
MLOPS
VOLVO
MLOps
community

LeonardAukea

Driving ML Engineering & Operations

VOLVO

Purpose 📝

> Transform the company to become a machine learning powerhouse where, machine learning & AI is a major differentiator and a core component in  's business 🧬

MLOPS

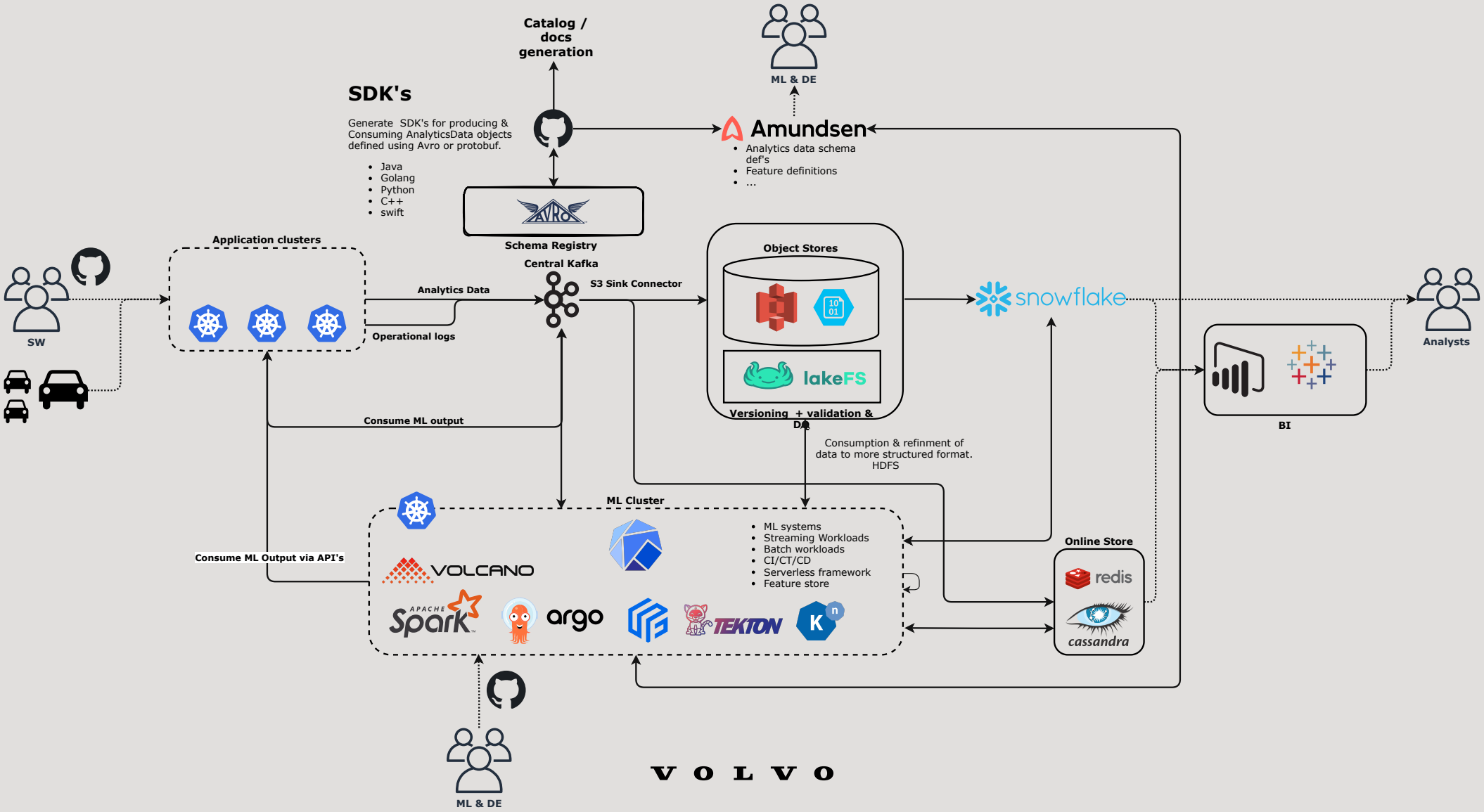
Objective: Shorten the development cycles and increase deployment velocity, in order to reduce friction and delays in the AI value stream. Doing so, in a robust and secure manner. (roughly speaking)

VOLVO

Approach 🤔

- Cloud Native Infra
- Glue infra with gitflow
- Template's & reusability
- Centralized knowledge platform
- Some SW expectations from ML practitioners
- Abstract ML practices into design patterns
- Evangelize and follow up adoption of good principles

