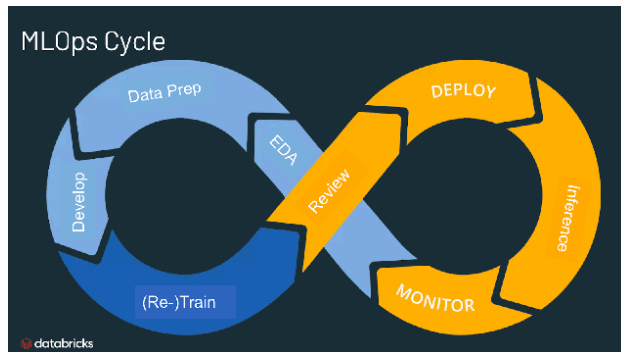


## O que é MLOps?

MLOps significa Machine Learning Operations. No centro da engenharia de machine learning, os MLOps se concentram em colocar modelos de machine learning em produção e simplificar o processo de manutenção e monitoramento. Os MLOps geralmente são responsáveis pela colaboração entre equipes de data scientists, engenheiros de DevOps e TI.



## Para que servem os MLOps?

Os MLOps são uma abordagem poderosa para criar e melhorar soluções de machine learning e IA. A adoção da abordagem MLOps permite a implementação de práticas de CI/CD (integração e implantação contínuas) com monitoramento, validação e governança adequados de modelos de ML, facilitando a colaboração entre data scientists e engenheiros de machine learning para acelerar o desenvolvimento e a produção de modelos.

## Por que precisamos de MLOps?

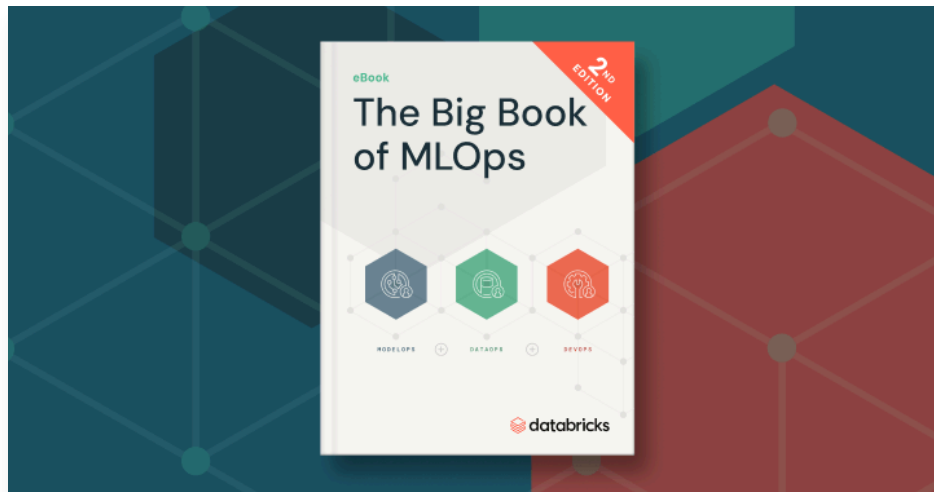
A produção do machine learning é difícil. O ciclo de vida do machine learning consiste em muitos elementos complexos, como ingestão e preparação de dados, treinamento, ajuste, implantação e monitoramento de modelos, explicabilidade e muito mais. Também requer colaboração e transferências entre equipes de data engineering, data science e engenheiros de ML. Operações rigorosas são essenciais para manter todos esses processos sincronizados e trabalhando juntos. O conceito MLOps abrange a fase de experimento, iteração e melhoria contínua no ciclo de vida do machine learning.

## Quais são os benefícios dos MLOps?

Os principais benefícios dos MLOps são eficiência, escalabilidade e mitigação de riscos. Eficiência: os MLOps permitem que as equipes de dados desenvolvam modelos mais rapidamente, forneçam modelos de ML de alta qualidade e acelerem a implantação e a produção. Escalabilidade: os MLOps permitem ampla escalabilidade e gerenciamento, supervisionando, controlando, gerenciando e monitorando milhares de modelos para integração, entrega e implantação contínuas.

Especificamente, os MLOps garantem a repetibilidade dos pipelines de ML e permitem uma colaboração mais próxima entre as equipes de dados, reduzindo o atrito com DevOps e TI e acelerando os ciclos de lançamento. Mitigação de riscos: os modelos de machine learning geralmente exigem verificação regulatória e detalhada. Os MLOps permitem maior transparência e resposta mais rápida a essas solicitações, garantindo a conformidade com as políticas da organização ou do setor.

