

Automating Machine Learning and Deep Learning Workflows

Information

- Name: Mourad Mourafiq
- Author of an open source platform: Polyaxon
- twitter: @mmourafiq
- GitHub: mouradmourafiq

What is Polyaxon

- Solves the machine learning life cycle
- Can be deployed on premise or on any cloud platform
- Is open source
- Works with any library or framework
- Can be used by single users or large organizations
- Provides compliance, auditing, and security

Why you need a tool to manage your ML operations?

- Software development is mature
- Why not use the same tools?
- What is the difference between software development and ML development?
- What is the difference between software deployment and ML deployment?

Difference between software development and ML development

- Development objectives
- Vetting and quality assurance
- Development stack

Difference between software deployment and ML deployment

- ML deployment needs a Feedback Loop
- Iteration and refinement
- People involved in the deployment cycle

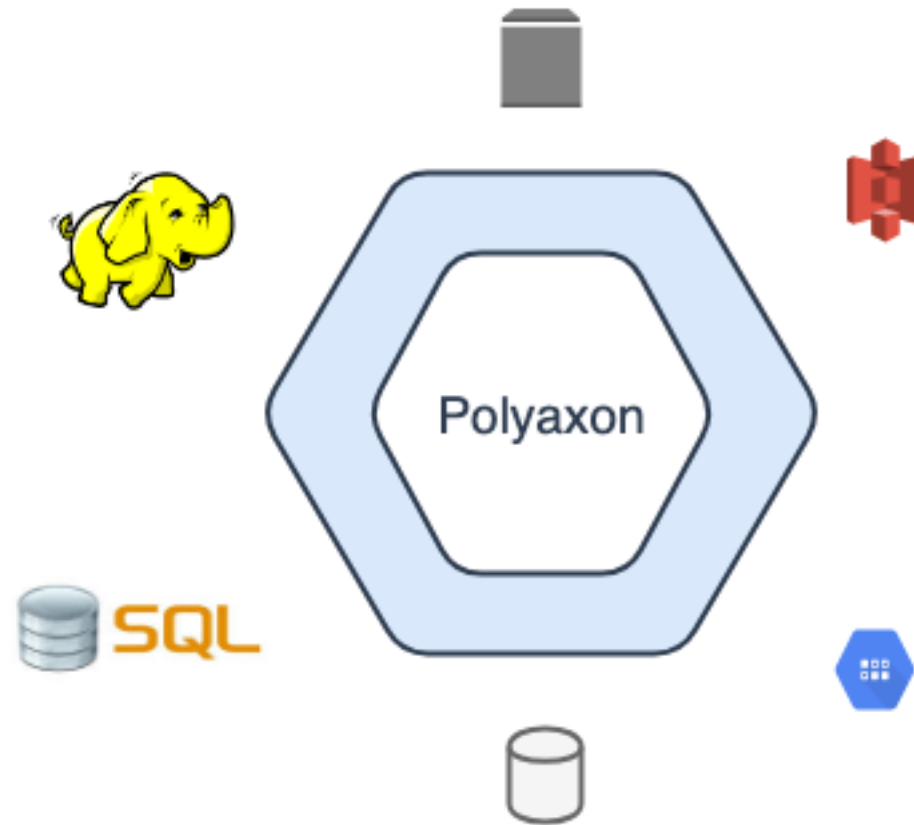
What should a ML platform answer

- Should be flexible to support open source initiatives
- Provides different deployment options
- Ideally open source
- Works with any library or framework
- Scales with users
- Provides compliance, auditing, and security

