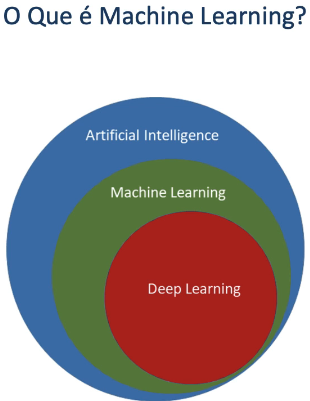
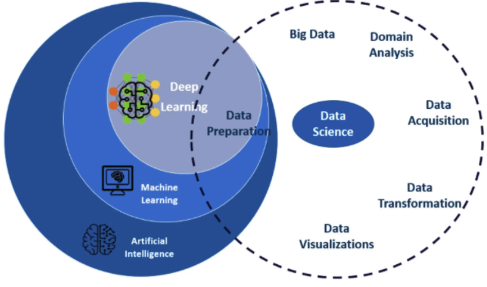
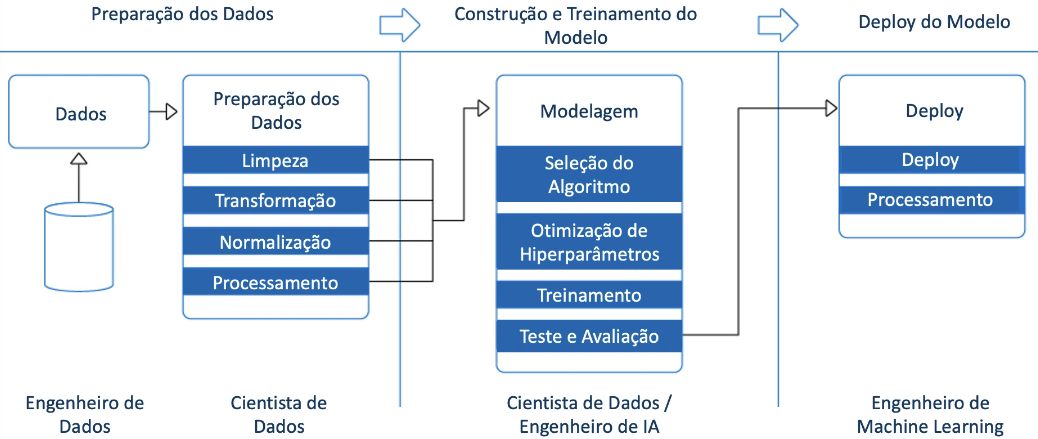
**O QUE É MACHINE LEARNING?**

é uma sub-área da inteligência artificial ( IA ) e da ciência da computação que se concentra no uso de dados e algoritmos para imitar a forma como os humanos aprendem, melhorando gradativamente sua precisão

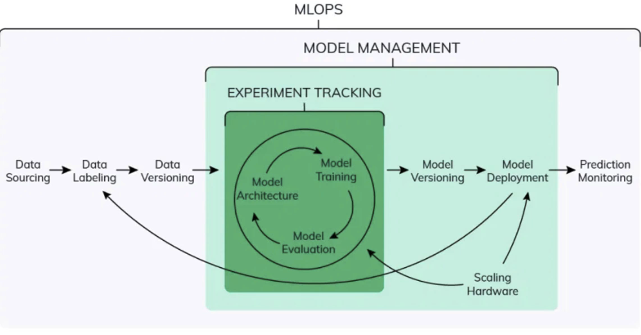
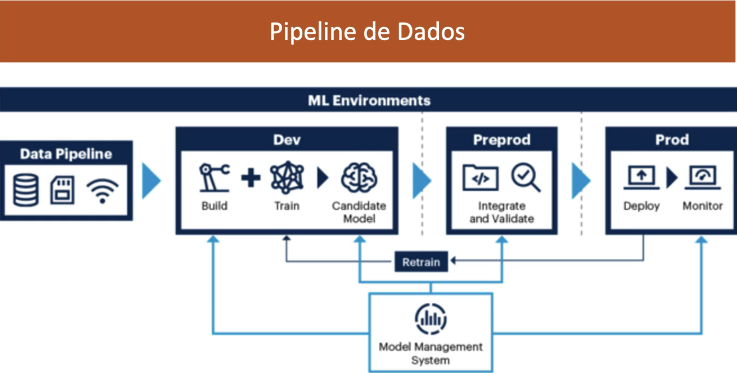
* o que fazemos em machine learning é alimentando por Big Data