## Continue explorando



### The Big Book of MLOps (2ª edição)

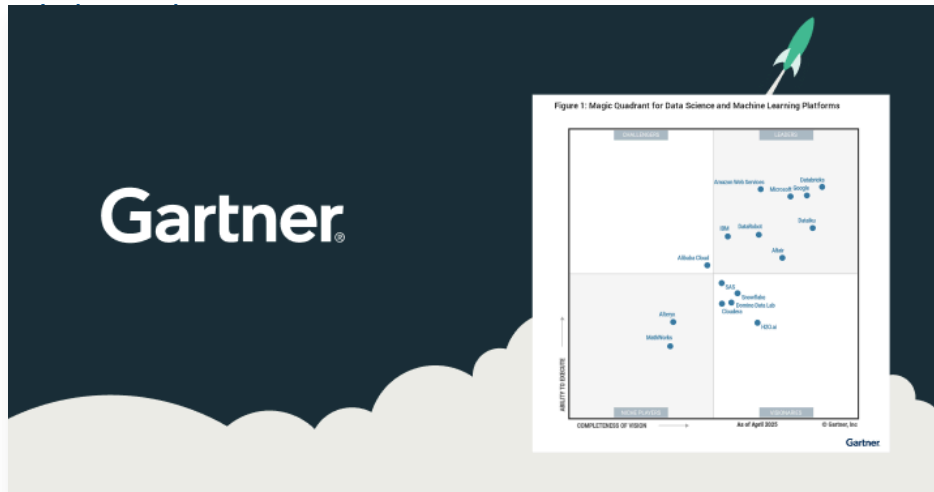
Aprenda novas estratégias de IA generativa e LLMOps.

[Obter o e-Book](#)



## The Big Book of Machine Learning Use Cases

Seu guia completo de instruções para colocar o machine learning em prática, além de casos de uso, exemplos de código e notebooks.

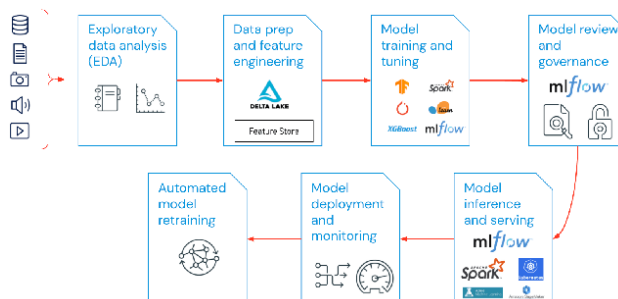


## Databricks classificada como #1 em Execução e Visão

Quadrante™ Mágico do Gartner de 2025 para plataformas de ciência de dados e aprendizado de máquina

[Leia o artigo](#)

## Quais são os componentes dos MLOps?



MLOps em projetos de machine learning podem ser restritos ou extensos, dependendo do projeto. Por exemplo, alguns podem abranger tudo, desde pipelines de dados até a criação do modelo, enquanto outros projetos podem precisar apenas implementar o processo de implantação do modelo. Muitas empresas implementam princípios de MLOps nas seguintes áreas:

- Análise exploratória de dados (EDA)
- Preparação de dados e engenharia de recursos
- Treinamento e ajuste de modelos
- Revisão e governança de modelos
- Inferência e disponibilização de modelos
- Monitoramento de modelos
- Retreinamento automatizado de modelos

## Quais são as práticas recomendadas para MLOps?

As práticas recomendadas de MLOps podem ser diferenciadas pelo estágio em que os princípios de MLOps são aplicados.

- **Análise exploratória de dados (EDA):** criar conjuntos de dados, tabelas e visualizações reproduzíveis, editáveis e compartilháveis para explorar, compartilhar e preparar iterativamente dados no ciclo de vida do machine learning.
- **Preparação de dados e engenharia de recursos:** criar recursos sofisticados transformando, agregando e deduplicando dados de forma iterativa. Sobretudo, aproveitar os armazenamentos de recursos para visualizar e compartilhar recursos entre as equipes de dados.
- **Treinamento e ajuste de modelos:** usar bibliotecas populares de código aberto, como scikit-learn e Hyperopt, para treinar e melhorar o desempenho de modelos. Uma alternativa mais simples é usar uma ferramenta automatizada de machine learning como o AutoML para executar seus testes e produzir código que possa ser revisado e implantado.
- **Revisão e governança de modelos:** rastrear a linhagem do modelo, as versões do modelo e gerenciar os artefatos e as transições do modelo durante todo o ciclo de vida. Além disso, descobrir, compartilhar e colaborar em modelos de ML usando plataformas MLOps de código aberto, como MLflow.
- **Inferência e disponibilização de modelos:** gerenciar a frequência de atualização do modelo, tempo de solicitação de inferência e especificidades de produção semelhantes em testes e controle de qualidade. Automatizar pipeline de pré-produção com ferramentas de CI/CD, como repositórios e orquestradores que incorporam princípios de DevOps.
- **Implantação e monitoramento de modelos:** automatizar a configuração de permissões, a criação de clusters e a preparação de modelos registrados para produção. Ativar o endpoint para seu modelo de API REST.

