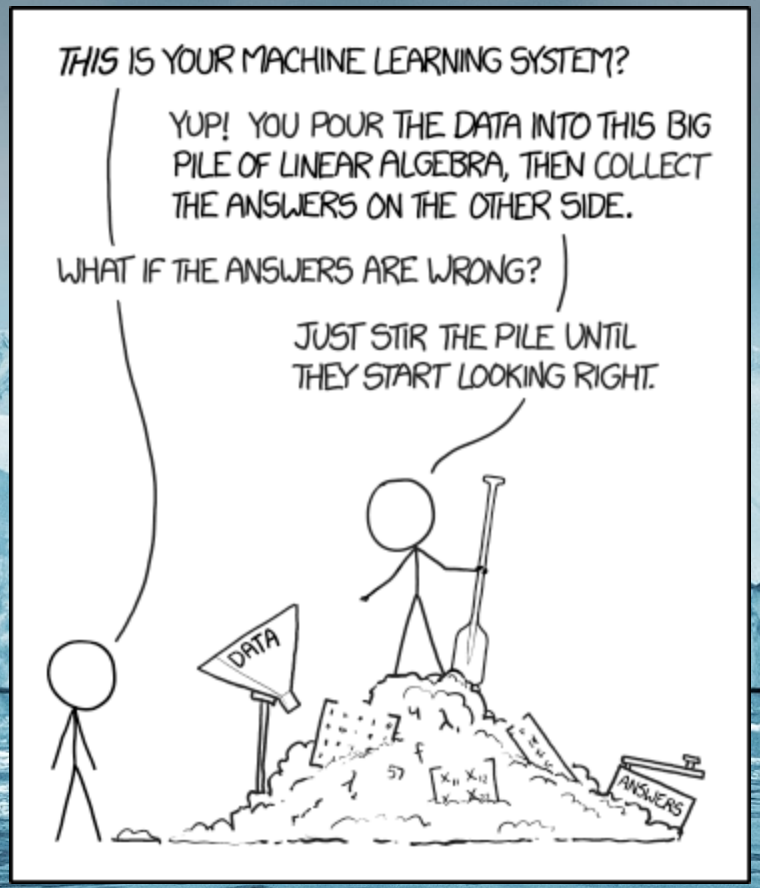
ZeroOps to MLOps



VOLVO

Driving ML Engineering & Operations

 LeonardAukea

VOLVO

Purpose

“Develop and operate tools and processes that accelerate and support all teams using advanced analytics and artificial intelligence at Volvo Cars.”

VOLVO

ZeroOps 2019

- 💻 Company laptop + On-prem vm
- 🕰️ Ask IT to install tools
- 0 Infrastructure for ML Workloads