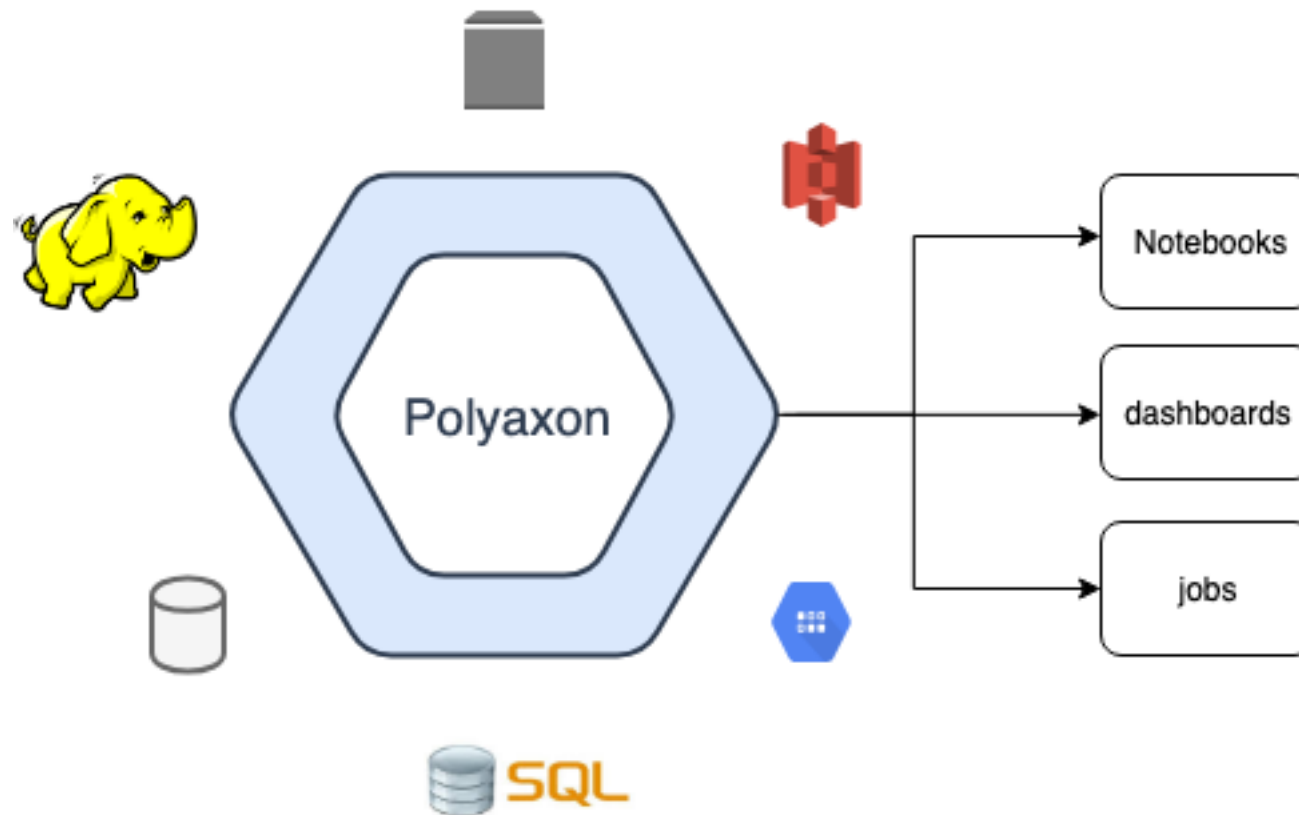
ML development lifecycle

- Data access
- Data exploration and Feature engineering
- Experimentation: iteration, packaging, reusability, reproducibility.
- Scaling: Scheduling, orchestration and optimization
- Tracking: code, data, params, artifacts, metrics
- Insights, reporting, and knowledge distribution
- Model management: packaging, deployment, and distribution
- Compliance, auditing, and access management.
- Automation, events, and workflows
- User experience

- Data access



- Data exploration & Feature engineering



- Experimentation
- Different environments: local, remote, cluster
- Portability and reusability
- Reproducibility

