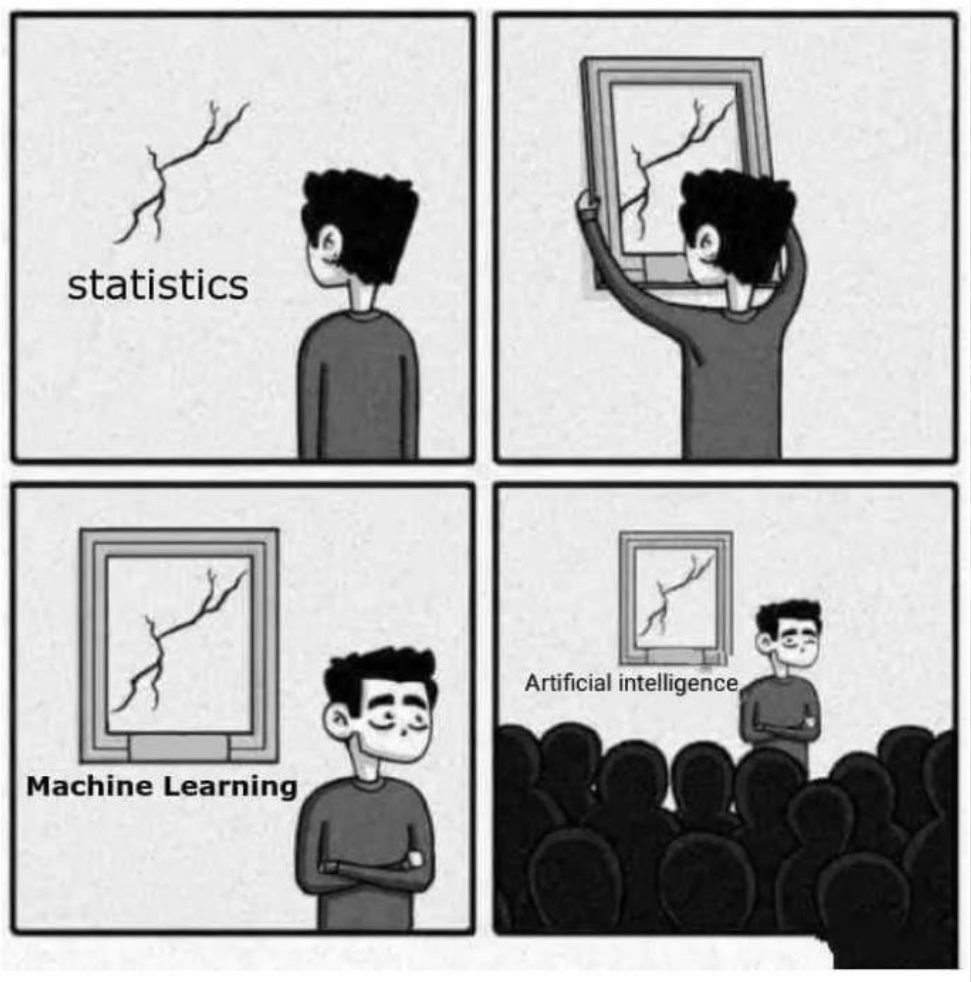
Productionising Machine Learning Models

PyData: Johannesburg
Feb 2024

Shaun De Ponte

Some Light Humour

No, Machine Learning is not just glorified Statistics



@sandserifcomics

Shaun De Ponte

Education

MSc in Business Analytics (Aston University)

