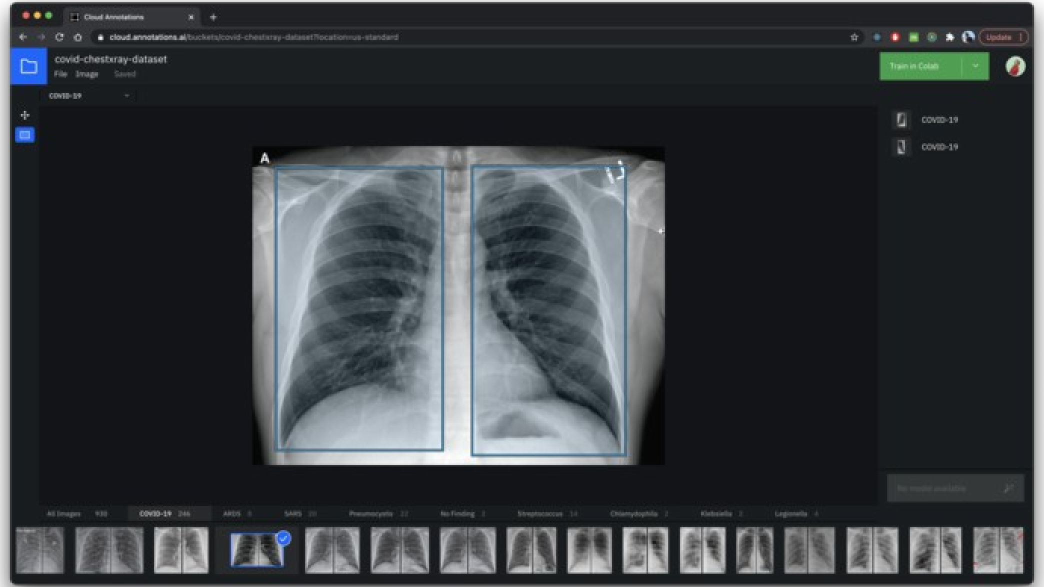
```yaml
---
apiVersion: slide/v1
author:
  -
    affiliation: "IBM Center for Open Source Data and AI Technologies (CODAIT)"
    name: "Romeo Kienzler"
    role: "Data Scientist"
  -
    affiliation: "University Hospital Basel"
    name: "Ivan Nesic"
    role: "ML Engineer"
  -
    name: "The Elyra, JupyterLab, Kubernetes and KubeFlow authors"
kind: Titlepage
published:
  -
    type: video
    url: "https://www.youtube.com/watch?v=H8WskMEUI74"
  -
    type: conference
    url: "http://conference.scipy.org/proceedings/scipy2021/pdfs/ivan_nesic.pdf"
title:
  - "CLAIMED – Component Library for AI, Machine Learning, ETL and Data Science"
  - "A Kubeflow no code / low code architecture"
```

Requirements

- Rapid prototyping using visual editing and notebooks
- Seamless scaling during development and deployment
- GPU support
- ML tools: PyData stack, TensorFlow, PyTorch, …
- life science tools: DICOM input, DICOM output, ...
- Reproducibility
- Data lineage
- Reference implementation in open source
- Collaboration support

Cloud annotations provides...

Browser based image labeling: Classification / Object recognition training data



IBM Cloud Annotations

...in Open Source

Docker provides...

OS-level "virtualization" to deliver and run software in packages called containers

Lightweight "virtualization"

Security and isolation

Super-fast startup/teardown

...on top of Linux

Kubernetes provides...

Some prominent users: Adidas, Booking.com, Box, Google, Huawei, IBM, The New York Times, ING, ricardo.ch, Spotify, Wikimedia, Zalando

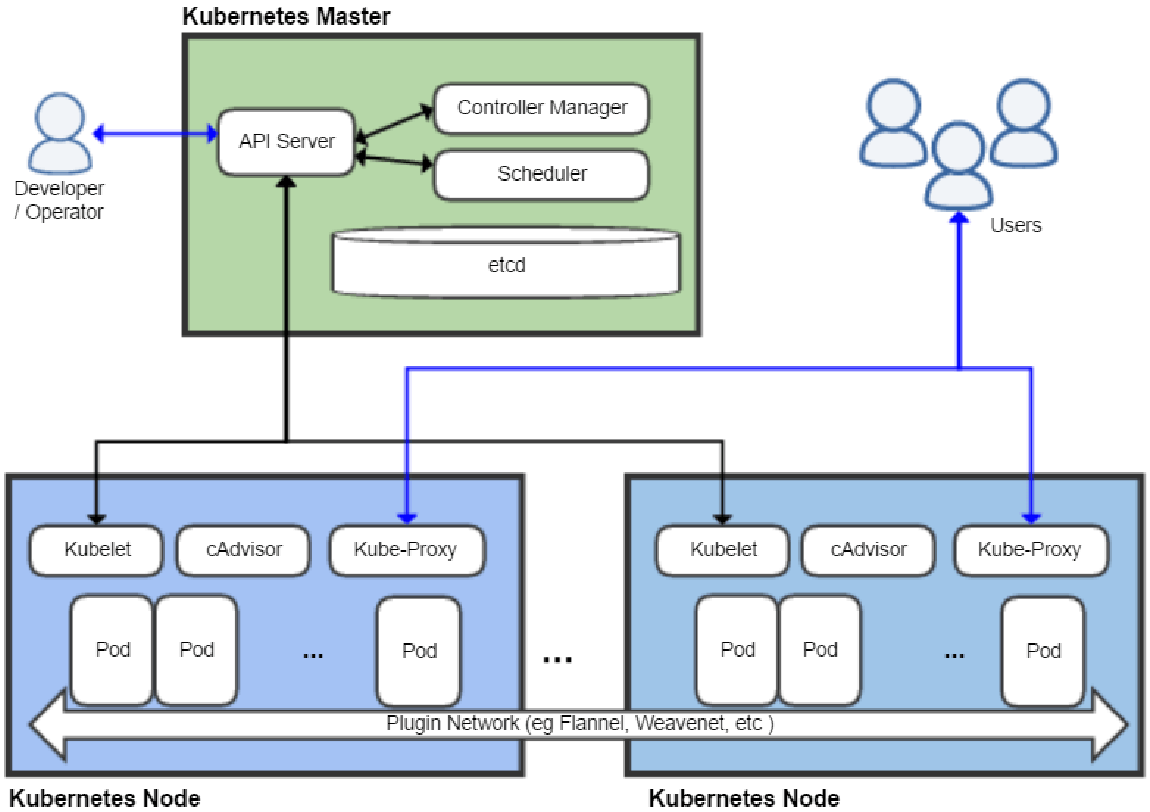
Container Orchestration

Deployment, scaling and management

High availability

...on top of Linux Clusters

Kubernetes architecture...