WHERE THE #%& IS MY AI?
VOLVO

VOLVO

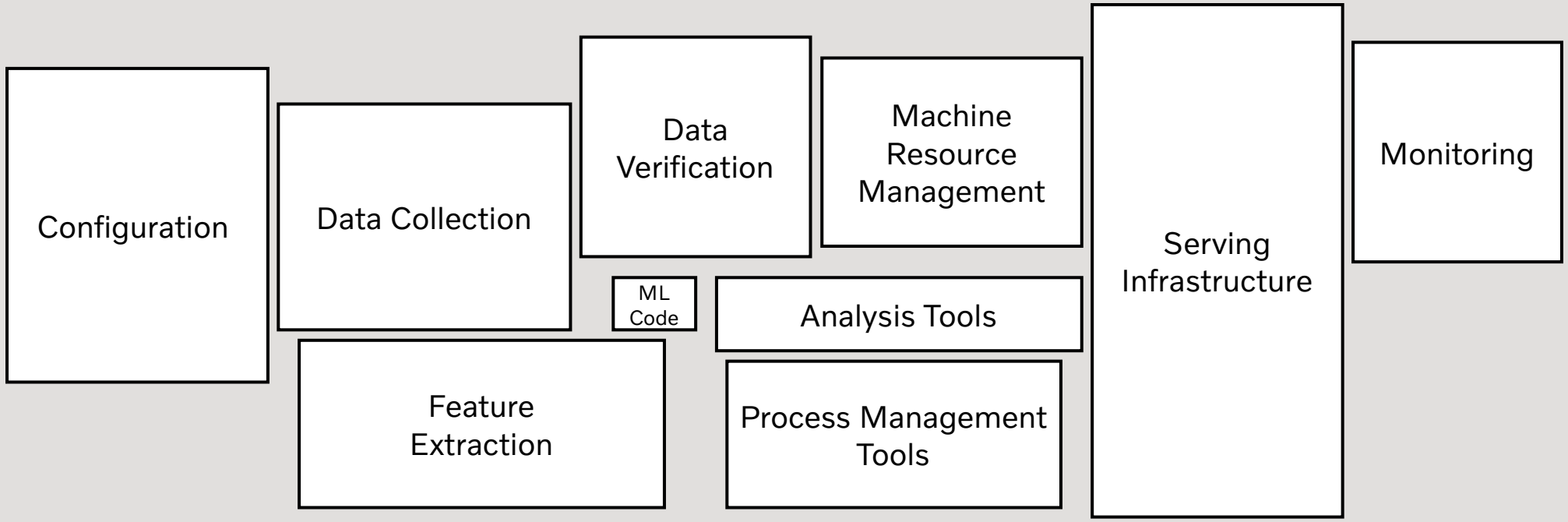
Hello, we are data scientists    

VOLVO

VOLVO


BACK TO THE
90's
ML Community
Researchers trying to build software
VOLVO

ML Model ≠ ML System



[1]Hidden Technical Debt in Machine Learning Systems

> The complexity of machine learning systems doesn't subsist in the complexity of individual components, but rather in their orchestration, connections, and interactions. We can thus think of machine learning systems as special cases of general distributed systems.

[2] Mapping the territory for MLOps

ML; why is it hard?

- Uncertainty
- Dependencies
- Reproducibility
- Cross functional teams
- etc.

Having the right skills?

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

Collaboration?

VOLVO


n
git

Basic software skills?

VOLVO

Pillars of MLOps

- Reproducibility ✗
- Scalability ✗
- Robustness ✗
- Collaboration ✗
- CI/CT/CD Automation ✗

VOLVO

Approach

- Cloud Native Stack
- Templating & reusable components
- Glue infra with gitflow
- Expect SW basics from ML practitioners
- Evangelize good principles

