CLOUDERA

WHITE PAPER

4 ESSENTIAL PLATFORM FACTORS FOR ENTERPRISE ML

Cloudera Machine Learning (CML) delivers faster time to value

AI AND MACHINE LEARNING

Executive Summary

Artificial Intelligence (AI) and Machine Learning (ML) are profoundly transforming how organizations engage with stakeholders. Across many industries, high value transformative use cases are rapidly emerging. The economic impact of AI/ML is immense with an estimated $13 trillion added to the global economy over the next decade.

However, deploying and scaling AI/ML across the enterprise requires implementing complex, iterative workflows from data to models to outcomes. This is not easy. In fact, currently only 35% of organizations indicate that analytical models are deployed in production. In addition, as the number of AI/ML projects and models multiply, production ML can be slow, cumbersome and fraught with "false starts" that make it even more difficult and expensive.

What's needed is an open, unified, collaborative, secure and governed enterprise-grade environment to run and manage all AI/ML models with transparency, consistency, trust and high-performance.

Cloudera, the Enterprise Data Cloud Company, delivers a platform that helps you overcome key production AI/ML challenges: monitoring, deployment, security, governance, scalability and infrastructure. Cloudera Machine Learning (CML) is the only end-to-end ML platform that enables standards-driven model and feature monitoring, cataloging, and ongoing governance at enterprise scale.

Learn how you can overcome your production ML challenges with a flexible open unified data platform that delivers faster time to value for production AI/ML for your enterprise.

Table of Contents

Higher-value probabilistic predictive analytics and prescriptive capabilities growing in importance

Artificial Intelligence/ Machine Learning (AI/ML) drive better outcomes

AI/ML workflows are complex and iterative

Move from research to production to maximize business value

Enterprises have spent trillions of dollars on IT infrastructure, collected troves of data and leveraged analytics to get better deterministic insights (What happened? or What is the case?) into their businesses. Now, to outflank competitors and sustain an astounding rate and pace of data-driven business transformation and innovation, organizations want highervalue probabilistic predictive insights (What is likely to happen? What is likely the case?) and even prescriptive capabilities (What should happen?) to drive better outcomes.

Data Science, Machine Learning (ML) and Artificial Intelligence (AI) can significantly improve insights with better predictions, more automation and lower costs. This helps drive optimizations of business processes in every organization – making them even more efficient, competitive and innovative. It is estimated that AI/ML will add over 13 trillion dollars to the global economy in the next decade.[1]

However, deploying and scaling AI/ML can be long and cumbersome with many obstacles along the way. Many projects don't make it into production because of model inefficiencies that slow down or halt the entire process. Or, in many cases, organizations fail to adequately adopt production models because of a lack of internal knowledge on how they work and other cultural/business impediments.

Clients need to rapidly implement and scale ML models across their entire organization, spanning a large spectrum of use cases. This sense of urgency and growing regulatory scrutiny, create new and unique challenges to move AI/ML from research to production.[2] In fact, currently only 35% of organizations indicate that analytical models are fully deployed in production. These organizations must also ensure that models continue to operate and perform as expected or better – throughout the entire lifecycle and workflow.

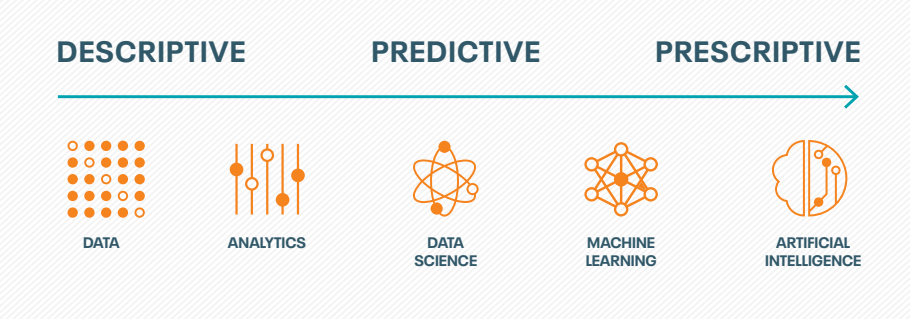

Figure 1: From Data to AI and Machine Learning and AI

Typical ML data science workflow is complex and highly iterative

The various phases of a typical Data Science workflow (Data to Models to Outcomes to Business Value) are shown in Figure 2 with more emphasis on the Production (Operational) phase and Governance.

