CLOUDERA

Expand Production Machine Learning with MLOps

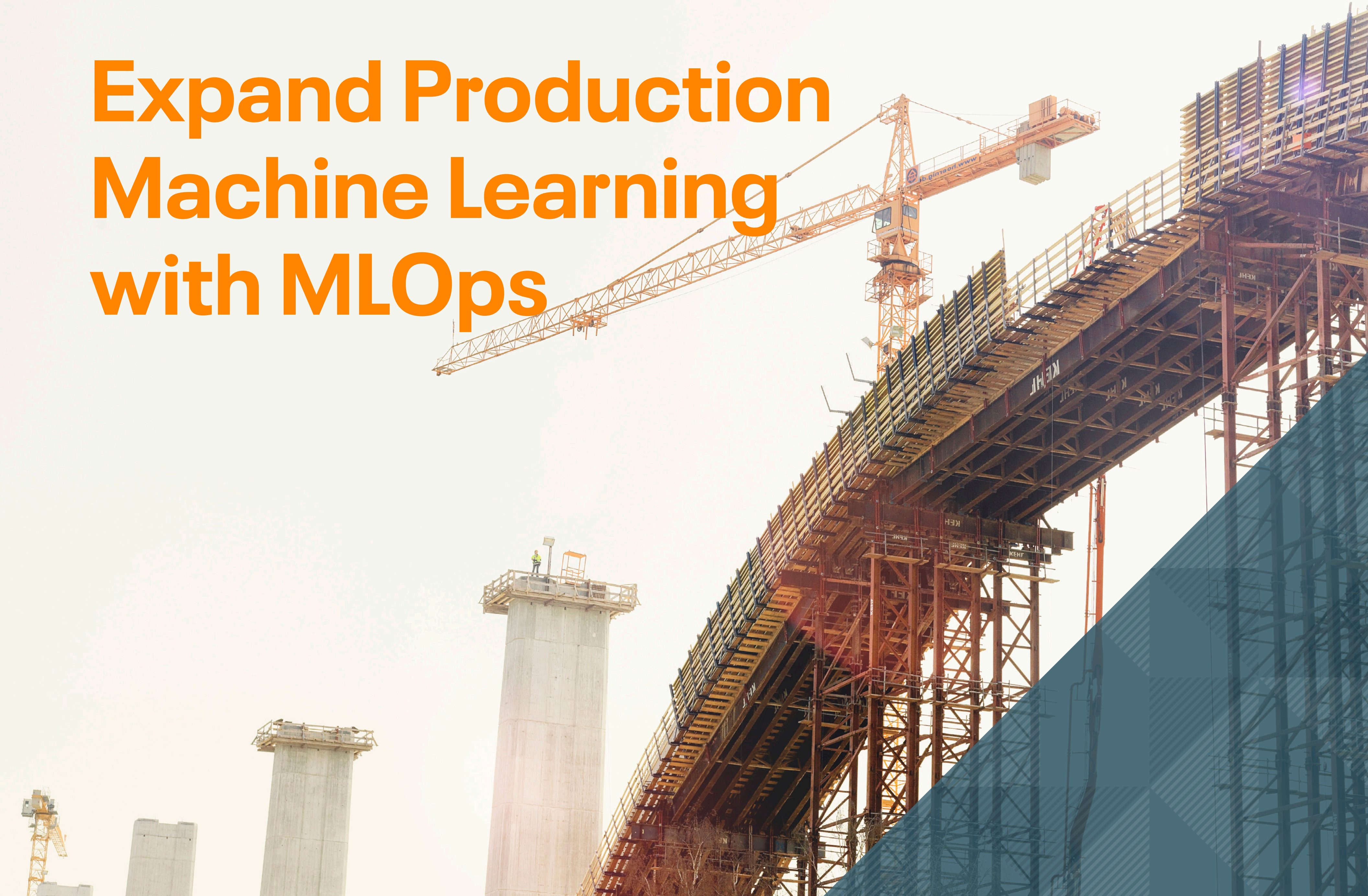


Table of Contents

Introduction: Why Operationalize Machine Learning? 3

Chapter 1: Why Production ML Should be a Critical Part of Your Strategy 4

Chapter 2: Understanding Tooling that Helps Scale Production ML 6

Chapter 3: What You Need to Know About Overcoming Technical Debt 7

Chapter 4: Managing the Risks of Security and Governance 9

Conclusion: Put ML Into Action 10

Introduction: Why Operationalize Machine Learning?

It's challenging enough to launch Artificial Intelligence (AI) and Machine Learning (ML) applications. But after that, ongoing success with production ML depends on continuous maintenance and management. You've likely felt that keeping multiple ML models up to date takes significant resources. So as more and more applications prove their value and come online, how can you keep your AI efforts delivering results at scale?