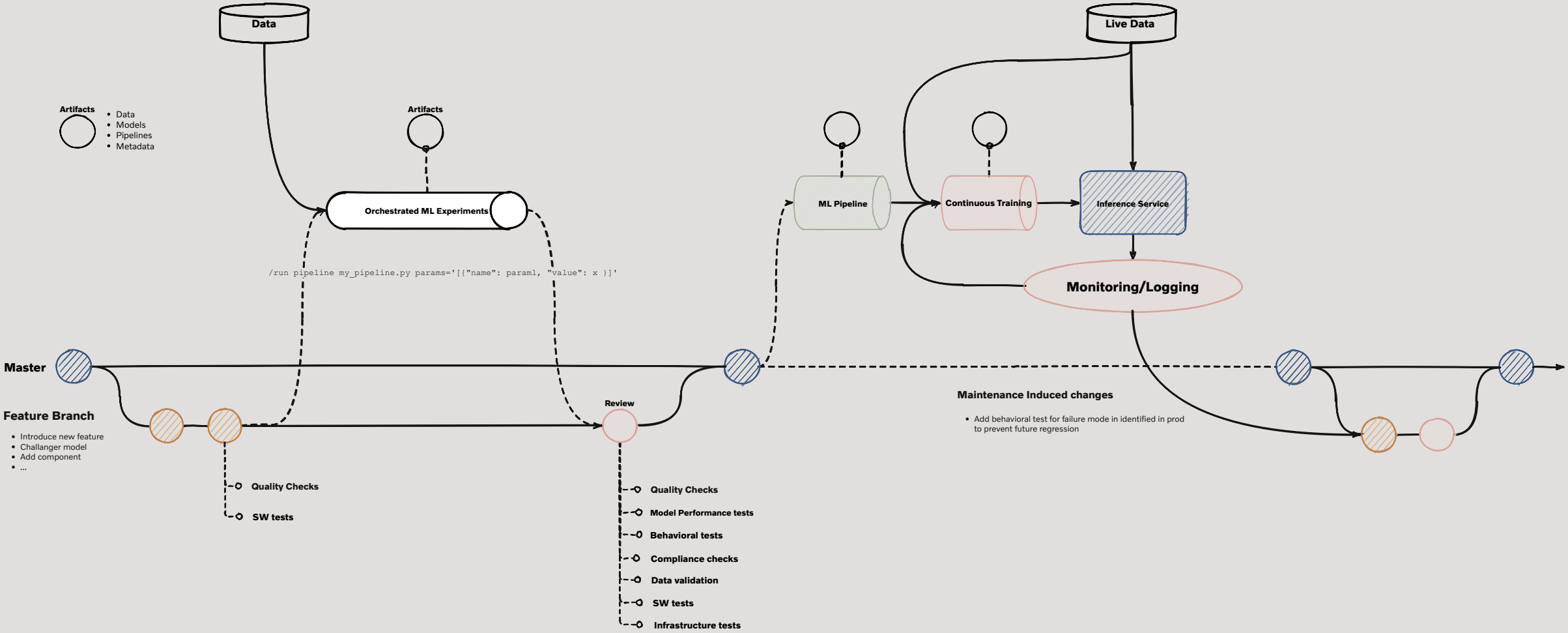
The Glue

VOLVO

Automation 

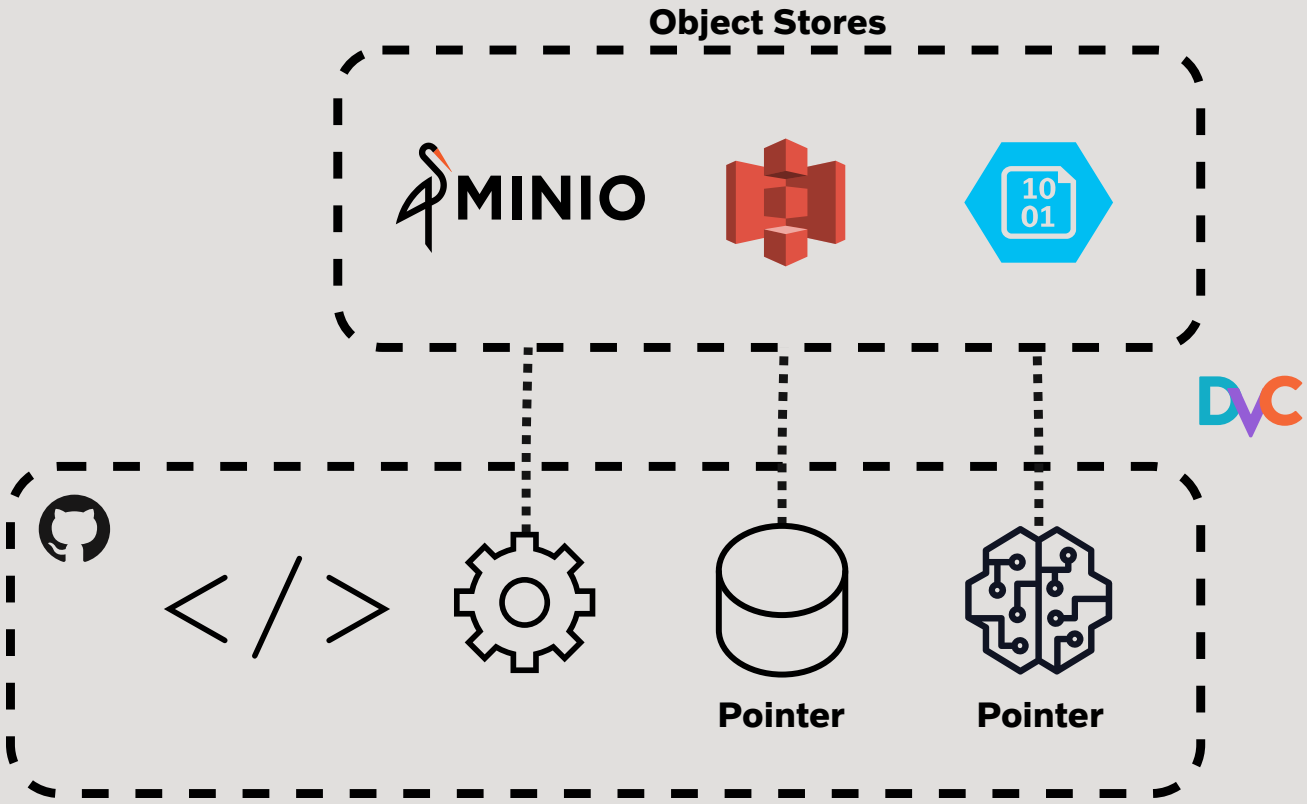
> The level of automation of the Data, ML Model, and Code pipelines determines the maturity of the ML process.

