Prof. Jean Carlos Alves

DataOps e Implantação de Sistemas de Machine Learning

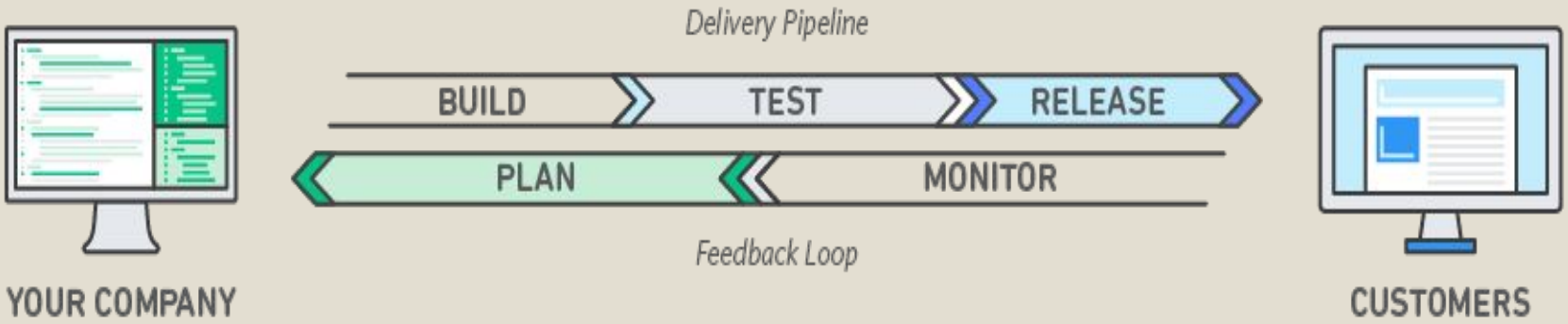
PUC Minas

Devops

- "DevOps é uma metodologia de desenvolvimento de software que traz entrega contínua para o ciclo de vida de desenvolvimento de sistemas, combinando equipes de desenvolvimento e equipes de operações em uma única unidade responsável por um produto ou serviço."

https://cio.com.br/tendencias/o-que-e-dataops-analytics-colaborativo-e-multifuncional/

Devops

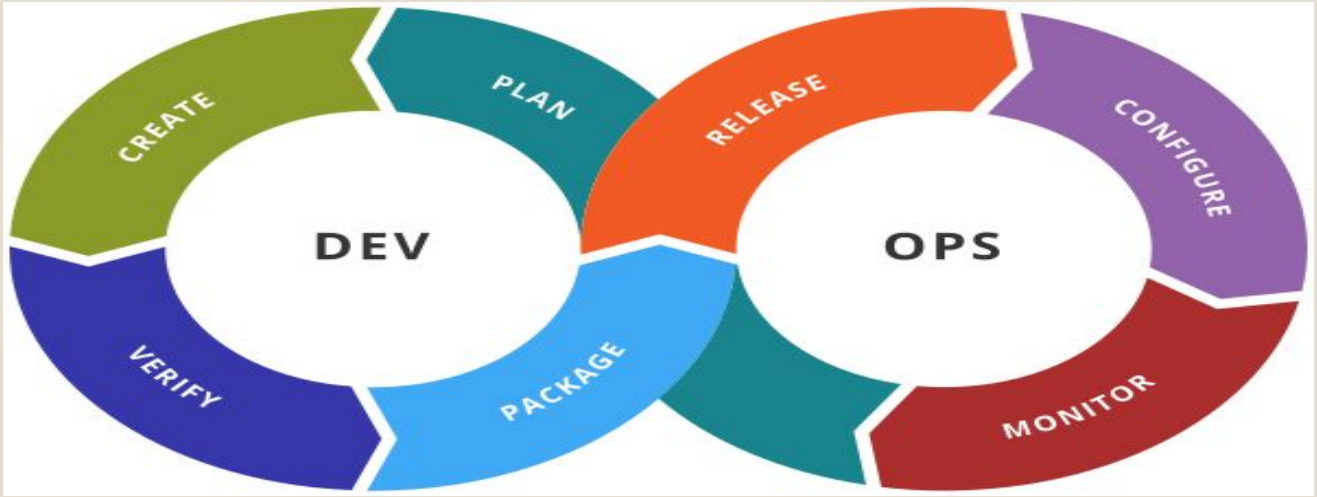


https://aws.amazon.com/pt/devops/what-is-devops/

Benefícios Devops

- Velocidade
- Entrega rápida
- Confiabilidade
- Escala
- Equipes colaborativas
- Segurança

Ciclo de vida Devops



https://medium.com/data-ops/dataops-is-not-just-devops-for-data-6e03083157b7

Ciclo de vida Devops

- Código
- Compilação
- Testes automatizados
- Artefatos
- Release
- Configuração
- Monitoramento

Práticas Devops

- Integração Contínua
- Entrega Contínua
- Microsserviços
- Infraestrutura como código
- Monitoramento
- Equipe Colaborativa

