Cognizant

MLOps Workshop - Kickoff

Kyllian Broers & Bram van Meurs

October 8th 2020 – Version 2.0

Workshop Goals

1. Understand the key concepts of MLOps and DevOps
2. Be able to explain when and when not to use MLOps

You will:

- Get hands-on experience with setting up a DevOps pipeline
- Productionalise a proof of concept ML model, including development and deployment
- Monitor data and detect drift

The Challenge

Develop/Experiment

75% of data science projects never make it to production

https://towardsdatascience.com/ml-ops-challenges-solutions-and-future-trends-d2e59b74dc6b

Agenda

1. What is DevOps?
2. Why MLOps?
3. Example within Cognizant
4. Workshop contents

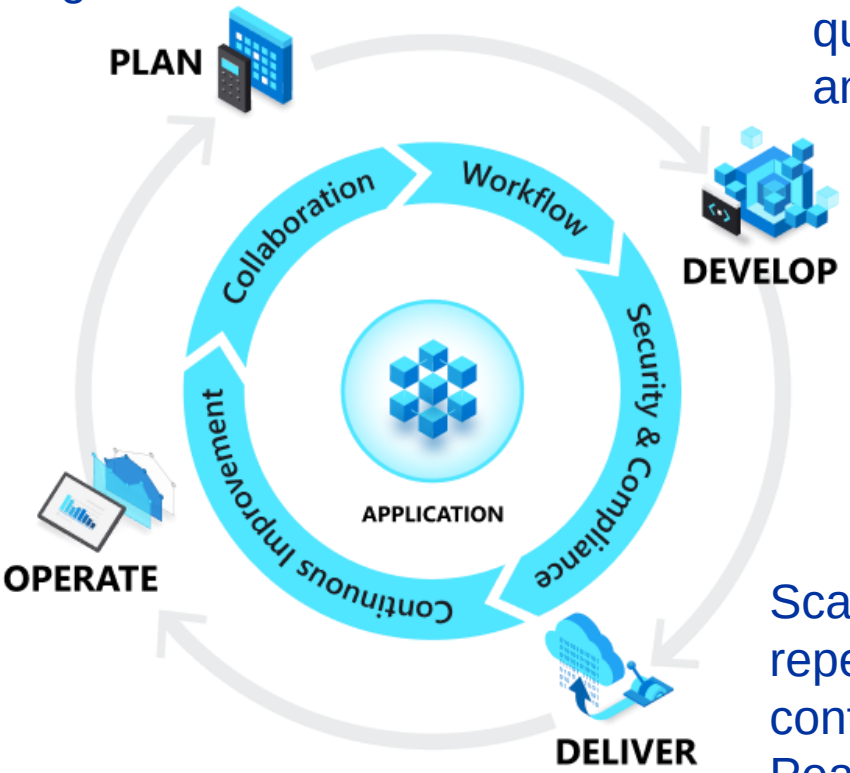
Cognizant

What is DevOps?