1. Use template https://github.com/volvo-cars/ds-project-template
2. Commit `profile.yaml` =>
   - Creates webhook & eventlistener for CICD
   - Application workspace on DSP, with added contributors
   - Minio bucket created
   - Project CI added


Data
Live Data
Artifacts
Data
Models
Pipelines
Metadata
Artifacts
Orchestrated ML Experiments
/run pipeline my_pipeline.py params='[{"name": param1, "value": x }]'
ML Pipeline
Continuous Training
Inference Service
Monitoring/Logging
Master
Feature Branch
Introduce new feature
Challanger model
Add component
...
Quality Checks
SW tests
Review
Quality Checks
Model Performance tests
Behavioral tests
Compliance checks
Data validation
SW tests
Infrastructure tests
Maintenance Induced changes
Add behavioral test for failure mode in identified in prod to prevent future regression

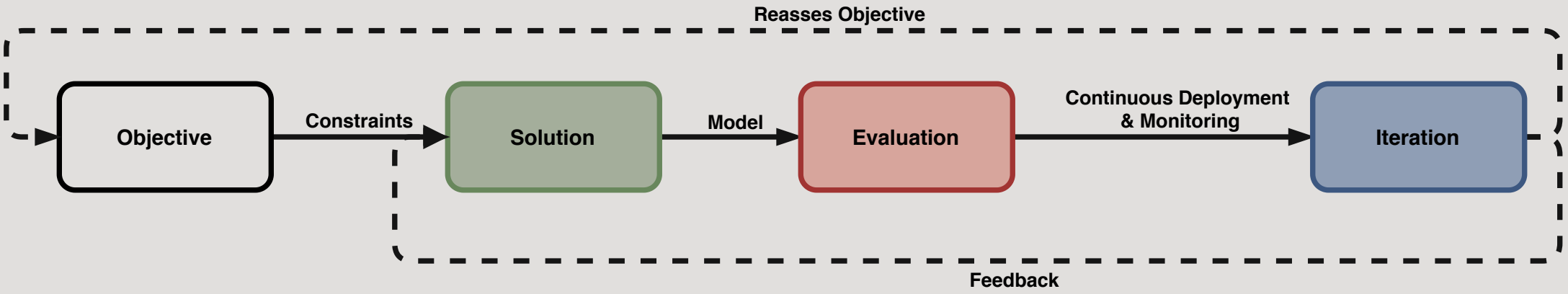
VOLVO

Reproducibility



VOLVO

Continuous feedback ♻



Demo ⏰

VOLVO

Roadmap

- Not 100% happy with DVC
- Extend chatbot functionality
- Better AD Integration
- Tighter integration with internal data stores
- Feast feature store

VOLVO

Pillars of MLOps

- Reproducibility ✅
- Scalability ✅
- Collaboration ✅
- CI/CT/CD Automation ✅
- Robustness ⭕

VOLVO

Robustness 

VOLVO

Treat ML like our cars

VOLVO

> (...) training data needs testing like code, and a trained ML model needs production practices like a binary does, such as debuggability, rollbacks and monitoring

VOLVO

Model evaluations metrics are by definition summary statistics

- **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.

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

The Red Team 

VOLVO

So what is MLOps?

VOLVO

So what is MLOps?

> MLOps is an ML engineering practice; not a tool or a platform

VOLVO

Example

> It is important to create team cultures that reward deletion of features, reduction of complexity, improvements in reproducibility, stability, and monitoring to the same degree that improvements in accuracy are valued.

Another example

> Aim for a “neutral” first launch: a first launch that explicitly deprioritizes machine learning gains, to avoid getting distracted. A simple model provides you with baseline metrics and a baseline behavior that you can use to test more complex models

Objective: Shorten the development cycles and increase deployment velocity, in order to reduce friction and delays in the AI value stream (roughly speaking)

VOLVO

References & Links

- [1]Hidden Technical Debt in Machine Learning Systems
- [2]Mapping the territory for MLOps
- [3]Made with ML
- [4]Engineering best practices for ML
- [5]Rules of Machine Learning: Best Practices for ML Engineering

Have you ever?

- Deployed a model to production? ✅
- Collaborated on a large complex codebase? ✅
- Written a unit test, regression test? ✅

We are Hiring 💼

📫 Leonard Aukea

LeonardAukea

VOLVO

Q&A 🙏

VOLVO