

The 4P's of Successful Data Science

People, Patterns, Process, Products

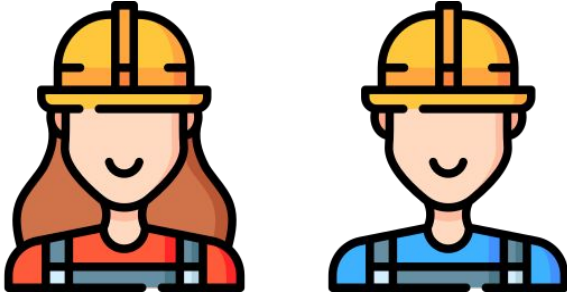
Andy McMahon

ML Engineering Lead

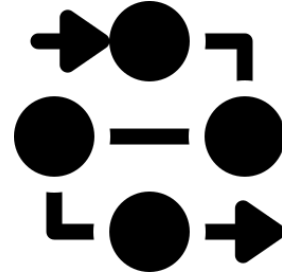
NatWest Group, Edinburgh, Scotland

RSS Conference, 2021

The 4 P's



People



Process



Product

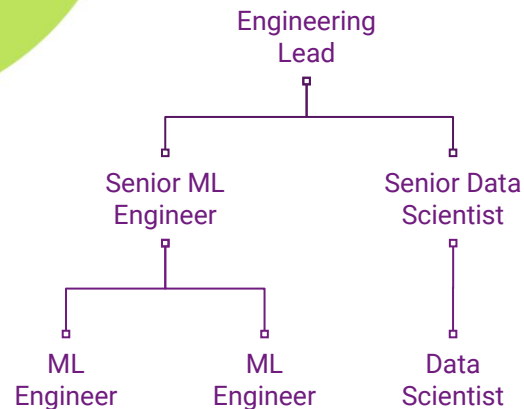
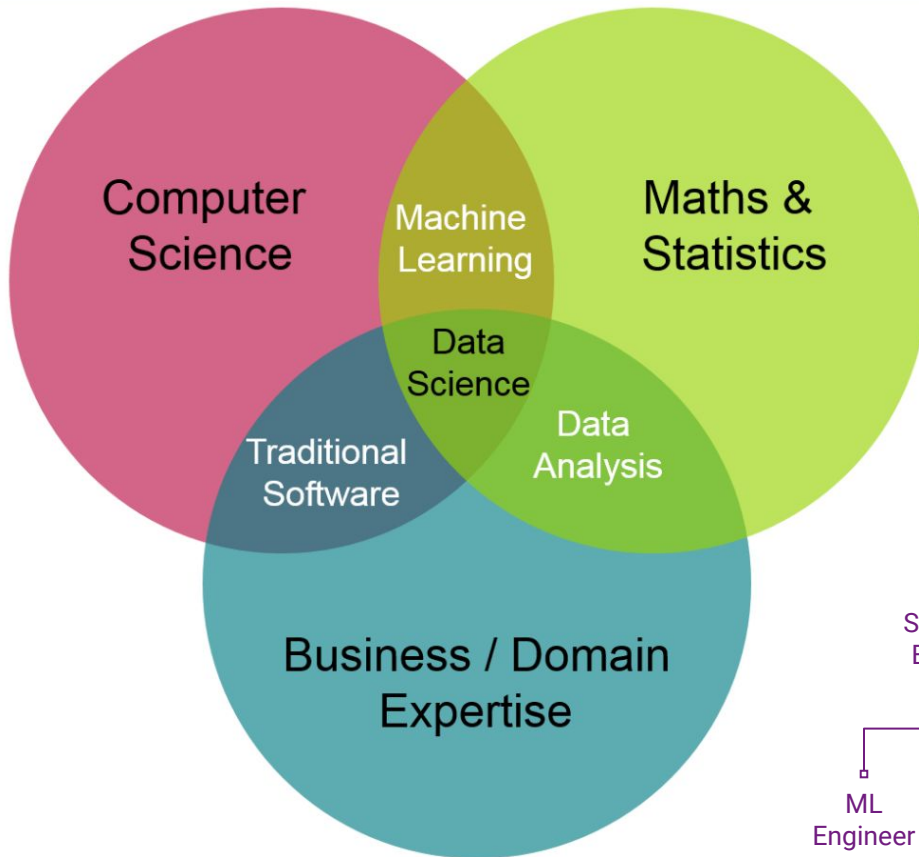
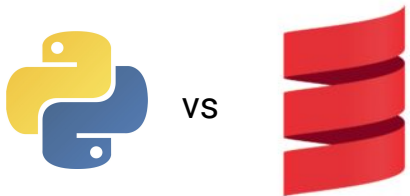


Pattern



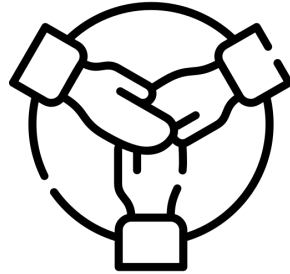
People

Blended Teams > Unicorns



The Vision Should Be ...

