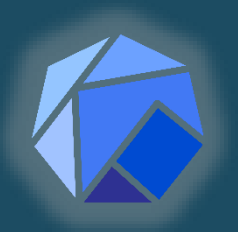
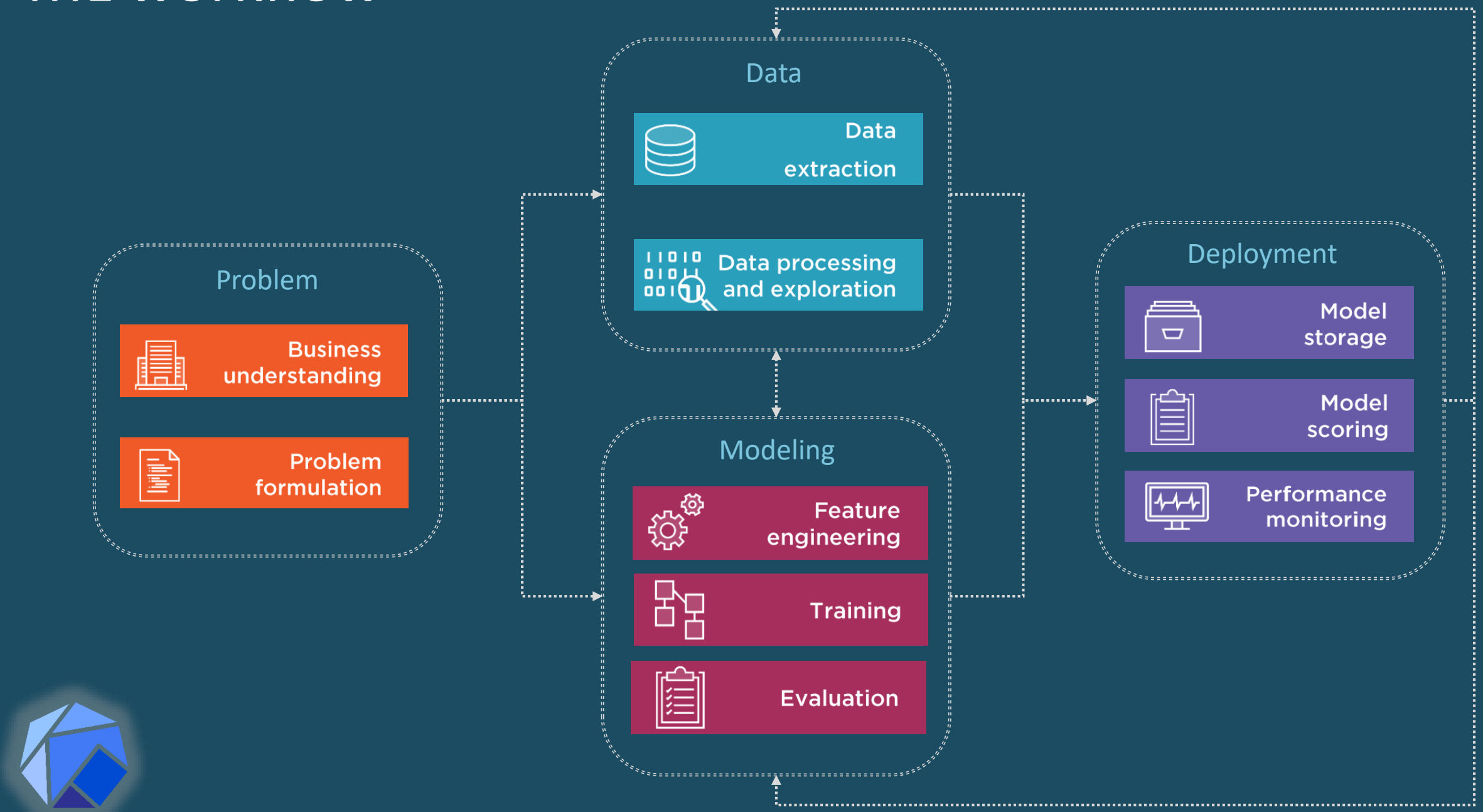
