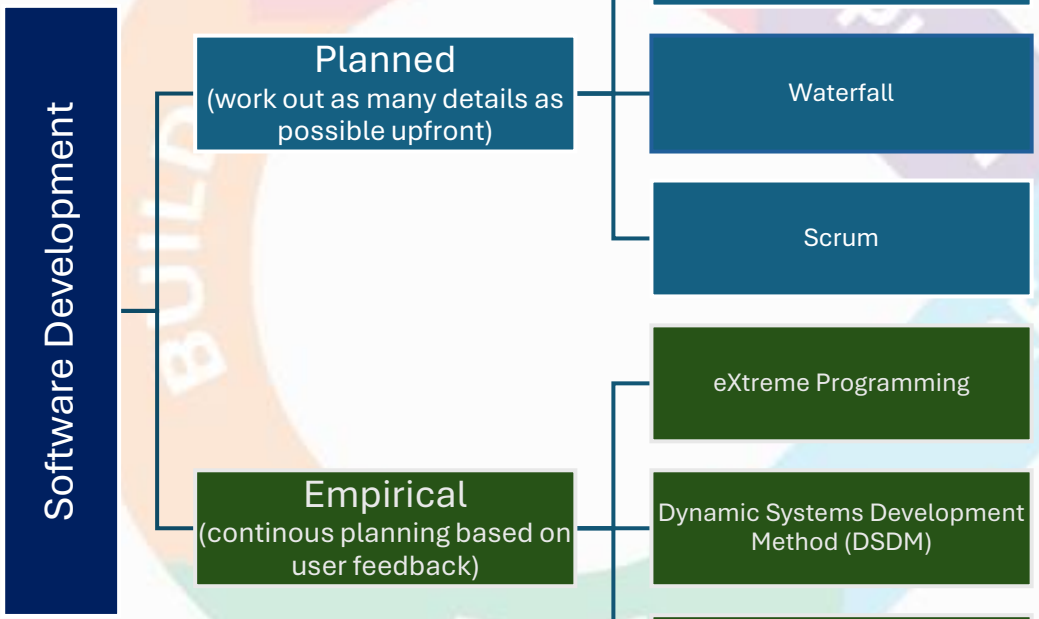
EduML

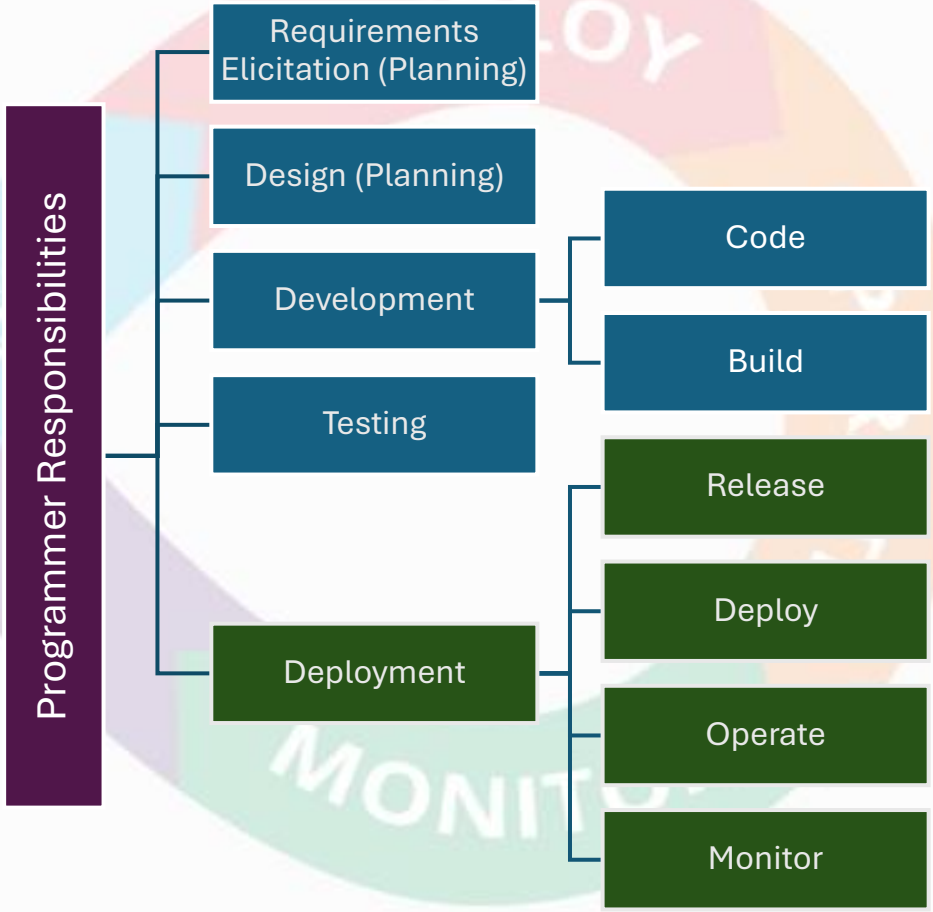
MLOps on Kubernetes

Andrei Olar

<andrei.olar@ubbcluj.ro>

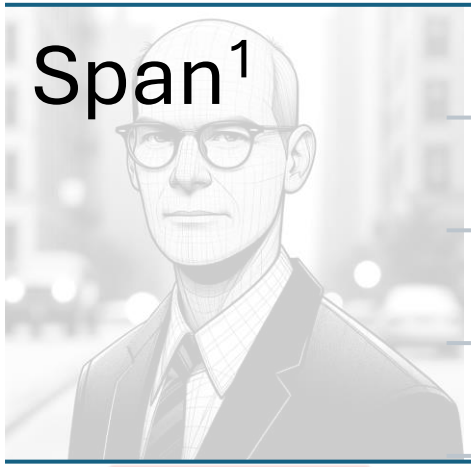
DevOps[1]





[1]Kim, G., Humble, J., Debois, P., Willis, J., & Forsgren, N. (2021). *The DevOps handbook: How to create world-class agility, reliability, & security in technology organizations*. It Revolution.
Image courtesy of https://mia-platform.eu/blog/devops/

Machine Learning



Span[1]	Narrow AI
	General Purpose AI
	AGI
	ASI
Process[2,3]	Level 0 – Manual (solitary development)
	Level 1a – Automated Training (team collaboration)
	Level 1b – Automated Serving (learning from evaluation)
	Level 2 – Model monitoring (automated feedback loop)

[1]Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
[2]https://mlops-for-all.github.io/en/docs/introduction/levels/
[3]Sculley, D et. al, *Hidden Technical Debt in Machine Learning Systems*, Advances in Neural Information Processing Systems, 2015.
Image credits: OpenAI + Google

ML Ops

Goals

DevOps extension for ML professionals[1]

Generic process for releasing ML artifacts[2]

Automate testing of ML artifacts[2]

Use established software development best practices with ML artifact[2]

[1]https://github.com/cdfoundation/sig-mlops/blob/main/roadmap/2020/MLOpsRoadmap2020.md#what-is-mlops
[2]https://ml-ops.org/content/mlops-principles

ML Ops

Traits

Framework, language and method agnostic

Reproducible

Automated

Image credit: https://ml-ops.org/content/mlops-principles

ML Ops

Stages

- Data Preparation
- Training / Experimentation
- Deployment / Serving
- Monitoring

Image: https://ubuntu.com/blog/what-is-mlops

References:

- https://blog.ml.cmu.edu/2020/08/31/3-baselines/
- https://cloud.google.com/blog/products/ai-machine-learning/key-requirements-for-an-mlops-foundation

