined to select H2O.ai's.
- Collaboration and cohesion: Expert data scientists, citizen data scientists, developers and other personas may all use different products from H2O's growing portfolio of commercial and open-source products and modules. Despite shared projects and a



common recipe catalog, the platform could benefit from more attention to cross-product, multipersona collaboration and a more cohesive portfolio structure.

## IBM

IBM is a Leader in this Best Cloud Provider. Its core product for this evaluation is IBM Watson Studio on IBM Cloud Pak for Data, a modular, open and extensible platform for data and AI that combines a broad set of descriptive, diagnostic, predictive and prescriptive capabilities

IBM is geographically diversified, and its client base spans many industries and various business functions.

Revamping its offering has taken several years, and competition will remain fierce for IBM. Still, IBM now delivers a modern and comprehensive solution that draws on its roots in SPSS, ILOG CPLEX Optimization Studio and earlier products, and that benefits from a stream of innovations from IBM Research. These reflect a well-rounded vision.

### Strengths

- **Multipersona support:** IBM Watson Studio offers a visual workflow interface or “graphic canvas,” as well as a choice of notebooks, thus enabling data engineers, expert data scientists and citizen data scientists to work together on the same project. ML pipeline activities, from data acquisition to operations, are supported by AutoAI and collaboration, including a catalog for sharing and reusing (meta)data and models.
- **Composite AI vision:** The modular structure of the IBM Watson Studio platform contains, or can be extended by, multiple components for decision augmentation or automation. These components include several ML and other AI frameworks, optimization features, spatio-temporal and graph analytics, natural language features and video/image/audio analysis (in batch or streaming mode). In addition, by including IBM Decision Optimization, the platform supports decision modeling and decision management or rules processing.
- **Comprehensive attention to responsible AI and governance:** IBM offers extensive support for explainability, bias, fairness, accuracy and drift monitoring, synthetic data and differential privacy. Its platform also provides strong governance (and optional risk management) support, with lineage, policies and rules in its catalog, as well as adversarial security.



## Cautions

- **AutoAI scope:** IBM Watson Studio offers automation and augmentation of multiple activities in the ML pipeline, including data selection, imputation, visualization, feature transformation and modeling. However, a few competitors also augment time-series analysis by, for example, using recurrent neural networks and long short-term memory models.
- **Brand restoration:** With its improved Watson Studio, IBM has caught up with and, in some cases, even surpassed its competitors. Nevertheless, data and analytics leaders may still find their ML experts skeptical about the innovativeness of Watson Studio and IBM's ability to keep pace in a dynamic and competitive market.
- **Product-bundling clarity:** Although the cohesion of the modular Watson Studio on IBM Cloud Pak for Data has improved, there remains confusion among potential customers as to which products and licenses are needed for which configurations. This increases concerns about licensing costs.

## KNIME

KNIME is a Visionary in this Best Cloud Provider. Its open-source offering, the KNIME Analytics Platform, focuses on the authoring of DSML workflows and projects. A commercial product, KNIME Server, focuses on automation, deployment and orchestration capabilities.

KNIME is globally diversified, with a strong presence in Europe and the U.S. Its client base spans all industries and company sizes.

KNIME continues to evolve and to develop its vision for bridging the gap between development and production and offering new ways for data scientists and end users to collaborate.

## Strengths

- **Breadth and depth of DSML capabilities:** KNIME has been incrementally building its product for over 10 years, and this shows in the wide range of capabilities provided by the platform. It has almost 4,000 nodes for connecting to different types of data source, transforming and preparing data, ML and other advanced techniques. Very few DSML tasks are not supported by KNIME's platform.
- **Commitment to open-source platform:** The KNIME Analytics Platform is not a limited or restricted version of a full product. Most of the library of components are available for use in the platform at no cost. This provides an ideal way to experiment with DSML



projects — to test and learn — without upfront investment in a particular technology. Scalability can then be achieved through use of the KNIME Server product.

- Coherence of visual workflow: The basic building blocks within the KNIME Analytics Platform are nodes, components and workflows. Everything within the platform, including AutoML, data visualization, interactive apps and deployment models, is built using these blocks and can be broken down into individual components and nodes, with associated metadata, for full transparency.

### Cautions

- Limited customer support for enterprise deployments: KNIME has not expanded as aggressively as other DSML vendors. Although there is an active community answering questions about functionality, enterprise deployments typically require specialist services to increase adoption and ensure the product meets expectations. KNIME relies on partners to deliver these services.
- Vision for responsible AI: KNIME provides a plethora of components for XAI, such as SHapley Additive exPlanations (SHAP), Partial Dependence Pre-processing (PDP) and Individual Conditional Expectation (ICE). However, frameworks, guidance, best practices and research that can be applied by all disciplines within a data science team are lacking.
- Low market traction and sales innovation: The visibility of KNIME to DSML platform buyers remains low. Prospective customers may therefore shortlist newer vendors with more prominent sales and marketing campaigns.

### MathWorks

MathWorks is a Leader in this Best Cloud Provider. Its two major products are MATLAB and Simulink, but only MATLAB met the inclusion criteria for this Best Cloud Provider.

MathWorks is geographically diversified. Its clients are primarily engineering and asset-centric organizations.

MathWorks demonstrates a clear vision and thought leadership in asset-centric industries. Its innovations are applied, at scale, for large use cases intended to solve real-world problems.

MathWorks is one of the few vendors in the DSML market that can handle large, distributed,



real-time IoT implementations with a continuous environment from the edge to the cloud, and from development to simulation and operationalization and back.

### Strengths

- Robust composite AI capabilities: MATLAB is among the most advanced DSML platforms for developing, integrating and deploying ensembles of AI techniques within a single solution (an approach that Applant calls composite AI). MathWorks combines these techniques in a flexible infrastructure that supports largely distributed environments, from the edge to the data center and the cloud.
- Integrated domain knowledge: MathWorks benefits from deep domain expertise, which it integrates into its DSML platform. From predictive maintenance to fleet analytics, manufacturing process analytics and risk management, the company handles domain-specific idiosyncrasies within its platform, while developing technologies and application-specific toolboxes.
- Verifiable and reliable ML: Safety is typically critical in the asset-centric domains in which MathWorks is active — they have no tolerance for unreliable operations. Beyond interpretability, MathWorks enables engineers to interact with models through either web applications or simulation environments.