What is DevOps

- Way of working
- Application lifecycle

Agile



Increased quality, stability and productivity

Improved reliability, availability, security

Scalable, repeatable, controllable. Peace of mind

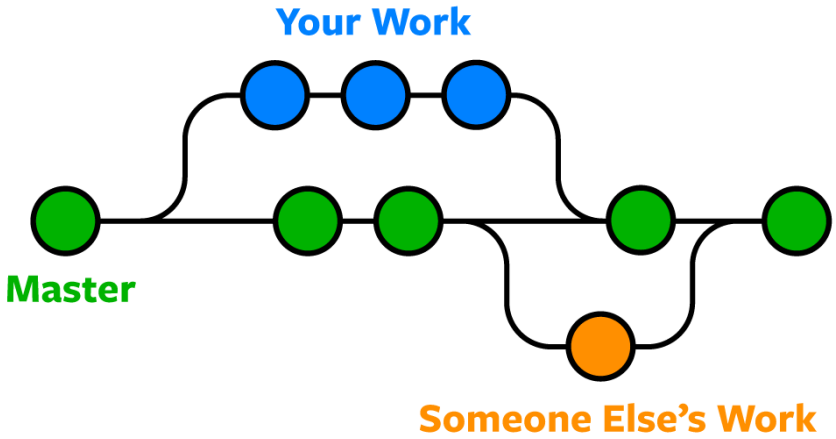
Cognizant

Develop

Programming: Python

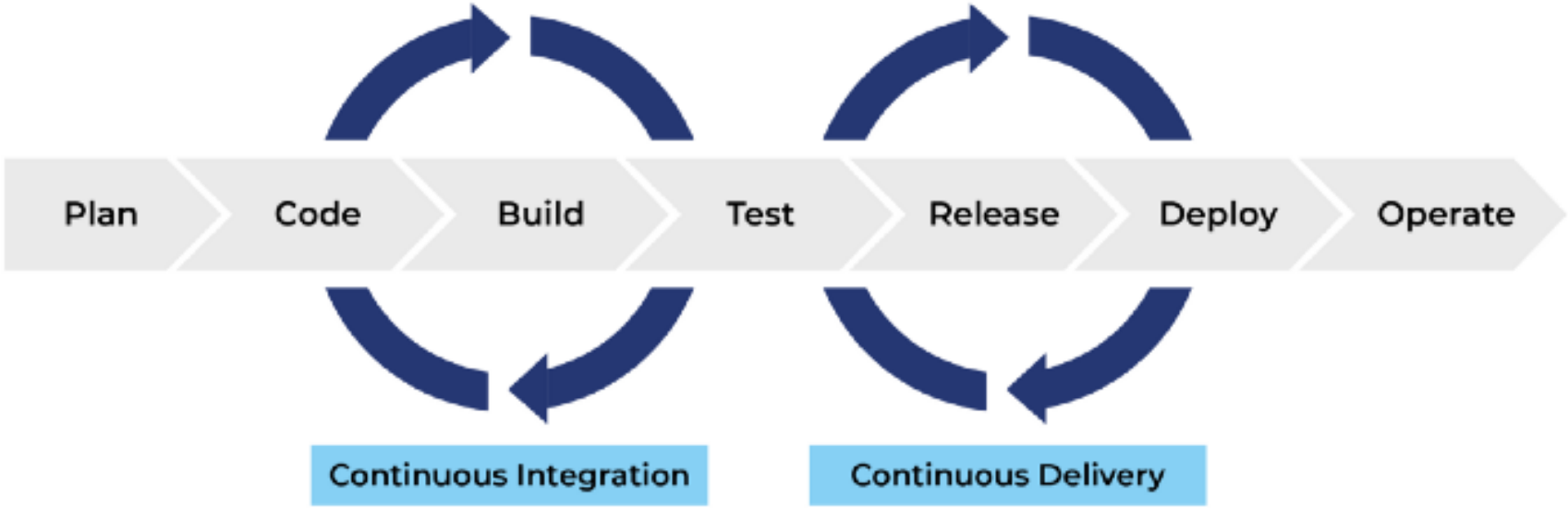
Collaboration: Git

- Github
- Gitlab



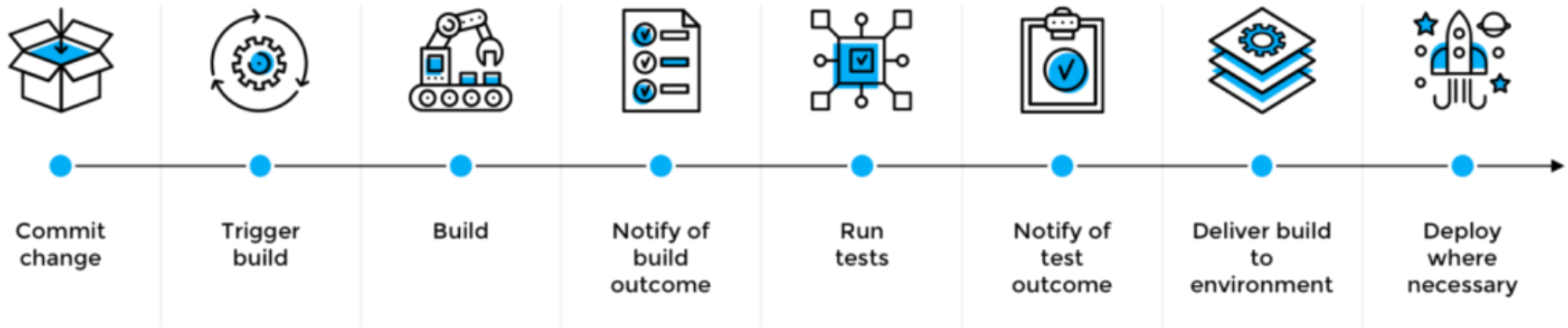
Cognizant

DevOps CI/CD



Cognizant

DevOps CI/CD