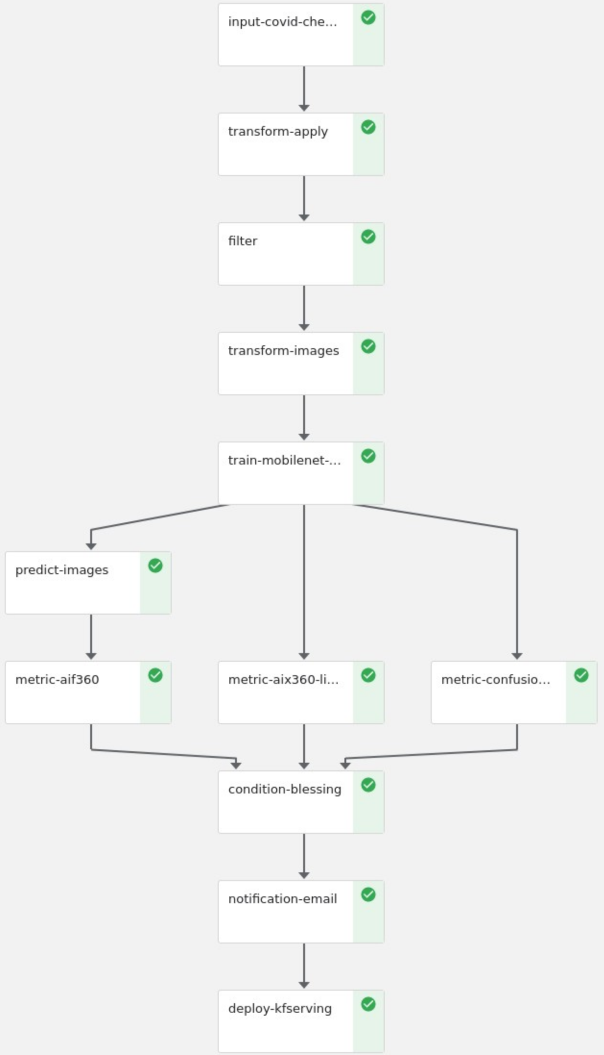
Kubeflow provides...

Some prominent users:
Amazon Web Services
IBM Watson Services
IBM's top clients

AutoML,
Deployment,
Reproducibility
Notebooks,
Pipelines, Serving,
Training, Scale

Kubeflow

...on top of Kubernetes



Elyra provides...

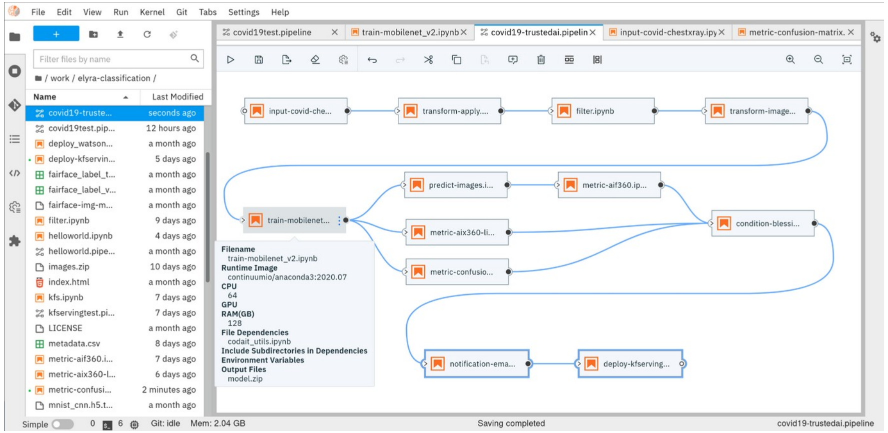
Some prominent users: Three IBM clients, one Fortune 500 company

No Code / Low Code ML Pipeline Design

Re-usable pipeline components

Interchangeability of Engines

(Kubeflow, Airfow, ...)



...on top of JupyerLab, VSCode, ...

CLAIMED...

Component **L**ibrary for ***AI***, **M**achine Learning, **E**TL and **D**ata Science

Current users:
University Hospital Basel
Motion.AI

Portability

No Code / Low Code

Pipeline Components

Jupyter Notebooks

Sample Pipelines

```
component-library
├── analyze
│   └── spark-ts-trends.ipynb
├── checkpoint
│   └── store_asset.ipynb
├── claimed_utils.py
├── CONTRIBUTING.md
├── deploy
│   ├── condition-blessing.ipynb
│   ├── deploy-kfserving.ipynb
│   ├── deploy_watson_machine_learning.ipynb
│   └── README.md
├── filter
│   ├── filter.ipynb
│   ├── README.md
│   └── spark-sample.ipynb
├── input
│   ├── claimed_utils (1).py
│   ├── claimed_utils.py
│   ├── defunct-ray-input-climate-copernicus.ipynb
│   ├── input-climate-copernicus.ipynb
│   ├── input-covid-chestxray.ipynb
│   ├── __pycache__
│   │   └── claimed_utils.cpython-38.pyc
│   └── README.md
├── LICENSE
├── metric
│   ├── metric-aif360.ipynb
│   ├── metric-aix360-lime.ipynb
│   ├── metric-confusion-matrix.ipynb
│   └── README.md
├── monitoring
│   ├── notification-email.ipynb
│   └── README.md
├── predict
│   ├── predict-images.ipynb
│   └── README.md
├── README.md
├── train
│   ├── README.md
│   └── train-mobilenet_v2.ipynb
├── transform
│   ├── README.md
│   ├── spark-csv-to-parquet.ipynb
│   ├── spark-parquet-to-csv.ipynb
│   ├── transform-apply.ipynb
│   └── transform-images.ipynb
└── visualize
    └── map-from-coordinates.ipynb
```

...on top of Elyra and Kubeflow

Other workflow managers..

