Digital House > Coding School

ML em produção

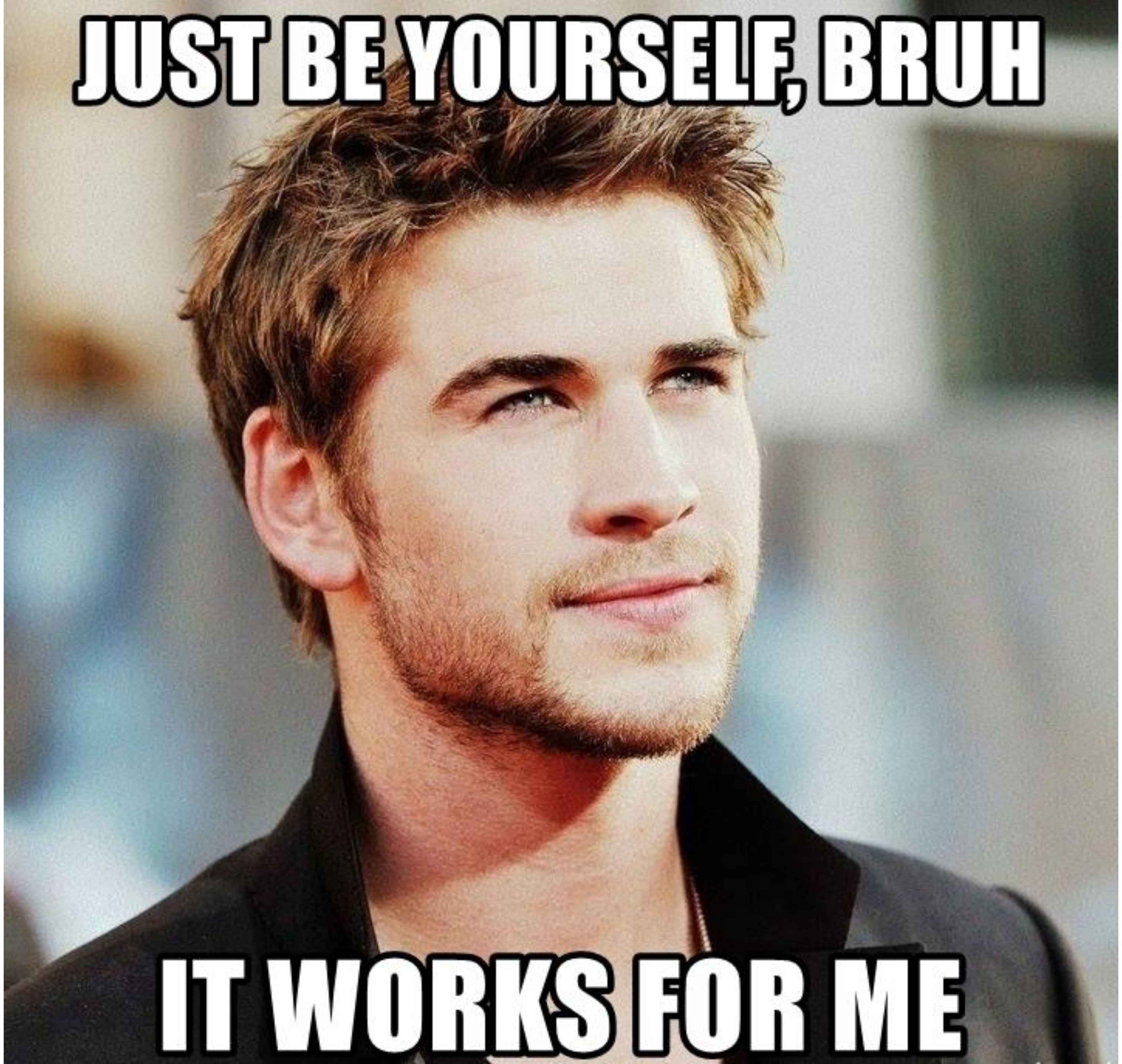
Fev/2021

Você conseguiu! E agora?

- Suponha que trabalha em uma empresa de petróleo. Você desenvolveu um modelo que antecipa falhas em soldas de tubulações a partir de sensores de ultrassom com 95% de precisão!
- Em seu computador com 32Gb de RAM, seu modelo responde em 100ms e consome 3Mb de memória RAM para cada coleta. Contudo, a aplicação requer mais 2Gb para executar a infraestrutura necessária (sistema operacional, interpretador de código, conexões de rede, etc) Portanto, a cada 100ms seu computador é capaz de analisar ~10.000 coletas (suponha que não há outras limitações de hardware)
- Sua empresa decide adotar a sua solução. Os sensores de campo coletam dados a cada 100ms e as soldas são executadas a cada 100m em uma rede de tubulação total de 100 km. Ou seja, são 10.000 coletas, capacidade do seu computador, no mesmo intervalo de execução.
- Como você sugere disponibilizar?


MY CODE
LUCK
OLD REPOSITORIES
INDIAN GUY ON YOUTUBE
STACKOVERFLOW
GOOGLE


JUST ASK NICELY
IT WORKS FOR ME


JUST BE YOURSELF, BRUH
IT WORKS FOR ME

Você conseguiu! E agora?