Current Work

Senior Data Engineer at Calybre

Other Projects and Interests

Open-Source Software Engineer
Building and Innovating Things

in shaundeponte

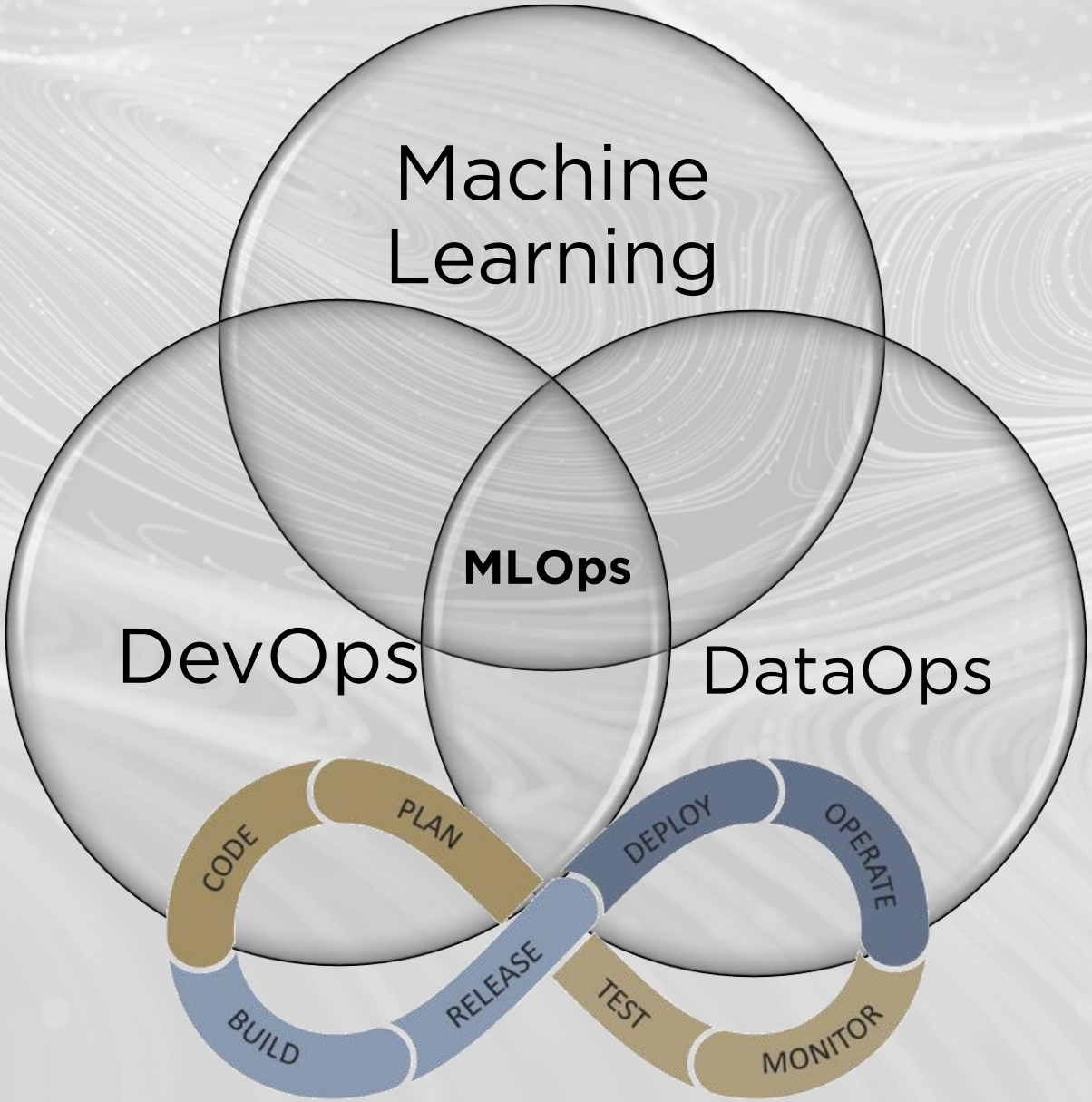
nawtybean

Agenda

- Introduction to Machine Learning Operations
- Example Architecture for Machine Learning in Production
- Quick Introduction to Kedro
- Quick Demo using Kedro and MLFlow
- Q and A