DataOps

- "Capacidade de habilitar soluções, desenvolver produtos de dados e ativar dados para valor de negócios em todas as camadas de tecnologia, da infraestrutura à experiência"

  Michele Goetz, Vice-Presidente e Analista Principal da Forrester

https://cio.com.br/tendencias/o-que-e-dataops-analytics-colaborativo-e-multifuncional/

Príncipios DataOps

- Satisfação Cliente
- Análise do trabalho
- Mudanças
- Trabalho em equipe
- Interações diárias
- Auto organizar
- Reduza o Heroísmo

https://www.dataopsmanifesto.org/

Príncipios DataOps

- Reflita
- Analytics é código
- Orquestrar
- Resultados reproduzíveis
- Ambientes descartáveis
- Simplicidade
- Análise é fabricação

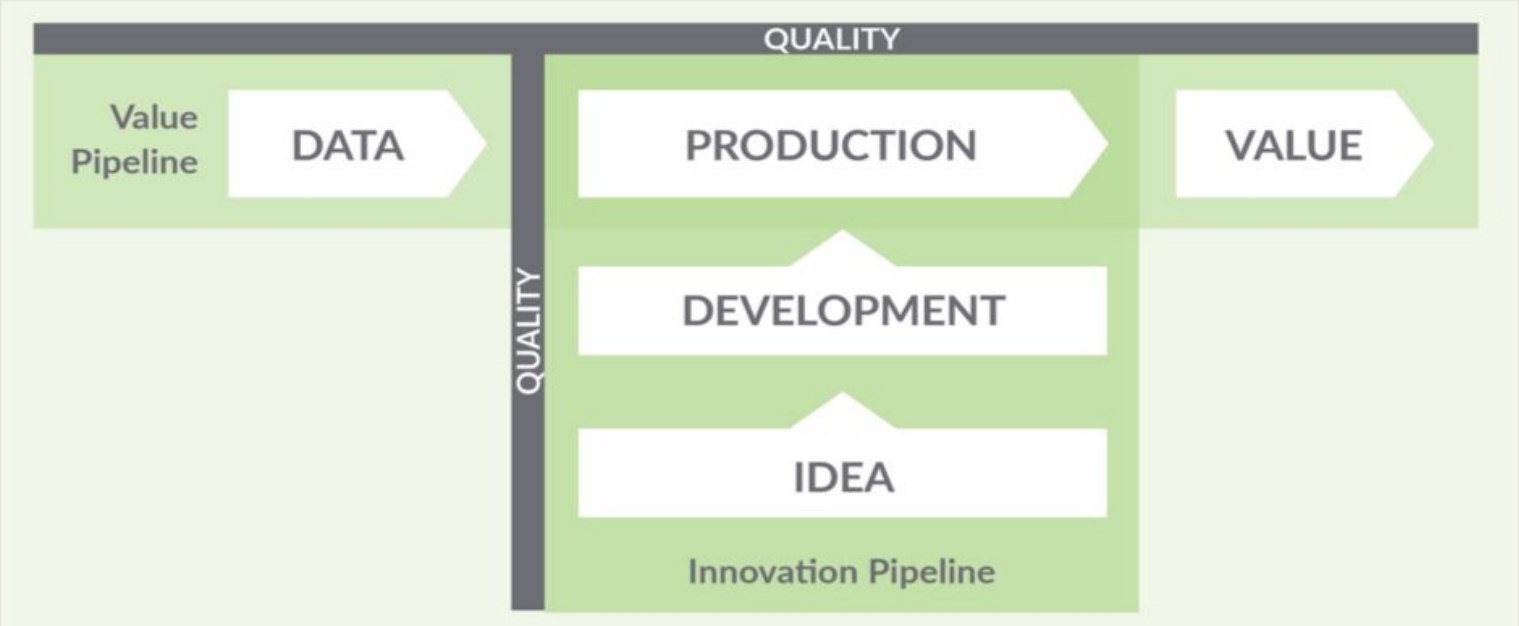
https://www.dataopsmanifesto.org/

Príncipios DataOps

- Qualidade é fundamental
- Monitorar qualidade e desempenho
- Reutilizar componentes
- Otimizar tempo para implantação de soluções