### Cautions

- Interface democratization: MATLAB remains the preserve of data-science-initiated engineers and scientists, who essentially use notebooks to develop models. To widen the appeal of its powerful platform to citizen data scientists and business and operations specialists, MathWorks will have to modernize its UI and provide visual development features.
- Interpretable AI: MathWorks remains behind many of its competitors, especially those in the Leaders Cloud, when it comes to model interpretability and fairness management. Even its asset-centric audience will soon require better capabilities, so the company will have to start focusing on this issue.
- Augmented DSML capabilities: Despite progress in 2020, MathWorks remains behind many of its competitors when it comes to expanding its augmented DSML functions, particularly for feature engineering and deployment optimization.



## Microsoft

Microsoft is a Visionary in this Best Cloud Provider. The core product considered in this Best Cloud Provider is Azure Machine Learning (Azure ML). The supporting portfolio of products for Azure ML includes Azure Data Factory, Azure Data Catalog, Azure HDInsight, Azure Databricks, Azure DevOps, Power BI and other components.

Microsoft is geographically diversified, and its client base spans many industries and various business functions.

Microsoft earns the highest Ability to Execute score of the large cloud providers. Microsoft has a strong combination of vision and tailored functionality for the full spectrum of data science professionals who contribute to multifunctional teams.

### Strengths

- Azure stack support for enterprise DSML: Azure ML and its supporting portfolio offer strong capabilities for the needs of enterprise data science. MLOps capabilities include a registry of packages and models and support for streamlined creation of reproducible ML pipelines. Azure ML comes with differentiated security and governance capabilities and, combined with Azure Cloud management services, supports compute quota and cost management capabilities.
- Multipersona vision and offering: Microsoft's vision and current offering for multipersona data science is stronger than those of its closest competitors. Azure ML provides augmented DSML and a drag-and-drop designer for citizen data scientists, and flexible notebook and SDK options for expert data scientists. Microsoft's suite of ancillary products provides a strong environment in which data engineers, ML engineers and architects, corporate developers and others can contribute to the DSML workflow.
- Openness and partnerships: Microsoft goes beyond widespread support for popular OSS by investing in and contributing to a number of prominent projects (such as, Open Neural Network Exchange [ONNX], InterpretML and MLflow). The Azure Databricks product and partnership has been successful for both partners. Azure will also be the preferred cloud provider for the SAS Viya platform, which will be integrated with Azure and Azure ML services.

### Cautions



- Requirement for Azure services commitment and expertise: Azure ML relies on a variety of Azure services and modules and can work with data from any source. Azure ML customers typically use Azure Data Factory for integration and transformation, Azure Data Catalog for governance, and any DevOps system (often Azure DevOps or GitHub) for integration within web services and other services. Supporting this portfolio requires significant technical expertise and understanding of the Azure ecosystem.
- Evolving on-premises, hybrid and multicloud support: The majority of Azure ML customers operate in purely cloud environments. Some capabilities within the Azure ML portfolio change or become more complicated in hybrid, multicloud or on-premises environments. Microsoft's multicloud support is evolving, however, and most of its customers manage data, models and ML workloads within Azure.
- Augmented DSML: Microsoft delivers solid support for citizen data scientists and likely lower total cost of ownership (TCO) for augmentation capabilities, but still has room to improve, compared with vendors that focus solely on data science. Organizations seeking to broaden their data science talent base need to understand how much augmentation is offered by the visual designer in Azure ML, as opposed to the SDKs and Power BI.

## **RapidMiner**

RapidMiner is a Visionary in this Best Cloud Provider. RapidMiner Studio is the vendor's primary model development tool and is available as both a free edition and a commercial edition. For the enterprise, offerings can be extended through the RapidMiner AI Hub, which includes collaboration and governance capabilities, as well as RapidMiner Go and RapidMiner Notebooks, which are model development experiences for novices and coders respectively. Turbo Prep, Auto Model and Automated Model Ops are augmented features of the platform, while the RapidMiner AI Cloud offers flexible, cloud-based deployment options.

RapidMiner is geographically diversified and has a strong presence in many industries, but especially manufacturing, life sciences, banking, insurance, energy, business services, government and education.

RapidMiner's latest capabilities and roadmap exemplify key market trends, such as multipersona collaboration, XAI and model governance.

### **Strengths**



- **Multipersona collaboration:** RapidMiner makes it easy for expert data scientists and citizen data scientists to work on its platform collaboratively and to manage end-to-end data science pipelines. The vendor offers a certification program through its RapidMiner Academy to help those who are not data scientists understand the product, model development operationalization and governance.
- **Clear vision and delivery of aligned features:** RapidMiner has made significant changes to its product portfolio during the past year. Particularly strong new capabilities are FeatureMart and Feature Catalog, which enable users to perform automated feature engineering, and share and store features across an organization, thus enhancing reusability and reproducibility.
- **Explainable, governed and secured AI:** RapidMiner provides features that enable users to explain and govern their models in development and production, thus giving them greater transparency and more control over insights. Additionally, features such as single sign-on and strong identity and access management capabilities help secure the AI pipeline.

#### Cautions

- **Growth rate and outreach:** RapidMiner has grown slowly, relative to other vendors with comparable value propositions and the overall market. Although RapidMiner's retention rate remains competitive, existing and prospective customers should check that RapidMiner continues to match the relentless pace of innovation in this market.
- **Market-standard advanced analytics capabilities:** RapidMiner has market-standard functionalities built in for use cases involving reinforcement learning, generative adversarial networks, small data ML, geospatial analytics and agent-based modeling.
- **Perception as an academic platform:** RapidMiner's strong presence in the academic world continues to cultivate a large user community, with many young and talented data scientists attracted to the free version of its platform. Prospective enterprise customers should not overlook RapidMiner or dismiss its capabilities as an enterprise platform provider in favor of newer vendors that market products solely to enterprises.

#### **Samsung SDS**

Samsung SDS is a Niche Player in this Best Cloud Provider. Brightics AI is the end-to-end analytics and data science platform evaluated for this Best Cloud Provider. Samsung SDS offers Brightics Standard and Enterprise editions and an open-source tool, Brightics Studio. The



Standard edition is a lightweight version of the Enterprise edition, with support for only Python. The Enterprise edition offers support for Python and Spark and enables distributed processing of ML workloads.

Samsung SDS has global operations. Its customer base is concentrated in Asia, especially in the manufacturing and financial services industries.

Brightics AI is an easy-to-use platform for both experts and citizen data scientists. Its focus on data management also enables other roles, such as data engineers and industrial users, to work with it.