1. **Data Engineering** consists of data acquisition, processing and governance.

- **Data Processing:** Raw data is typically not in a convenient format for a developer to run analysis, so it must be cleansed and prepared.
- **Data Governance:** As organizations use data (and analytics) more, and for more important questions, the need to govern those assets increases. Every organization should be concerned about data quality in their source systems, but often these concerns are isolated and not visible across departments. Security, privacy and regulatory compliance are important elements of Governance.

Need open platform to connect Data Engineering to Traditional Analytics & BI to Data Science Exploration to Production

Data governance across the workflow critical for Production

Governance also key to successful deployment of production ML at scale

2. Traditional analytics and Business Intelligence (BI) includes data wrangling, data visualization and data modeling.

- **Data Wrangling** is the process of transforming and mapping data from one "raw" data form into another format with the intent of making it more appropriate and valuable for a variety of downstream analytics.
- **Data Visualization** helps identify significant patterns and trends in the data. One can gain better insights through simple charts like line charts or bar charts.
- **Data Modeling** is the process of producing a descriptive diagram of relationships between various types of information that are stored in a database.

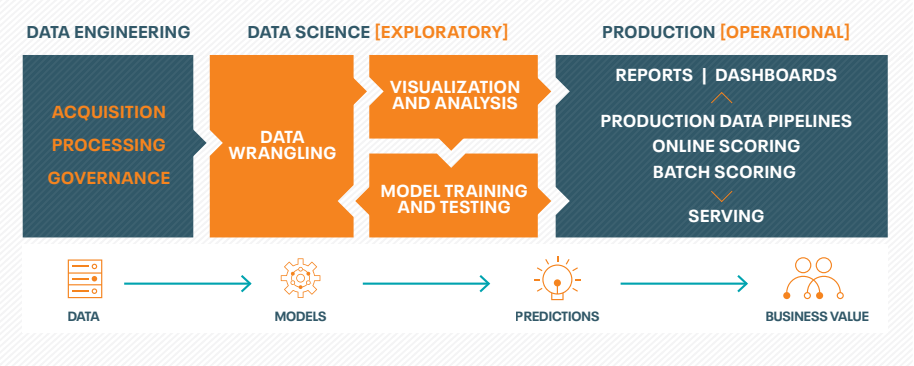

Figure 2: A Typical Data Science Workflow with Emphasis on Production and Governance

3. Exploratory Data Science and ML includes probabilistic modeling and ML model development (model training and testing).

- **Probabilistic Modeling** helps understand the probability of what could happen based on a variety of inputs and data.
- **ML Model Development** to automate processes or make ongoing predictions and learn/change based on new data. This can be at a product level – for example learning what a person regularly watches on Netflix and suggesting programming they will likely enjoy; or on a business level – for example detecting anomalies or patterns from incoming data for fraud prevention.

4. Production/Deployment is the process of delivering the outcomes (better automation, predictions, innovations, etc.) to stakeholders (customers, internal business, etc.). There are several different ways to deploy a model and understanding the end user (customer) intention helps determine the technology required. The deployment phase can be as simple as generating a report or as complex as implementing a repeatable data science process.

Successful production ML requires streamlined, frictionless and predictable deployment, serving and ongoing governance of ML models at scale. Purpose built end-to-end governance on a unified platform empowers clients to build data pipelines, train and productionize machine ML models that optimize business processes and products to gain a competitive edge.

To maximize value from AI/ML investments, the many production model deployment challenges must be overcome.