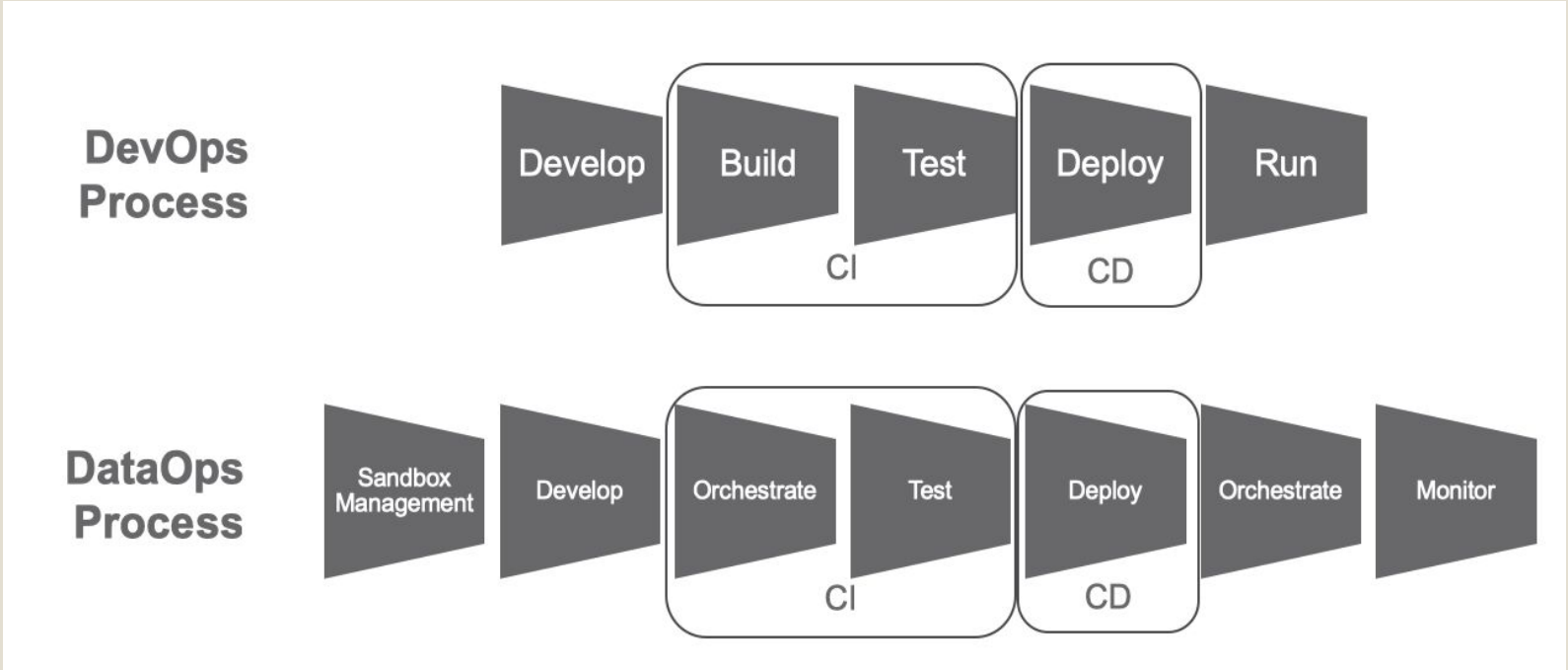
https://www.dataopsmanifesto.org/

Ciclo de vida DataOps



https://medium.com/data-ops/dataops-is-not-just-devops-for-data-6e03083157b7

Devops X Dataops



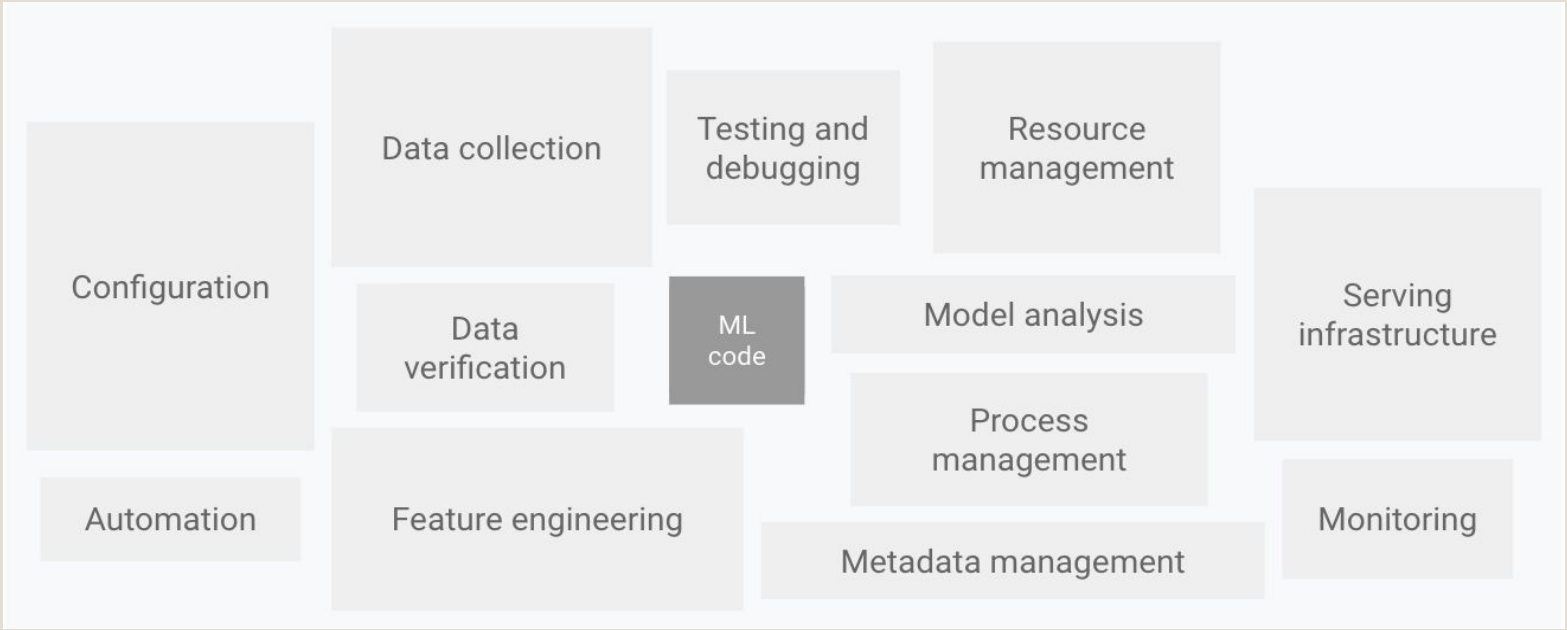
https://medium.com/data-ops/dataops-is-not-just-devops-for-data-6e03083157b7