Introduction to Machine Learning Operations

Introduction to Machine Learning Operations

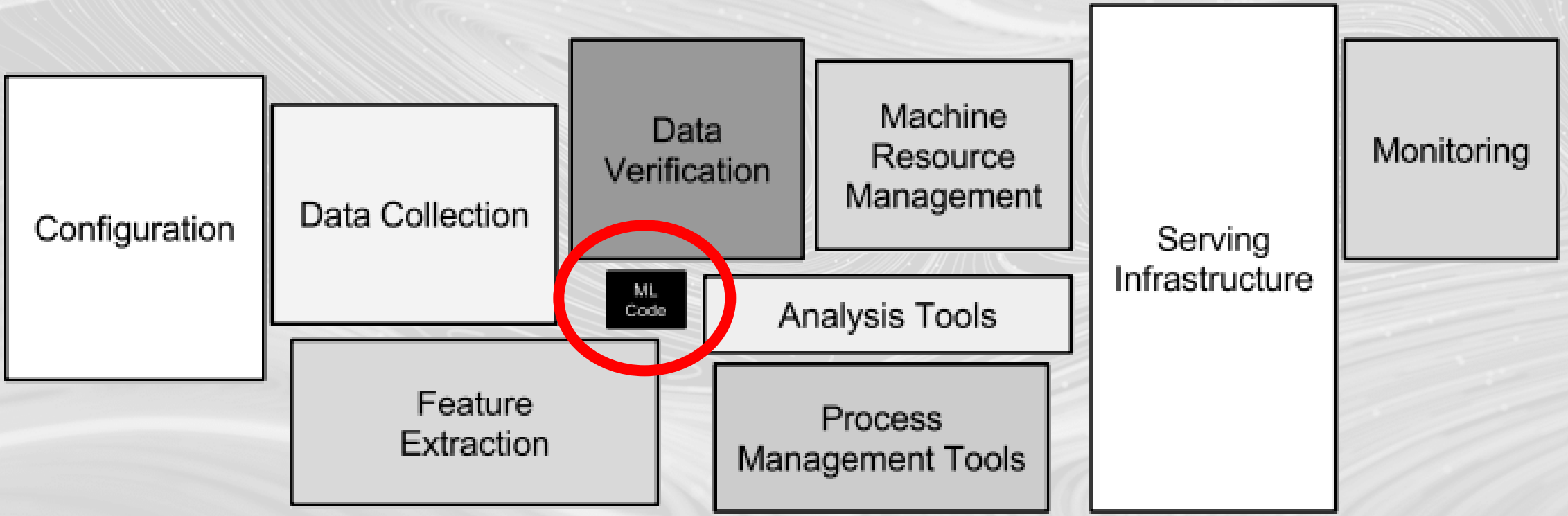