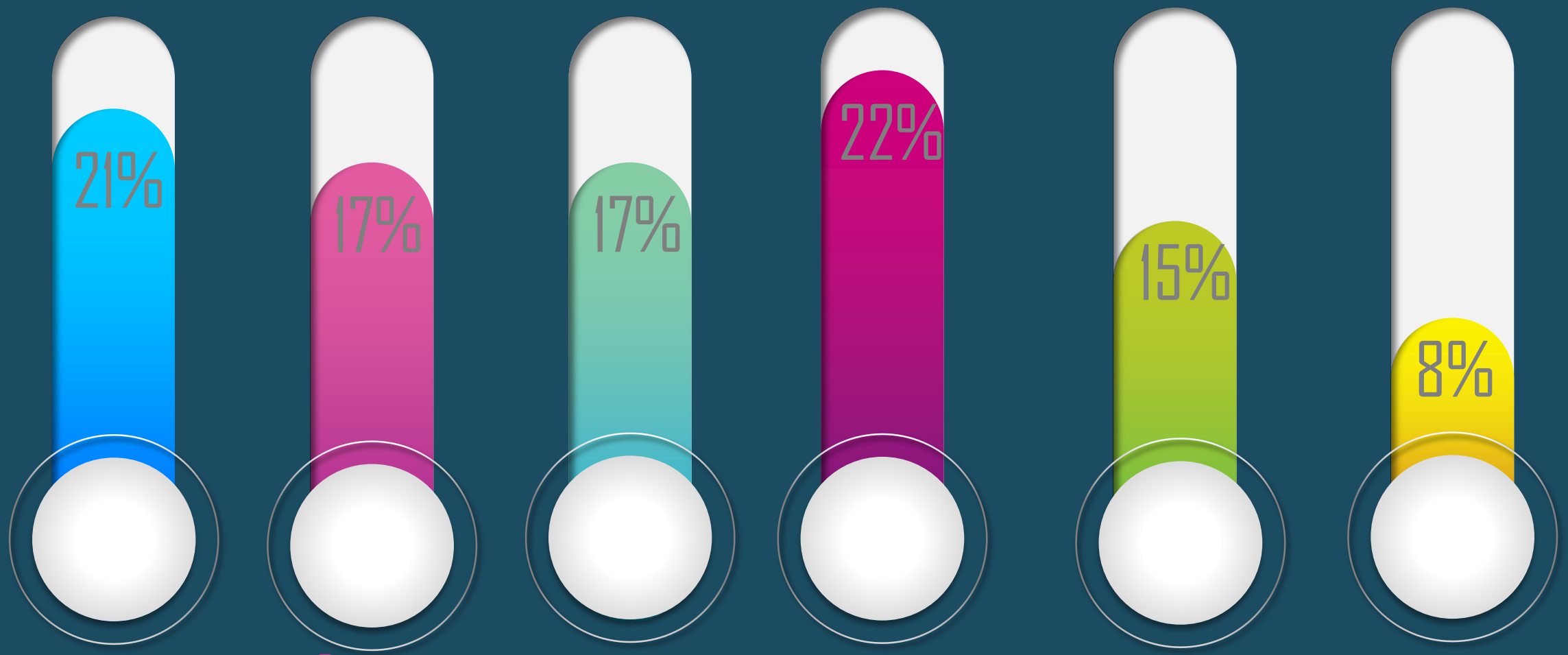
Managing Machine Learning with MLOps and Kubeflow