MlOps

- "MLOps (um composto de aprendizado de máquina e “operações de tecnologia da informação”) é uma nova disciplina / foco / prática para colaboração e comunicação entre cientistas de dados e profissionais de tecnologia da informação (TI) enquanto automatiza e produz algoritmos de aprendizado de máquina. Por meio da prática e das ferramentas, o MLOps visa estabelecer uma cultura e um ambiente onde as tecnologias de ML podem gerar benefícios de negócios ao construir, testar e lançar a tecnologia de ML de forma rápida, frequente e confiável." Nisha Talagala

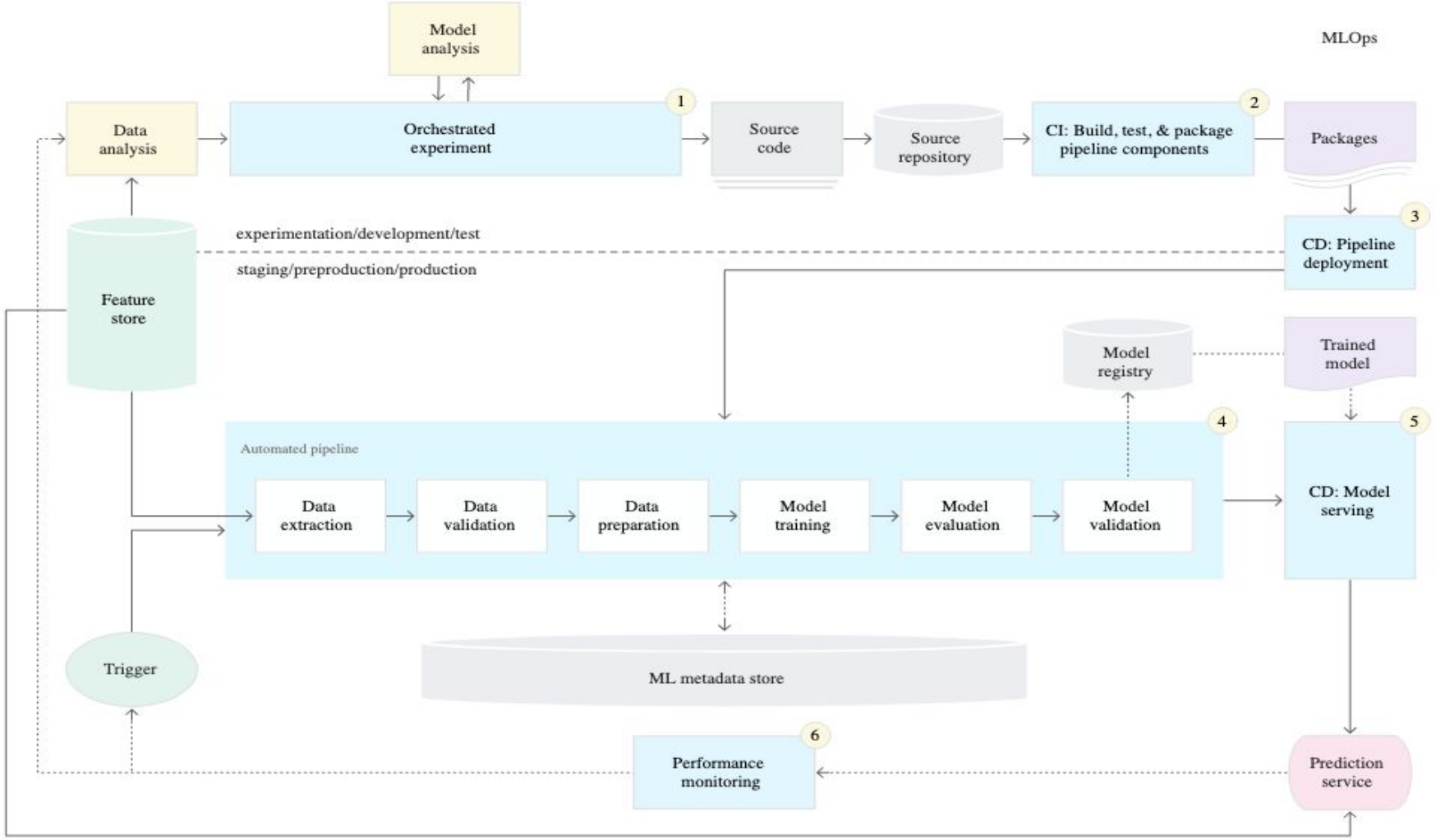
https://www.aitrends.com/machine-learning/mlops-not-just-ml-business-new-competitive-frontier/