Requirement	KF	AF	Slurm	SM	Qsub	HTCondor	Reana
Kubernetes Support	X	X		X		X	X
GPU support	X	X	X	X	X	X	X
Component Library	X						
Reproducibility	X	X		X		X	X
Data Lineage	X						X

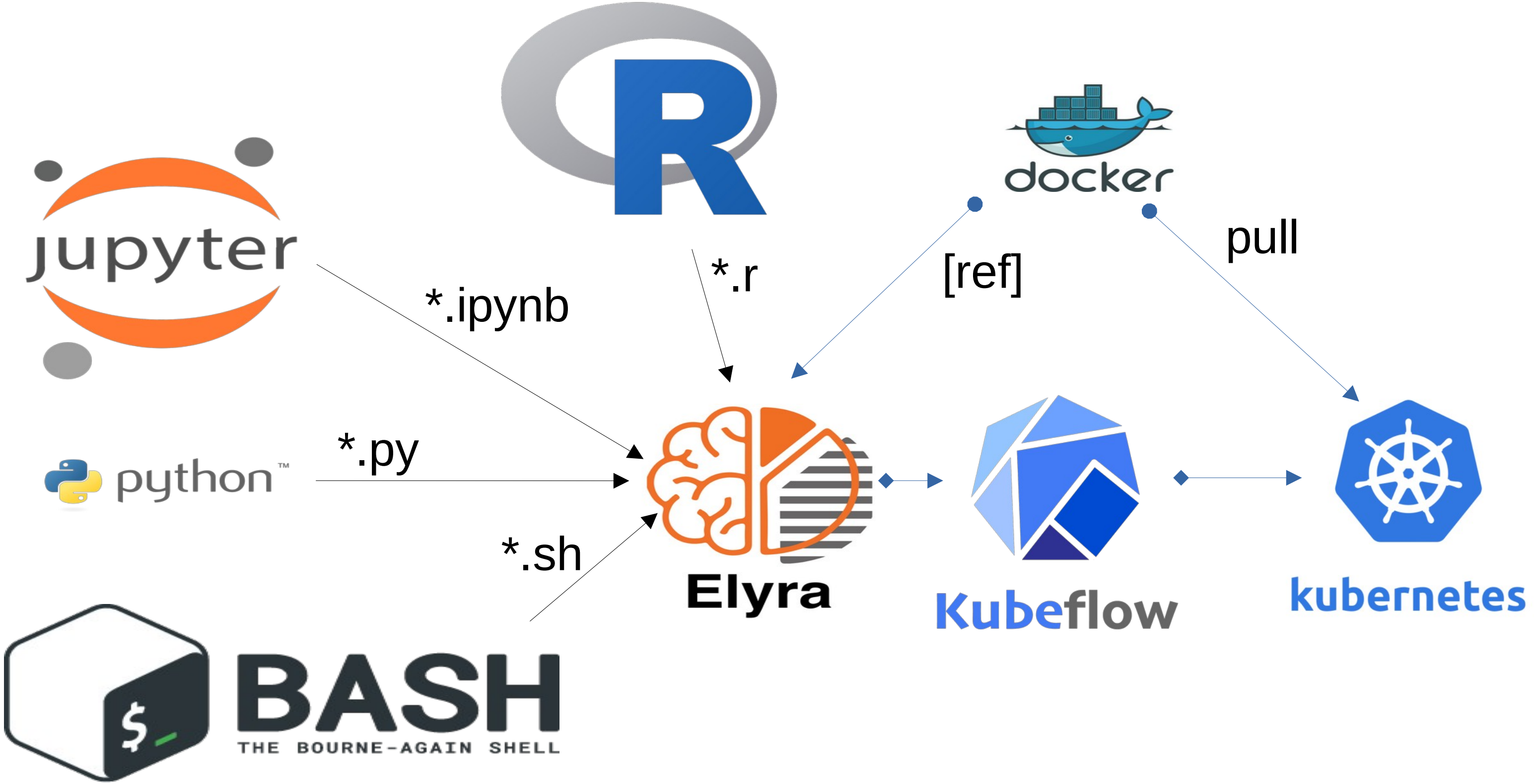
vs. Kubeflow

Other low code / no code tools..

Requirement	Nifi	NodeRED	KNIME	Galaxy	Elyra
Kubernetes Support				X	X
GPU support				X	X
Component Library	X	X	X	X	X
Reproducibility	X		X	X	X
Data Lineage	X			X	X
Visual Editing	X	X	X	X	X
Jupyter Notebooks					X