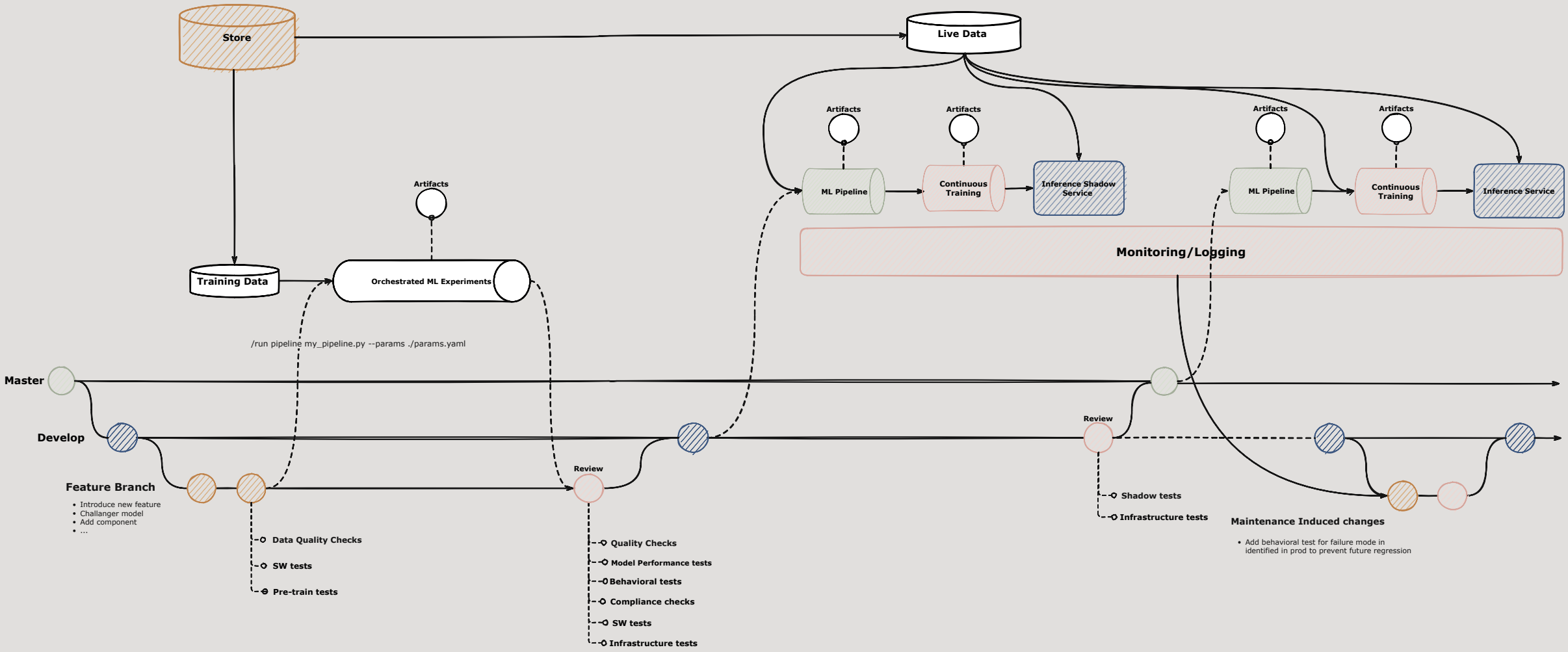
Dedicated ML Clusters



Automation 

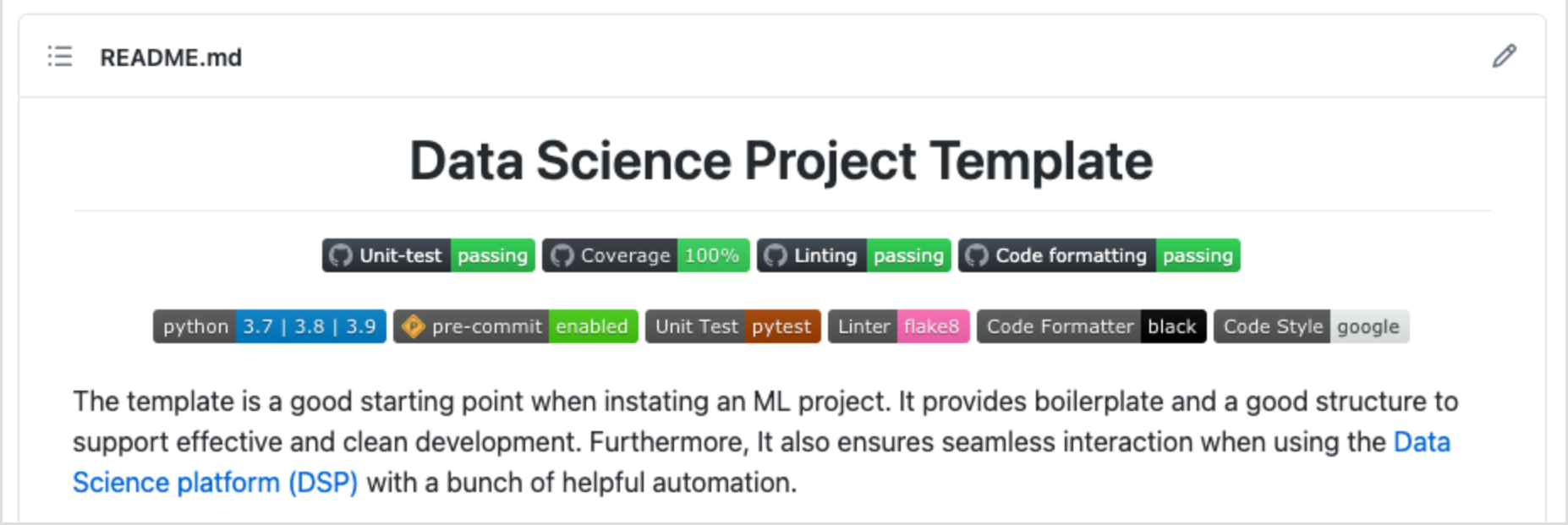
> The level of automation of the Data, ML Model, and Code pipelines determines the maturity of the ML process.

VOLVO


Store
Live Data
Artifacts
Artifacts
Artifacts
Artifacts
ML Pipeline
Continuous Training
Inference Shadow Service
ML Pipeline
Continuous Training
Inference Service
Monitoring/Logging
Artifacts
Training Data
Orchestrated ML Experiments
/run pipeline my_pipeline.py --params ./params.yaml
Master
Develop
Feature Branch
• Introduce new feature
• Challanger model
• Add component
• ...
Data Quality Checks
SW tests
Pre-train tests
Review
Quality Checks
Model Performance tests
Behavioral tests
Compliance checks
SW tests
Infrastructure tests
Review
Shadow tests
Infrastructure tests
Maintenance Induced changes
• Add behavioral test for failure mode in identified in prod to prevent future regression