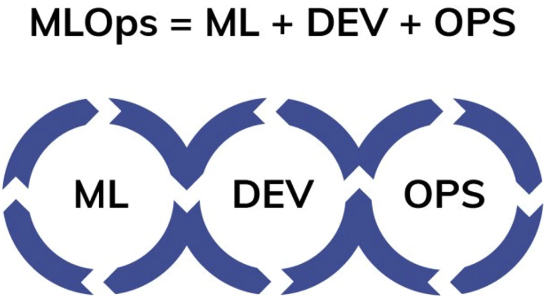


* 
* **O Pipeline de Machine Learning**



* **O que é machine learning Ops?**





* MLOps é um conjunto de práticas para colaboração e comunicação entre cientistas de dados e profissionais de operações

“a criação do processo de MLops é tarefa do eng de machine learning e a criação do modelo de machine learning é tarefa do cientista de dados”

* A aplicação dessas práticas aumenta a qualidade, simplifica o processo de gerenciamento e automatiza a implantação de modelos de aprendizado de máquina em ambientes de produção em grande escala. *é mais fácil alinhas os modelos ás necessidades de negócios,* bem como aos requisitos regulamentares
* MLOps visa unificar o desenvolvimento de sistemas de ML (dev) e a implantação de sistemas de ML (ops) para padronizar e agilizar a entrega contínua de modelos de alto desempenho em produção

**PDF MLOPs**

* [5. Quais Problemas Podem Ser Resolvidos com MLOps.pdf](https://drive.google.com/file/d/1pg-O8i-iWyh9F3HU-KkkcLvPXQeSlJDX/view?usp=sharing)
* [5. O Engenheiro de Machine Learning.pdf](https://drive.google.com/file/d/11Hs6E9VTYuPbrz4Sjmdytm2SvgtRcJTx/view?usp=sharing)
* **DevOps, MLOps, AIOps e DataOps**

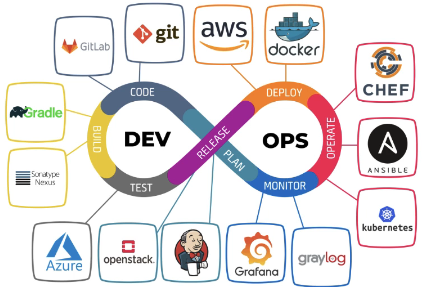
**MLOps = Operação do fluxo de trabalho em machine learning**

**AIOps = Operação do fluxo de trabalho em IA**

**DataOps = conceito mais recente que abrange toda a operação de dados de uma empresa**

* **DevOps**
* é uma abordagem para desenvolvimento de software que acelera o ciclo de vida de construção usando automação. O devOps se concentra na implantação contínua de software, aproveitando os recursos de TI sob demanda e automatizando a integração, o teste e a implantação de código. Essa fusão de desenvolvimento de software (‘dev’) e operações de TI (‘ops’) reduz o tempo de implantação, diminui o tempo de lançamento no mercado, minimiza defeitos e diminui o tempo necessário para resolver problemas

* Usando DevOps, empresas conseguiram reduzir o tempo do ciclo de lançamento de software de meses para literalmente segundos. Essa descoberta permitiu o crescimento e liderança em mercados emergentes e em ritmo acelerado. Empresas como Google, Amazon e muitas outras agora lançam software muitas vezes por dia. Ao melhorar a qualidade e o tempo de ciclo dos lançamentos de código, o DevOps merece muito crédito pelo sucesso dessas empresas



* **DataOps ( Operação de Dados )**
* é uma metodologia ágil e orientada a processos para desenvolver e entregar análises.
* DataOps fornece as ferramentas, processos e estruturas organizacionais para apoiar a empresa focada em dados.



* DataOps é a capacidade de habilitar soluções, desenvolver produtos de dados e ativar dados para valor comercial em todas as camadas de tecnologia, da infraestrutura á experiência do usuário final.
* O objetivo do DataOps é agilizar o design, o desenvolvimento e a manutenção de aplicativos com base em dados e análise de dados. Busca melhorar a forma como os dados são gerenciados e os produtos são criados e coordenar essas melhorias com os objetivos do negócio.
* As equipes de DataOps também buscam orquestrar dados, ferramentas, código e ambientes do início ao fim, com o objetivo de fornecer resultados reproduzíveis.
* As equipes de DataOps tendem a ver os pipelines analíticos como análogos ás linhas de produção de uma fábrica, sendo que aqui a matéria-prima é o Big Data.

Semelhante a um Engenheiro DevOps no campo de desenvolvimento de software, os

Engenheiros DataOps são profissionais técnicos que se concentram principalmente ou

exclusivamente no ciclo de vida de desenvolvimento e implantação, e não no produto em si.