vs. Elyra

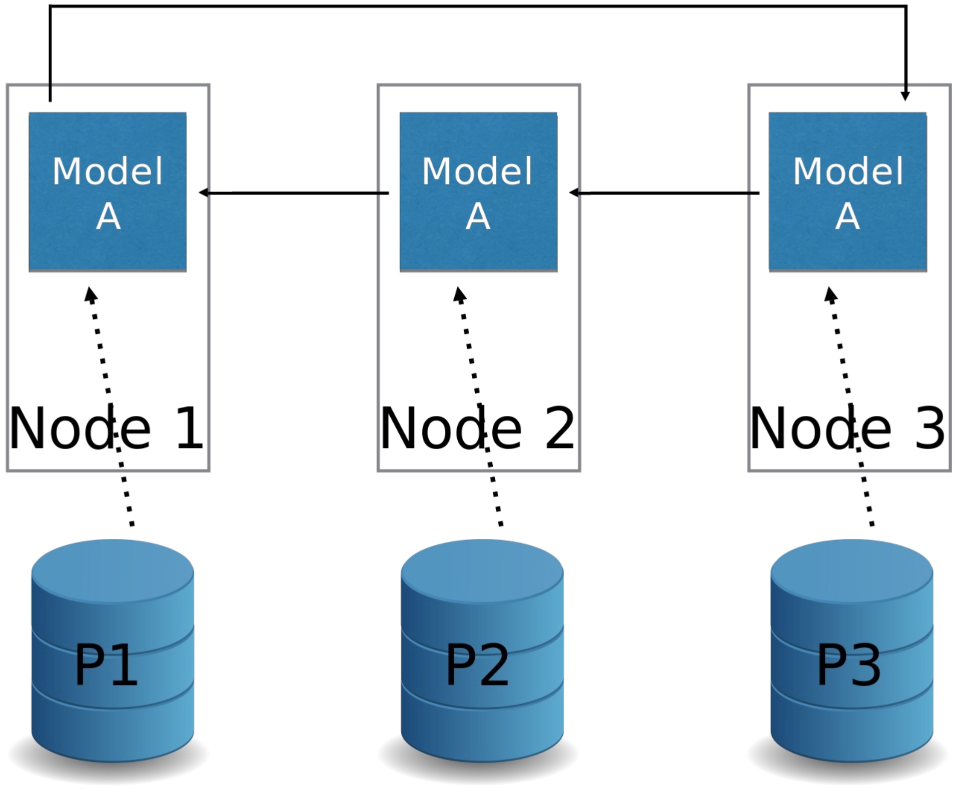
Full (source code) round-trip visual editing…



… and Kubeflow deployment using the Elyra Pipeline Editor

Example Pipeline Components

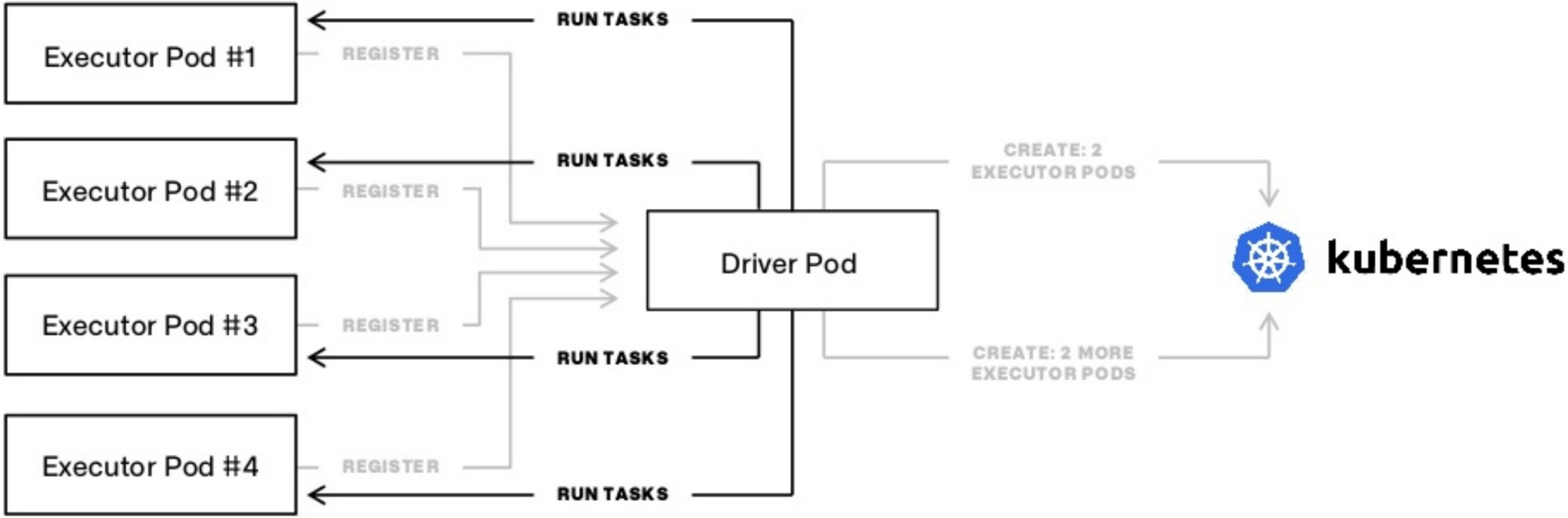
Category: Training **Group**: Distributed **Name**: TFJob