- **Retreinamento automatizado de modelos:** tomar medidas corretivas com alertas e automação quando ocorrer um desvio do modelo devido a diferenças nos dados de treinamento e inferência.

## Qual é a diferença entre MLOps e DevOps?

MLOps são um conjunto de práticas de engenharia especializadas em projetos de machine learning que se baseia nos princípios de DevOps amplamente adotados na engenharia de software. DevOps, por outro lado, oferece uma abordagem iterativa rápida e contínua para produzir aplicações. O MLOps se baseia nos princípios do DevOps para colocar modelos de machine learning em produção. Ambos resultam em melhor qualidade de software, correções e lançamentos mais rápidos e maior satisfação do cliente.

## Diferenças entre treinar grandes modelos de linguagem (LLMOps) e MLOps tradicionais

Muitos dos conceitos de MLOps se aplicam, mas há outras coisas a serem consideradas ao treinar grandes modelos de linguagem (LLMs) como [Dolly](#). Vejamos alguns dos principais pontos em que o treinamento de LLMs pode diferir das abordagens MLOps tradicionais:

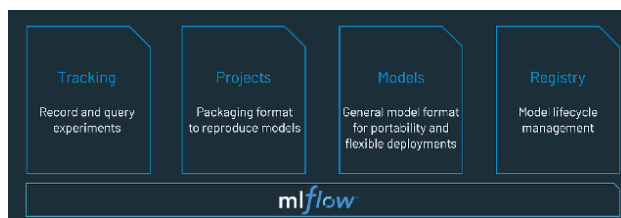
- **Recursos computacionais:** o treinamento e o ajuste de grandes modelos de linguagem geralmente requerem a execução de cálculos de ordens de magnitude em enormes conjuntos de dados. Para acelerar esse processo, hardwares especializados, como GPUs, são usados para executar cálculos paralelos de dados mais rápidos. O acesso a esses recursos computacionais especializados é essencial para treinar e implantar grandes modelos de linguagem. Técnicas de compressão e destilação de modelos também podem ser enfatizadas devido ao custo de inferência.
- **Aprendizado por transferência:** ao contrário dos modelos tradicionais de ML, que são criados e treinados do zero, muitos grandes modelos de linguagem começam com um modelo básico e são ajustados com novos dados para melhorar o desempenho em domínios mais específicos. O ajuste nos permite oferecer desempenho de ponta para aplicações específicas com menos dados e recursos computacionais.
- **Feedback humano:** o aprendizado por reforço com feedback humano (RLHF) é um dos maiores avanços no

treinamento de grandes modelos de linguagem. De maneira mais geral, as tarefas de LLM costumam ser bastante abertas. Portanto, o feedback humano dos usuários finais do aplicativo é importante para avaliar o desempenho do LLM. A incorporação desse loop de feedback no pipeline LLMOps pode melhorar o desempenho de grandes modelos de linguagem treinados.

- **Ajuste de hiperparâmetro:** no ML tradicional, o ajuste de hiperparâmetro geralmente gira em torno de melhorar a precisão e outras métricas. Para LLMs, o ajuste também é importante para reduzir o custo e o poder computacional necessário para treinamento e inferência. Por exemplo, ajustar o tamanho do batch e a taxa de aprendizado pode alterar drasticamente a velocidade e o custo do treinamento. Assim, embora o ML e LLMs tradicionais se beneficiem do rastreamento e da otimização do processo de ajuste, a ênfase é diferente.
- **Métricas de desempenho:** os modelos tradicionais de ML têm métricas de desempenho muito bem definidas, como precisão, AUC, pontuação F1 etc. Esses indicadores são muito fáceis de calcular. No entanto, quando se trata de avaliação de LLMs, são aplicados indicadores padronizados e pontuações bastante diferentes, como Bilingual Evaluation Understudy (BLEU) e Recall-Oriented Understudy for Gisting Evaluation (ROUGE).

## O que é uma plataforma MLOps?

A plataforma MLOps oferece aos data scientists e engenheiros de software um ambiente colaborativo que facilita a exploração iterativa de dados, rastreamento de experimentos, engenharia de recursos, recursos de edição colaborativa em tempo real para gerenciamento de modelos e migração, implantação e monitoramento controlados de modelos. O MLOps automatiza os aspectos operacionais e síncronos do ciclo de vida do machine learning.



Experimente a Databricks: um ambiente totalmente gerenciado para MLflow, a principal plataforma MLOps aberta do mundo. <https://www.databricks.com/try/databricks-free-ml>