Cognizant

Benefits of DevOps

- Shortening the development cycles
- Increasing deployment velocity
- Increasing quality
- Dependable releases
- Ease of rollback

DevOps and MLOps?

Cognizant

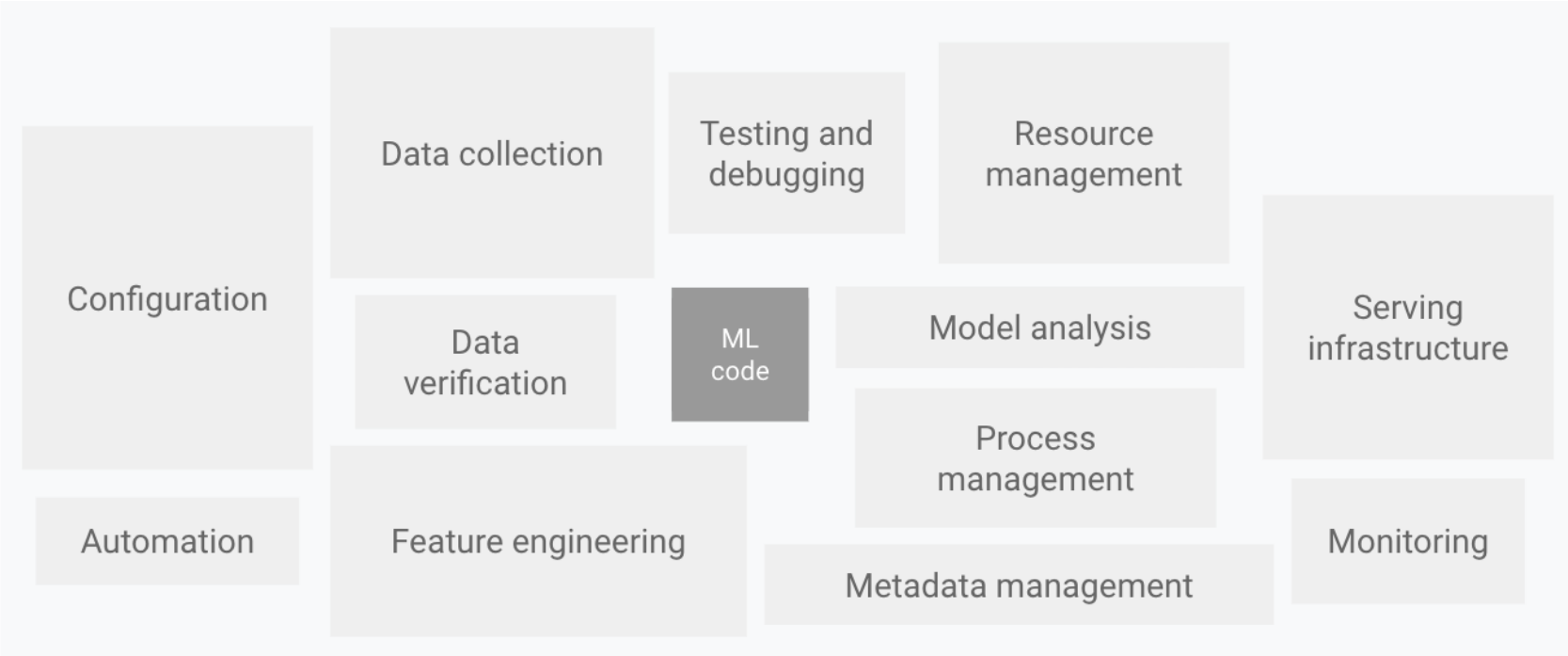
DevOps and MLOps

- Team skills: ML researchers are not experienced software engineers
- Development is experimental
- Testing is more involved
- Deployment includes retraining
- Production: handling evolving data profiles