The TFJob operator supports parallel training on multiple nodes and GPUs

Example Pipeline Components

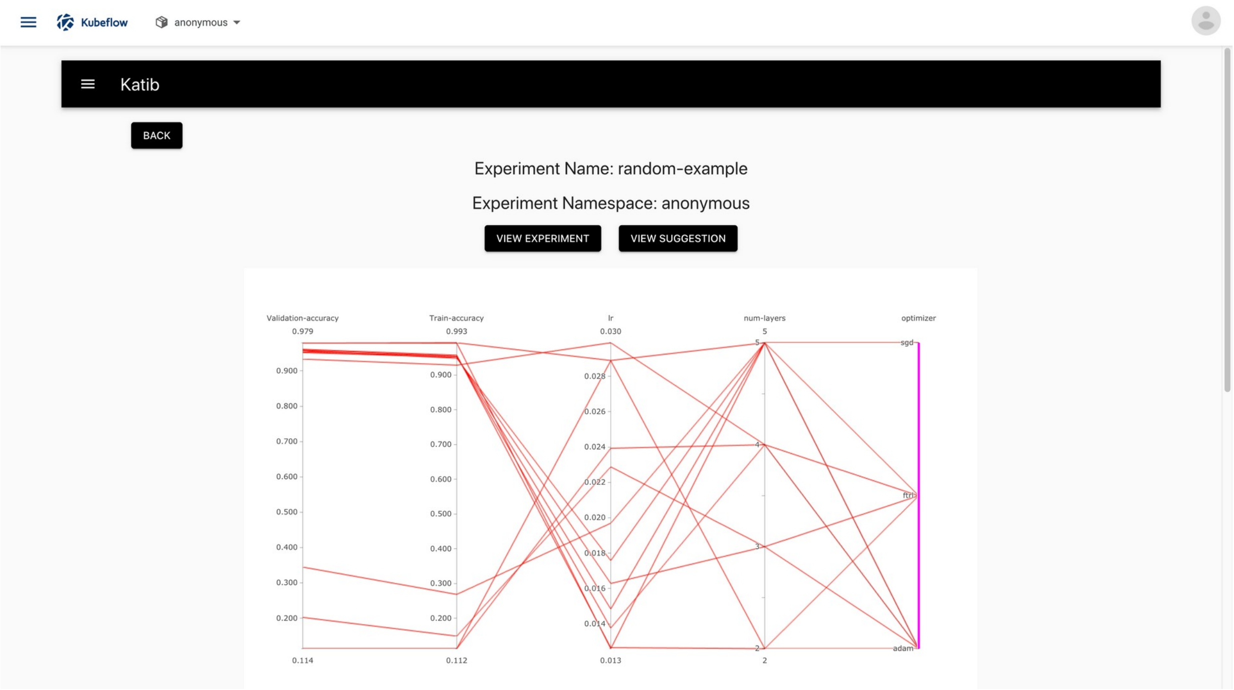
Category: Processing **Group**: Distributed **Name**: SparkJob



The SparkJob operator supports parallel processing on multiple nodes

Example Pipeline Components

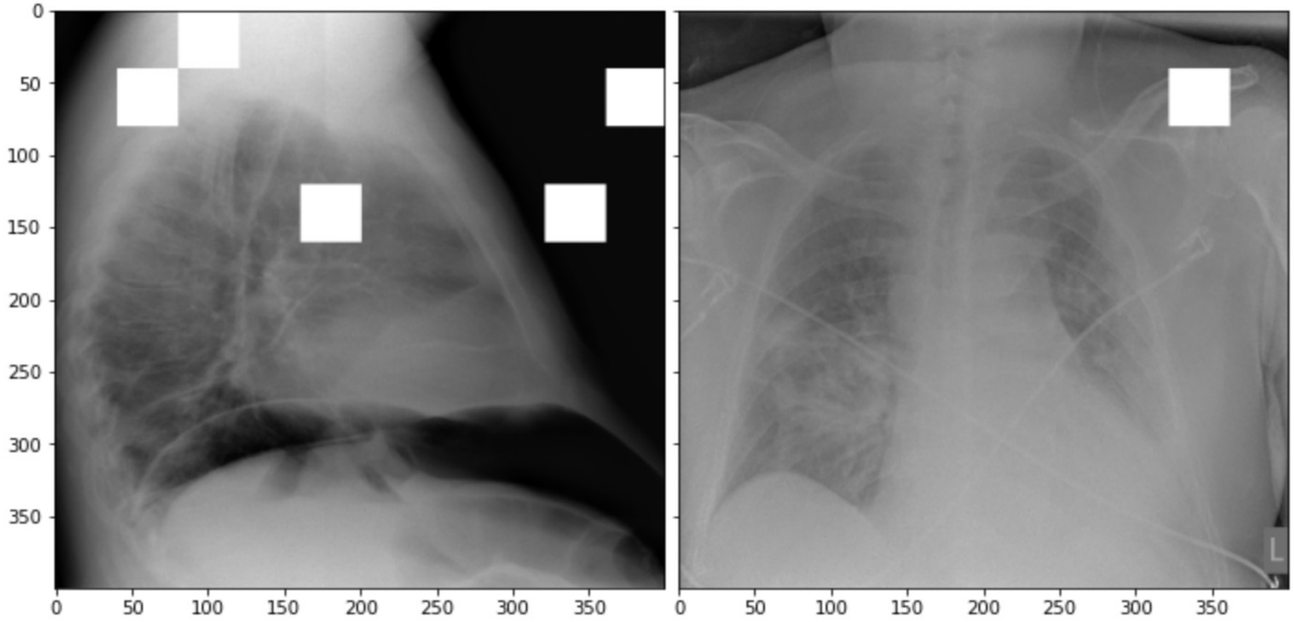
Category: Tuning Group: **Hyperopt Name:** Katib



Visualization of a hyper parameter optimization result

Example Pipeline Components

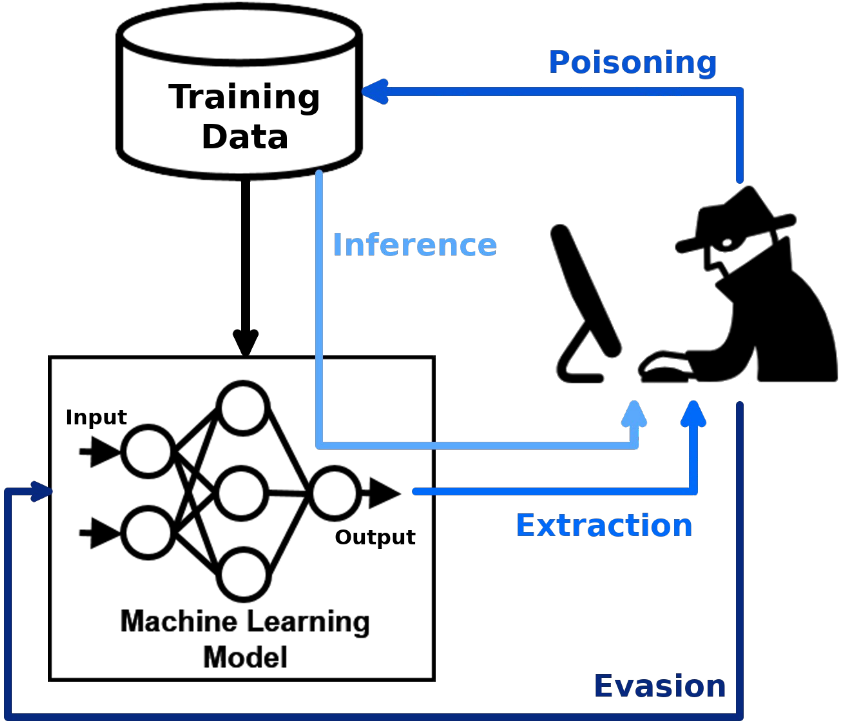
Category: Metric **Group:** Explainability Name: AIX360/LIME