Mas os Engenheiros DataOps apoiam o ciclo de fornecimento e utilização de dados

definindo o processo e as tecnologias que outros usam para originar, transformar e utilizar dados. O Engenheiro DataOps tem um perfil mais amplo que o Engenheiro de Dados.

**BIG DATA X SMALL DATA**

Big data = grandes volumes de dados com muita variedade e gerados em alte velocidade

Small data = Dados que estão disponíveis em quantidade mínima para compreensão huma

**PRINCIPAIS FERRAMENTAS DE MLOps**

* DVC

<https://dvc.org>

DVC, ou Data Version Control, é um sistema de controle de versão de código aberto para projetos de aprendizado de máquina. É uma ferramenta de experimentação que ajuda a definir o pipeline, independentemente da linguagem usada.

* Pachyderm

<https://www.pachyderm.com/>

Pachyderm é uma plataforma que combina linhagem de dados com pipelines de ponta a ponta. Ele está disponível em três versões, Community Edition (código aberto, com capacidade de ser usado em qualquer lugar), Enterprise Edition (plataforma completa com versão controlada) e Hub Edition (ainda uma versão beta, combina características das duas versões anteriores).

* Airflow

<https://airflow.apache.org/>

Airflow é uma plataforma de código aberto que permite monitorar, agendar e gerenciar osfluxos de trabalho usando app web. Ele fornece uma visão sobre o status das tarefas concluídas e em andamento, juntamente com uma visão dos logs.

* Neptune

<https://neptune.ai/product>

O Neptune é um repositório de metadados desenvolvido para equipes de pesquisa e produção que realizam muitos experimentos.

* MLflow

<https://mlflow.org/>

MLflow é uma plataforma de código aberto que ajuda a gerenciar todo o ciclo de vida do aprendizado de máquina que inclui experimentação, reprodutibilidade, implantação e um registro de modelo central. O MLflow é adequado para indivíduos e equipes de qualquer tamanho. A ferramenta é independente de biblioteca. Você pode usá-lo com qualquer biblioteca de aprendizado de máquina e em qualquer linguagem de programação.

**PRINCIPAIS FERRAMENTAS DE DATAOps**

* DataKitchen

<https://datakitchen.io/>

Uma das ferramentas DataOps mais populares, DataKitchen é a melhor para automatizar e coordenar pessoas, ambientes e ferramentas em análise de dados de toda a organização. O DataKitchen cuida de tudo - do teste à orquestração, ao desenvolvimento e à implantação. O DataKitchen permite que as organizações criem ambientes de trabalho em questão de minutos para que as equipes possam experimentar sem interromper os ciclos de produção. O pipeline de qualidade do DataKitchen é baseado em três seções principais; dados, produção e valor. É essencial entender que, com esta ferramenta, você pode acessar o pipeline com o código Python, transformá-lo via Linguagem SQL, projetar o modelo em R, visualizar na pasta de trabalho e obter relatórios no Tableau.

* Genie

<https://netflix.github.io/genie/>

Desenvolvida pela Netflix, essa ferramenta DataOps é um mecanismo de código aberto que oferece serviços de orquestração de trabalhos distribuídos. Essa ferramenta fornece APIs para desenvolvedores que desejam executar uma ampla variedade de trabalhos com Big Data, usando Hive, Hadoop, Presto e Spark. Genie também fornece APIs para gerenciamento de metadados em clusters de processamento distribuído.

.

* Piper

<https://www.piperr.io/>

Piper é um pacote de ferramentas de DataOps baseadas em aprendizado de máquina que permite que as organizações leiam dados de maneira mais suave e eficiente. Esta solução expõe os dados por meio de um conjunto de APIs que se integram facilmente aos ativos digitais da organização.

* Airflow

<https://airflow.apache.org/>

Apache Airflow é uma plataforma de DataOps (e também MLOps) de código aberto que gerencia fluxos de trabalho complexos em qualquer organização, considerando os processos de dados como DAG (Directed Acyclic Graphs). Projetado pelo Airbnb para agendar e monitorar seus fluxos de trabalho, agora as empresas podem utilizar essa ferramenta de código aberto para gerenciar seu processo de dados no MacOS, Linux e Windows.

REFERÊNCIAS 👍

MLOps: Continuous delivery and automation pipelines in machine learning

<https://cloud.google.com/architecture/mlops-continuous-delivery-and-automation-pipelinesin-machine-learning>

Machine Learning Operations

<https://ml-ops.org/>

The Complete Guide to an Enterprise DataOps Transformation

<https://datakitchen.io/recipes-for-dataops-success-guide-to-dataops-transformation/>