VOLVO

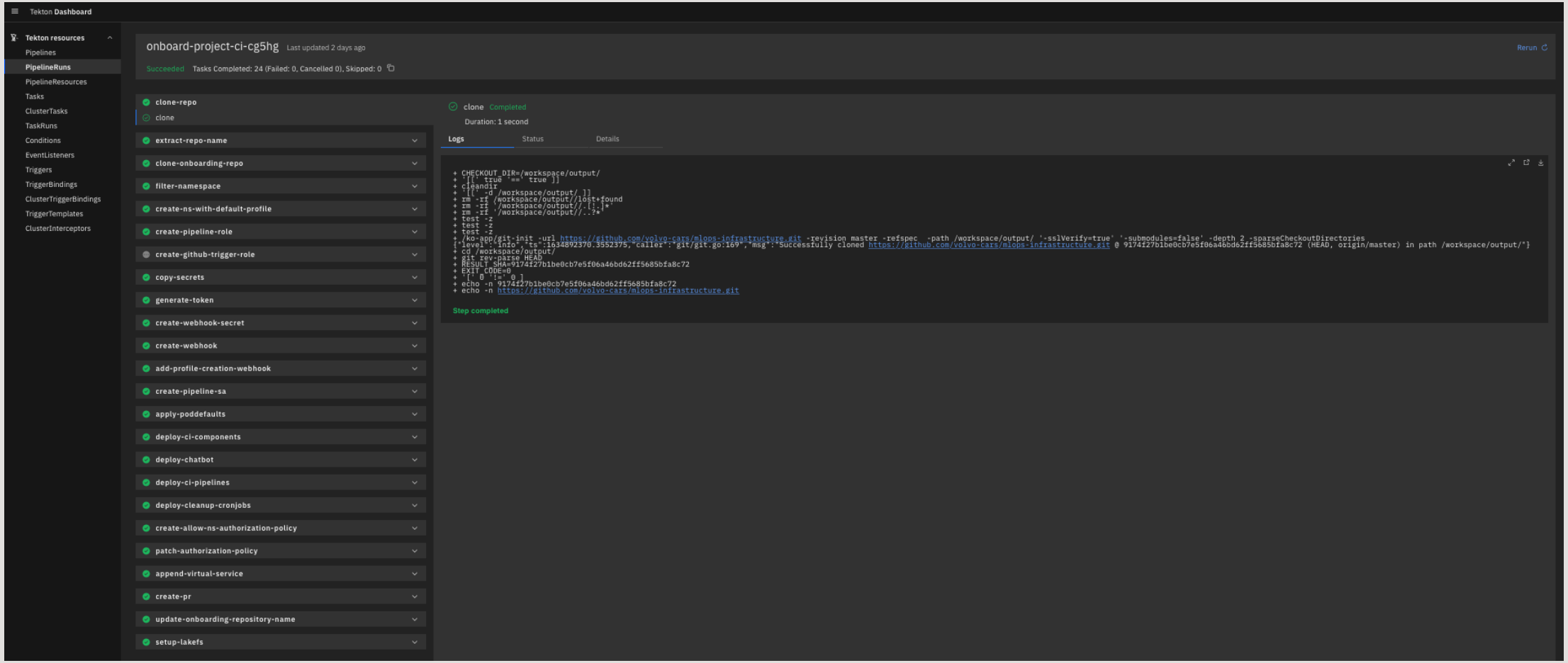
The template



profiles.yaml

```yaml
profiles:
  name: project-name
  admins:
    - john.doe@volvocars.com
  contributors:
    - jane.doe@volvocars.com
    - john.smith@volvocars.com
```

Onboarding CI

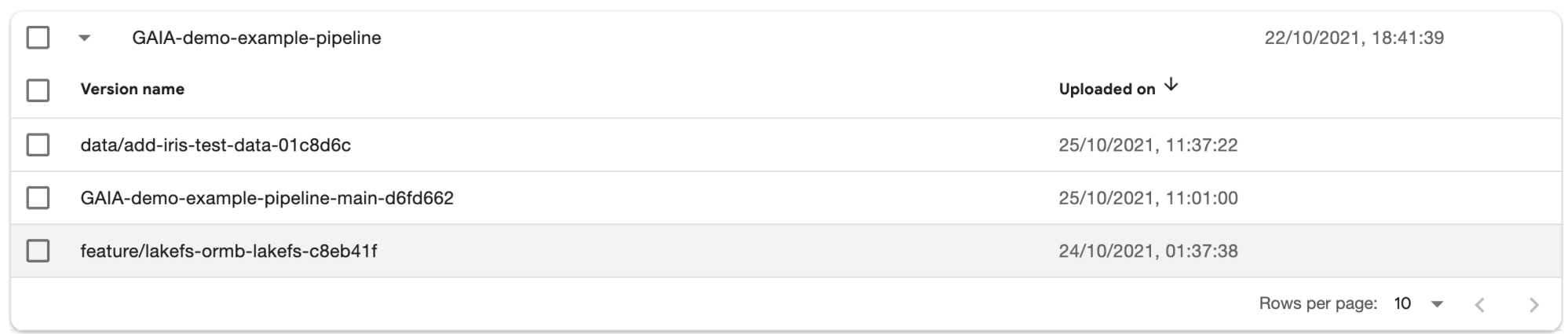


VOLVO

/workflows

```
.
├── README.md
├── components
│   ├── docker-compose.yaml
│   ├── lakefs-download-file
│   │   ├── Dockerfile
│   │   ├── component.yaml
│   │   ├── requirements.txt
│   │   └── src
│   │       └── download.py
│   └── train-ormb-component
│       ├── Dockerfile
│       ├── README.md
│       ├── component.yaml
│       ├── requirements.txt
│       └── src
│           └── train.py
└── iris_pipeline.py
```