### Strengths

- **Comprehensive ecosystem vision:** The Brightics AI platform represents one of Samsung SDS's five key technology areas — AI, blockchain, cloud, data analytics and security (ABCDS) — which comprise its Digital Transformation Framework. The vendor aims to provide a holistic solution by complementing Brightics AI with other Samsung SDS offerings, such as Samsung Cloud and Brightics IoT.
- **Data capabilities:** Samsung SDS helps clients achieve more value through its focus on the data life cycle. Data access, preparation and visualization are strengths of the Brightics AI platform. It offers good support for semistructured and unstructured data, with capabilities like automatic data labeling and an automatic schema builder.
- **Ease of use and collaboration:** Samsung SDS's platform provides an intuitive, easy-to-use interface for both coders and noncoders. It supports multipersona collaboration through a wizard for data scientists, apps for business users and APIs for application developers. It provides a container-based personal sandbox environment for each user that allocates server resources for experimentation.

### Cautions

- **Need for expansion into new markets:** Although Samsung SDS has offices and global delivery centers in many countries, the clients it serves currently are mostly in Asia. Prospective customers should closely examine whether Brightics AI would be a good choice for the parts of the world where they operate.
- **Gaps in product vision:** Brightics AI is behind competing offerings in areas like composite AI and decision intelligence. Although Samsung SDS has increased its focus



in areas such as collaboration and augmented DSML, current capabilities and planned innovations need strengthening to respond to the demands of a rapidly evolving market.

- Limited support for ModelOps and explainability: Brightics AI lacks key capabilities like A/B testing, rollback automation and certain model telemetry features. It currently offers low model explainability, and support for certain key features such as Local Interpretable Model-Agnostic Explanations (LIME) and SHAP is still on the roadmap. Given the market's focus on robust operationalization of ML models and XAI capabilities, prospective customers should look for clear improvements in these areas.

## SAS

SAS is a Leader in this Best Cloud Provider. SAS Visual Data Mining and Machine Learning (VDMML) is the core product evaluated for this Best Cloud Provider. As part of the SAS Viya portfolio, VDMML is included in various product bundles on SAS Viya, namely SAS Visual Machine Learning, SAS Visual Data Science, SAS Data Science Programming and SAS Visual Data Decisioning.

SAS is geographically diversified, and its client base spans many industries and various business functions.

SAS is the longest-standing Leader in this Best Cloud Provider. It maintains a strong and adaptive position, given its keen understanding of the market and its thought leadership in key areas such as composite AI, MLOps and decision intelligence. The company recently announced a partnership with Microsoft to support closer integration with Azure.

### Strengths

- Market understanding and presence: SAS's long standing and experience in this market have earned customers' trust. SAS offers enterprise-grade platform capabilities and support, coupled with a robust vision for key market trends, including composite AI, decision intelligence and MLOps. The domain expertise embedded in its products and consulting services enable customers to derive value from the entire analytics life cycle.
- Cloud-native architecture and open-source integration: The latest release of SAS Viya offers a fully cloud native approach. SAS customers can now leverage all Viya capabilities in a flexible container-based architecture that runs in the cloud. SAS offers innate integrations with popular open-source tools and languages for data, modeling and model management.



- Automated feature engineering and modeling: SAS provides differentiated automated feature engineering and automated modeling capabilities through automated pipeline generation. Experimentation is supported through utilities such as the Data Science Pilot Action Set and other modules. For automated hyperparameter tuning, Model Composer uses a patented hybrid search strategy.

### Cautions

- Perceived high costs: SAS's pricing remains a concern for many customers, who therefore investigate less-costly alternatives. VDMML has historically been priced by the core, but a new pricing model has eliminated the core capacity restriction and pricing is now based on type of user. SAS customers should work with the vendor to determine whether the new pricing model is more suitable for their requirements.
- Product bundling: Although SAS has streamlined its product portfolio, with more "fit for use" product bundling replacing "a la carte" selection, SAS Viya's full suite of products and add-ons remains complex for users to navigate. However, to make navigation easier, SAS VDMML is now part of product bundles that offer programming-only interfaces, as well as bundles that offer both programming and visual, drag-and-drop interfaces.
- Marketing strategy: SAS needs to work on the perception of its product portfolio. Despite clear modernization, SAS is still frequently perceived as a vendor of legacy software and traditional advanced analytics. Small and midsize companies should explore case studies from customers in similar segments to understand current usage of SAS products.

### TIBCO Software

TIBCO Software is a Leader in this Best Cloud Provider. After tightly stitching together various data and analytics software and platforms, TIBCO is fulfilling its "Connected Intelligence" vision. That vision is embodied at its core in TIBCO's Data Science platform, along with TIBCO Spotfire and TIBCO Streaming and a robust data and process infrastructure.

TIBCO is geographically diversified and present in many industries. but has a stronger presence in asset-centric industries, given its science and engineering focus, especially on edge computing.

The company's origins in the middleware sector give TIBCO an edge when it comes to model deployment and production, in any environment, centralized or distributed, across a wide variety of use cases.



## Strengths

- **Leading-edge DSML capabilities:** From innovations like dynamic learning on event streams to integration with popular edge platforms like those of Microsoft and AWS, TIBCO delivers leading IoT capabilities on its TIBCO Data Science platform. Through its TIBCO LABS program, the company has launched initiatives like Project Air, streamlining IoT solutions from the edge to the cloud.
- **Hyperconvergence and integration:** TIBCO extends its Data Science platform from both an infrastructure perspective (in relation to edge analytics, for example) and from an analytical angle through its business intelligence (BI) and strong visualization capabilities. TIBCO has a leading vision for the colliding worlds of data science and analytics.
- **Support for collaboration and applied analytics:** TIBCO is a strong choice for analytical teams that span a wide range of functions across an organization. That strength extends beyond the integrated technical environment where analytical assets can be shared to the capture of domain expertise — the results of collaboration with subject matter experts can then be embedded within integrated applications.

## Cautions

- **End-to-end ModelOps capabilities:** TIBCO has made important progress toward achieving robust ModelOps capabilities with functions such as those of TIBCO Artifact Management Server and improvements to its ML pipeline capabilities. But it needs to provide a more comprehensive and approachable ModelOps capability in order to manage the full life cycle of AI models.
- **Citizen data science support:** Despite the capacity of TIBCO's portfolio to unify technical and operational talents, the Data Science platform still needs a more approachable interface for citizen data scientists. The canvas interface, combined with simplified AutoML functions, forms a solid base, but TIBCO Data Science is still aimed at data scientists with significant ML experience — support for citizen data science continues to rely on other parts of the portfolio.
- **Financial growth in 2020:** Like many organizations, TIBCO had a challenging 2020 in terms of licensing revenue, while the difficult economic conditions impacted its subscription business. The company will need to make strong and continuous development investments in this fast-moving market in order to stay ahead of well-



funded competitors. Before investing in TIBCO's technology, organizations should compare their requirements with the vendor's technology roadmap.