- Experimentation: Different environments



- Experimentation: Packaging

```
version: 1

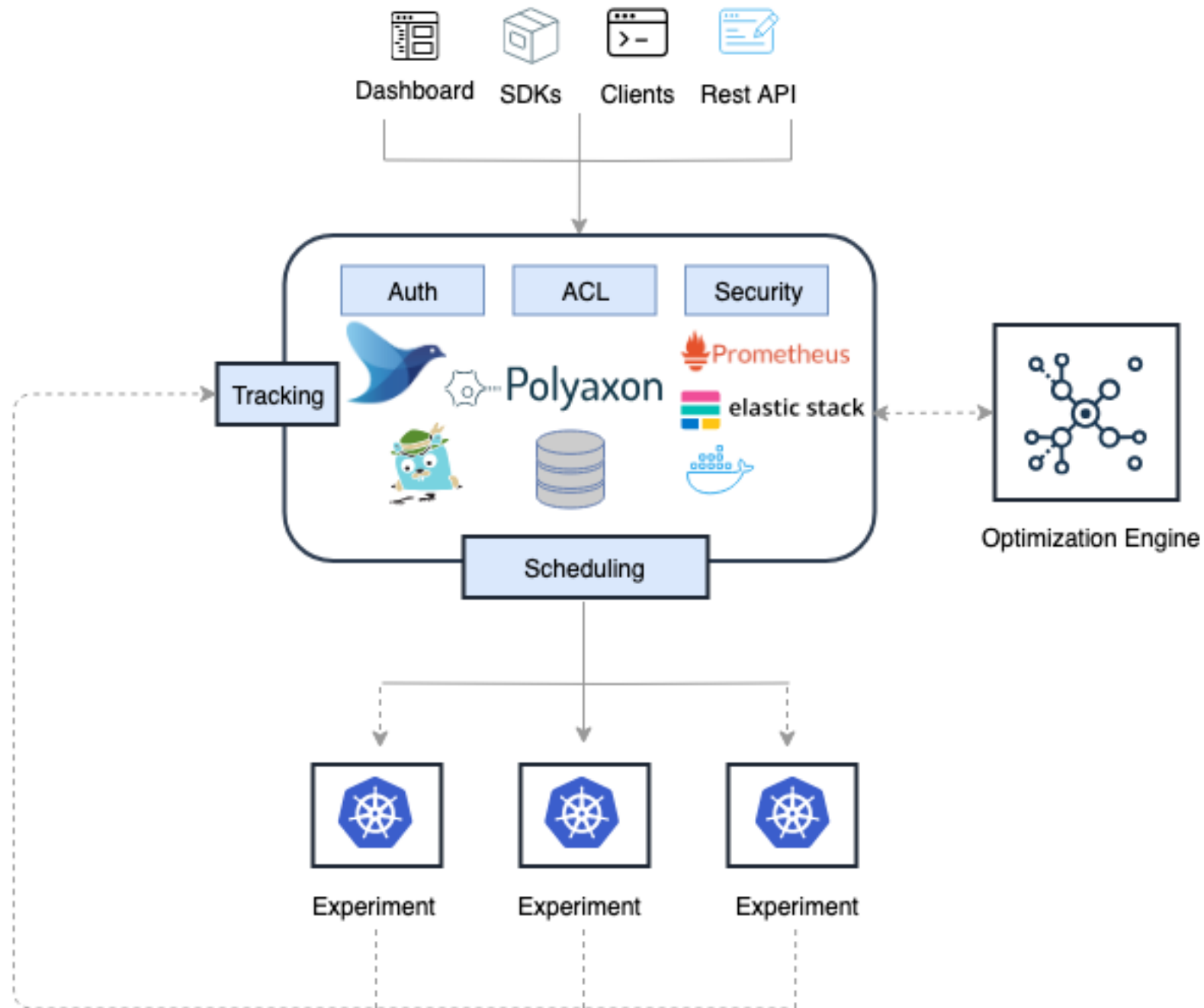
kind: experiment

build:
  image: tensorflow/tensorflow:1.4.1-py3
  build_steps:
    - pip3 install polyaxon-client

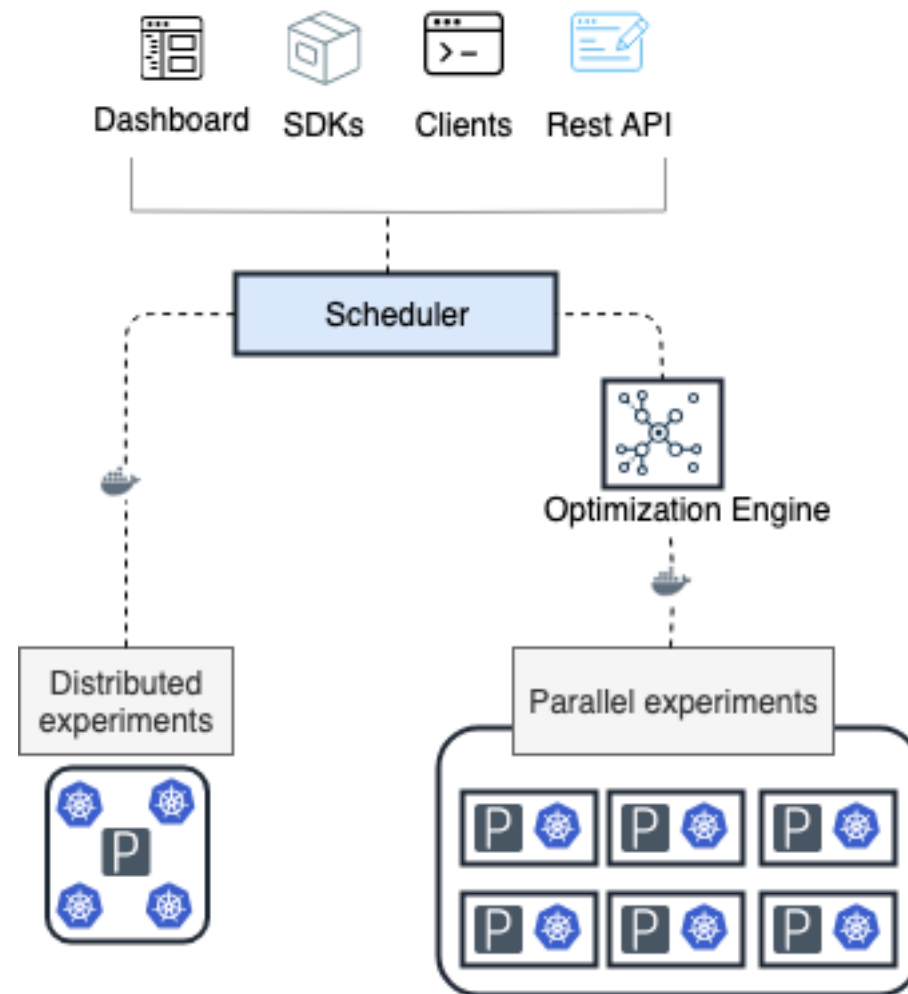
run:
  cmd: python model.py
```