ML Ops - Components

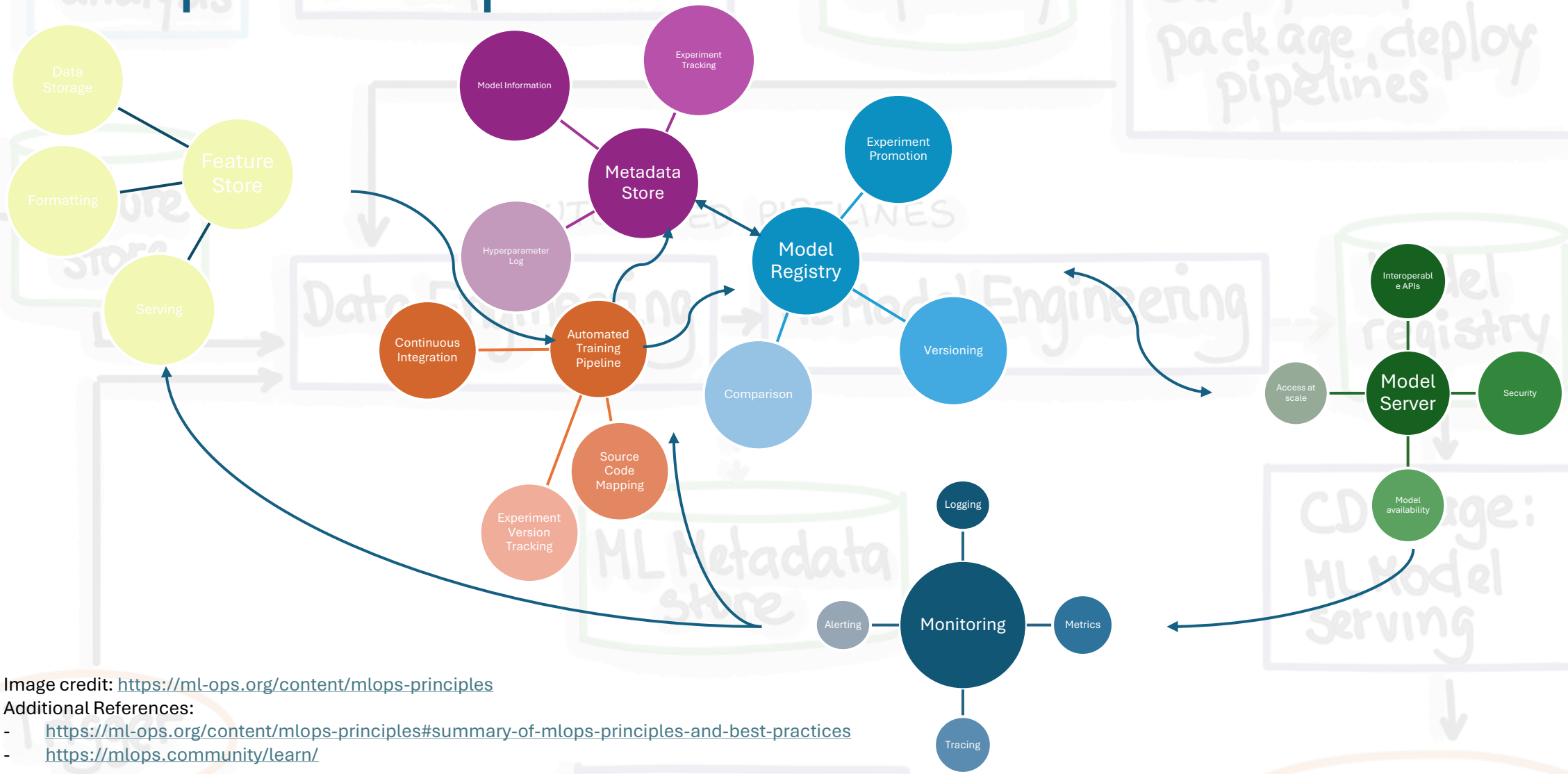


Image credit: https://ml-ops.org