These challenges require a Machine Learning Operations strategy, or MLOps. MLOps establishes an organized, efficient way of deploying and maintaining ML applications. Like an enterprise AI factory, MLOps establishes a production line for getting ML applications to production, establishing repeatable processes, automating whatever can be automated.

MLOps helps you capitalize on early successes and scale up, following steps to keep existing models current, and establishing controls to maintain data security and governance through the entire production ML lifecycle.

A true end-to-end development to production workflow empowers all your data stakeholders. MLOps leads to improvements in transparency, collaboration, and ROI. It means a quicker, more orderly process for turning raw data into valuable insights.

In this ebook we'll be examining the common challenges that data practitioners face in scaling production ML initiatives, and how an MLOps strategy can turn these challenges into core competencies.

2x

Organizations that document and enforce production ML processes are approximately two times as likely to achieve their goals.[1]

[1] Deloitte, "Deloitte's State of AI in the Enterprise, 4th Edition," 2020

Chapter 1: Why Production ML Should be a Critical Part of Your Strategy

Adopting an organized approach to production ML can quickly become a necessity as your organization becomes dependent on ML. This requires a suite of capabilities to deploy, manage, monitor, retrain, and govern machine learning models at scale.

Scaling ML takes tools and skills above and beyond just data science. ML engineers, operations engineers, and governance managers all play a critical role in keeping ML a consistent and growing part of the business.

Challenges for Enterprise ML Stakeholders

ML engineers

- Managing a backlog of ML models
- Understanding how the business uses ML models
- Maintaining ML models over the long term

• Operations engineers

- Improving visibility and transparency
- Monitoring ML model accuracy
- Ensuring applications perform optimally

• Governance managers

- Understanding data lineage
- Tracking model building and delivery
- Ensuring compliance with data regulations

MLOps affords data professionals the flexibility for production ML deploying and governing models where they perform best, whether that's on-premises, in the cloud, or in a hybrid environment. This flexibility and access to multiple environments is how to enable your enterprise to scale to hundreds or thousands of ML models.

It's not just about scale. It's also about speed. By moving machine learning models into production across your environments—and by making the management and maintenance of those models easier—MLOps can unlock the business value of machine learning faster. Accurate, scalable machine learning models are brought to production quickly, without sacrificing visibility and control.

For these capabilities to succeed, MLOps requires a powerful platform. With Cloudera Machine Learning (CML) on the Cloudera Data Platform, you can run workloads in your data center through CDP Private Cloud or work in any public or multi cloud environment. Teams can train and manage models with total transparency and control on a single end-to-end platform.

3x

Using a MLOps methodology, fast-growing firms are 3X more likely to get models into production while maintaining capabilities like monitoring and governance.[2]

[2] Forrester, "MLOps in Enterprise Production Machine Learning," May 2020

Chapter 2: Understanding Tooling that Helps Scale Production ML

As we've seen, production ML can only be successful with the proper, integrated tools to support a machine learning pipeline. These tools should support efficient data workflows, along with security and governance, by letting teams collaborate without requiring movement of data.

MLOps requires tools that support all the steps across an end-to-end data lifecycle:

- Data ingest
- ETL workloads
- Automation of data engineering pipelines
- Building and deploying models
- Maintenance and monitoring of models
- Retrain models

That last step, monitoring, is both important and complex. There are technical metrics, like uptime and throughput, to make sure ML apps are serving their users. And there are model performance metrics, using scoring to make sure models remain accurate.

Monitoring tools—including visualizations and SDKs for analysis using custom code—help teams sift through this unending flow of metrics, to spot mission-critical issues and address potential problems without disrupting models in production.

MLOps is about setting teams up for success, and following through with monitoring and repeatable steps for improvement in production—so all the work teams put in at the start of the process continue to deliver valuable benefits in the long term.

Key Capabilities of Production ML on Cloudera Machine Learning

Deployment and serving

IT administrators and data scientists can ensure high availability serving of machine learning models at scale.

Monitoring

Data scientists and machine learning engineers can track technical and prediction metrics, and operators should be able to track model performance against KPIs.

Collaboration and governance

Teams can employ repeatable workflows, streamline model operations, and deliver models into production with inherited security and unified authorization.

Chapter 3: What You Need to Know About Overcoming Technical Debt

Work with data science at scale and you're certain to encounter technical debt—legacy code and systems that cannot be sustained over the long term.

Consider the vast array of expertise that each member of a data science team brings. It's natural that everyone prefers different tools. But the result can be an amalgam of solutions pieced together that clash and don't all work in sync. Keep in mind only a few of these solutions might be specific to ML, but all of them are necessary for ML to work.