In MLOps

- CI: includes testing and validating data, data schemas and models
- CD: it is about deploying a system
- CT: continuous training

Parts of a real ML project



Cognizant

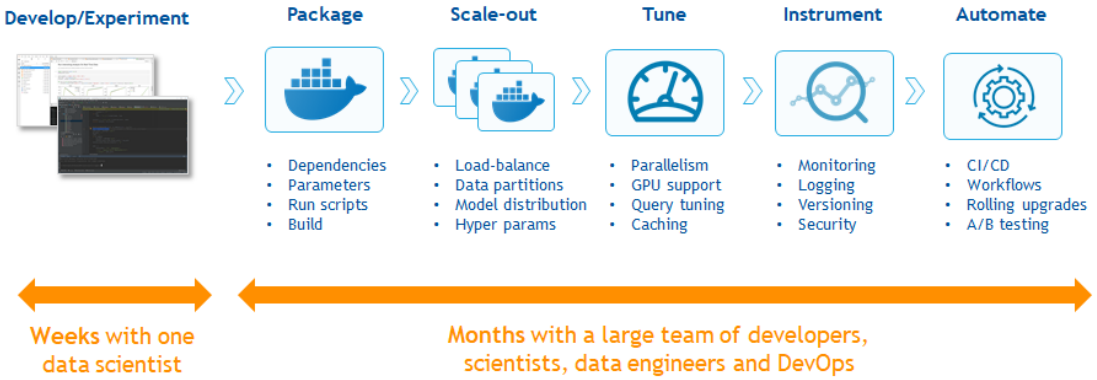
Why MLOps?

Cognizant

MLOps Benefits

- Reproducible model creating, tracking development
- Continuous Training
- Automatic deployment
- Continuous improvement