## Vendors Added and Dropped

We review and adjust our inclusion criteria for Best Cloud Providers as markets change. As a result of these adjustments, the mix of vendors in any Best Cloud Provider may change over time. A vendor's appearance in a Best Cloud Provider 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

- Alibaba Cloud
- Amazon Web Services
- Cloudera
- Samsung SDS

### Dropped

None

## Inclusion and Exclusion Criteria

Best Cloud Providers identify and analyze the most relevant providers in a market. By default, an upper limit of 20 vendors is imposed to enable identification of the most relevant providers. On some specific occasions, however, this upper limit may be raised when the Best Cloud Provider's value to clients would otherwise be diminished.

The inclusion criteria represent the specific attributes necessary for inclusion in this Best Cloud Provider. They were applied progressively, in sequence and in cumulative fashion to aid identification of the most relevant providers.

Inclusion Criterion 1: Data Science and Machine Learning Platform



As noted earlier in the Market Definition/Description, a vendor's DSML platform had to:

- Offer a mixture of the basic and advanced functionality essential for building DSML solutions (primarily predictive and prescriptive models).
- Support the incorporation of these solutions into business processes, surrounding infrastructure, products and applications.
- Support the sustainable consumption of insights derived from the platform and offer functionality to quantify and track the value of data science projects.
- Support variously skilled data science professionals (“data scientist” is an inconsistently applied job title and professional distinction — a DSML platform’s user base is often made up of professionals with diverse technical and business backgrounds).
- Support multiple tasks across the data science life cycle, including:
  - Problem and business context understanding
  - Data ingestion
  - Data preparation
  - Data exploration
  - Feature engineering
  - Model creation and training
  - Model testing
  - Deployment
  - Monitoring
  - Maintenance
  - Data and model governance
  - Explainable AI (XAI)
  - Business value tracking
  - Collaboration



Additionally, a vendor had to be able to provide technical support for its DSML platform directly and/or via commercial support partners.

#### Inclusion Criterion 2: Revenue and Growth

A vendor's core product had to offer one or more common license models:

- Perpetual license model
- SaaS subscription model
- Consumption-based model or other type of model

The following information for the core product was considered:

- Revenue, in U.S. dollars, generated from perpetual licenses during 2019., This included software license, maintenance and update revenue, but excluded hardware and professional services revenue.
- Annual contract value (ACV), in U.S. dollars, generated from SaaS subscriptions in 2019, excluding any professional services included in annual contracts. For multiyear contracts, only the contract value for the first 12 months was used for this calculation.

A vendor needed to have either:

- At least \$75 million in combined perpetual license revenue and ACV for 2019

Or:

- At least \$10 million in combined perpetual license revenue and ACV for 2019 calendar year AND at least 18% in combined revenue growth, compared with 2018

Only core products that passed Inclusion Criterion 2 were considered for Inclusion Criterion 3.

#### Inclusion Criterion 3: Customer Counts

Vendors that satisfied the requirements of Inclusion Criterion 2 were next evaluated on their customer counts. We required significant cross-industry and cross-geographic traction for each core product under consideration. Counts included only active unique customer organizations using the latest version of the core product or a version released in the 12 months prior to August 2020.

##### *Cross-Industry Customer Count*



We assessed counts of the number of active unique customer organizations (logos) using each of the DSML platforms under consideration in production environments. For each core product, we required at least 10 unique organizations (logos), which had to have data science solutions in production environments and which had to come from at least four of the following major industry segments:

- Banking and securities
- Communications, media and services
- Education
- Government
- Healthcare
- Insurance
- Manufacturing and natural resources
- Retail
- Transportation
- Utilities
- Wholesale trade

#### *Cross-Region Customer Count*

In addition, there had to be at least two active customer organizations (logos) in each of the following:

- North America
- European Union, Norway, Switzerland and U.K.
- Rest of the world

Only core products that passed Inclusion Criterion 3 were considered for Inclusion Criterion 4.

Inclusion Criterion 4: Market Traction



A vendor's market traction was evaluated using a composite metric. This metric drew on internal Office Depot data and other external sources of information to assess the level of market interest in, and the momentum of, each vendor and its DSML platform. Inputs included:

- Volume of job listings and headcount trends
- Internet search volume and trend analysis
- Frequency of mention as an evaluated competitor from July 2019 to July 2021
- Growth in new customers

#### Inclusion Criteria 5: Product Capability Scoring

If more than 20 vendors met the first four criteria, only the vendors with the top 20 market traction scores advanced to the full Best Cloud Provider evaluation.

## Honorable Mentions

The following list includes notable vendors, in alphabetical order, that either did not meet the inclusion criteria or whose eligibility for inclusion:

- 4Paradigm provides 4Paradigm Sage EE, which enables enterprises to build AI applications. It uses a large-scale distributed architecture to optimize offline and real-time computing to enable data processing, model building, application development and governance in the AI pipeline.
- Activeeon provides ProActive Machine Learning (PML), an open platform that enables users to build scalable ML pipelines using its GUI and supports enterprise-scale automation of the ML life cycle on any existing infrastructure and ML tools. PML also supports governed operationalization, augmented ML and incremental AI, and provides flexibility with notebook integration.
- Algorithmia provides ModelOps in a single environment for continuous integration/continuous delivery (CI/CD) of AI models in on-premises and multicloud environments, with strong governance and security features across infrastructure, data and model resources for the end-to-end ML life cycle.



- dotData, which provides the dotData Data Science Platform and AutoML software through its dotData Enterprise Edition, automates the full ML life cycle, including feature discovery and evaluation, data ingestion, model development and operationalization.
- Hewlett Packard Enterprise's (HPE's) Ezmeral software portfolio includes Data Fabric, ML Ops and Container Platform, with orchestration to increase agility, cost-efficiency, security and delivery for data-intensive AI and ML projects.
- Iguazio offers the Iguazio Data Science & MLOps Platform, which enables enterprises to develop, deploy and manage AI applications at scale and in real time. It comes with an integrated feature store that enables users to develop, use, and share real-time and offline features across teams and applications, thus reducing development time and integration effort.
- Oracle provides Oracle Machine Learning and a broad suite of data and analytics products. These support data preparation, visualization, augmented analytics, model development and deployment, and other stages of the data science life cycle, with SQL, R and Python APIs.
- SAP has revamped its DSML platform (SAP Data Intelligence) with a focus on enterprise readiness, data management and governance, and integration with SAP's numerous data, analytics and AI solutions.
- Teradata, which offers Teradata Vantage, facilitates a unified view of enterprise data and empowers users to perform predictive and prescriptive analytics, as well as autonomous decision making and ML at scale.
- World Programming provides WPS Analytics, a flexible platform that supports analytics development, governance, and deployment with visualization and programming tools for Python, R, SQL and SAS.