- É necessário definir como a solução será integrada ao processo atual de manutenção. Como são feitas (inspeção visual no local, análise manual, etc) e com que frequência são feitas as análises.
- Antes de colocar em operação, rodar em paralelo o processo atual e calcular a performance efetiva.
- Uma máquina seria suficiente, mas e se ocorrer algum problema com ela?
- O que fazer quando for necessário instalar atualizações críticas?
- Devido a instabilidades na rede de dados, o tempo de entrega das medições tem um desvio-padrão de 20ms. Você precisa de cada medição ou pode fazer uma amostragem inferior?
- Como seu programa irá responder 10.000 requisições simultâneas? Consolida os dados no request?
- E se você expandir essa solução para outros países totalizando, digamos, 10.000km? Você consolidaria a execução ou seria melhor disponibilizar o código para os engenheiros de cada país?
- Um novo algoritmo aumentou a performance em teste para 99%, mas utiliza dados de sensores de pressão e temperatura, aumentando a necessidade de hardware.
- A auditoria informou que será necessário manter o log dos dados por 5 anos.
- A área de engenharia perguntou se poderia adotar a mesma solução para os seus silos.
- A performance em operação degradou para 75%. O que pode ter ocorrido?
- As condições de operação da rede são diferentes das condições dos dados utilizados (sensores diferentes, ruído, fluidos diferentes, tempo de vida, etc), descobriu um erro no código, etc

Why do 87% of data science projects never make it into production?

https://venturebeat.com/2019/07/19/why-do-87-of-data-science-projects-never-make-it-into-production/

businesses. Many firms say they are uncertain of the business case or return on investment. A review of more than 160 use cases shows that AI was deployed commercially in only 12 percent of cases.

https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx

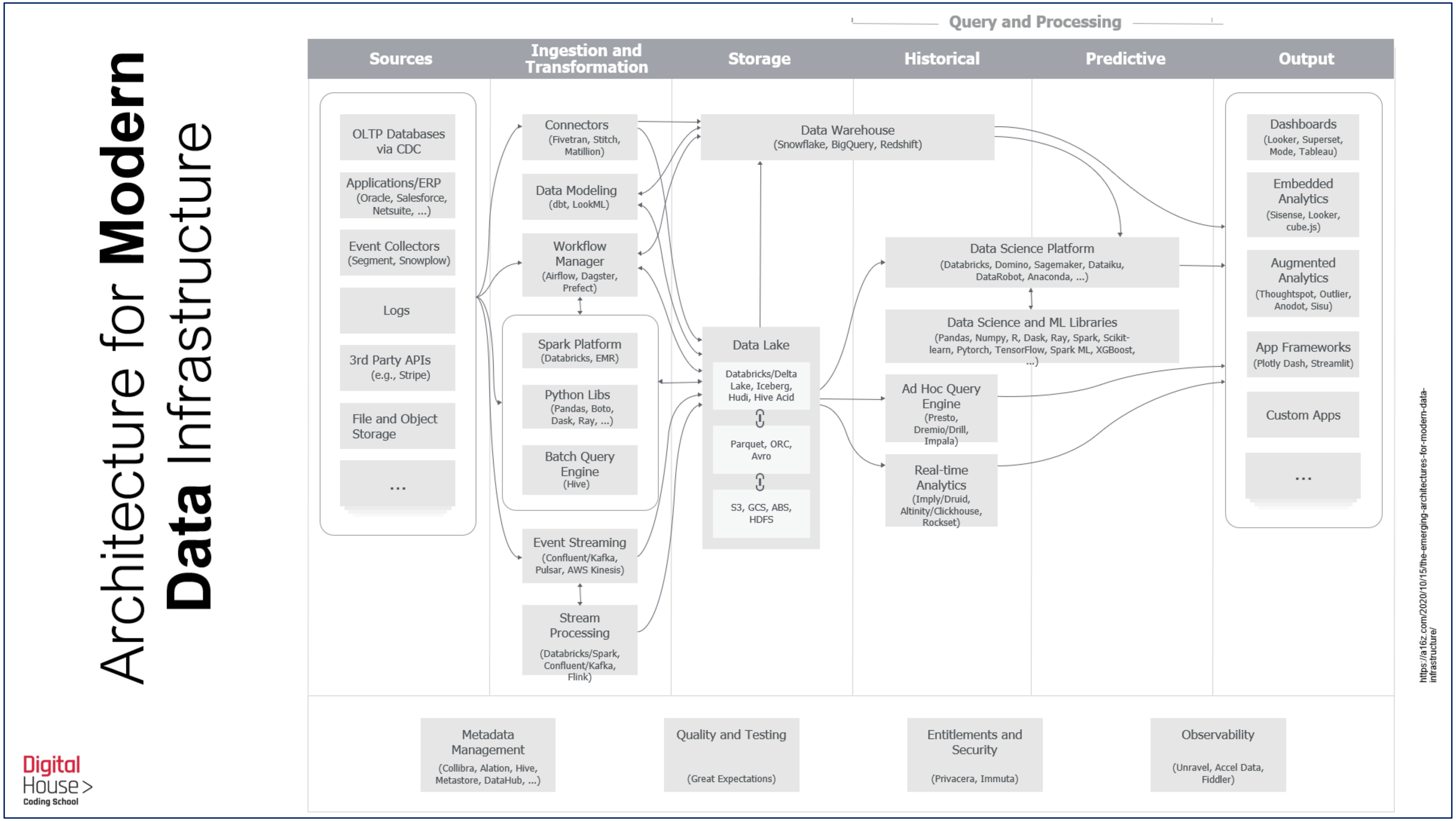
Why Is It so Hard to Put Data Science in Production?