Elementos sistemas ML



https://cloud.google.com/solutions/machine-learning/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning?hl=pt-br

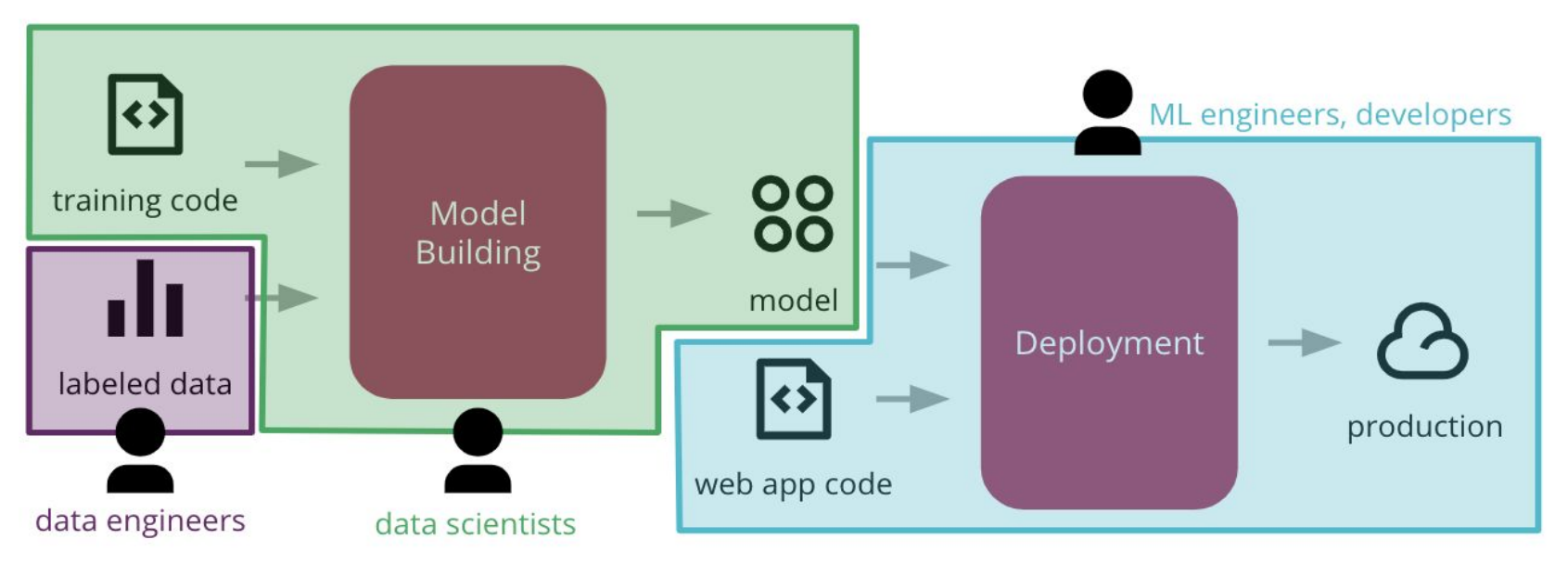
Exemplo pipeline ML


MLOps
Model analysis
Data analysis
Orchestrated experiment
1
Source code
Source repository
CI: Build, test, & package pipeline components
2
Packages
3
CD: Pipeline deployment
experimentation/development/test
staging/preproduction/production
Feature store
Model registry
Trained model
4
5
Automated pipeline
Data extraction
Data validation
Data preparation
Model training
Model evaluation
Model validation
CD: Model serving
Trigger
ML metadata store
6
Performance monitoring
Prediction service