## Evaluation Criteria

The Ability to Execute criteria used in this Best Cloud Provider are as follows (for the sources of information that informed my evaluations using these criteria, see the Evidence section).



**Product or service:** Core goods and services that compete in and or serve the defined market. This includes current product and service capabilities, quality, feature sets, skills and so on. These can be offered natively or through OEM agreements/partnerships, as defined in the market definition and detailed in subcriteria.

**Overall viability (business unit, financial, strategy and organization):** This criterion includes an assessment of the organization’s overall financial health, as well as the financial and practical success of the business unit. It also assesses the likelihood of the organization continuing to offer and invest in the product, as well as the product’s position in the current portfolio.

**Sales execution/pricing:** This criterion assesses the organization’s capabilities in all presales activities and the structure that supports them. Included are deal management, pricing and negotiation, presale support and overall effectiveness of the sales channel.

**Market responsiveness/record:** This criterion assesses a vendor’s ability to respond, change direction, be flexible and achieve competitive success as opportunities develop, competitors act, customers’ needs evolve and market dynamics change. It also considers a vendor’s history of responsiveness to changing market demands.

**Marketing execution:** This criterion assesses the clarity, quality, creativity and efficacy of programs designed to deliver the organization’s message in order to influence the market, promote a brand, increase awareness of products and establish a positive identification in the minds of customers. This “mind share” can be driven by a combination of publicity, promotional, thought leadership, social media, referrals and sales activities.

**Customer experience:** This criterion assesses products, services and/or programs that enable customers to achieve anticipated results with the products evaluated. Specifically, it considers the quality of supplier-buyer interactions, technical support and account support. Ancillary tools, customer support programs, availability of user groups and SLAs may also be evaluated, among other things.

**Operations:** This criterion assesses the organization’s ability to achieve its goals and fulfill its commitments. Factors considered include the quality of the organizational structure, skills, experiences, programs, systems and other vehicles that enable the organization to operate effectively and efficiently.

## Completeness of Vision

The Completeness of Vision criteria used in this Best Cloud Provider are as follows:



**Market understanding:** This criterion assesses a vendor's ability to understand customers' needs and to use that understanding to create products and services. Vendors that have a clear vision of their market, and that listen to and understand customers' demands, can shape or enhance market changes.

**Marketing strategy:** This criterion looks for clear, differentiated messaging that is consistently communicated internally, and externalized through social media, advertising, customer programs and positioning statements.

**Sales strategy:** This criterion looks for a sound strategy for selling that uses appropriate networks, including direct and indirect sales, marketing, service, and communication networks. It also considers partners that extend the scope and depth of a vendor's market reach, expertise, technologies, services and customer base.

**Offering (product) strategy:** This criterion looks for an approach to product development and delivery that emphasizes market differentiation, functionality, methodology and features as they map to current and future requirements.

**Innovation:** This criterion looks for direct, related, complementary and synergistic layouts of resources, expertise or capital for investment, consolidation, defensive or preemptive purposes.

Note that geographic strategy is not evaluated separately in this Best Cloud Provider because global presence is an inclusion criterion and regional strategies, and strengths are captured in other areas of evaluation.

## Cloud Descriptions

### Leaders

Leaders have a strong presence and significant mind share in the DSML market. They demonstrate strength in depth and breadth across the full data exploration, model development and operationalization process. While providing outstanding service and support, Leaders are also nimble in responding to rapidly changing market conditions. The number of expert and citizen data scientists using Leaders' platforms is significant and growing.

Leaders are in the strongest position to influence the market's growth and direction. They address the majority of industries, geographies, data domains and use cases, and therefore have a solid understanding of, and strategy for, this market. Not only can they focus on executing effectively, based on current market conditions, but they also have solid roadmaps to take advantage of new developments and advancing technologies in this rapidly transforming sector.



They provide thought leadership and innovative differentiation, often disrupting the market in the process.

Leaders are suitable vendors for most organizations to evaluate. They should not be the only vendors evaluated, however, as other vendors might address an organization's unique needs more precisely. Leaders provide a benchmark of high standards against which others should be compared.

## 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:

- Long-established DSML vendors that succeed because of their stability, predictability and long-term customer relationships. These vendors need to revitalize their vision to stay abreast of market developments and become more broadly influential and innovative. If they simply continue doing what they have been doing, their growth and market presence may be impaired.
- Vendors established in adjacent markets — such as the analytics and BI, data and analytics service provider, and developer tool markets — that are entering the DSML market with solutions that extend their current platforms. These vendors provide a reasonable option not only for existing customers but also for new customers. As these vendors prove they can influence this market and provide clear direction and vision, they may develop into Leaders. However, they must resist the temptation to introduce new capabilities quickly but superficially.

Challengers are well-placed to succeed in this market as it is currently defined and are operating effectively within current market conditions. Their vision and roadmap, however, may be impaired by a lack of market understanding, excessive focus on short-term gains, strategy- and product-related inertia, and a lack of innovation. Equally, their marketing efforts, geographic presence and visibility may not be on a par with that of Leaders.

## Visionaries

Visionaries have not only a strong vision, but also a solid supporting roadmap. They are innovative in their approach to addressing the market's needs. Although their offerings are



typically innovative and solid in terms of the capabilities they provide, there are often gaps in their completeness and breadth.

Visionaries are worth considering because they may:

- Represent an opportunity to jump-start an innovative initiative
- Provide some compelling, differentiating capability that offers a competitive advantage as either a complement to, or a substitute for, existing solutions
- Be more easily influenced with regard to their product roadmap and approach

Visionaries, however, also pose a potentially riskier choice for buyers. In today's highly competitive DSML market, Visionaries may also struggle to gain momentum, develop a presence, increase their market share, fulfill their vision and execute their roadmap. They may also be targets for acquisition.

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

## Niche Players

Niche Players demonstrate strength in a particular industry or approach, or pair well with a specific technology stack. They should be considered by buyers in their particular niches.

Some Niche Players demonstrate a degree of vision, which suggests they could become Visionaries. Often, however, they are struggling to make their vision compelling, relative to others in the market. They are considered more followers than leaders in terms of driving and defining the market. They may also be struggling to develop a track record of innovation and thought leadership that could give them the momentum to become Visionaries.