https://blogs.oracle.com/datascience/why-is-it-so-hard-to-put-data-science-in-production

Why Majority Of Data Science Projects Never Make It To Production

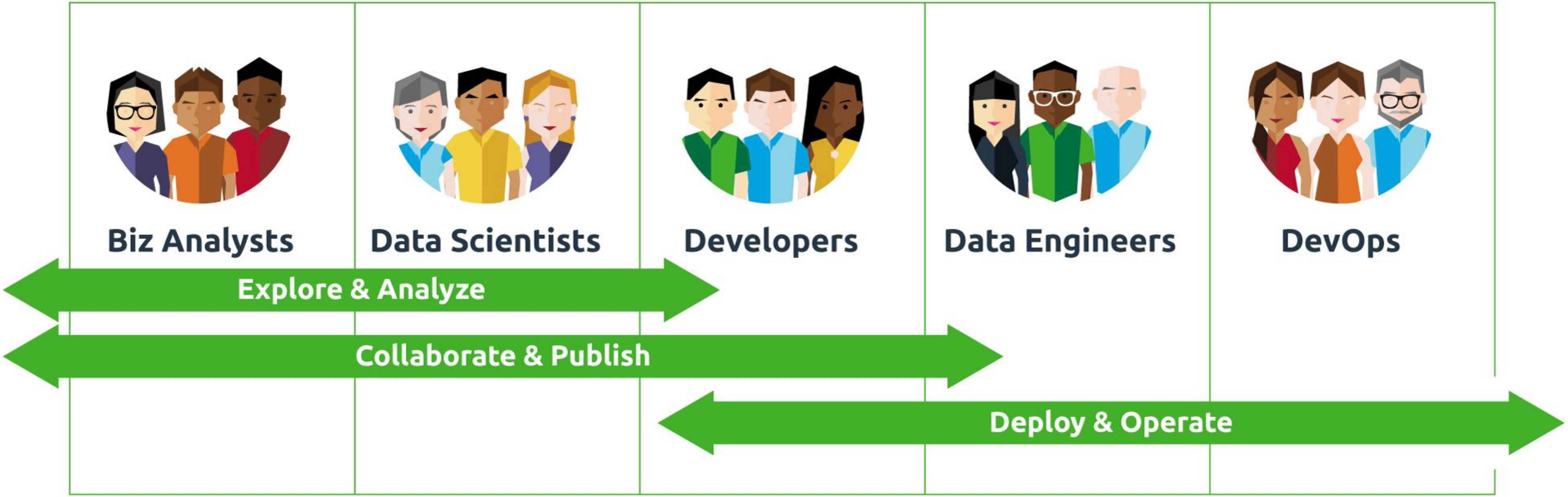
https://analyticsindiamag.com/why-majority-of-data-science-projects-never-make-it-to-production/

Porque é complexo construir e manter software



Digital House > Coding School

Processo integrado para “*data products*”



O unicórnio

Job Description

This position is a contract to hire, fully remote position for candidates in Latin America.

*Resume must demonstrate English ability

Must be within NYC Timezone

Blue Orange Digital is looking for a Machine Learning Engineer to join our awesome multi-disciplinary team. We build data analytics platforms for our clients that incorporate machine learning to solve business problems. Blue Orange Digital works across multiple industries, this role provides an exciting set of experiences across a wide range of domains.

Your primary focus will be architecting and developing systems that include data ingestion, data processing, algorithm development, and ML model development & deployment for Recommendation Systems.

Major technologies involved include AWS, Azure, Python 3, Spark, Pandas, Tensorflow. **The ideal candidate for this position has a mixture of experience in Machine Learning model development, Cloud Engineering, and Data Engineering.**

Core Responsibilities & Skills

At least 6 years of experience in an ML senior engineer or technical lead role

Experience with SageMaker required and experience with other AWS tools p

Experience **developing and designing scalable distributed systems** with
design, modern program languages, and design patterns.

Excellent communication skills and ability to work effectively with engineers, p
analysts, and business stakeholders.

Experience developing and designing recsys and production quality ML models fo

Algorithm and model development experience for large-scale applications

Experience distilling informal customer requirements into problem definitions, dealing with ambiguity, and competing objectives.

Architecting, building, and maintaining modern, scalable data architectures on AWS

Building resilient ETL pipelines using workflow orchestration tools such as Airflow, Prefect, Luigi

Data exploration, analysis, and reporting with an eye towards developing a narrative using Notebooks.