- `polyaxon run -f polyxonfile.yaml`
- `polyaxon run -f polyxonfile.yaml --local`

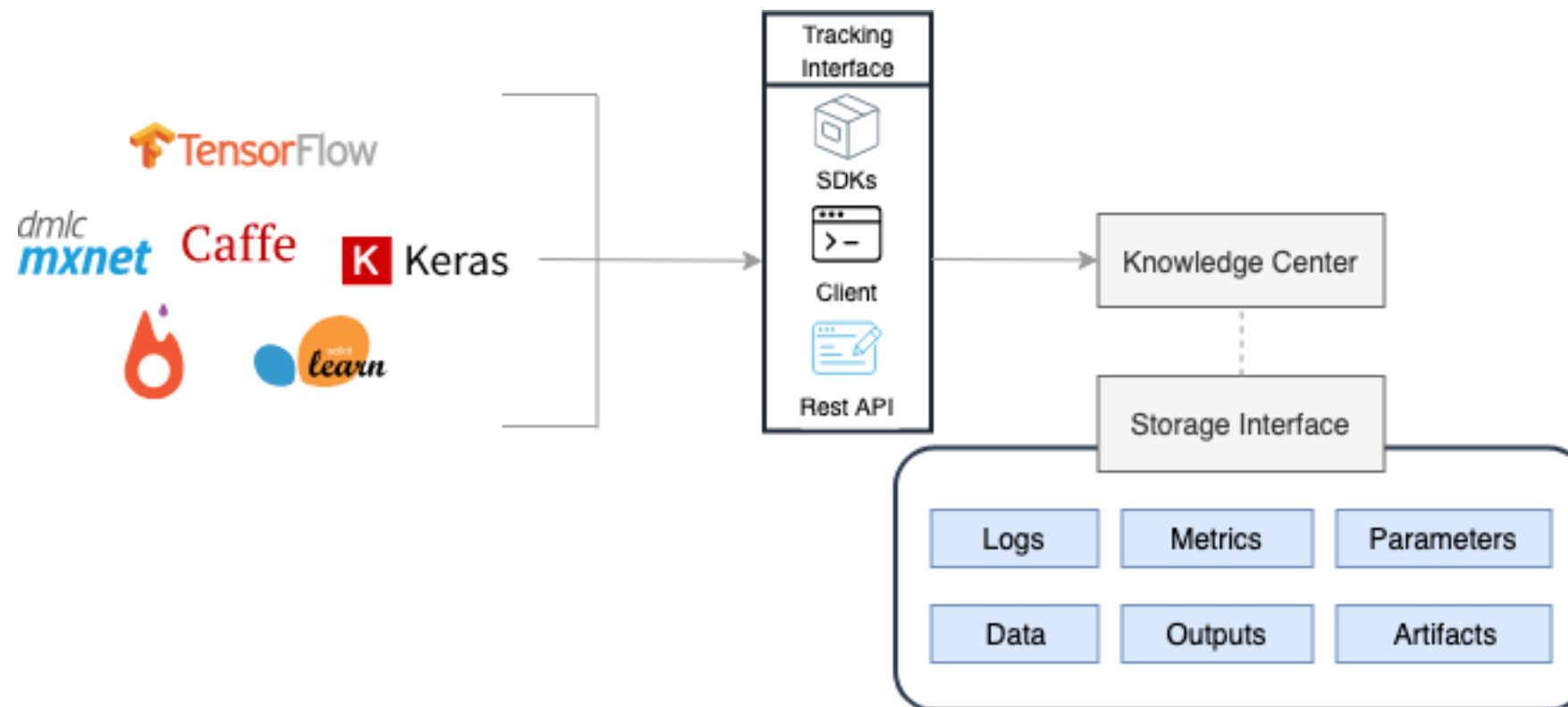
- Scheduling & Orchestration



- Hyperparams tuning & distributed training



- Experiments tracking



- Experiments tracking

```
# Polyaxon
experiment = Experiment('iris')
experiment.create(framework='xgboost', tags=['examples'])
experiment.log_params(log_learning_rate=args.log_learning_rate,
                    max_depth=args.max_depth,
                    num_rounds=args.num_rounds,
                    min_child_weight=args.min_child_weight)

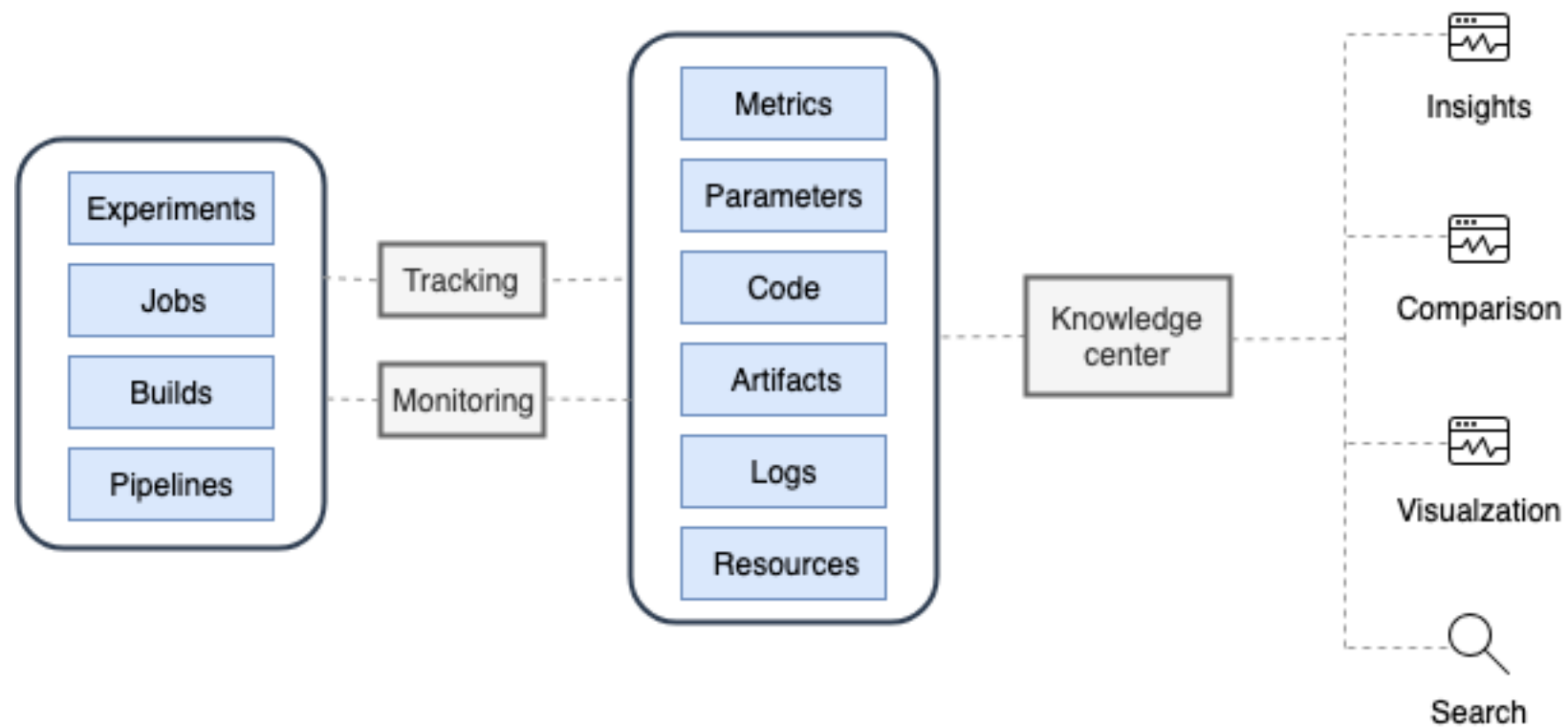
iris = load_iris()
X = iris.data
Y = iris.target

X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.2)

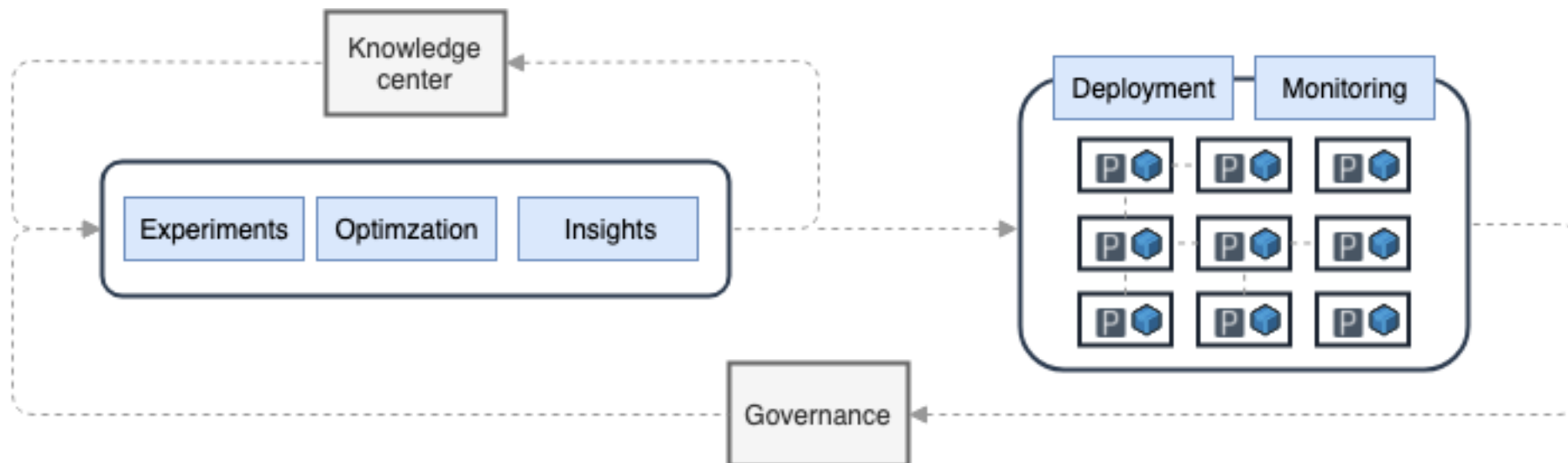
# Polyaxon
experiment.log_data_ref(data=X_train, data_name='x_train')
experiment.log_data_ref(data=y_train, data_name='y_train')
experiment.log_data_ref(data=X_test, data_name='X_test')
experiment.log_data_ref(data=y_test, data_name='y_train')

logger.info('Train model...')
accuracy = model(log_learning_rate=args.log_learning_rate,
                max_depth=args.max_depth,
                num_rounds=args.num_rounds,
                min_child_weight=args.min_child_weight)
experiment.log_metrics(accuracy=accuracy)
```

- Insights, reporting, and knowledge distribution



- Model Management



- Compliance & Governance
- Manage model development and deployment
- Rigorous and auditable workflows

- Automation & Events
- Simple yet effective specification to create workflows and automation
- Integration with other pipelining tools, e.g. airflow
- Events and triggers based on data, code, metrics, ...

mourad@polyaxon.com

twitter: @mmourafiq

GitHub: mouradmourafiq

<https://polyaxon.com>

twitter: @polyaxonai

GitHub: polyaxon