Example on how LIME helps to identify classification relevant areas of an image

Example Pipeline Components

Category: Metric Group: **Adversarial Robustness Name:** ART



Example on how Adversarial Attacks happen

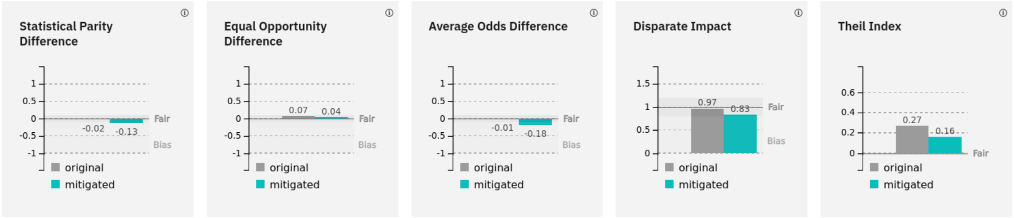
Example Pipeline Components

Category: Metric **Group:** AI Fairness Name: AIF360

Dataset: German credit scoring
Mitigation: **Adversarial Debiasing algorithm applied**

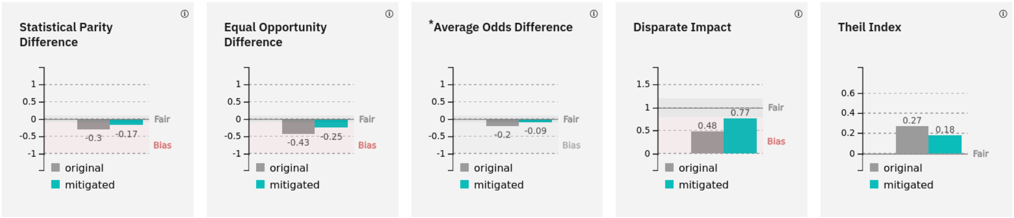
Protected Attribute: Sex

Privileged Group: ***Male***, Unprivileged Group: ***Female***
Accuracy after mitigation changed from 75% to 70%
Bias against unprivileged group unchanged after mitigation (0 of 5 metrics indicate bias)



Protected Attribute: Age

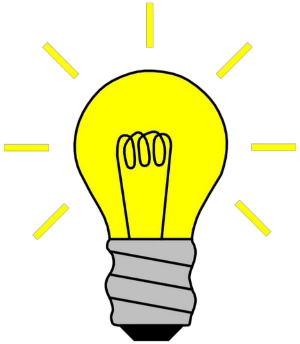
Privileged Group: ***Old***, Unprivileged Group: ***Young***
Accuracy after mitigation changed from 75% to 69%
Bias against unprivileged group was reduced to acceptable levels* for 1 of 4 previously biased metrics (3 of 5 metrics still indicate bias for unprivileged group)



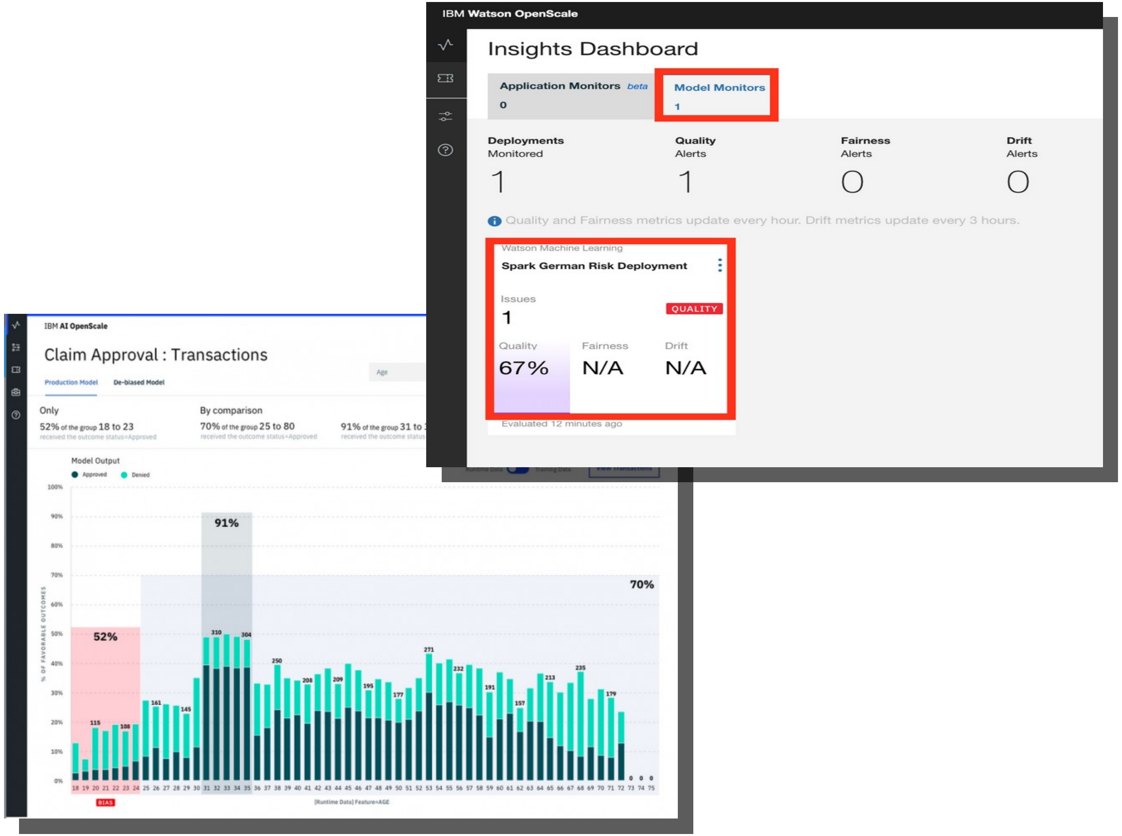
Example on how the AIF360 toolkit computes fairness metrics and mitigates bias