ML workflow


Problem
Business understanding
Problem formulation
Data
Data extraction
Data processing and exploration
Modeling
Feature engineering
Training
Evaluation
Deployment
Model storage
Model scoring
Performance monitoring

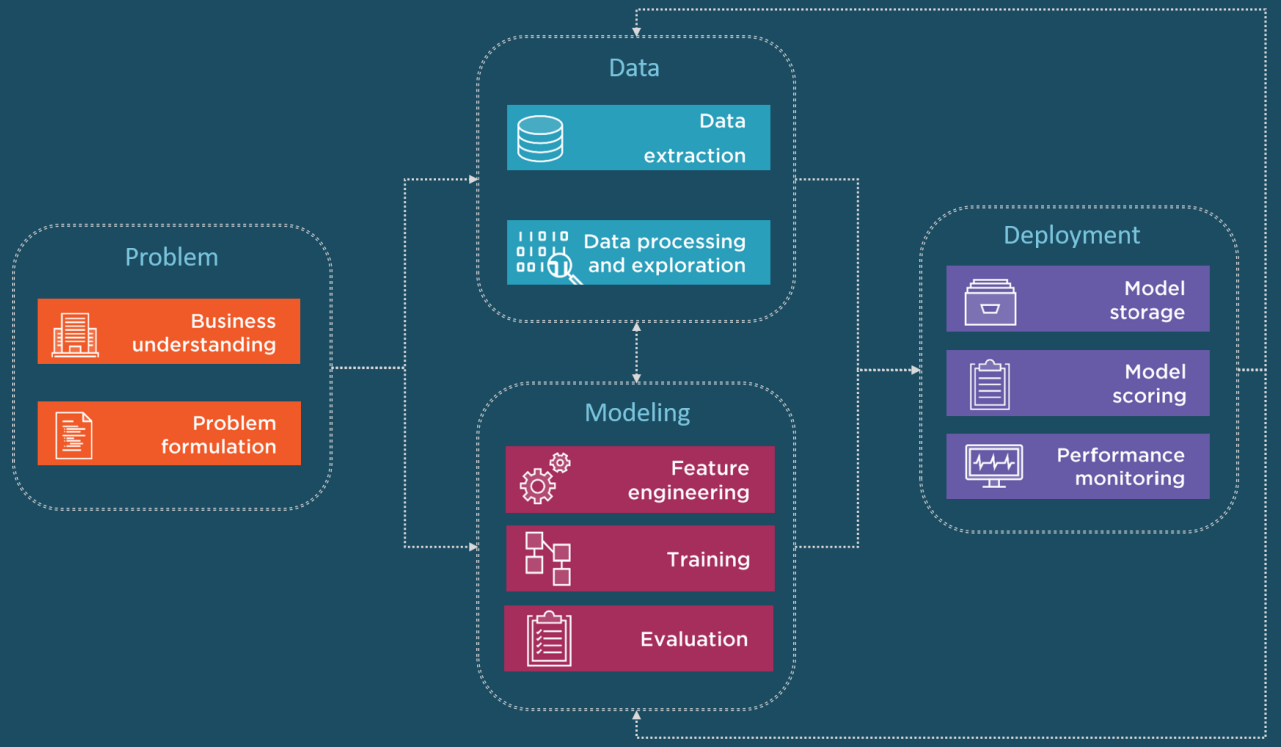
ML maturity levels - 2020



2020 state of enterprise machine learning - https://info.algorithmia.com/hubfs/2019/Whitepapers/The-State-of-Enterprise-ML-2020/Algorithmia_2020_State_of_Enterprise_ML.pdf