- Faster time to market
- Higher quality product



Cloud services & open source tools make MLOps easier

Cognizant

When not to use MLOps

High setup cost

- Develop once, deploy once
- Overhead should not exceed the benefits

Tools and example

Cognizant

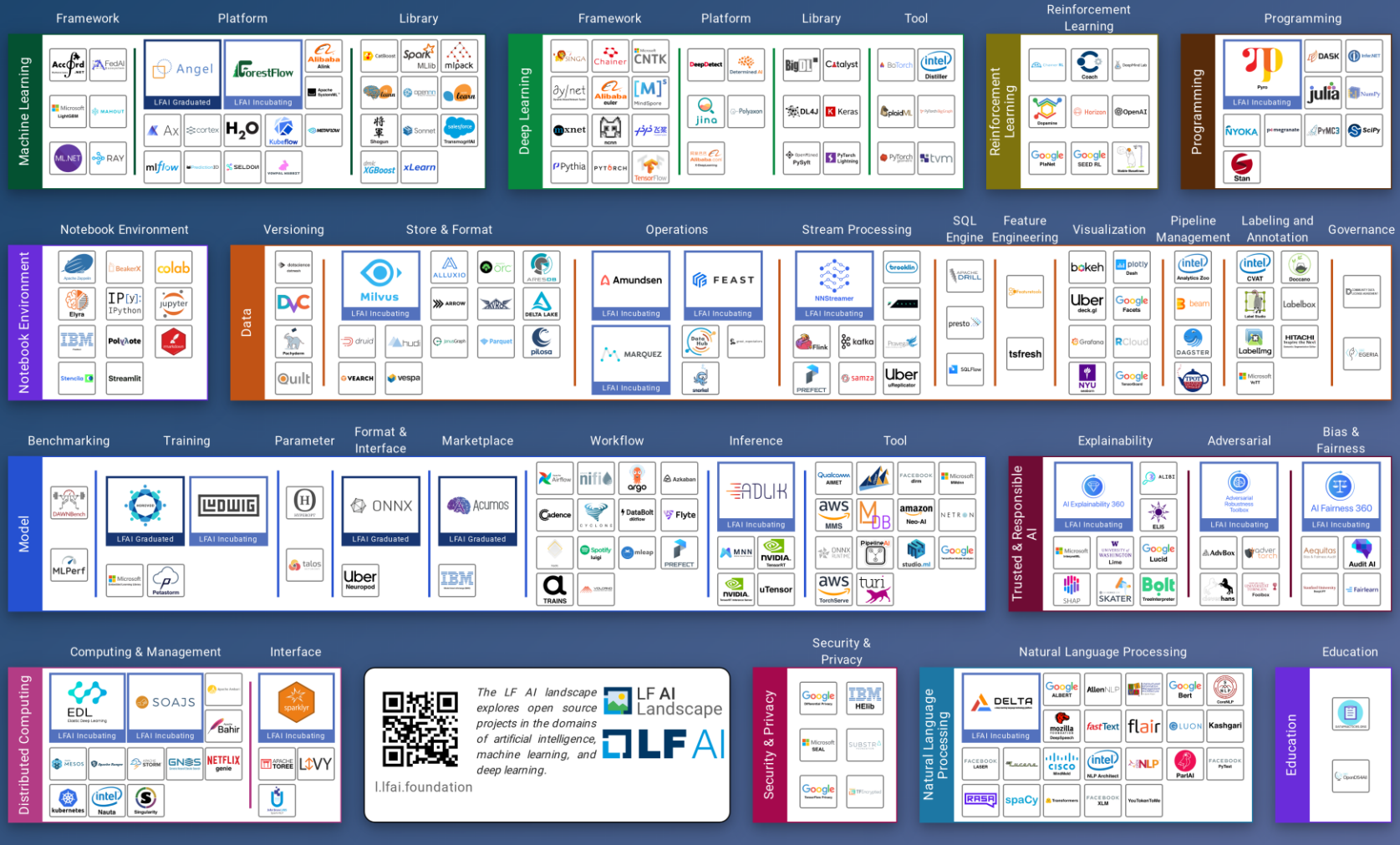
https://landscape.lfai.foundation/

Linux Foundation AI Landscape

2020-10-08T02:39:48Z 5361fd2

See the interactive landscape at l.lfai.foundation

Greyed logos are not open source



DevOps CI/CD

Deployment service
- Docker
- Kubernetes / Kubeflow
- Managed service: AWS/Azure

Data storage

Version control
- Git

Build and test service
- Jenkins
- Travis
- Azure Pipelines
- Github
- Gitlab

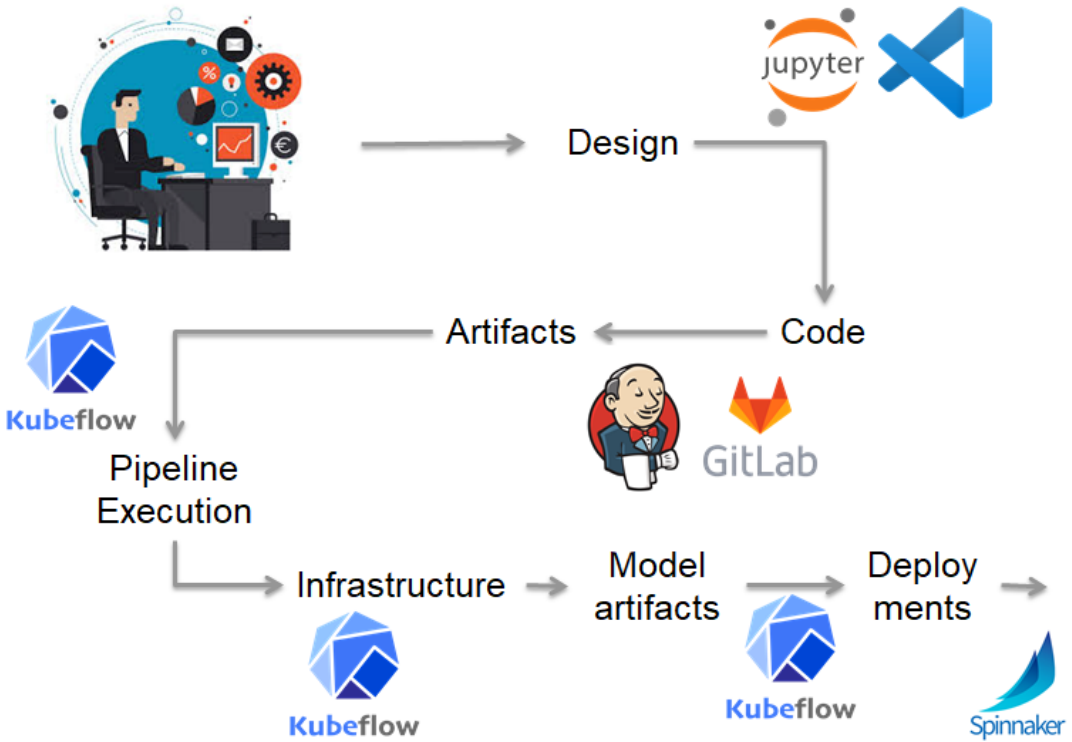
Monitoring tools
- Managed services

ML Lifecycle tools/services

- ML Studio (Azure)
- Sagemaker (AWS)
- Cnvrg.io
- Kubeflow
- ML Flow
- Seldon

Cognizant

Macy's



Summary:

1) A data scientist **designs** a model (workflow)
2) The design is **coded** using JupyterHub or an IDE and stored in GitLab
3) A **Continuous Integration** (CI) pipeline builds the artifacts automatically with each commit using Jenkins
4) One of the artifacts is **a Kubeflow pipeline** that executes all steps of the workflow
5) Kubeflow **orchestrates** executing of all steps and deploys the steps on the required infrastructure
6) Kubeflow stores **model artifacts**
7) A **Continous Delivery** (CD) pipeline updates the services when the model artifacts are refreshed (Spinnaker)

6

Cognizant

Workshop

Cognizant

Workshop

Completely in Azure

- As a test to see what azure offers
- Focus is on concepts and general coding
- Azure specific code is avoided as much as possible

Assignments

1. Getting familiar with Git and DevOps
2. Bringing a model from POC to production