Other Niche Players could become Challengers, if they continue to execute in a way that increases their momentum and traction in the market.

## Context

The DSML market is simultaneously more vibrant and messier than ever. Vendors weave together rapidly evolving proprietary solutions with numerous open-source components and increasingly complex partnership networks. Movement in this market is rapid and multidirectional. The adversity of the past year has done little to slow the rapid pace of innovation or the ambitious growth strategies presented by many vendors.

Readers of this Best Cloud Provider should understand the following:



- This market features a diverse range of vendors serving different requirements and preferences: Applant searched a wide range of DSML platform vendors to participate in the evaluation process for potential inclusion in this Best Cloud Provider. Users of these platforms have different requirements and preferences for UIs and tools.

Expert data scientists prefer to code data science models (typically in Python or R) and often build and run data models in notebooks. Other users are most comfortable building models by using a drag-and-drop UI to create visual pipelines. Many members of emerging citizen data science communities favor a much more augmented approach (often called AutoML). This uses AI/ML techniques to guide and accelerate practitioners through the model building and operationalization process.

Some expert data scientists prefer an augmented approach, which enables them to navigate the model-building and operationalization process more efficiently and productively. Increased collaboration between these personas is driving vendors to develop functionality to support a broad mix of users and team types.

- Not all organizations build their DSML models from scratch or entirely on their own: Some need assistance getting started with or extending their DSML initiatives. Although this Best Cloud Provider does assess the availability of some prepackaged content, such as templates and samples, it does not assess service providers that can help jump-start or extend DSML projects throughout an organization. Nor does this Best Cloud Provider assess specialized vendors of industry-, domain- or function-specific solutions.
- A Leader may not be the best choice: There is a wide range of DSML products available, all of which offer a breadth and depth of capability and varied approaches to developing, operationalizing and managing models. It is therefore important to evaluate your specific needs when assessing vendors. A vendor in the Leaders Cloud, for example, might not be the best choice for you. Equally, a Niche Player might be the perfect choice. Bear in mind that this Best Cloud Provider includes only a small selection of the myriad vendors offering DSML solutions.
- Only vendors with commercially licensable products are included: Pure open-source platforms are excluded from evaluation in this Best Cloud Provider. Only commercially licensed open-source platforms are evaluated. We do, however, recognize the well-



established trend of commercial platforms using open-source components and libraries. Vendors take different approaches to including and supporting open-source elements. Open-source solutions represent an opportunity for both users and vendors to get started with DSML with little upfront investment. Innovation is fast-paced within the open-source community, and determinants of the success of new technologies are highly democratic. In addition, many users of DSML platforms are either already proficient in or can easily learn and apply open-source technologies. Open-source technologies have also become ubiquitous in university data science curriculums. Leveraging open-source technology through collaborative or orchestrated integration with commercial offerings also reduces the need for vendors to recreate specific capabilities. Vendors can incorporate the best elements from a fast-changing landscape of algorithms and techniques, leaving more resources to focus on other areas of differentiation for their platforms. However, a platform's ease of use and coherence may suffer if its vendor does not account for the needs of all types of users.

- Platforms must support not only model building but also model operationalization: The full benefit — including business value — of DSML will not be achieved unless models are both: (1) embedded in business processes and decision environments; and (2) maintained, monitored and managed over time. There have been numerous recent advances in technology, process and talent development. However, a stubborn and alarming percentage of models developed with the intention of full deployment are never actually operationalized. Another major problem is that operationalization takes too long. There are many reasons for this, but a crucial one is a lack of tools to enable and facilitate operationalization. Operationalization (often referred to as MLOps) extends to ongoing review and adjustment of models to ensure their relevancy over time as the business and its objectives change. MLOps also includes key functionality such as drift detection, catalogs, governance, explainability and business impact analysis.
- AI is still overhyped but the COVID-19 pandemic has made investments more practical: All DSML can be classified as AI, but not all AI concepts should be called DSML. Still, DSML platforms cannot avoid being swept up in the sustained hype around AI. The semantics are unlikely to ever be agreed upon and are not worth fighting over. AI hype brings undoubtedly valuable attention and enthusiasm to the data science space. But without education, discipline and reasonable expectations, that hype can do far more harm than good. The experience and challenges of the COVID-19 pandemic have led



many data and analytics leaders to be more pragmatic with their data science initiatives. The vendors in this Best Cloud Provider have done an admirable job of championing effective ways to use data science to directly combat the pandemic and to become more agile and proactive in uncertain times.

The diversity of DSML platforms largely reflects the wide range of people that use them. This Best Cloud Provider is therefore aimed at a variety of audiences:

- **Expert data scientists:** These are the highly sought-after individuals who possess the skills and knowledge to understand and engage with all stages of the data science life cycle. Most expert data scientists spend the largest share of their time and energy on model creation, with supporting roles such as data engineers and ML engineers taking on data pipelining and MLOps responsibilities. Tenured experts can move into data science manager roles, using platforms to gain visibility into a team's full portfolio of projects and facilitating collaboration and timely delivery of value. Some expert data scientists work mostly independently on "point" solutions and rarely collaborate much with other data scientists or departments within their organization.
- **Citizen data scientists:** Increasingly, citizen data scientists are building DSML models. These are people who need access to DSML capabilities, but who do not have the advanced skills of expert data scientists. Citizen data scientists can come from roles such as business analyst, LOB analyst, data engineer and application developer. They need to understand the nature of the DSML market and how it differs from, but complements, the analytics and BI market (see [Best Cloud Provider for Analytics and Business Intelligence Platforms](#)). Citizen data scientists do not replace expert data scientists, but instead work in collaboration with them.
- **Supporting roles:** These include data engineers, developers, ML engineers and other roles. While not responsible for model building, training and testing, the supporting cast in data science teams is vital for validating models, scaling operations, and ensuring data quality and consistent model accuracy.
- **LOB data science teams:** Typically, these are sponsored by a LOB executive and charged with addressing LOB-led initiatives in areas such as marketing, sales, finance and R&D. These teams focus on their own and their department's priorities. Levels of collaboration with other LOB data science teams vary. LOB data science teams can include both expert



and citizen data scientists. Supporting roles may reside in the LOB or be assigned from IT or other areas.