CD4ML - Continuous Delivery for Machine Learning

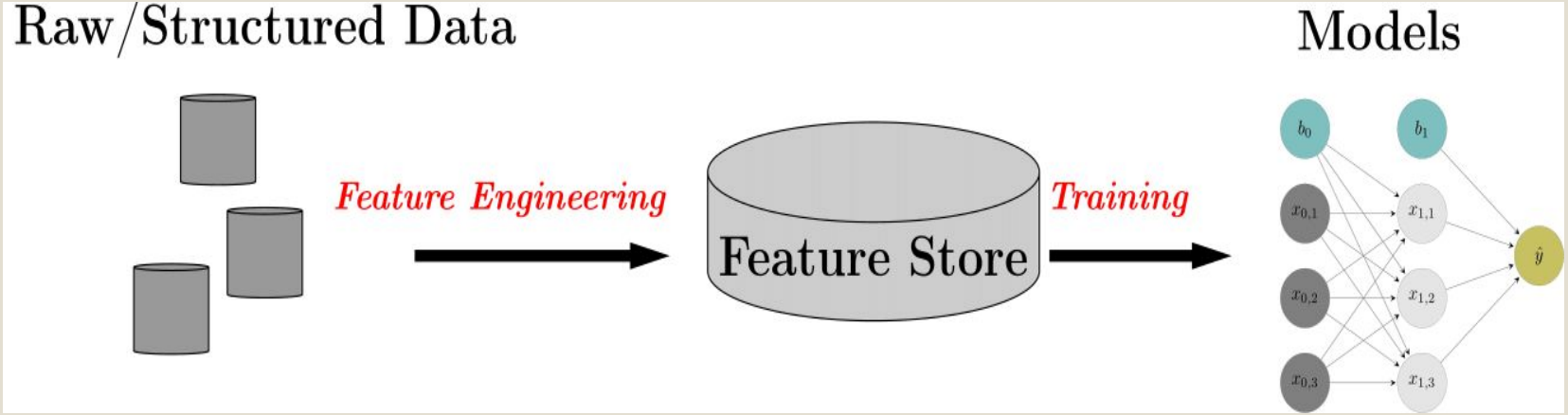
- "A entrega contínua é a capacidade de colocar mudanças de todos os tipos - incluindo novos recursos, mudanças de configuração, correções de bugs e experimentos - em produção ou nas mãos dos usuários com segurança e rapidez de forma sustentável." Jez Humble e Dave Farley

Etapas para criação de modelo ML



https://martinfowler.com/articles/cd4ml.html

Dados (Feature Store)



https://hopsworks.readthedocs.io/en/1.1/featurestore/featurestore.html

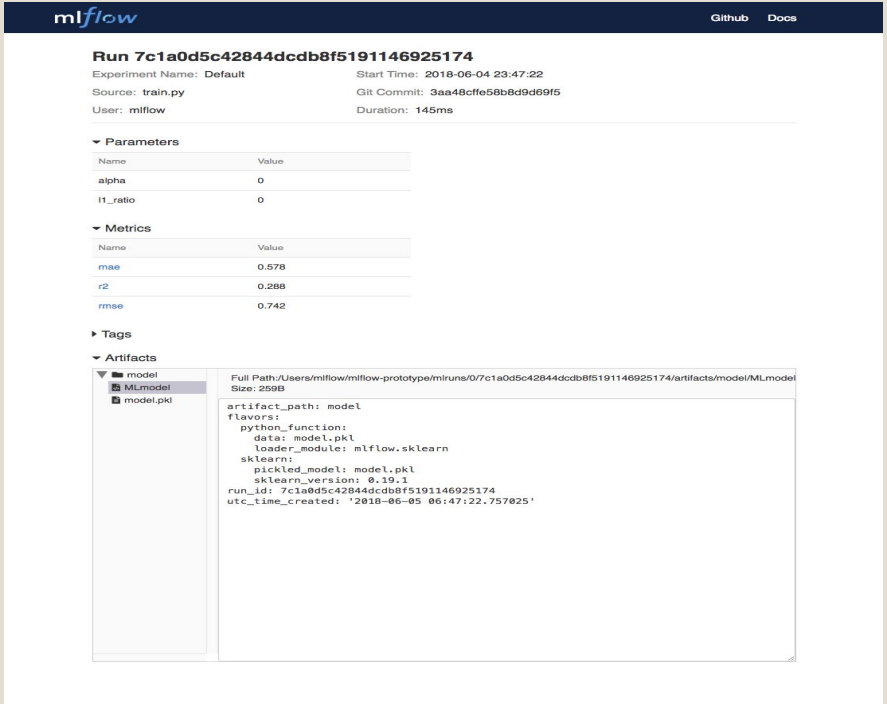
Dados (Feature Store)

- Reutilização
- Flexibilidade
- Disponibilidade
- Qualidade

Desenvolvimento (MLFLOW)