Model deployment in production is extremely challenging

Challenges in monitoring, deployment, governance, security, scalability and infrastructure

Monitoring must provide detailed visibility

Production ML challenges, requirements and solutions

Production model deployment is one of the most difficult processes to unlock ML value. It requires coordination between data scientists, IT teams, software developers, and business professionals to ensure the model works reliably in the organization's production environment. Often, there is a discrepancy between the programming language in which a ML model is written and the languages the production system can understand, and re-coding the model can delay model implementation.

Point solutions are not optimal since they break IT security and governance, require data scientists to move data which could create shadow IT and other issues. What's needed is a holistic solution to address these issues and address the many challenges associated with monitoring, deployment, governance, security, scalability and infrastructure.

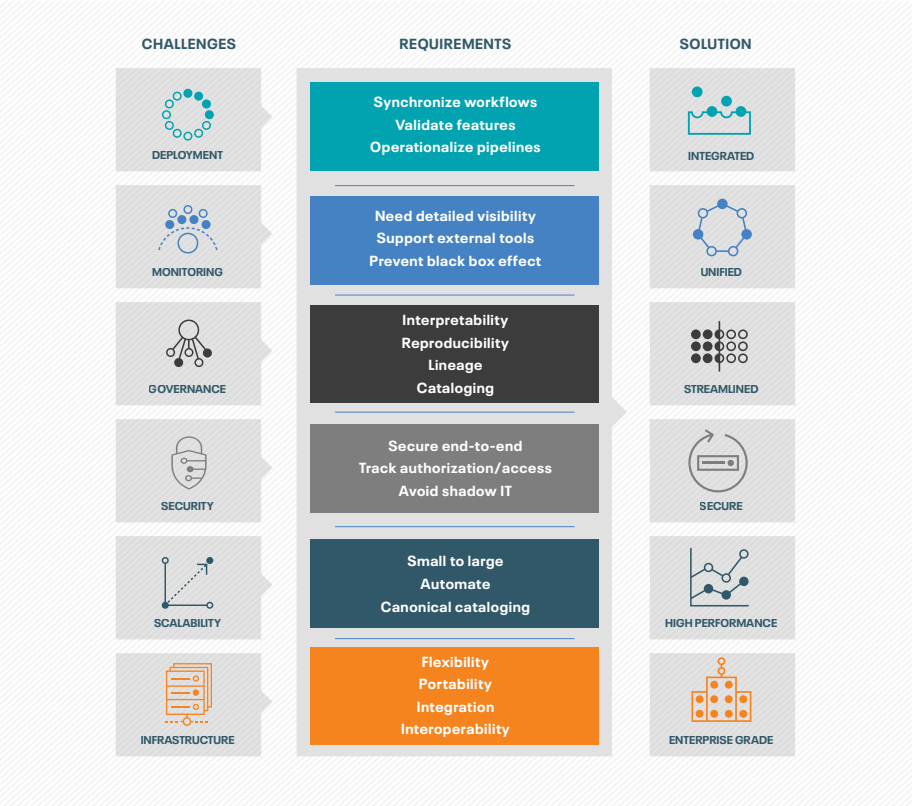

Figure 3: ML Deployment Challenges, Requirements and Solutions

Figure 3 depicts these **challenges** with detailed **requirements** and **solutions** as follows:

Monitoring

Monitoring is an essential element of production deployment as it provides visibility into its various phases. Poor visibility into mathematical metrics and to the external tools used for monitoring is a major challenge. This challenge can be mitigated by using a **unified** model monitoring platform for all the deployed models from a single pane of glass. Visual cues and alerts can be set to track both technical benchmarks and custom mathematical metrics.