Pipeline versioning

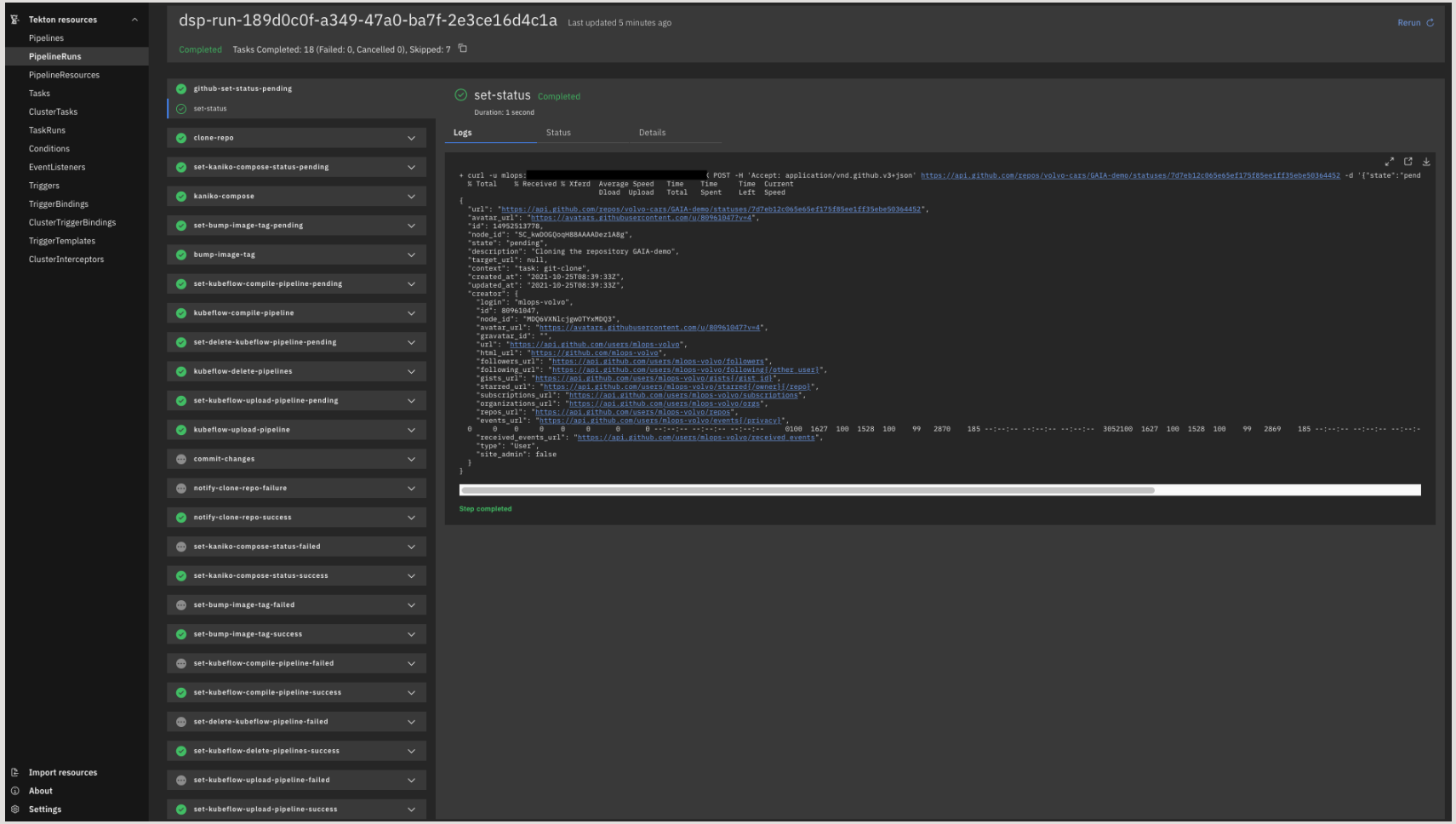


	GAIA-demo-example-pipeline	22/10/2021, 18:41:39
	Version name	**Uploaded on** ↓
	data/add-iris-test-data-01c8d6c	25/10/2021, 11:37:22
	GAIA-demo-example-pipeline-main-d6fd662	25/10/2021, 11:01:00
	feature/lakefs-ormb-lakefs-c8eb41f	24/10/2021, 01:37:38