3. Monitoring the model in production: detecting drift in the data

Cognizant

Assignment 1

Goal: getting familiar with Git and DevOps

- No ML included
- Create a pipeline that performs data analysis
- The pipeline will throw errors: unit tests and linting errors
- Working together with Git
  - Create your own development branch, create pull requests to the assignment-1 branch.

Cognizant

Assignment 2

Goal: going from POC to production using CI/CD

- Transform a notebook into production ready code
- Create a CI pipeline that performs code integration and model training
- Create a CD pipeline that deploys the latest model to container with API

Assignment 3

Goal: monitor a model in production

- Continues at a working assignment 2
- Collect inference data into a dataset
- Use the data monitor tool to detect data drift

Repo structure

- Each one branch for assignment 1 and 2
- Assignment 3 will branch of assignment 2

Graph	Description	Date	Author	Commit
	setup *origin* setup environment	27 Aug 2020 ...	Bram van Me...	81f48876
	assignment-2 *origin* assignment-2	27 Aug 2020 ...	Bram van Me...	423f36ab
	assignment-1 *origin* **assignment-1**	27 Aug 2020 ...	Bram van Me...	b9ea4e80
	documentation *origin* documentation	27 Aug 2020 ...	Bram van Me...	42551738

Cognizant

Workshop Communication

Cognizant

Communication

Use the Teams MLOps Workshop team

- Plenary meeting in general channel
- Team meetings in teams channels
  - Please do not use private calls
- For questions
  - Use mentioning @Bram or @Kyllian, can be both in general or a teams channel.
  - Or add one of us to a channel meeting

Schedule

09:00 - 10:00 Kickoff

12:30 - 13:00 Discussing Assignment 1

Groups

1: Laura, Vetle, Bart

2: Bas, Stan

3: Timo, Erik

4: Kristoffer, Haoyu

5: Aleksander, Jodie

https://dev.azure.com/MLOpsTraining/group1

Cognizant