ML current scenario

Typical software development workflow and toolkit



Machine Learning workflow

Increased time to production

Low return on investment

Doesn‘t work seamlessly

ML current scenario

Hardware utilisation

GPU Resource Consumption

Collaboration

Work in team with different enivornments

Track experiments

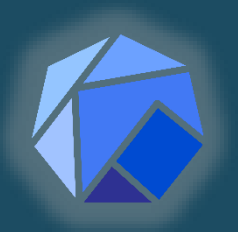
Track and compare experiments

Logging

Generate, analyse and act on monitoring logs

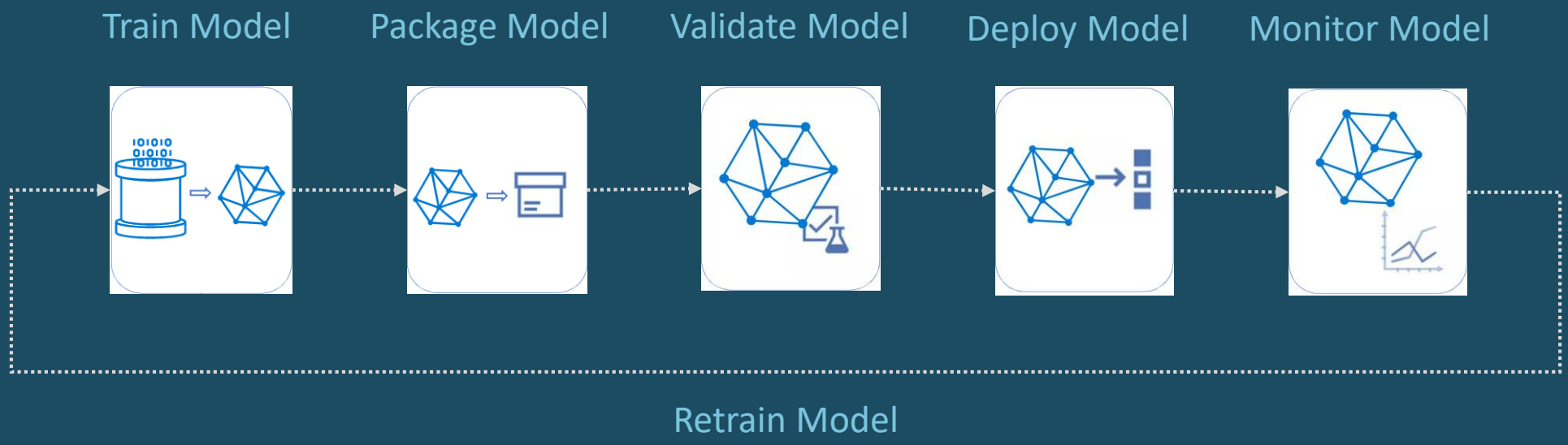
Scale

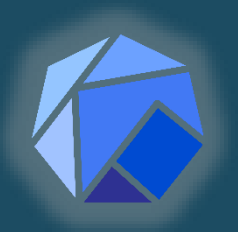
Deal with scale



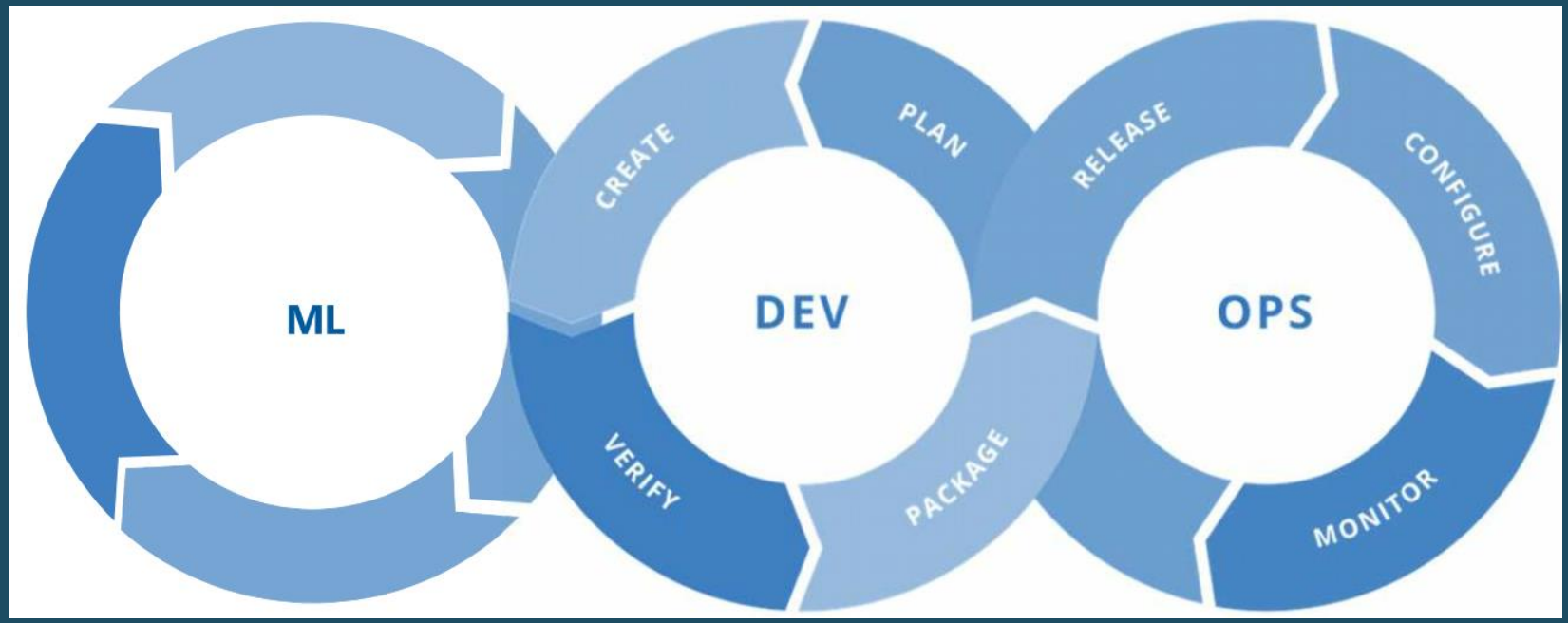
E2E ML lifecycle

- **Develop & train model** with reusable ML pipelines
- **Package model** using containers to capture runtime dependencies for inference
- **Validate model** behavior functionally, in terms of responsiveness, in terms of compliance
- **Deploy model** to cloud & edge, for use in real-time / streaming / batch processing