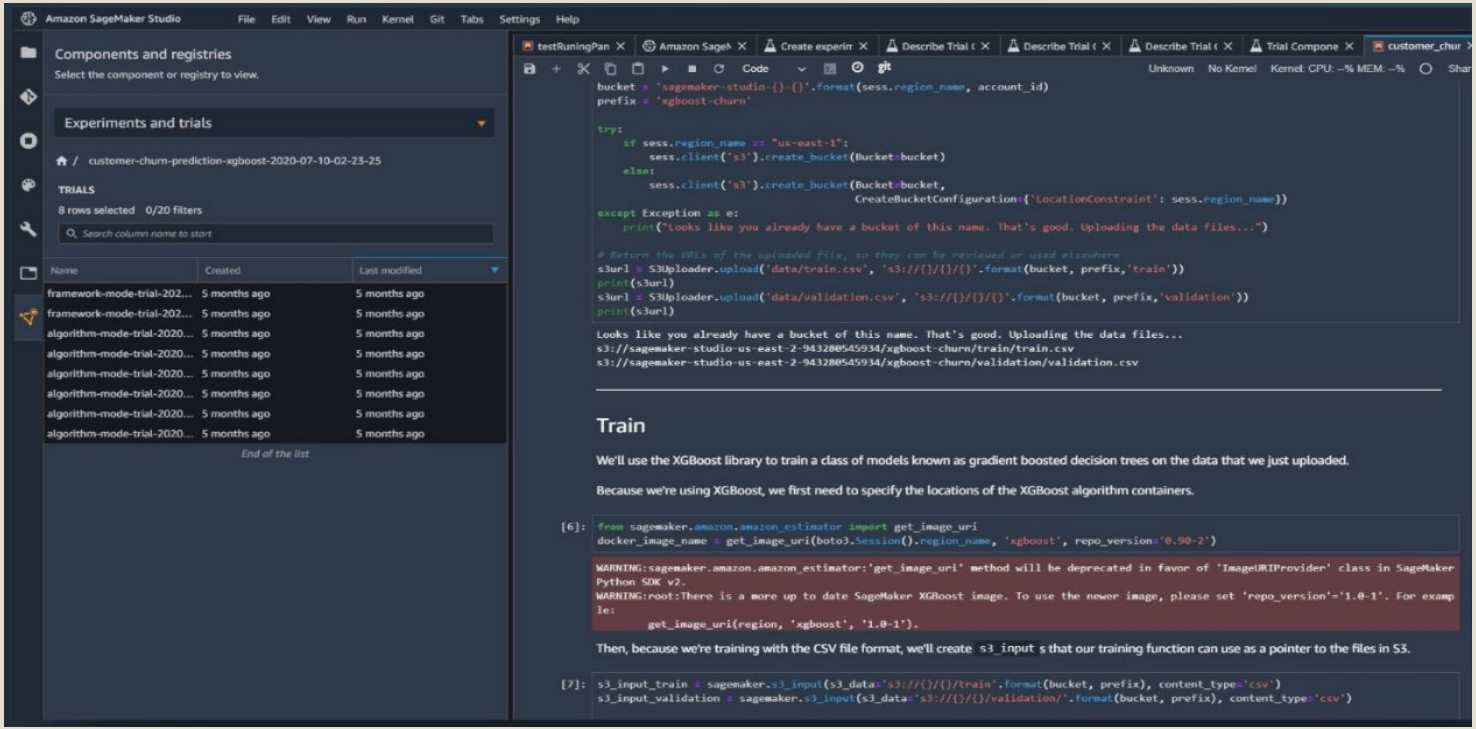
- Rastreamento
- Projetos
- Modelos
- Registros de modelos

Desenvolvimento (MLFLOW)