An end-to-end unified platform streamlines model deployment in production for faster time to value

By 2022, 65% of enterprises are expected to task CIOs to transform and modernize governance policies to unlock value and confront new risks posed by AI/ML

Better governnce provides greater model visibility, explainability, interpretability and reproducibility

Infrastructure must be scalable, flexible and high-performance

Deployment

Data scientists today use a variety of different tools to solve many critical business problems. This often results in models being recoded into different languages as the initial language may not be used in the production environment. This leads to longer cycle times and potential inconsistencies in the translated model. An end-to-end **unified** and **integrated** platform alleviates this challenge. It also avoids point solutions, synchronizes and operationalizes the pipelines, and helps validate/test numerous features.

Governance

As AI/ML moves to production, the need to govern all IT assets (data, models, infrastructure, etc.) and ensure security, privacy and regulatory compliance increases. In fact, by 2022, 65% of enterprises are expected to task CIOs to transform and modernize governance policies to unlock value and confront new risks posed by AI/ML, data privacy and ethics.[3]

Defining standards for ML Operations (MLOps) is essential for deploying and governing ML models at scale for enterprises. ML metadata definitions must be universal and standard in how data structures specify tables, columns, etc.

Visibility of models and features within teams and across organizations is another basic requirement for model governance. It enables teams to understand how ML is being applied in their organizations and requires a canonical catalog of models and features. In the absence of such a catalog, many organizations are unaware of their models and features, where they are deployed, what they are doing, etc. This leads to substantial rework, model inconsistency, recomputing features, and other inefficiencies.

Better governance ensures the workflow is streamlined and promotes **explainability**, **interpretability** and **reproducibility** of production models.

- **Explainability:** description of the internal mechanics of an ML model in human terms.
- **Interpretability:** understand the relationship between model inputs, features and outputs; and predict the response to changes in inputs.
- **Reproducibility:** reproduce the output of a model consistently for the same inputs.

Security

Need end-to-end governance and enterprise security from data to the production environment. The chosen platform must be capable of delivering models into production with inherited **security**, **unified** authorization and access tracking. This also helps mitigate unauthorized "Shadow IT".

Scalability

As the model moves forward to production, it is typically exposed to larger volumes of data and data transport modes. The platform must have the ability to scale from small to large volumes of data and automate model creation. Adopting a consistent, microservices based approach to production analytics also helps in scaling models.

Infrastructure

Challenges include the **flexibility** to run anywhere – on-premise or in multiple clouds—ability to leverage common security, privacy and governance standards across all workloads and data, and provide the **high-performance** and **data gravity** (bring data science to data) needed to accelerate ML workloads. The chosen platform must support different infrastructures (multi-cloud, on-premises), **portability** and **integrate** existing silos of information.

To maximize value of ML models, the **enterprise-grade** production platform must ensure that the entire workflow is **unified, integrated, streamlined, secure, scalable** and **high-performance**. In addition, it must eliminate barriers from research to production ML at scale.

Production ML needs an enterprise-grade platform to ensure the workflow is unified, integrated, streamlined, secure, scalable and high-performance

ML Ops needs key capabilities for technology infrastructure and tooling

Production ML Operations (MLOps) at scale

Figure 4 depicts the ML workflow highlighting functional capabilities to implement production MLOps at scale. It includes the technology infrastructure and tooling necessary to deploy ML algorithms and data pipelines reliably so as not to destabilize other parts of the workflow. MLOps extends from the data science tools used to select and train ML algorithms down to the hardware that those algorithms use to process data. It also includes the databases and message queues used to store, move, monitor, and track technical and mathematical metrics.

Functional capabilities include:

Packaging, deployment and serving: There are different approaches to deploy production models, with benefits that vary depending on the specific use case. Right packaging is necessary for automated deployment of production models and to address multiple deployment patterns such as Batch, Function-as-a-service and Edge. In addition, enterprise level deployments need high availability, autoscaling and strong authentication features. Serving makes a trained model available to other software components. Models can be served in batch or real-time modes.

Monitoring: It is an important element of operationalizing ML. Monitoring is done at various stages of the lifecycle: check input and output distribution, look for skew, drift and accuracy change, add custom thresholds, send emails with results and trigger pager systems as needed.