Introduction to Machine Learning Operations

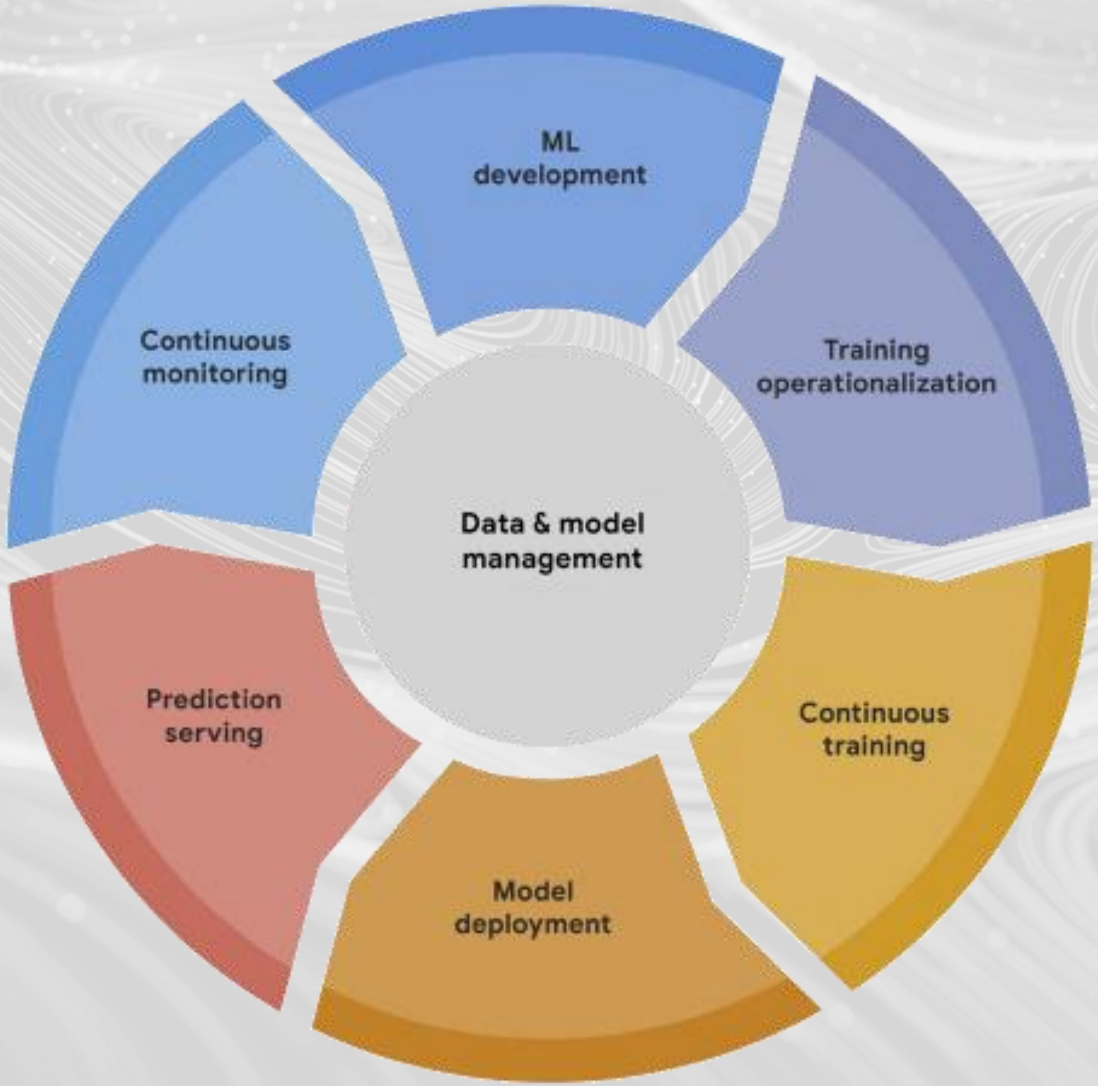
“90% of Machine Learning Models Never Make it to Production” – redapt.com