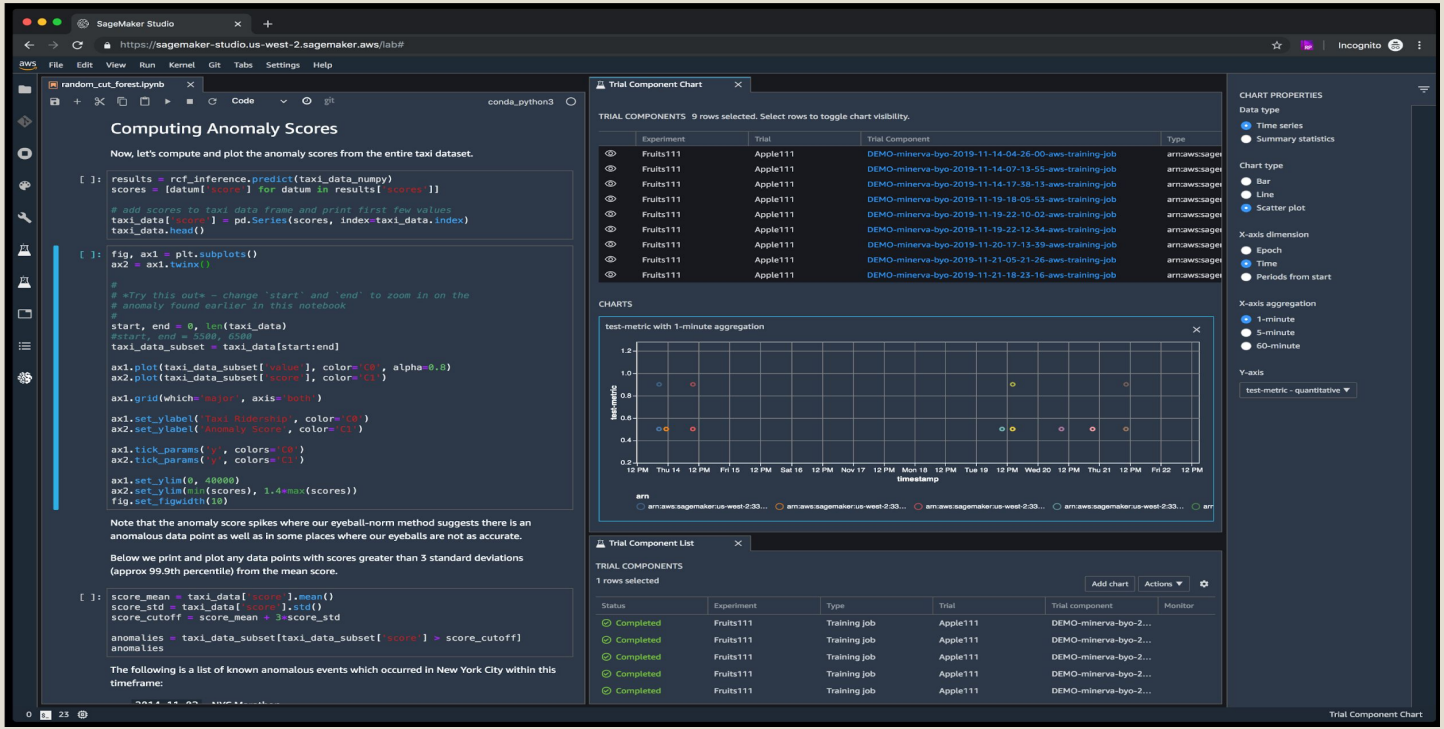


https://mlflow.org/docs/latest/tutorials-and-examples/tutorial.html

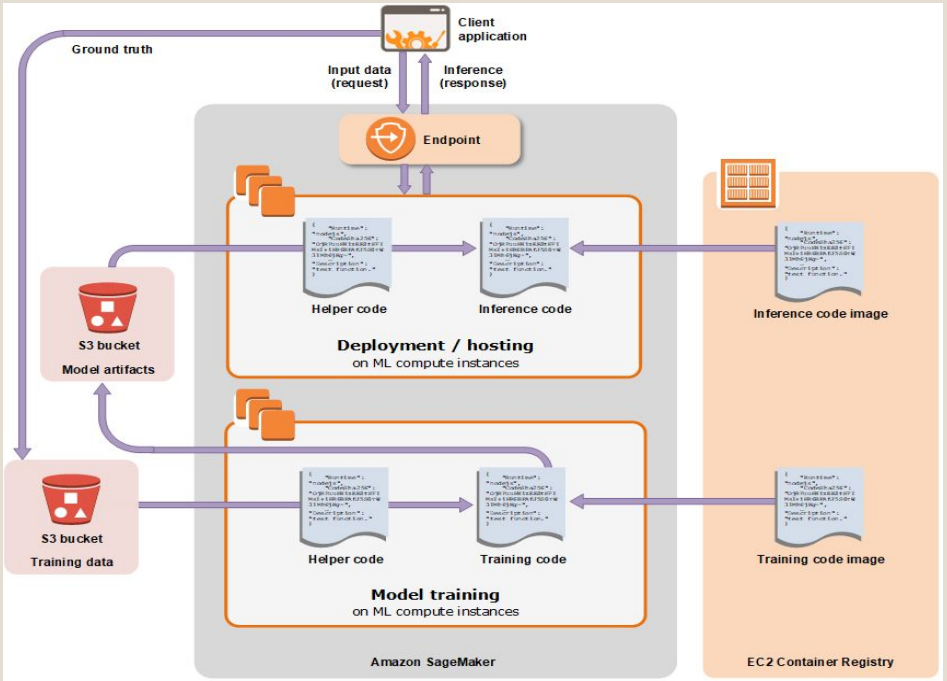
Desenvolvimento (SageMaker Stúdio)



Desenvolvimento (SageMaker Stúdio)



Desenvolvimento (SageMaker Stúdio)



https://docs.aws.amazon.com/sagemaker/latest/dg/how-it-works-training.html

Bibliografia

- https://cio.com.br/tendencias/o-que-e-dataops-analytics-colaborativo-e-multifuncional/
- https://aws.amazon.com/pt/devops/what-is-devops/
- https://medium.com/data-ops/dataops-is-not-just-devops-for-data-6e03083157b7
- https://aws.amazon.com/pt/devops/what-is-devops/
- https://www.dataopsmanifesto.org/
- https://medium.com/data-ops/dataops-is-not-just-devops-for-data-6e03083157b7
- https://www.aitrends.com/machine-learning/mlops-not-just-ml-business-new-competitive-frontier/
- https://cloud.google.com/solutions/machine-learning/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning?hl=pt-br
- https://www.aitrends.com/machine-learning/mlops-not-just-ml-business-new-competitive-frontier/
- https://martinfowler.com/articles/cd4ml.html
- https://hopsworks.readthedocs.io/en/1.1/featurestore/featurestore.html
- https://mlflow.org/docs/latest/tutorials-and-examples/tutorial.html
- https://docs.aws.amazon.com/sagemaker/latest/dg/how-it-works-training.html

OBRIGADO

Jean Carlos Alves