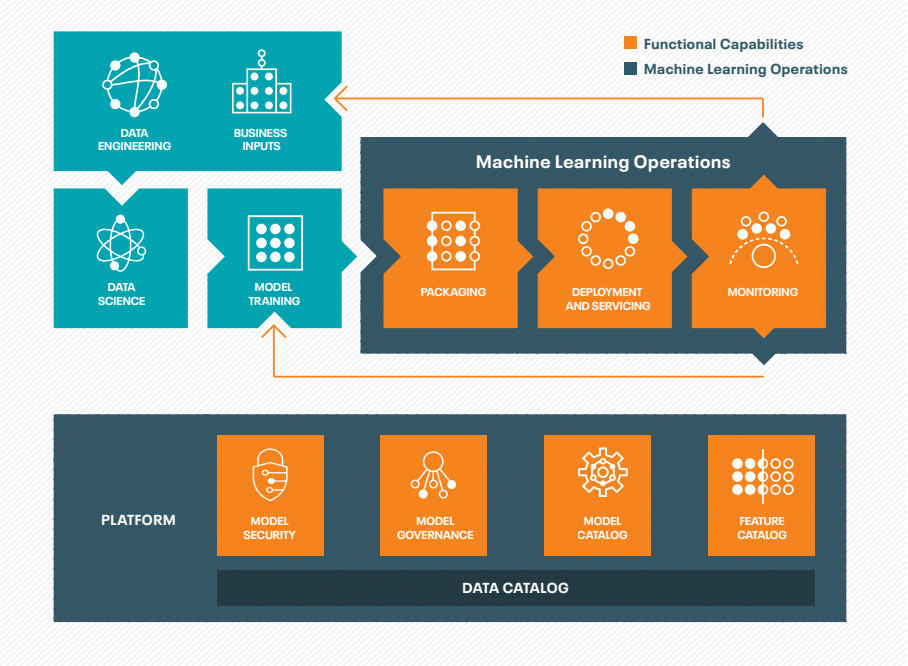

Figure 4: Production ML Workflow with Key Functional Capabilities

Model security, governance and cataloging: This is a basic requirement for model governance and enables teams to understand how ML is being applied in their organizations. It requires a centralized catalog of models and features which facilitate tracking models and their features throughout their lifecycle to understand these features and their relationship with the data catalog. In addition, catalogs facilitate authorization and tracking access to models thereby maintaining end-to-end security of the environment.

The Cloudera Machine Learning (CML) platform provides these functional capabilities to drive faster time to value for production AI/ML at scale.

CML is built on the open Cloudera Data Platform and enables collaboration and agile experimentation with governance

Proactive monitoring and governance enable seamless scaling

CML integrates data management with explainable, interoperable and reproducible MLOps

Cloudera Machine Learning (CML)

Built on Cloudera's open data platform (Cloudera Data Platform – CDP), CML drastically reduces time to value for production ML models. It enables data scientists, ML engineers, and operators to collaborate in a single unified platform that is purpose-built for agile experimentation and production ML workflows with enterprise-grade governance capabilities built in. Unlike ML point solutions, which compromise security and require complex, costly workflows for production models, CML is the only end-to-end ML platform that enables standards-driven model and feature monitoring, cataloging, and ongoing governance at enterprise scale (Figure 5).