- **Monitor model** behavior & business value, know when to replace / deprecate a stale model





MLOps = ML + DEV + OPS



Experiment

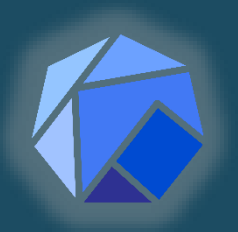
Data Acquisition
Business Understanding
Initial Modeling

Develop

Modeling + Testing
Continuous Integration
Continuous Deployment

Operate

Continuous Delivery
Data Feedback Loop
System + Model Monitoring



MLOps Benefits

Automation / Observability

- Code drives generation and deployments
- Pipelines are reproducible and verifiable
- All artifacts can be tagged and audited

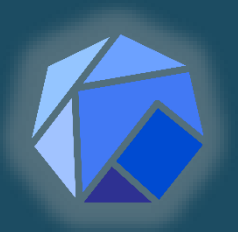
Validation

- SWE best practices for quality control
- Offline comparisons of model quality
- Minimize bias and enable explainability

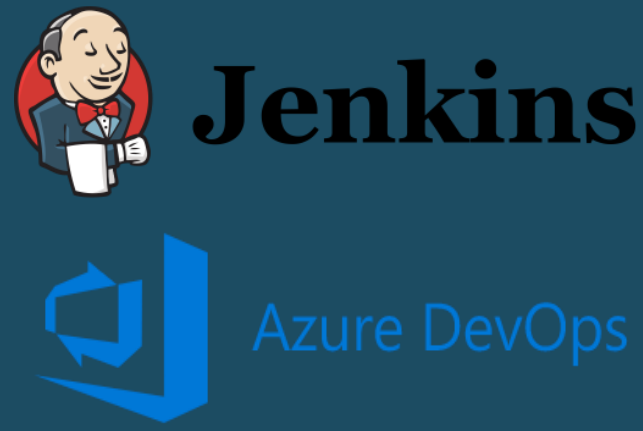
Reproducibility / Auditability

- Controlled rollout capabilities
- Live comparison of predicted vs. expected performance
- Results fed back to watch for drift and improve model