Demonstrable experience in one or more of the following specializations: Recommenders, NLP, pattern detection, anomaly detection, predictive modeling, and optimization.

Qualifications

BA/BS degree in Computer Science or a related technical field, or equivalent practical experience.

Advanced experience in Python with an excellent understanding of computer science fundamentals, data structures, and algorithms

Experience in Amazon AWS/Azure, DevOps, and Automation (Cloudformation)

Experience with distributed machine learning using tools like Dask, Tensorflow, Kubernetes, Airflow, Kubeflow.

Enjoys collaborating with other engineers on architecture and sharing designs with the team

Interacts with others using sound judgment, good humor, and consistent fairness in a fast-paced environment

https://stackoverflow.com/jobs/295632/machine-learning-engineer-latin-america-remote-blue-orange-digital

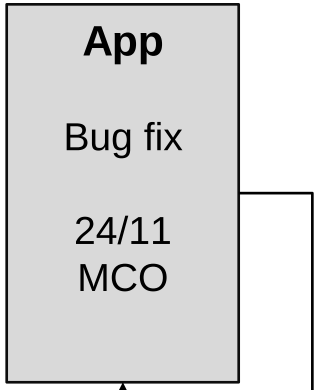
De onde viemos e para onde vamos

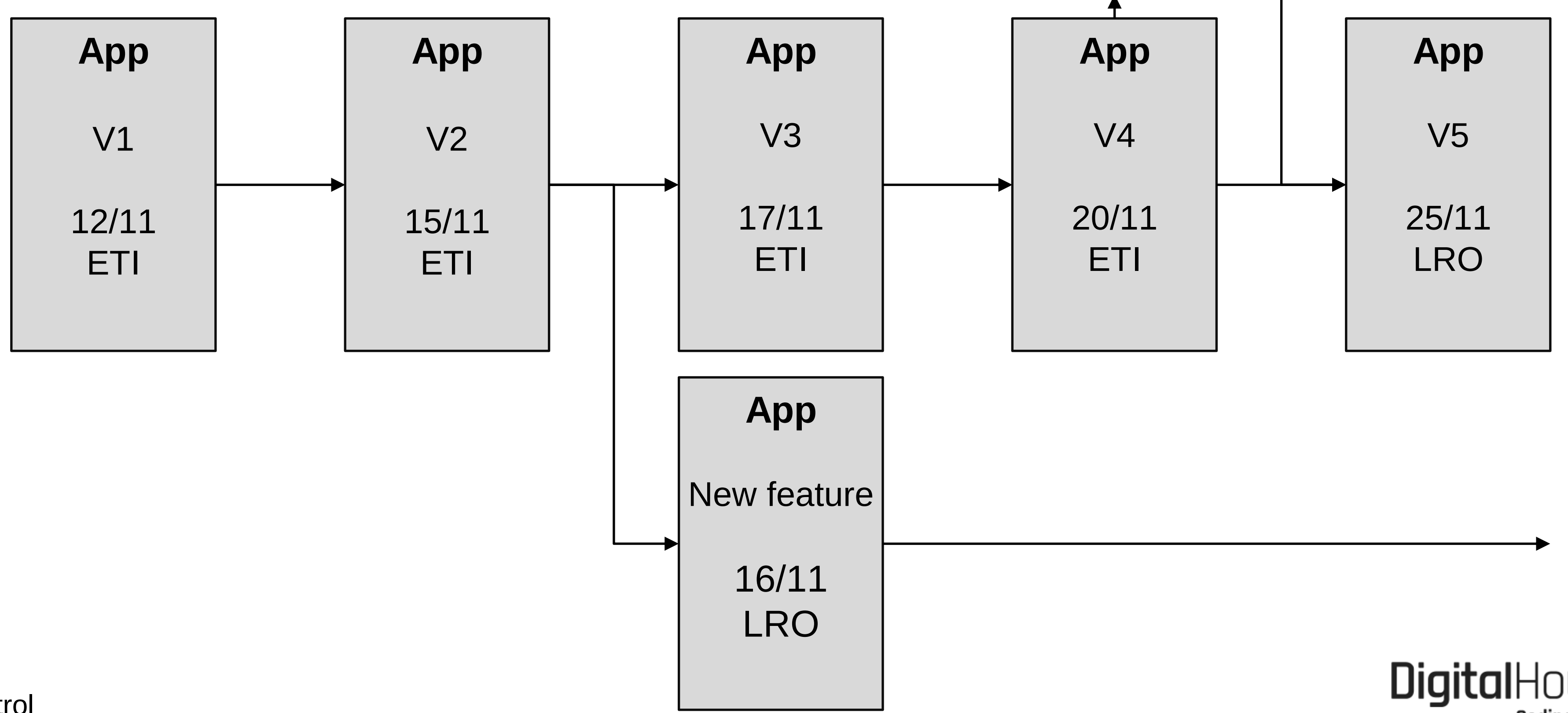
De onde viemos e para onde vamos

Version Control System (VCS)