Tech debt often happens when a series of point solutions are patched together. It only gets worse when some of these solutions reach obsolescence with no easy replacement. Vendor lock-in can contribute to this problem, when critical systems require an existing vendor relationship to keep operational.

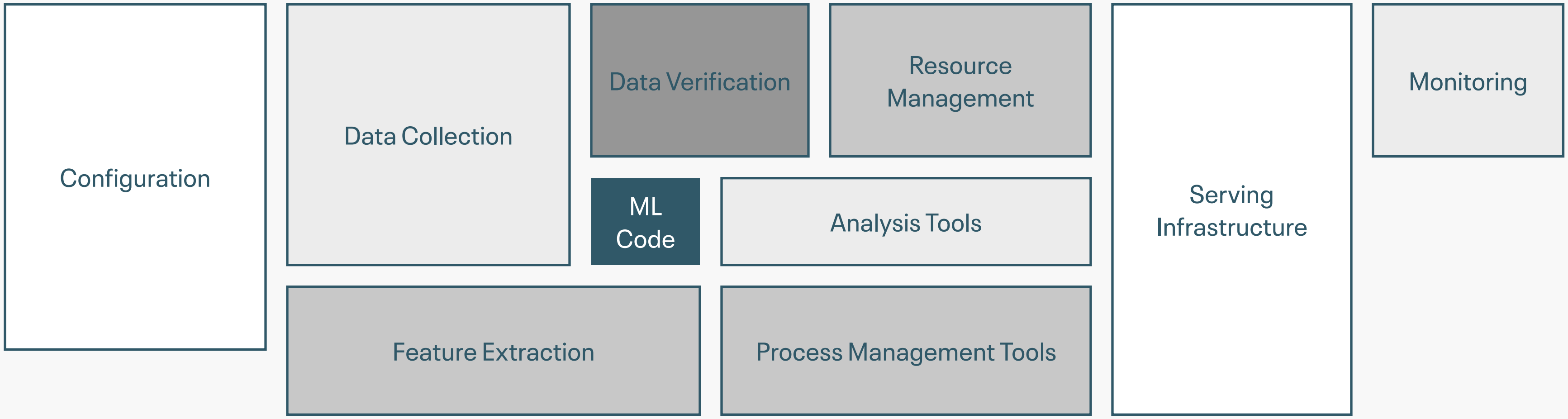
The best solution is to get ahead of the problem, first by recognizing that everyone on an ML team has a specific role and skills. Embrace the different functions of the team, and make sure data professionals have the flexibility to run the open-code tools they prefer. And next, ensure that tools are ready to scale when new ideas emerge.

MLOps strategies make your ML investments successful. Their goal is to sustain ML models and production environments over the long term, prioritizing accuracy and uptime, and mitigating the burden of technical debt.

CML delivers a true end-to-end development to production workflow with tooling that's designed to empower all users across the lifecycle. And because CDP is built on open-source technology, data workflow integrations are made easier and your enterprise can avoid vendor lock-in. All of which reduces the amount of technical debt your teams are creating as they build the future of ML.

How tech debt can surround ML

ML code (represented by the small box in the middle of this diagram) represents only a small fraction of a complete system.[3]



[3] Sculley, et al., "Hidden Technical Debt in Machine Learning Systems"

Chapter 4: Managing the Risks of Security and Governance

As production ML expands in influence and importance at your enterprise, so do the risks associated with data security and compliance. MLOps treats these considerations as an integral part of the strategy.

Data protection regulations—along with laws governing specific use cases like medical, financial, and public sector data—make it essential to control how data is processed and who has access to it. Compliance failures can lead to fines, penalties, and reputational damage.

As you scale your ML initiatives, you need to integrate security and governance as a built-in factor at every step. Users can analyze data in place, and ML models can be developed, deployed, and monitored without introducing additional risks.

As a fully integrated part of Cloudera Data Platform, the Shared Data Experience (SDX) provides an integrated set of security and governance technologies, run independently from compute and storage layers. SDX delivers consistent data context across deployments, through automatic model cataloging and lineage, along with governed and secure production workflows. SDX's built-in security, governance, lineage tracking, and management helps organizations scale effectively while maintaining compliance.

MLOps gives practitioners the freedom to build and share ML models that drive value, while keeping data in place, to reduce additional risks to security or compliance.

CDP Success: GlaxoSmithKline

Operating in the highly regulated pharmaceutical industry, GSK's R&D Information Platform uses Cloudera to manage more than 2,000 databases. GSK needs to confirm that their privacy and security standards meet the rigorous healthcare industry and internal requirements, including the Health Insurance Portability and Accountability Act (HIPAA). Across thousands of data sets, Cloudera SDX enables GSK to manage all the metadata and policy information in a centralized fashion.