== VELOCITY and SECURITY for ML

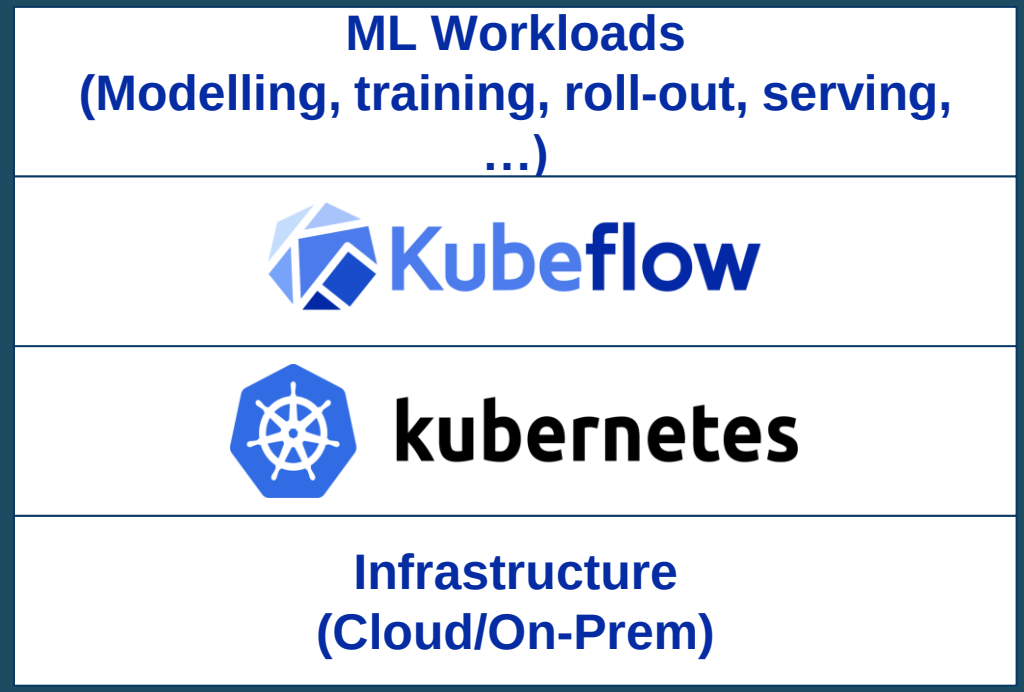


MLOps with Kubeflow + CI/CD

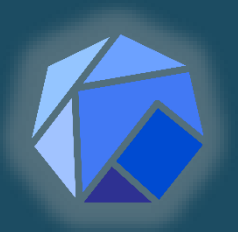


Kubeflow

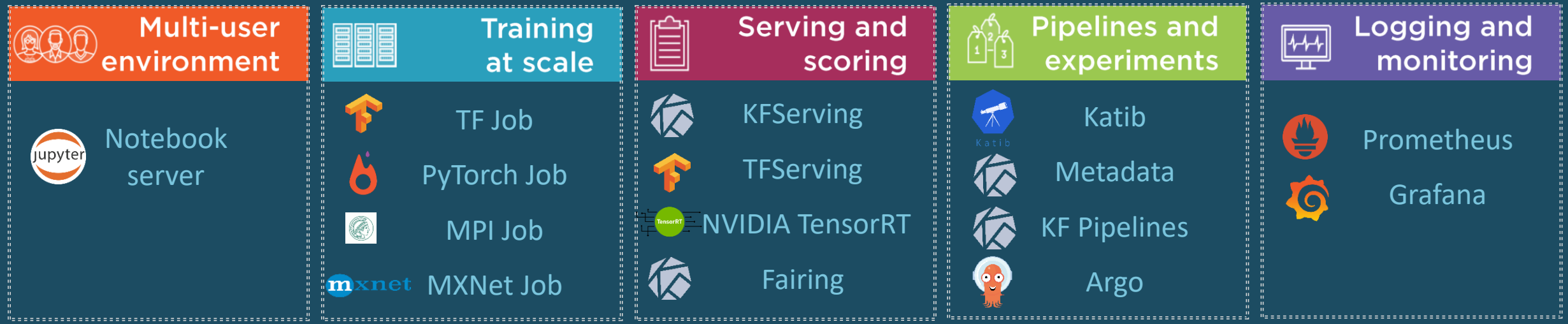
Kubeflow – Machine Learning toolkit for Kubernetes

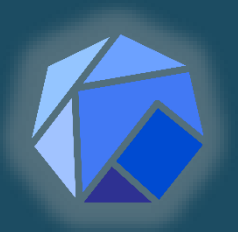


- Open source machine learning toolkit for Kubernetes
- Simple, portable and scalable workflow
- Adapted Kubernetes for Machine Learning
- Originated at Google



Kubeflow components





ML in real world multi-cloud