IBM Watson OpenScale

IBM Watson OpenScale uses the same Open Source components on top of Kubernetes, Kubeflow and KFServing

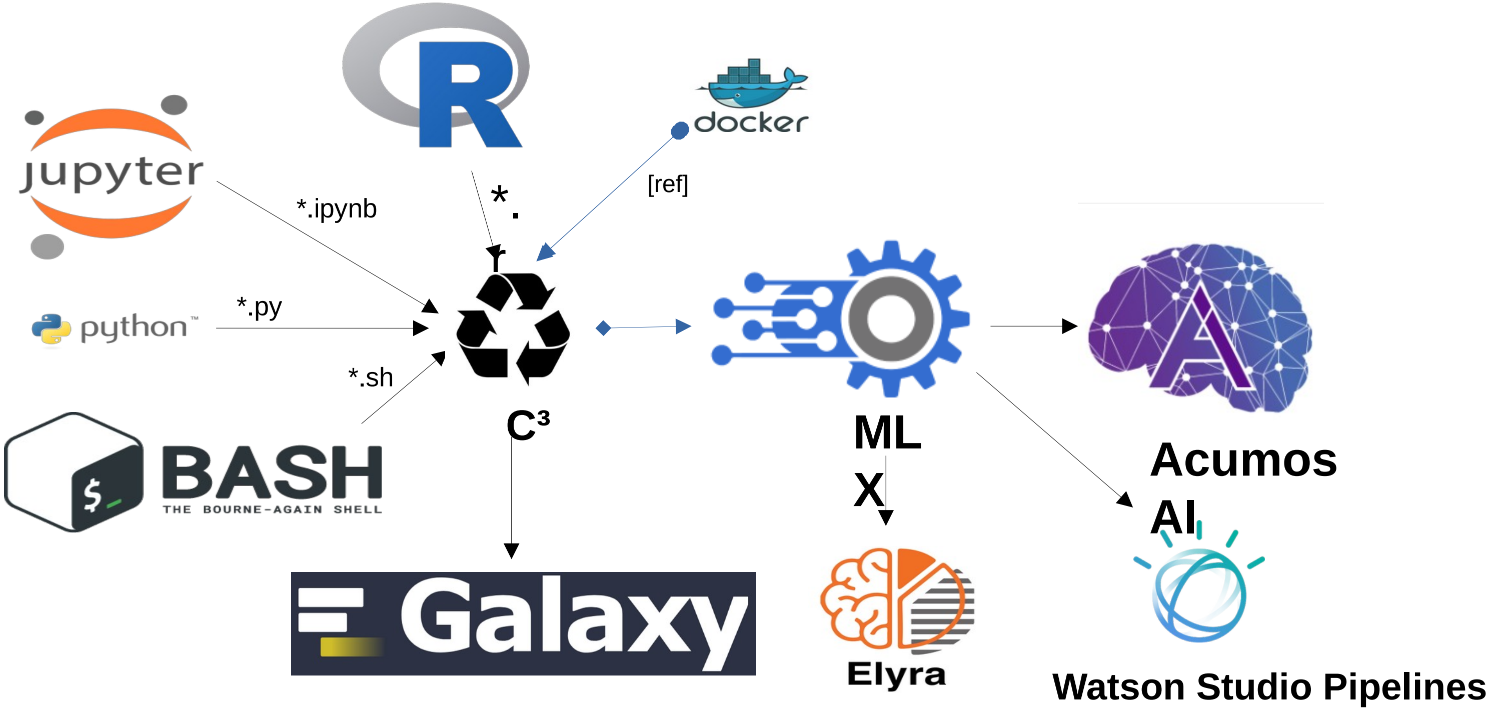


Did you know?



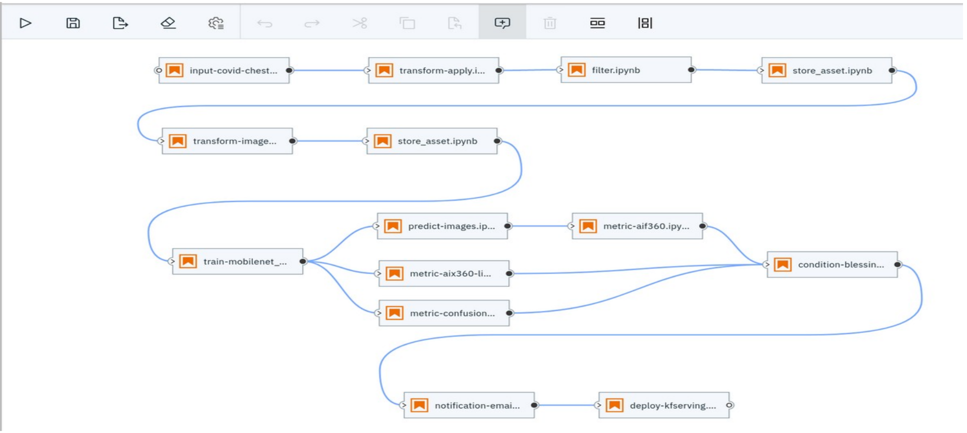
Model Monitoring and visualization of bias (IBM Watson OpenScale)