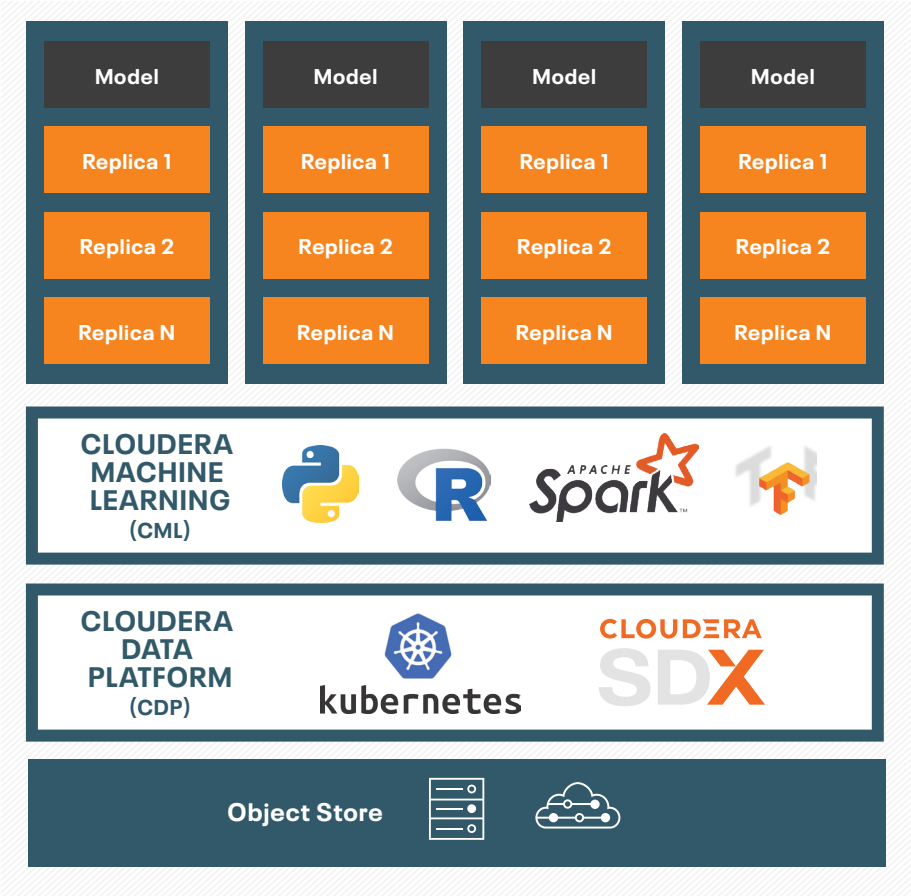

Figure 5: High Level Architecture of Cloudera Machine Learning

CML enables enterprises to proactively monitor technical metrics such as service level agreements (SLA) adherence, uptime, and resource use as well as mathematical metrics including model distribution, drift, and skew from a single governance interface. Users can set custom alerts and eliminate the model "black box" effect with native tools for visualizing model lineage, performance and trends.

All enterprise production ML workflows are securely contained in Cloudera's Enterprise Data Cloud. This enables seamless workflows for governing and quickly customizing models in production while maintaining complete visibility into the end-to-end data and model lineage. Clients can effectively maintain hundreds or thousands of models in production with resources that auto-scale to business needs and set model governance rules that enable fast response to mission critical changes in their production environments. Governing production ML workflows in CML enables enterprises to accelerate time to value and deliver ongoing results securely, from the edge to AI

CML integrates data management with explainable, interoperable and reproducible MLOps workflows. It also provides the capabilities to define ML metadata objects. These CML components are open, extensible, and have a pre-built governance feature.

Working with 1000s of organizations, Cloudera is at the forefront of developing best practices and solutions to empower customers to build, deploy and manage ML workflows across the enterprise and industrialize AI on an open, enterprise platform with their data, skills and intellectual property (IP). Only Cloudera offers an Enterprise Data Cloud, tools and expert guidance and services to help clients worldwide unlock business value from AI/ML.

Customer examples

Advanced analytics and AI/ML solutions from Cloudera are allowing customers to meet the challenges of managing big data and are delivering substantial benefits across various industries. Three representative client industries are highlighted here.

30%

reduction in losses due to fraud—significantly below the industry standard. This amounts to billions in cost savings for the business.