According to McKinsey up to 85% of Machine Learning Models Fail

Introduction to Machine Learning Operations



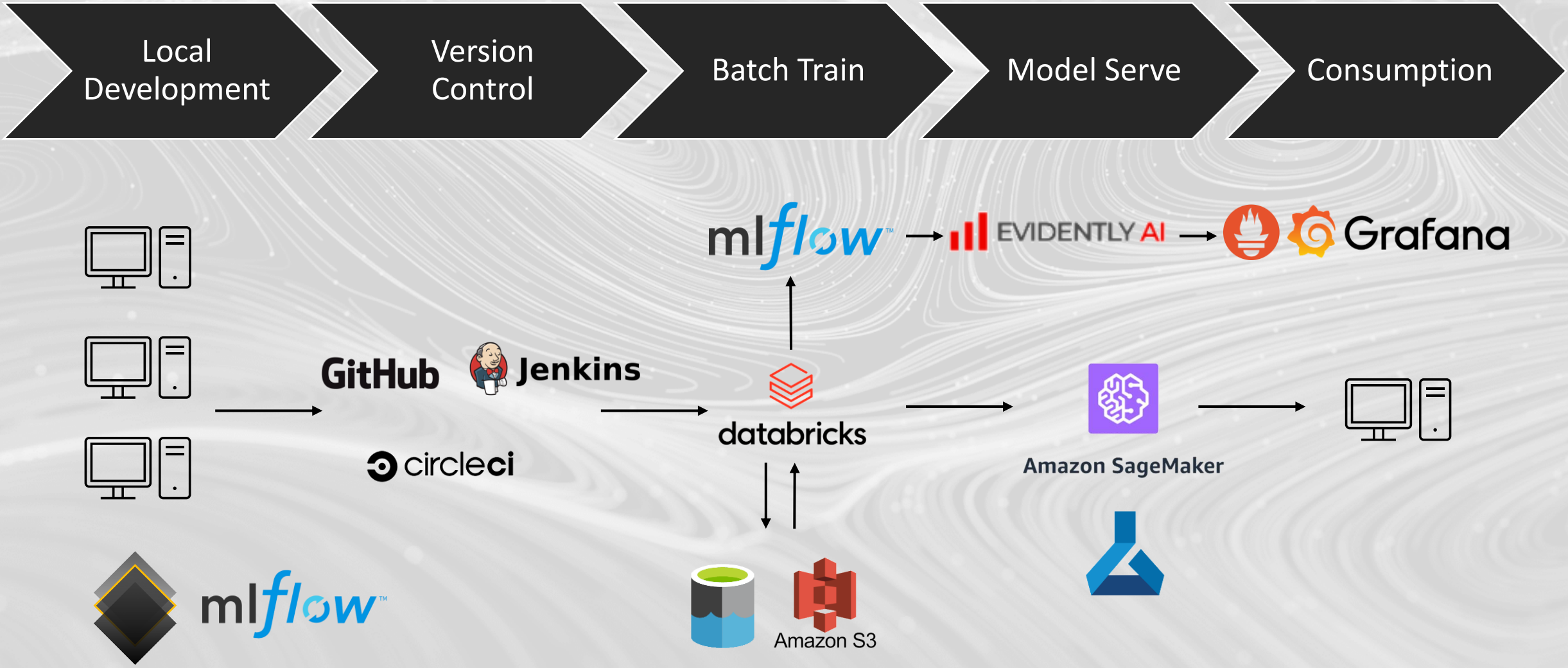
Introduction to Machine Learning Operations



Practitioners Guide To MLOps Whitepaper– Google Inc, 2021

Example Machine Learning Production Architecture

Example Machine Learning Production Architecture



Copyright © QuantumBlack, 2019

An Introduction to Kedro

Introduction to Kedro

- **Engineering Machine Learning**
  The basis for clean data science code is Kedro. It uses ideas from software engineering to machine-learning initiatives.

- **Adapts to Complexity**
  A Kedro project offers machine-learning and complicated data pipeline scaffolding. You focus on solving new challenges and spend less time on tedious "plumbing".

- **Uniformity**
  Kedro guarantees that teams work together to solve problems quickly and standardizes the creation of data science code.

- **Ready for Production**
  With exploratory code, you can easily move from development to production and create modular, repeatable, and maintainable experiments.

DataCatalog is the registry of all data sources that the project can use.