/content/mlops-principles

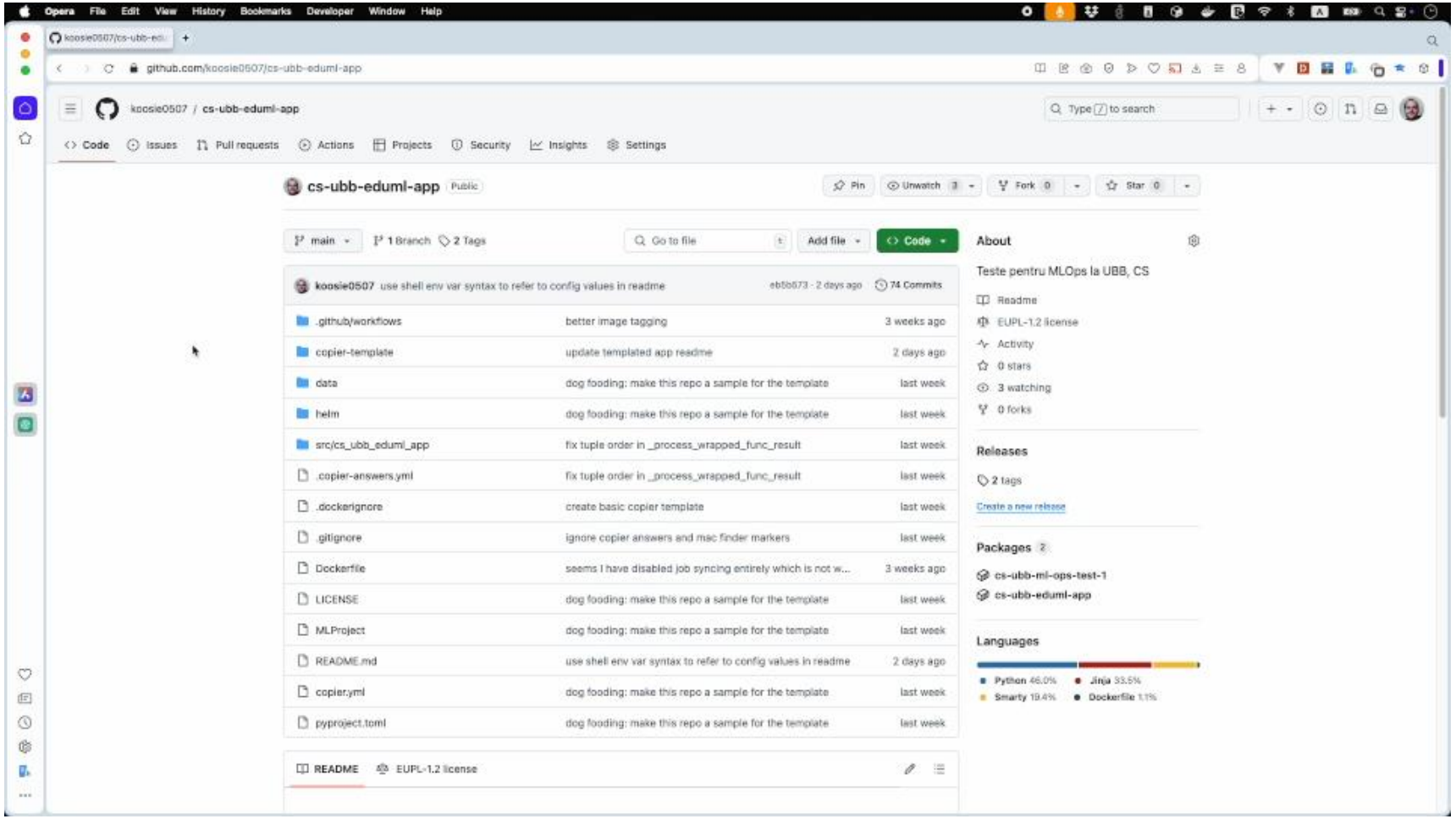
Additional References:

- https://ml-ops.org/content/mlops-principles#summary-of-mlops-principles-and-best-practices
- https://mlops.community/learn/

ML Ops

Tooling	MLFlow	✓
	Polyaxon	
	Kubeflow	
	MLRun	

Demo


Opera File Edit View History Bookmarks Developer Window Help
github.com/koosie0507/cs-ubb-eduml-app
koosie0507 / cs-ubb-eduml-app
Type / to search
Code Issues Pull requests Actions Projects Security Insights Settings
cs-ubb-eduml-app Public
Pin Unwatch 3 Fork 0 Star 0
main 1 Branch 2 Tags Go to file Add file Code
koosie0507 use shell env var syntax to refer to config values in readme ebbb673 · 2 days ago 74 Commits
.github/workflows better image tagging 3 weeks ago
copier-template update templated app readme 2 days ago
data dog fooding: make this repo a sample for the template last week
helm dog fooding: make this repo a sample for the template last week
src/cs_ubb_eduml_app fix tuple order in _process_wrapped_func_result last week
.copier-answers.yml fix tuple order in _process_wrapped_func_result last week
.dockerignore create basic copier template last week
.gitignore ignore copier answers and mac finder markers last week
Dockerfile seems I have disabled job syncing entirely which is not w... 3 weeks ago
LICENSE dog fooding: make this repo a sample for the template last week
MLProject dog fooding: make this repo a sample for the template last week
README.md use shell env var syntax to refer to config values in readme 2 days ago
copier.yml dog fooding: make this repo a sample for the template last week
pyproject.toml dog fooding: make this repo a sample for the template last week
README EUPL-1.2 license
About
Teste pentru MLOps la UBB, CS
Readme
EUPL-1.2 license
Activity
0 stars
3 watching
0 forks
Releases
2 tags
Create a new release
Packages 2
cs-ubb-ml-ops-test-1
cs-ubb-eduml-app
Languages
Python 46.0% Jinja 33.5%
Smarty 19.4% Dockerfile 1.1%

Specs

Kubernetes v1.31.1+k3s1

Cluster Nodes

- 3 control plane nodes: 8GB RAM, 4 core Intel Xeon@2.2GHz
- 2 GPU worker nodes: 32GB RAM, 16 core Intel Xeon@2.3GHz

Available GPUs

- One A100D-8C, CUDA Version: 12.2, 8GB RAM / GPU worker node
- Can’t share or use NVLink (can’t mine bitcoin either)

Recap

ML is complex both in terms of span and in terms of process

DevOps provides a solid framework to build upon automated ML processes

JupyterNotebooks – good Level 0 solution

MLFlow – good Level 1a solution

We have a nice, serial job running cluster

@Andrei Olar in MIRPR-2024-2025

@lauradiosan/MIRPR-UBB/2024-2025

@koosie0507/cs-ubb-eduml-app

andrei.olar [at] ubbcluj.ro