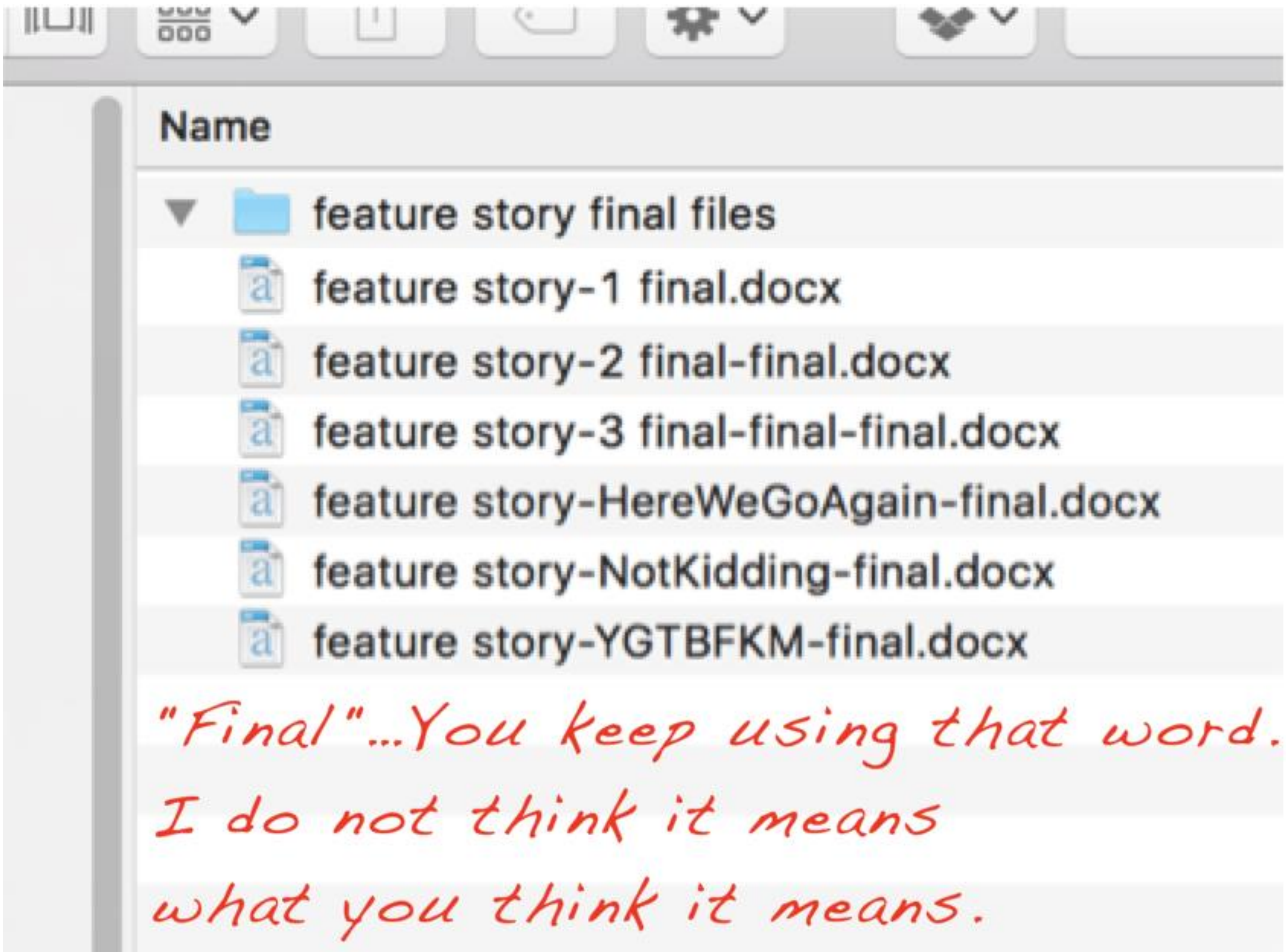
git (2005)

RCS (1982)