Rows per page: 10

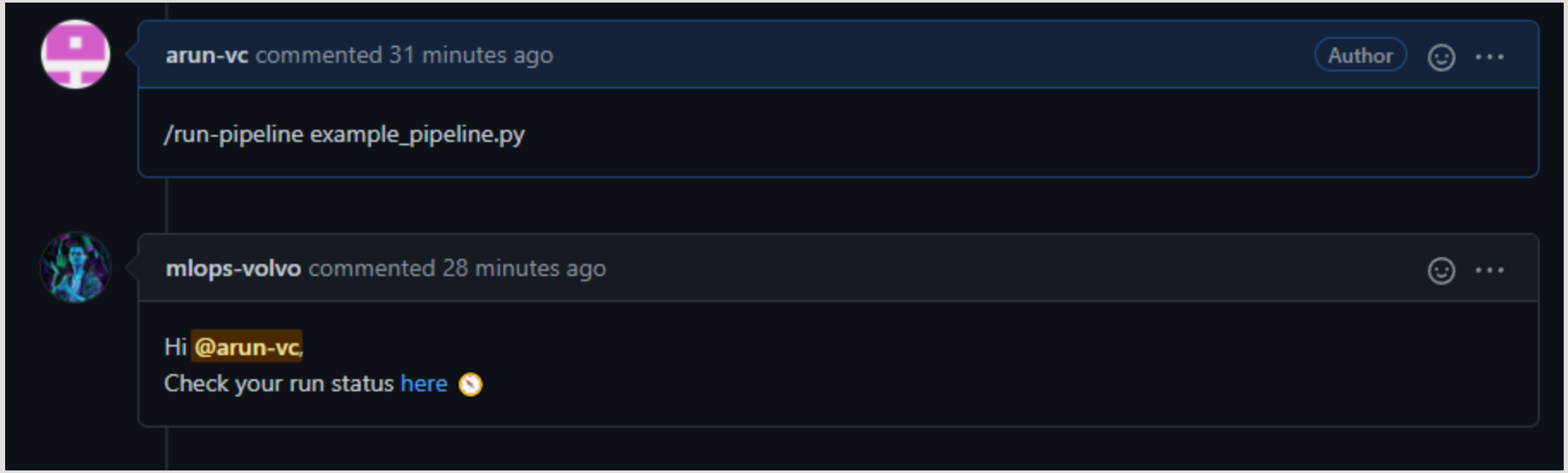
VOLVO

Commit CI



VOLVO

Trigger pipeline with github comment



`/run-pipeline example_pipeline.py`

Run an optionally parameterised `kubeflow` pipeline.

Parameters

- **kubeflow-pipeline-file** – Required
  `kubeflow` pipeline name.
  E.g. `/run-pipeline iris_pipeline.py`
- **params** – Optional
  Parameters to pass to the `kubeflow` pipeline.
  E.g. `/run-pipeline example_pipeline.py params={"name":"x","value":"206"}, {"name":"seed","value":"3007"}`

Once completed, a comment will be added to the PR which contains a 'run status' link for that particular `/run pipeline` command which will direct you to the relevant pipeline run.