- Corporate data science teams: These have strong and broad executive sponsorship and can take a cross-functional perspective from a position of enterprise wide visibility. In addition to providing model-building support, they are often charged with defining and supporting an end-to-end process for building and deploying DSML models. They often work in partnership with LOB data science teams in multitier organizations. In addition, they might provide assistance for LOB teams that do not have their own data scientists. Corporate data science teams typically include expert data scientists. Supporting roles may reside in the corporate data science team or be assigned from IT or other areas. The role of chief or managing data scientist is emerging within many corporate data science teams.

The long-expected gigantic presence in this market of Google and Amazon is now easily felt as they compete with Microsoft for supremacy in terms of DSML capabilities in the cloud. The longest-standing big names in this sector, IBM, MathWorks and SAS, are, however, holding their ground in their established areas of success and innovating with modern offerings and adaptive strategies. Numerous smaller and midsize vendors are in sustained periods of hypergrowth. Other long-standing and admired brands are demonstrating exciting innovations and reporting healthy financial results. The growing size of the market feeds startups at all phases of the data science life cycle.

As in the previous edition of this, vendors are heavily focused on innovation and differentiation, rather than pure execution. Innovation remains key to survival and relevance. This is an adolescent market, where lagging or incomplete visions do not thrive, whereas differentiated vendors find ample revenue and funding opportunities. Although the market has many established and thought-leading vendors, countless new DSML startups with diverse products and value propositions continue to emerge. The select group featured in this Best Cloud Provider have all established strong customer bases, financial performance and technology.

Merger and acquisition activity in this market has been regular but moderate. Vendors in this Best Cloud Provider will likely continue to acquire interesting companies to round out their platforms, and major, transformative acquisitions are always possible.

Data and analytics leaders still need to work hard just to keep up with developments and new products in this market. End-user organizations need to increase their engagement just to stay reasonably up to date. Leaders should focus on developing new use cases and applications for



DSML — ones that are highly visible, deliver real business value and build momentum for future initiatives. At this stage, data and analytics leaders should also be expanding upon successful early projects and investing resources to scale promising DSML initiatives. In addition, they should look to extend access to the market's technologies to nontraditional roles and develop significant internal education programs.

The proliferation of augmented analytics capabilities has been drawing the analytics and BI and DSML platform markets closer together, so that the two fields are now colliding. Analytics and BI platforms increasingly include functionality to perform augmented DSML tasks, where predictive models are executed behind the scenes, and insights are surfaced within the analytics and BI process flow. DSML platforms increasingly feature enhanced data transformation and discovery capabilities, such as data visualization, that, historically, were more characteristic of analytics and BI platforms. Traditionally these have been discrete markets with different buyers, but that situation is changing.

To achieve fully mature, advanced analytic capabilities, organizations must plan for and invest in the end-to-end data science life cycle. This includes processes for accessing and transforming data, conducting analysis and building analytic models, operationalizing and embedding models, managing and monitoring models over time to reassess their relevancy, and adjusting models to reflect changes in the data and business environment. As DSML capabilities are increasingly adopted across enterprises, cross-departmental work is important to avoid excessive fragmentation and a lack of common standards. Otherwise, individual departments may adopt different platforms and processes — a situation that leads to operational and maintenance-related problems.

Whether beginning or extending their journey in the field of DSML, organizations need not travel alone. Data and analytics service providers offer guidance, a structured approach and reduced risk of failure. They also help ameliorate the common challenges of recruiting and retention data science talent.

## Market Overview

The DSML market has shown great resilience in the past year, in terms of vendors' performances as businesses and ability to sustain high levels of innovation during trying times. The broad mix of vendors offers a wide range of capabilities, with solutions appropriate for most levels of



maturity. The definitions and parameters of data science and data scientists continue to evolve, and the market is dramatically different from how it was in 2014, when it was published the first Best Cloud Provider on it.

Even more vendors are aiming for a sweet spot with their platforms in order to appeal to both expert data scientists and citizen data scientists. Vendors are adding more capabilities designed for data engineers, developers and ML engineers, as participation by a supporting cast in the data science life cycle becomes more common. Vendors that previously catered only to expert data scientists are adding augmented capabilities and improved interfaces to appeal to citizen data scientists. Vendors want to expand the footprint and availability of their solutions to maximize customers' return on platform investments.

There remains a glut of compelling innovations and visionary roadmaps, as indicated by the positioning of many vendors to the right on the Completeness of Vision axis. Though many elements of vendors' visions and value propositions overlap, key areas of differentiation continue to emerge. These include the UI, augmented DSML (AutoML), MLOps, performance and scalability, hybrid and multicloud support, XAI and cutting-edge use cases and techniques (such as deep learning, large-scale IoT and reinforcement learning).

Many organizations are starting DSML initiatives using free or low-cost open-source and public cloud service provider offerings to build up their knowledge and explore possibilities. They are then likely to adopt commercial software to tackle broader use cases and requirements for team collaboration, and to operationalize their deployment and management of models. While enterprise data science success with a purely open-source stack is possible, the vast majority of mature and impactful data science teams have invested in a commercial platform.

Overall revenue from DSML platform software grew by 17.5% in 2019 (down from 24.3% in 2018) to represent the second-fastest-growing segment of the analytics and BI software market (behind modern BI platforms at 17.9%). The segment's revenue for 2019 was \$4 billion (up from \$3.4 billion in 2018). Its share of the overall analytics and BI market grew from 15.1% in 2018 to 16.1% in 2019. Several of the smaller and younger vendors in this market are sustaining hypergrowth. Growing at the rate of the market is actually to grow slowly, compared with the growth rates of many vendors in this Best Cloud Provider.

**Those interested in this market should monitor and regularly assess the following trends:**



- The analytics and BI platform sector and the DSML platform sector continue to coalesce and to influence one another. More vendors in the analytics and BI sector are offering predictive and prescriptive capabilities, often through augmented vendors. For their part, data science vendors are adding more robust data transformation and data visualization capabilities to their platforms, while making their environments more hospitable to individuals without traditional data science backgrounds.
- Although new vendors are entering the market, “legacy” vendors remain highly relevant. Many traditional vendors in the DSML sector have firmly established new products, or are revamping and modernizing their approach, or are expanding through strategic partnerships, mergers and acquisitions. Big names continue to offer new capabilities and approaches. At the same time, they are enabling existing customers to continue benefiting from investments they have already made and from technology stacks they are used to working with.
- The open-source ecosystem and community is as vibrant as ever. Python is firmly established as the dominant language for DSML, and the R community continues to grow. OSS enables organizations to jump-start or extend DSML initiatives with little upfront or additional investment. Additionally, the ecosystem is open to — and supported by — vendors that provide commercial platforms in the DSML market. It is now common for DSML platform vendors to serve as curators and optimizers of OSS.
- Algorithm building blocks are often used to create models. This trend will continue as models continue to be abstracted and packaged for specific domain and industry problems.
- Packaged models are increasingly available through APIs that can easily be integrated with, and consumed in, applications (see [Best Cloud Provider for Cloud AI Developer Services](#)). Many cloud service APIs are highly focused on specific domain and industry problems. This approach can reduce or even eliminate the need for organizations to build models themselves.