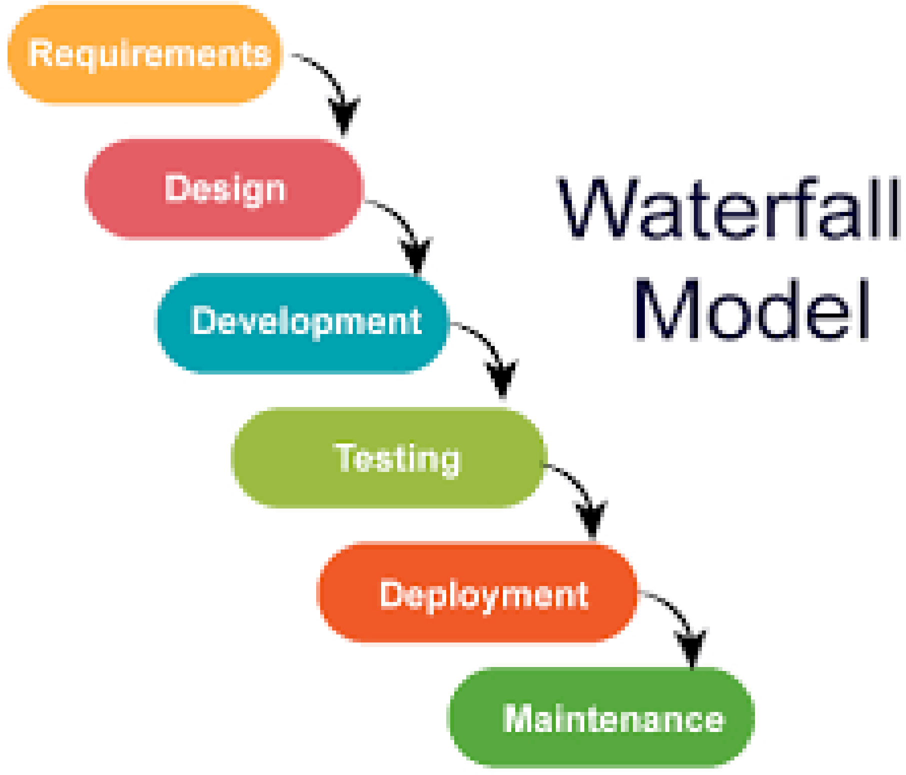


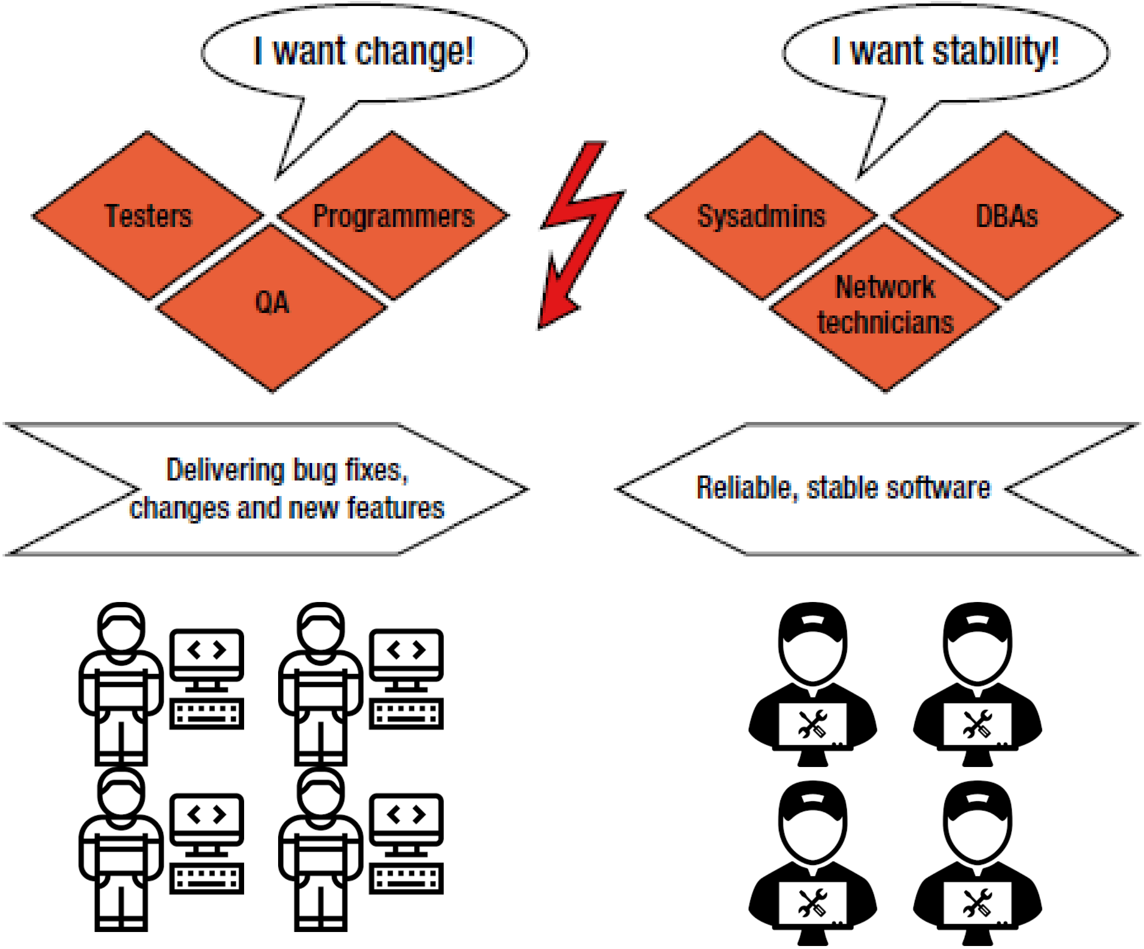


https://en.wikipedia.org/wiki/Version_control

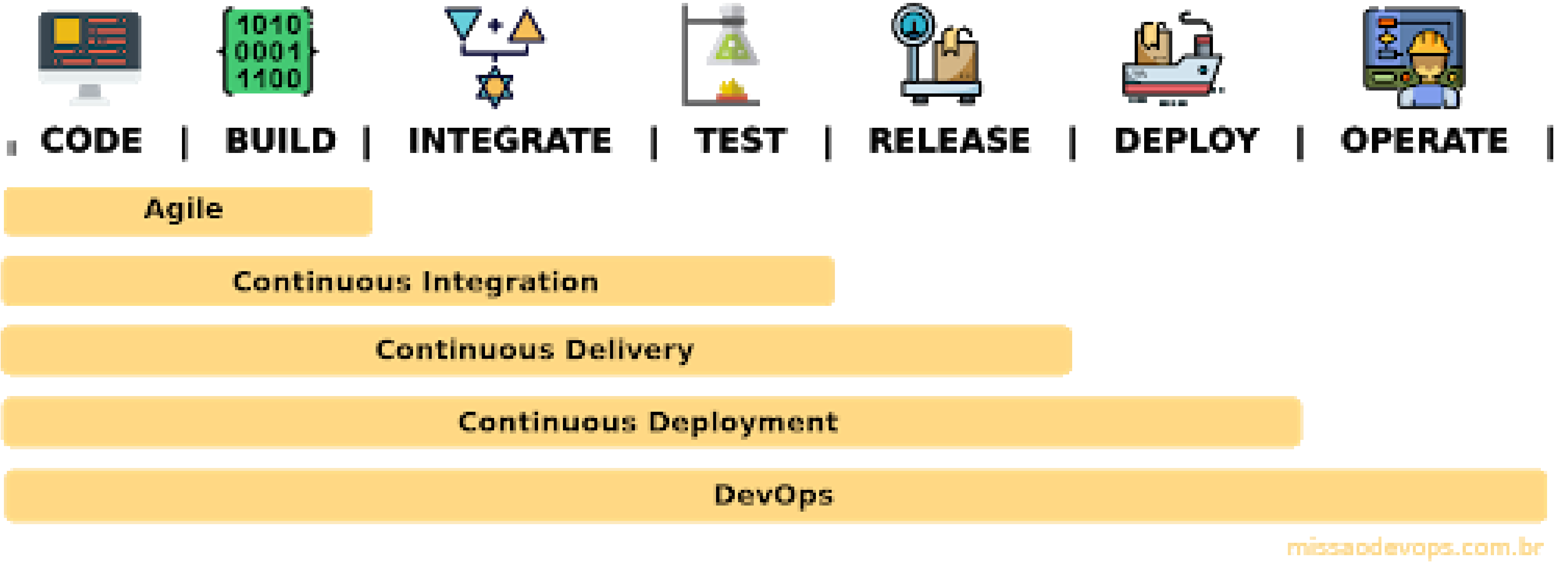

Name
feature story final files
feature story-1 final.docx
feature story-2 final-final.docx
feature story-3 final-final-final.docx
feature story-HereWeGoAgain-final.docx
feature story-NotKidding-final.docx
feature story-YGTBFKM-final.docx
"Final"...You keep using that word.
I do not think it means
what you think it means.