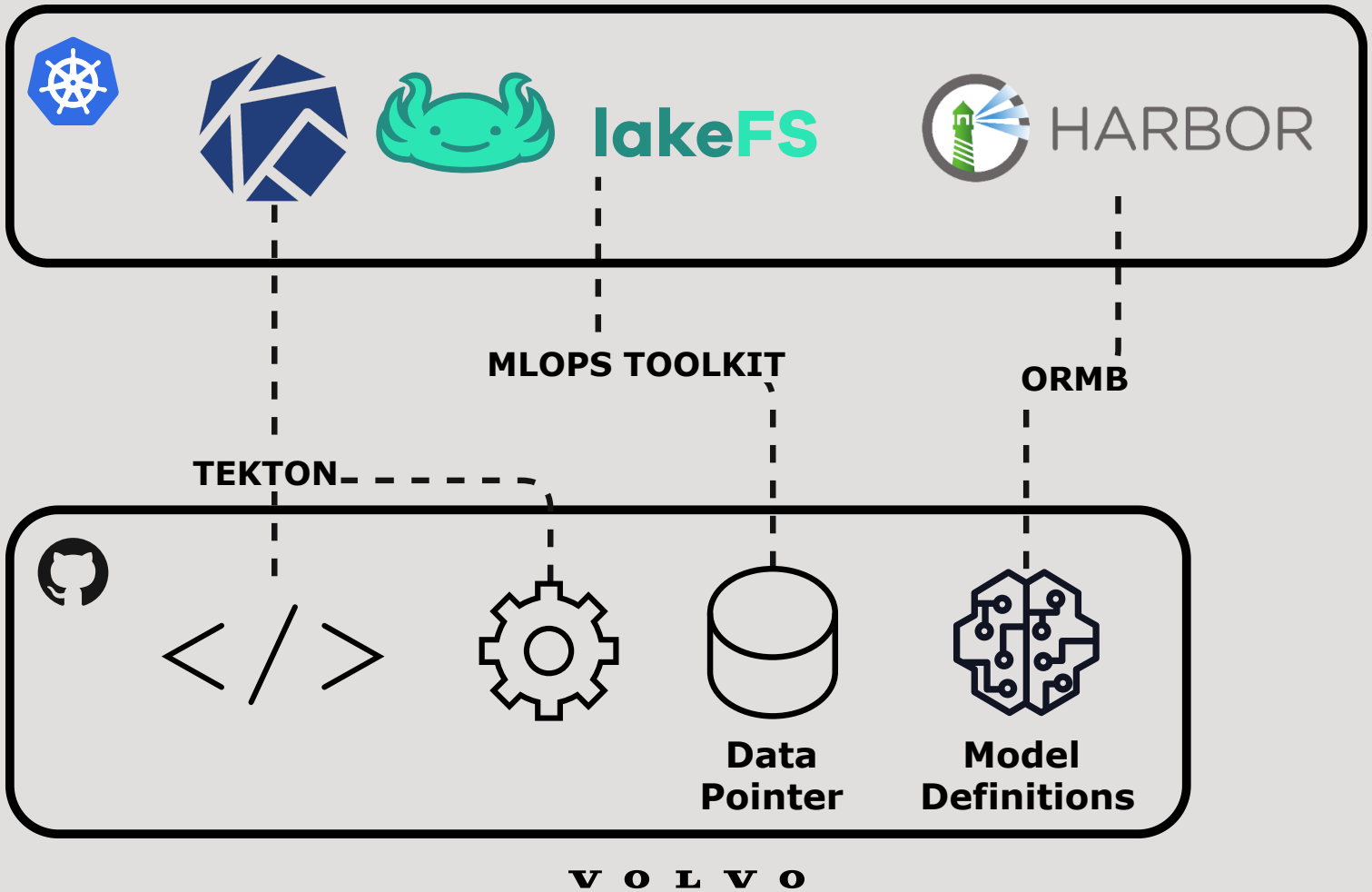
Versioning Models/Data

ORMB +  + 

 lakeFS + 

VOLVO

Versioning


lakeFS
HARBOR
MLOPS TOOLKIT
ORMB
TEKTON
Data Pointer
Model Definitions

Continuous Monitoring

VOLVO

Is it enough for ML?

VOLVO

Roadmap

- Improved monitoring capabilities
- Feature store integration (Feast)
- Feast + LakeFS ??
- Extend **mlops-tookit**
  - Generate pipeline & components from `/src`
- Federated learning approaches

VOLVO

Challenges

VOLVO

Having the right skills?

> You can’t be an AI expert these days and not have some grounding in software engineering. – Grady Booch


n
git

VOLVO

Data Scientist+

- Git
- CI/CD
- Container basics
- SW/ML testing
- SW/ML design patterns

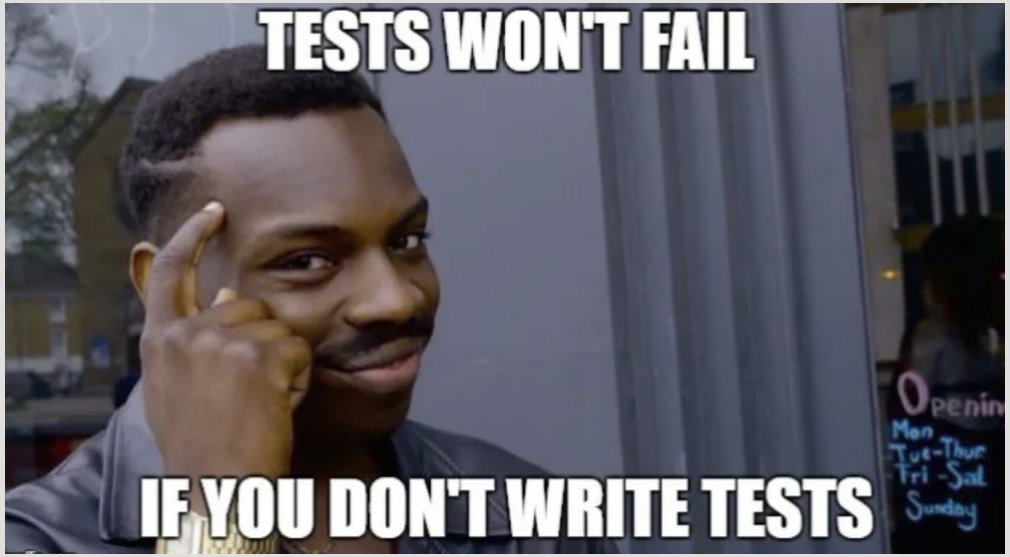
VOLVO

“Do machine learning like the great engineer you are, not like the great machine learning expert you aren’t. – Martin Zinkevich”