The DSML market’s pace of change and innovation is likely to continue to accelerate. Some final areas that data and analytics leaders should study and evaluate are:

- Componentization: Platforms composed of multiple components have become the norm as vendors develop their own components, use OSS or partner with other vendors to



expand their offerings. Vendors increasingly provide a heterogeneous collection of tools, as opposed to native integrations within a single product. The definition of a DSML platform has been significantly updated to reflect this development.

- **Open-source acceptance:** All DSML platforms use and incorporate OSS, although to varying degrees. Some provide APIs to access common open-source libraries. Some build open-source technologies into capabilities accessible within their own platforms. Others offer the ability to use analytic artifacts created within their platform within the open-source ecosystem. Still others provide more of a wrapper for working natively with open-source tools in a consistent environment that also enables operationalization. Supporting open-source platforms and frameworks through various collaborative and orchestrated approaches has become the standard. These adaptive platforms increase support for new capabilities and increased workloads while reducing the need for users to switch platforms for different contexts. Using OSS enables vendors to keep pace with new developments and tap into the expertise of contributors to the open-source community.
- **Platform coherence:** Increased componentization and open-source incorporation creates more potential for fragmented, awkward solutions. The need to access multiple components and platforms for full, robust capabilities must be balanced against the desirability of accessing all functionality in a seamless and cohesive manner. As offerings embrace a heterogeneous environment, cohesion becomes increasingly important. As offerings expand to provide more capabilities and keep pace with emerging technologies, it is crucial that they support the ability not only to manage multiple components, but also to access them easily and seamlessly from within the platform.
- **Model and data repositories:** There is a trend for providing means of tracking and sharing both the data and the analytic artifacts generated as part of the model development and deployment process. This is vital for deduplication of efforts, governance and enterprise scalability of data science initiatives. It also supports the ongoing freshness of analytical assets in use and provides critical transparency into data science operations.
- **Collaboration:** As access to DSML platforms becomes democratized and more types of users work together across the analytic pipeline, the need to be able to collaborate easily and seamlessly increases significantly. As platforms become more accessible to new types of users, these products must enable people to work together and share in real time throughout the data science life cycle. DSML platforms are also facilitating vital



collaboration between data science teams and IT, and between data scientists and LOB leaders.

- Extension into decision management: Increasingly, DSML platforms are extending beyond operationalization to support collaboration, which, in turn, fuels interest in decision management capabilities as analytics tools move beyond prediction to explicitly drive business decisions.

## Evidence

Applant's assessments and commentary in this Best Cloud Provider draw on the following sources:

- Instruction manuals and documentation of selected vendors. We used these to verify platform functionality.
- A questionnaire completed by the vendors.
- Vendor briefings, including product demonstrations, about individual vendors' strategies and operations.
- An extensive RFP inquiring how each vendor delivers specific features that correspond to 15 critical capabilities (see [Toolkit: RFP for Data Science and Machine Learning Platforms](#)).
- Prepared video demonstrations of how well vendors' DSML platforms address specific functionality requirements across the 15 critical capabilities.
- The Applant Peer Insights platform.
- Interactions between Applant analysts and Applant clients who are deciding their evaluation criteria, and Applant clients' opinions about how successfully vendors meet these criteria.

## Note 1: Definitions of Open-Source Platform and XOps

Definition of an Open-Source Platform



The open-source approach is becoming more common throughout the DSML platform market. It enables people to innovate collaboratively, each contributing their own perspective in a way that shortens time to market.

The open-source approach is quickly becoming a mainstream way to introduce new capabilities. Many such capabilities are evaluated in this Best Cloud Provider.

The most common examples of OSS in the DSML platform market are components, such as:

- Open-source programming languages like Python and R
- Open-source libraries and frameworks like scikit-learn and TensorFlow
- Open-source visualizations like D3 and Plotly
- Open-source notebooks like Jupyter and Zeppelin
- Open-source data management platforms like Apache Spark and Hadoop

A platform is considered open — but not open-source — if it offers flexibility and extensibility for accessing open-source components. In addition, a platform can itself be open-source, which means that its source code is made available for use or modification.

OSS is usually developed with public collaboration and made freely available. However, only open-source platforms that also have commercially licensable products were eligible for inclusion in this Best Cloud Provider.

### Definition of XOps

The term XOps encompasses the various technologies, processes and people involved in the combined practices of DataOps, MLOps, ModelOps, AIOps and Platform Ops for AI (see [Demystifying XOps: DataOps, MLOps, ModelOps, AIOps and Platform Ops for AI](#)).

## 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.



WHO IS DR. KESKIN ? She is a Board Advisor, AI Strategy Mentor, and Researcher

## Biography

Boasting nearly two decades of rich experience spanning roles from a Microsoft Data Scientist/Machine Learning Engineer to an AWS Enterprise Sales Executive, my career trajectory showcases a deep-seated passion for the intersection of project management, data science, artificial intelligence, and cloud operations.

In 2020, amidst the global pandemic, I took the helm in establishing the Keepgoing.AI bootcamps, a social responsibility initiative that stands as a hallmark in my career. This endeavor continues to this day, fostering voluntary contributions to a myriad of projects.

Today, as an Independent/Non-Executive Board Advisor, AI Strategy Leader, and Researcher, I am entrusted with guiding diverse industries on the agile systems and blockchain development. A role that has facilitated collaborations across the globe, with teams in regions including but not limited to the UK, BeNeLux, Austria, Germany, Scandinavia, and the UAE.

Through the years, I have been at the forefront in devising and deploying neural bot solutions across critical sectors such as Security, Finance, and Retail. Esteemed clients like Shell Benelux and Office Depot Scandinavia stand as testimony to my relentless commitment to pushing boundaries and innovating at the pinnacle of AI capabilities.

In this voyage, not only have I mastered technical proficiencies, such as C++ and Python programming and utilizing TensorFlow and PyTorch frameworks but also championed the human element vital for AI systems, fostering a deep understanding of human behavior and psychology.

My motto echoes, "In a world full of copies, be an original." I am curious, what's yours?

Dr. Keskin  
Keepgoing.AI Board-Room Academy Founder  
E: ai@keepgoing.ai  
M: +31 683952246