O ciclo “tradicional” de eng. de software

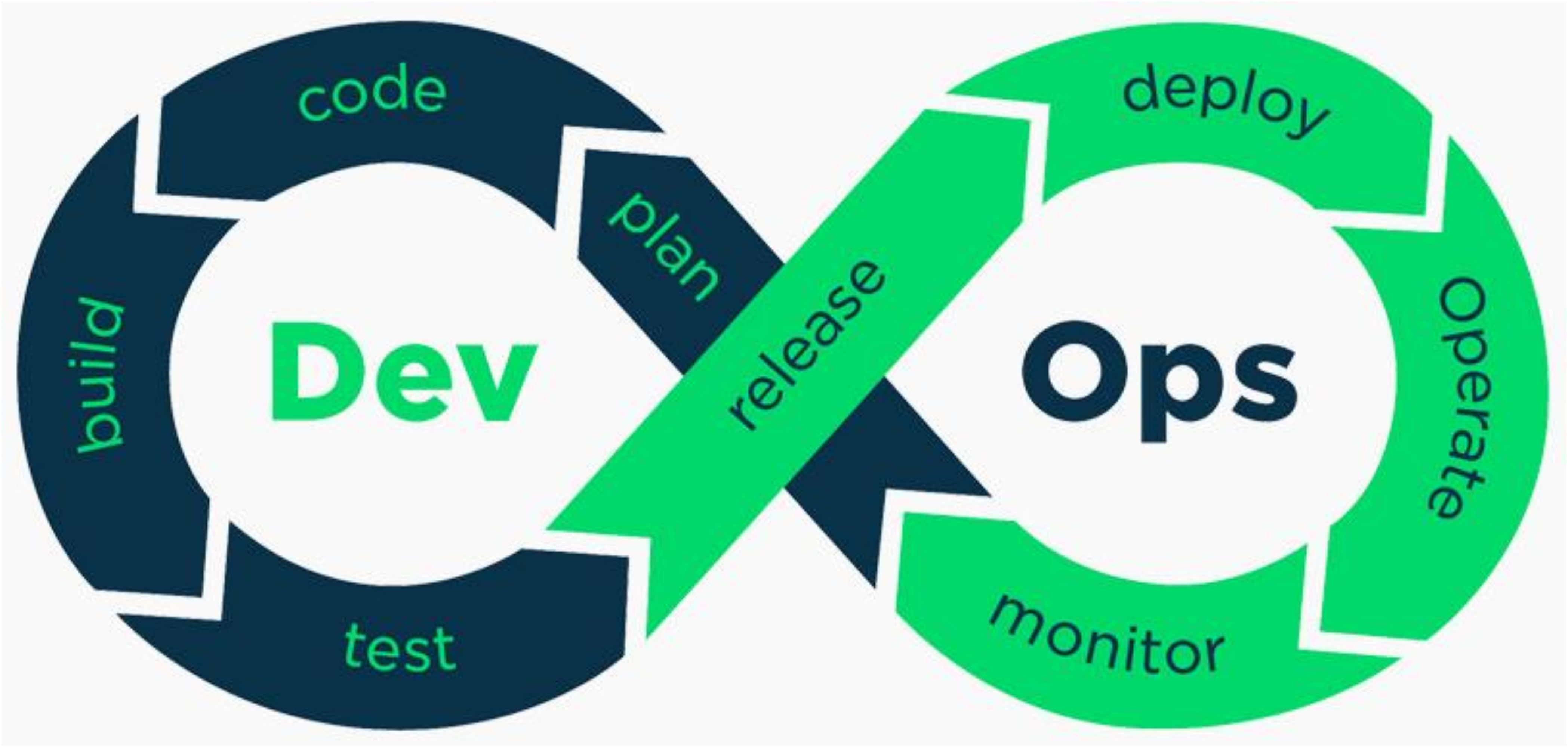




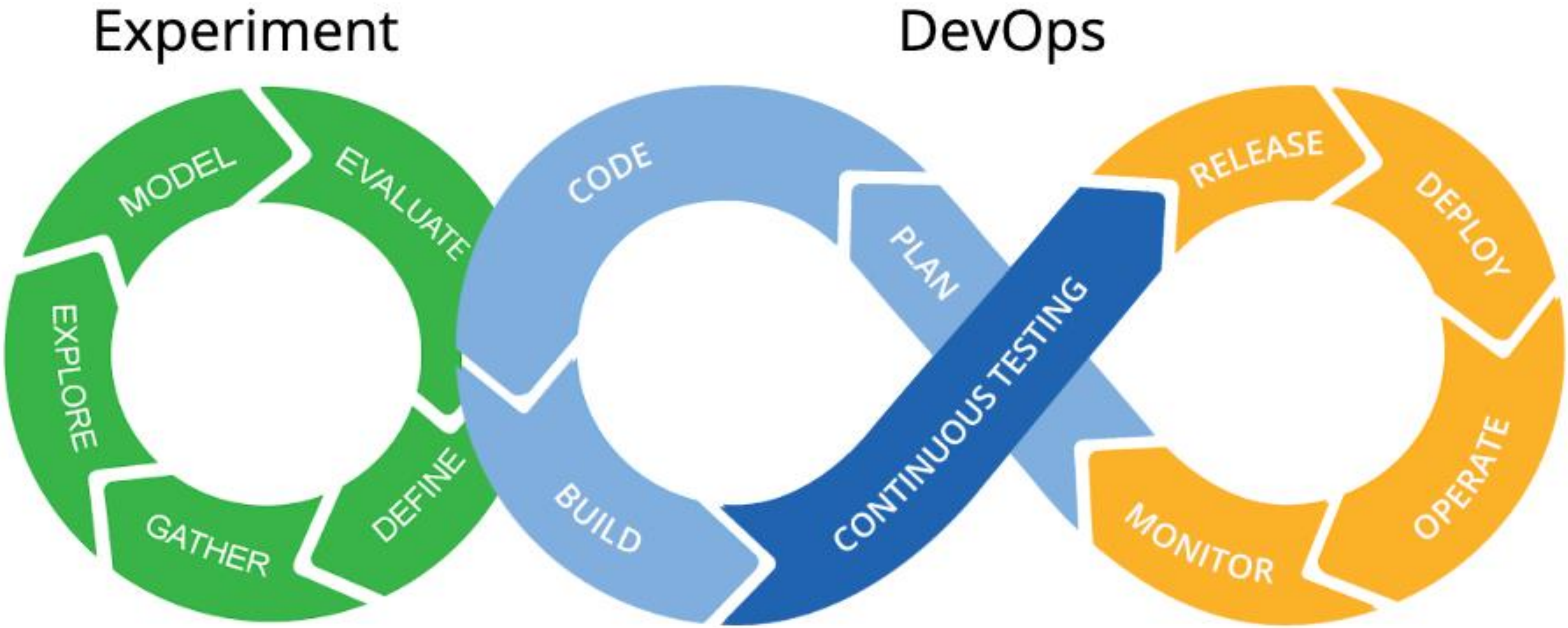
O ciclo “moderno” de eng. de software



O ciclo “DevOps”



MLOps



https://blog.jdriven.com/2019/10/machine-learning-systems-require-paradigm-shift-in-software-engineering/

Mão na massa!