VOLVO

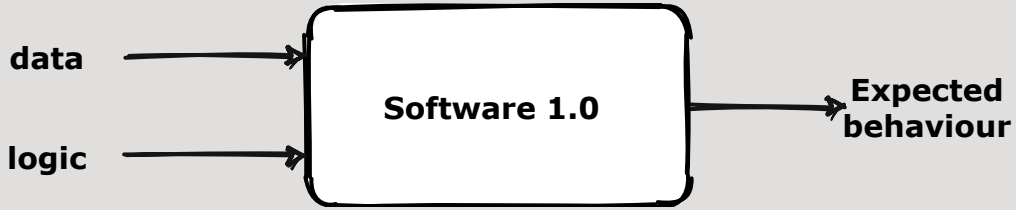

Treat ML like our cars
VOLVO

What are the key differences between SW & ML?



Software 1.0

Software tests help ensure that this **written logic** aligns with the actual expected behavior.



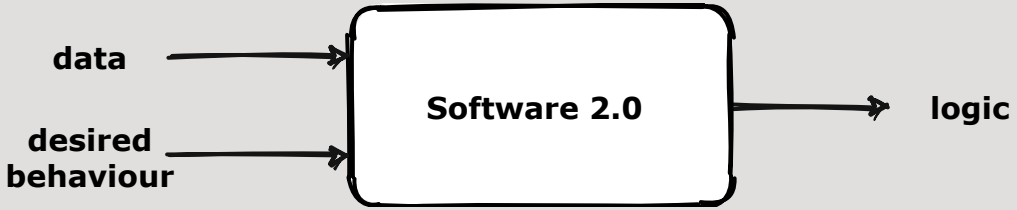
VOLVO

Conventions

- Don't merge code unless all tests are passing
- Always write tests for newly introduced logic when contributing code
- When contributing a bug fix, be sure to write a test to capture the bug and prevent future regressions.

Software 2.0

Objective: Learned logic consistently reflecting the desired behavior.



VOLVO

Evaluation

- Performance of an established metric on a validation dataset
- Plots
- Operational statistics
- Examples where the model was most confidently incorrect

Compare performance between models and make relative judgments

VOLVO

Conventions

- Save hyper-params used in training
- Promote models which offer an improvement over the existing model (or baseline)

Testing vs. Evaluation

- **Model evaluation** covers metrics and plots which summarize performance on a validation or test dataset.
- **Model testing** involves explicit checks for behaviors that we expect our model to follow.

Terminology

- **Failure modes** - Scenarios where model fails
- **Behavioral Regression** Decrease in performance on specific examples

Revised Conventions

- Save hyper-params used in training
- Promote models which offer an improvement over the existing model (or baseline) **and passes existing model regression tests.**
- Don't merge unless all tests are passing **(code, model, data)**

Revised Conventions cont.

- Always write tests for newly introduced (logic, **data, models)** when contributing code
- When contributing a bug fix, be sure to write a test to capture the bug and prevent future regressions.
- **When a failure mode in your ML system is identified, be sure to write a behavioral regression test for the specific examples.**

The Red Team 🚨

- Break the system
- Monitor production ML systems
- Identification of failure modes
- Track/prevent behavioral regressions for specific failure modes

MLOps is an engineering practice

VOLVO

The Bibles

- [1]Rules of Machine Learning: Best Practices for ML Engineering - Martin Zinkevich
- [2]Engineering best practices for ML - SE[4]ML

[1]Rules of Machine Learning: Best Practices for ML Engineering - Martin Zinkevich

VOLVO

Svet Penkov 🙏

> Rule #4: Define the **operational domain** for your ML system

VOLVO

“Doing ML without MLOps is like trying to win a formula 1 race without a pit crew”

VOLVO

Thanks for listening and don't be a stranger 🙏

VOLVO

We are Hiring 💼

📪 leonard.aukea@volvocars.com

LeonardAukea

LeonardAukea

VOLVO

References

- [1]Rules of Machine Learning: Best Practices for ML Engineering - Martin Zinkevich
- [2]Engineering best practices for ML - SE[4]ML
- [3]Effective testing for machine learning systems - Jeremy Jordan

Links

- https://tekton.dev/
- https://github.com/kleveross/ormb
- https://lakefs.io/
- https://www.kubeflow.org/
- https://github.com/kserve/kserve
- https://github.com/google/ml-metadata
- https://feast.dev