Western Union | Finance

Deliver a differentiated omni-channel customer experience and fight fraud

Challenge

Wiring more than $300B yearly across 200 countries leads to billions lost to fraudsters each year

Solution

Cloudera Enterprise Data Hub and Cloudera Data Science Workbench empowers 360-degree view of data as well as security of information and event management

Outcomes

- Better fraud detection accuracy
- Saw their loss drop by at least 30%
- Approval rate increased by 5% for new customers

4X

increased prediction accuracy for clinical trial patient eligibility to help patients gain access to life-saving clinical trials sooner.

IQVIA™

IQVIA | Lifesciences and Healthcare

Deliver a differentiated omni-channel customer experience and fight fraud

Challenge

Data siloed across 250 different data warehouses. Took days to copy data from silos for analysis. Limited by performance and scalability restraints.

Solution

Cloudera brings together more than two petabytes (PB) of data from 250 data warehouses for self-service data science across ~2000 users at scale.

Outcomes

- Accelerates query responses from days to seconds
- Reduced development and innovation costs
- 4X increase in prediction accuracy

20%

reduction in revenue loss through improved fraud detection, CRM, network quality and operational efficiency with a Cloudera.

Deutsche Telekom

Deutsche Telekom | Telecommunication

Improve fraud detection, CRM, network quality and operational efficiency

Challenge

Preventing network fraud is a major challenge. Huge data captured in silos, made machine learning at scale impossible.

Solution

By applying ML and AI, the company identifies network problems by detecting fraud patterns and real-time threats before the business is affected.

Outcomes

- Reduced revenue loss from fraud by 20%
- Reduced churn by 5-10%
- Improved operational efficiencies by 50% overall.

AI/ML is a C-suite initiative, but organizations are challenged to unlock value

CML drastically reduces time to value for production AI/ML

Only Cloudera is providing holistic solutions for the entire ML workflow to industrialize AI

Cloudera is trailblazing the path to production AI/ML

Many clients are implementing high-value AI/ML use cases in several industries. For this they need a reliable partner with deep expertise to overcome the many challenges with deploying and scaling AI/ML in production.

Cloudera, the Enterprise Data Cloud Company, provides Cloudera Machine Learning (CML) which is an open, unified, collaborative, secure and governed enterprise-grade production platform to run and manage all AI/ML models with transparency, consistency, trust and high-performance.

CML drastically reduces time to value for production ML models by enabling data scientists, ML engineers, and operators to collaborate in a single unified platform; purpose-built for agile experimentation and production ML workflows with enterprise-grade governance capabilities built in.

About Cloudera
At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.

Learn more at cloudera.com

Connect with Cloudera
About Cloudera:
cloudera.com/more/about.html

Read our VISION blog:
vision.cloudera.com
and Engineering blog:
blog.cloudera.com

Follow us on Twitter:
twitter.com/cloudera

Visit us on Facebook:
facebook.com/cloudera

See us on YouTube:
youtube.com/user/clouderahadoop

Join the Cloudera Community:
community.cloudera.com

Read about our customers' successes:
cloudera.com/more/customers.html

Unlike ML point solutions, which may compromise security (moving data from a secure cloud/data management system and creating shadow IT) and require complex, costly workflows for production models, CML is the only end-to-end ML platform that enables standards-driven model and feature monitoring, cataloging, and ongoing governance at enterprise scale.

This is how Cloudera is trailblazing the journey to industrialize AI and empowering clients, worldwide, to build their AI factory on an open, enterprise platform with their data, skills and IP. Clients can now not only make internal processes faster and cheaper but also build better products and services, create brand new products or completely reinvent processes.

Sources

[1] Tim Fountaine, Brian McCarthy and Tamim Saleh, "Building the AI-Powered Organization", Harvard Business Review, July-August 2019.

[2] IDC's Advanced and Predictive Analytics survey and interviews, n = 400, 2017 – 2019 from https://www.sas.com/en_us/news/press-releases/2019/october/model-ops-last-mile.html

[3] https://www.idc.com/getdoc.jsp?containerId=prUS44420918

Contributor

Cabot Partners

CLOUDERA

Cloudera, Inc. 395 Page Mill Road Palo Alto, CA 94306 USA cloudera.com