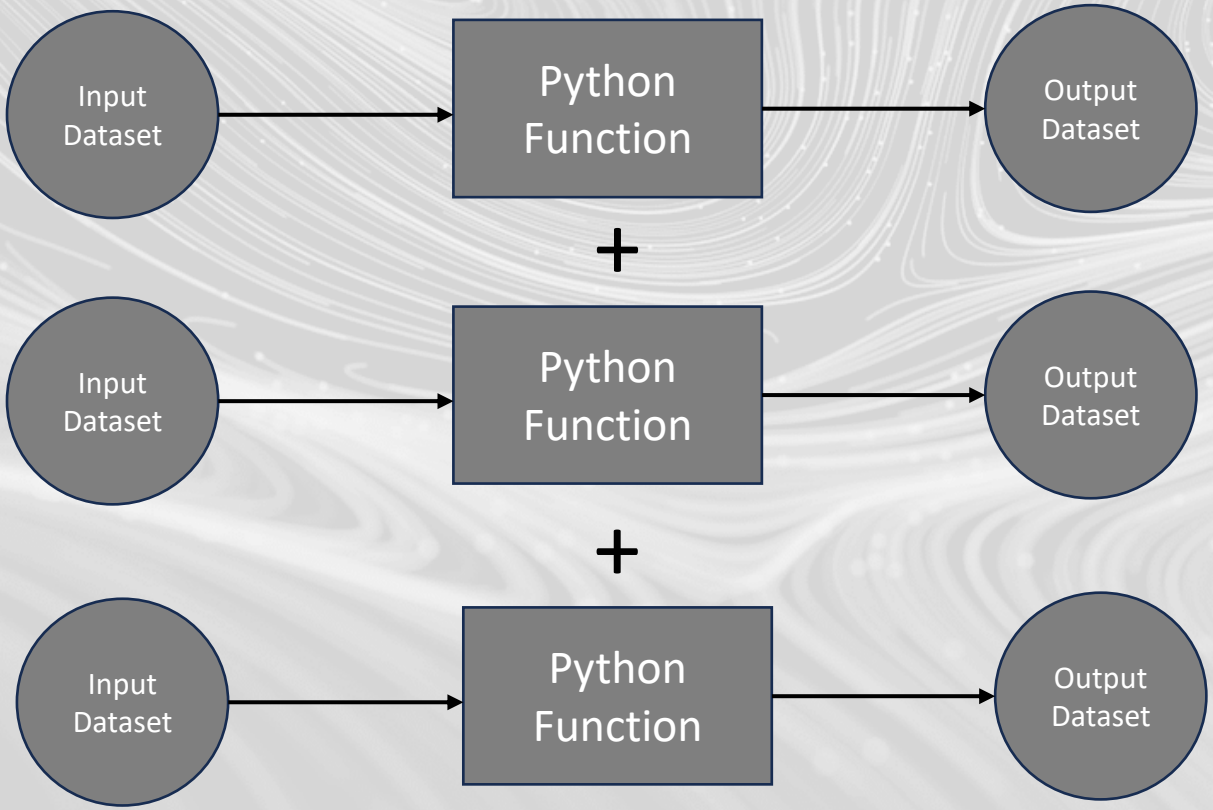
Kedro Node

Node in Kedro is a python function wrapper that names the inputs and outputs of that function.

We can link one *Node* to another by setting the output of one node as the input to another node.

Kedro Pipeline

A pipeline organizes the dependencies and execution order of your collection of nodes and connects inputs and outputs while keeping your code modular.



Kedro Visualizations

Kedro Viz visualises the pipelines in a Kedro project by showing data, nodes, and the connections between them.

Kedro Supports The Tools You Love

https://docs.kedro.org/en/latest/extend_kedro/plugins.html

Demo

Questions and Answers Time!

References

- Google MLOps White Paper
  (practitioners_guide_to_mlops_whitepaper.pdf (google.com))

- Hidden Technical Debt in Machine Learning System
  (Hidden Technical Debt in Machine Learning Systems (neurips.cc))

- Kedro
  (Kedro | An open-source framework for data science code)

- DataEngineerOne
  ((129) Introduction to Data Pipelines and Kedro - Writing Data Pipelines With Kedro 1 - YouTube)

- Kedro-MLFlow
  (Goal of the tutorial — kedro-mlflow 0.12.1 documentation)