Read more here

Conclusion:
Put ML Into Action

It's a story that happens again and again. Early experiments with ML prove their value, and the result is steady demand for ML applications across an enterprise. At that moment, it takes the right strategy to keep those successes going.

MLOps is about supporting production ML and beyond. Your enterprise should be able to effectively use and trust ML predictions in your everyday decision making for better business outcomes—and to continue the process into the future.

When it's time for production ML to scale up, MLOps is the answer. It provides standard processes for working with data and models, along with monitoring and maintenance to make sure models in production are truly delivering value. And by building ML solutions atop a unified platform, as with CML, you'll transform your toughest data challenges into powerful new capabilities.

Take Your Next Step

Learn how a unified data platform offers the tools you need to make MLOps a part of your enterprise.
Read more

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 | US: +1 888 789 1488 | Outside the US: +1 650 362 0488

Sources

1 Deloitte, "Deloitte's State of AI in the Enterprise, 4th Edition," 2020

2 Forrester, "MLOps in Enterprise Production Machine Learning," May 2020

3 Sculley, et al., "Hidden Technical Debt in Machine Learning Systems"

Privacy Policy | Terms of Service

For the most optimized experience leverage AMD CPUs on Dell hardware

Cloudera Data Platform Private Cloud Base

Pod Network:

PowerSwith S5248F-ON series switch

Cluster Aggregation Network:

PowerSwitch Z9432F-ON series switch

Infrastructure Nodes:

PowerEdge R6515

(3) Master nodes

(1) Utility node

(1) Edge node

(3+) Worker Nodes:

PowerEdge R6515 (Configuration 1)

or PowerEdge R7515 (Configuration 2)

GPU Accelerated Worker Node Option:

PowerEdge R7525

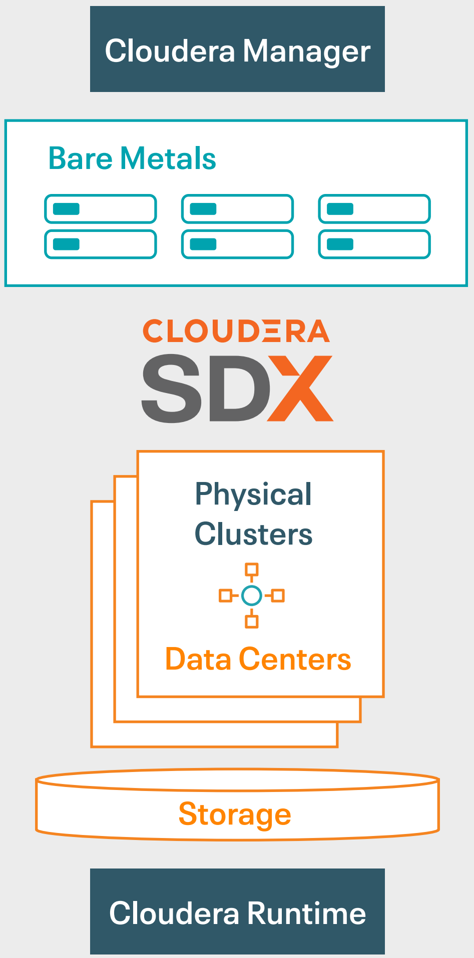
HDFS:

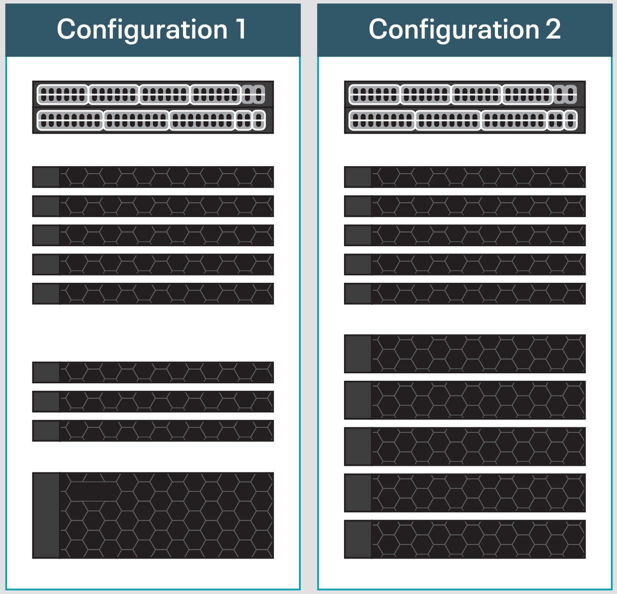
Powerscale H5600 (Configuration 1)

or Additional Worker Nodes (Configuration 2)

CDP Data Center

Installable Software





Independent Compute & Storage

RECOMMENDED

Combined Compute & Storage

AMD

DELL