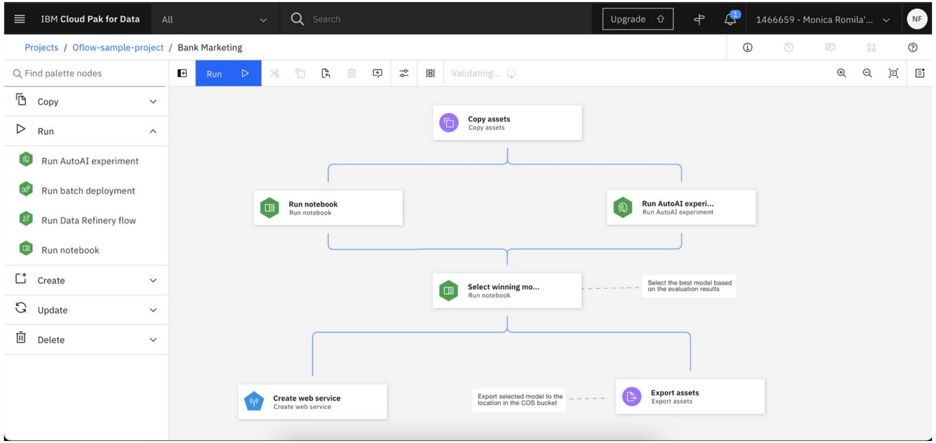
Component creation and cataloging...



...using Elyra and ML Exchange (MLX)

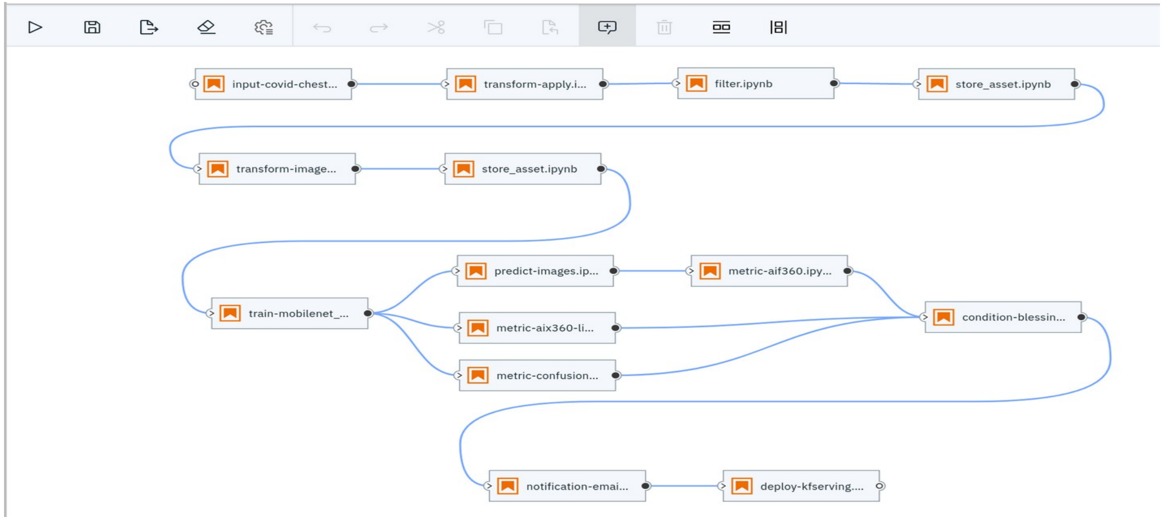
Component consumption…

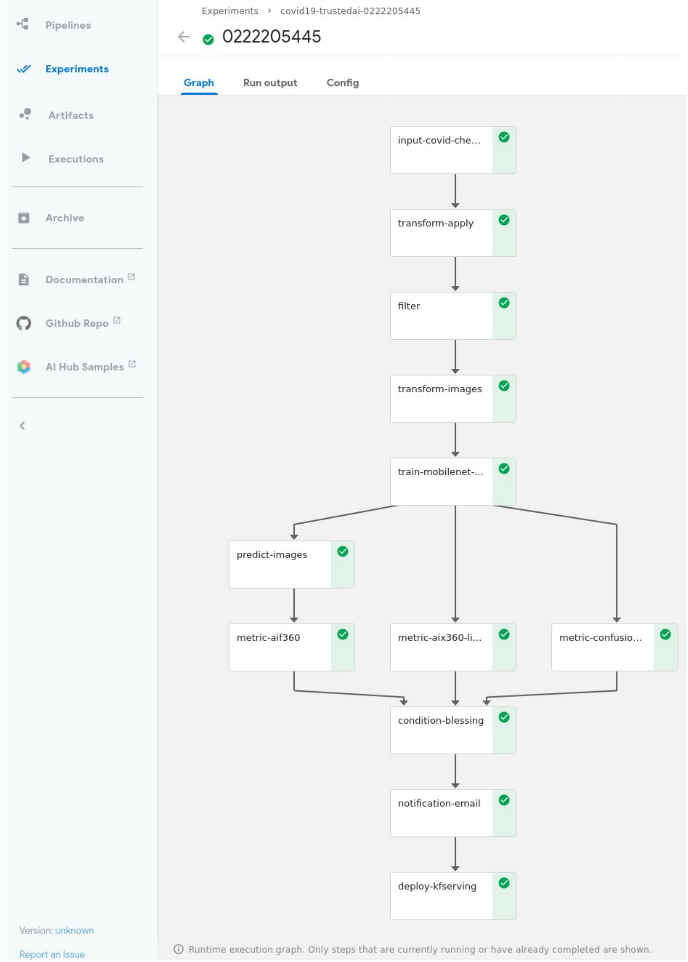




…using Elyra or Watson Studio Pipelines

The final pipeline in Elyra..





...and KubeFlow

Summary

- Rapid prototyping using visual editing and notebooks ✓
- Seamless scaling during development and deployment ✓
- GPU support ✓
- ML tools: PyData stack, TensorFlow, PyTorch, … ✓
- life science tools: DICOM input, DICOM output, … ✓
- Reproducibility ✓
- Data lineage ✓
- Reference implementation in open source ✓
- Collaboration support ✓