Focussed



Relevant

Exciting

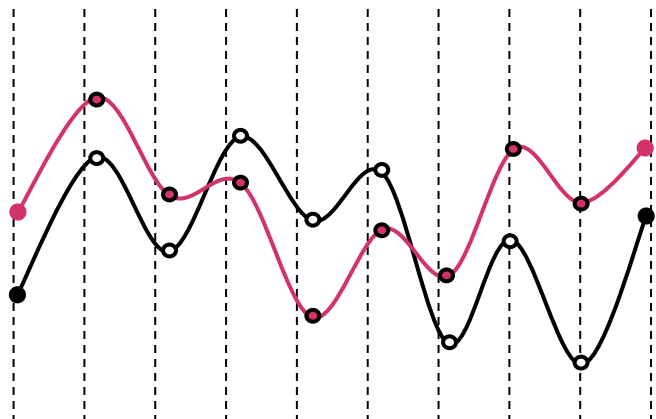


Know Your Customer

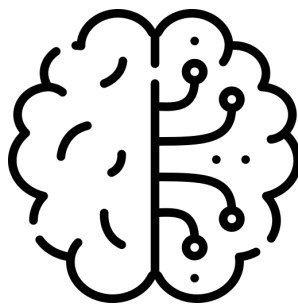
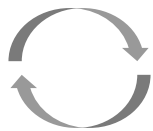


Patterns

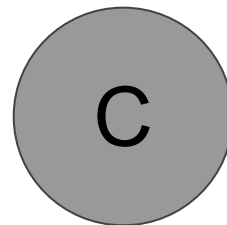
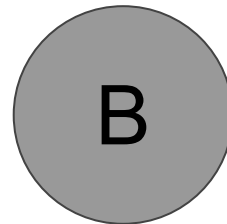
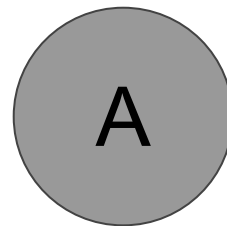
Decisions, decisions ...



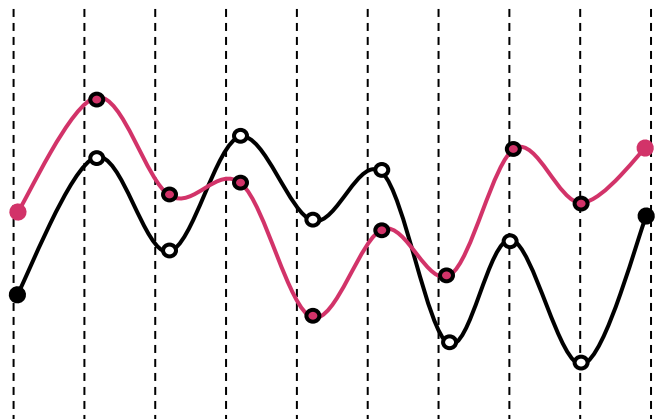
DATA



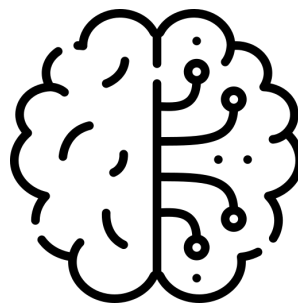
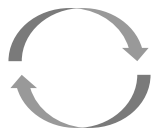
MODEL



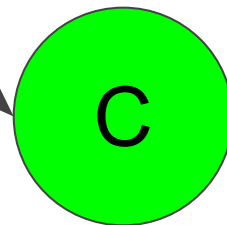
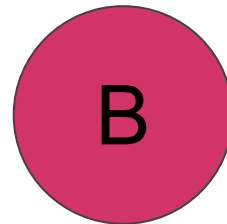
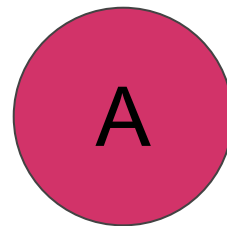
Decisions, decisions ...



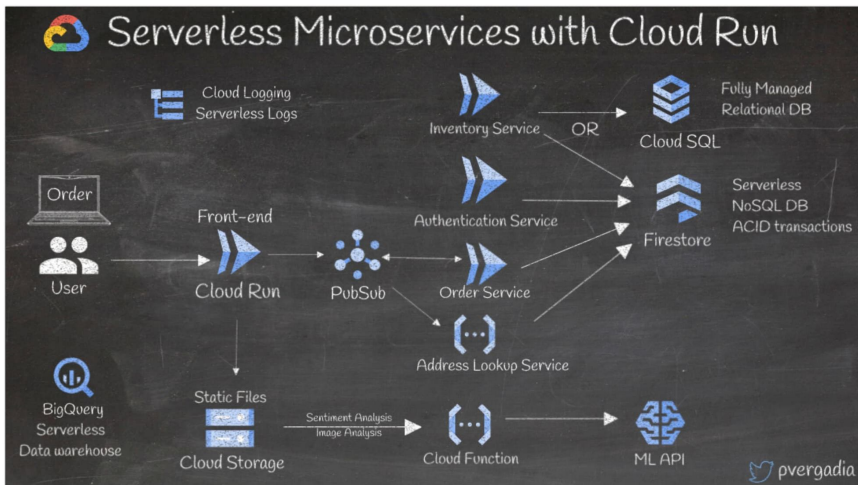
DATA



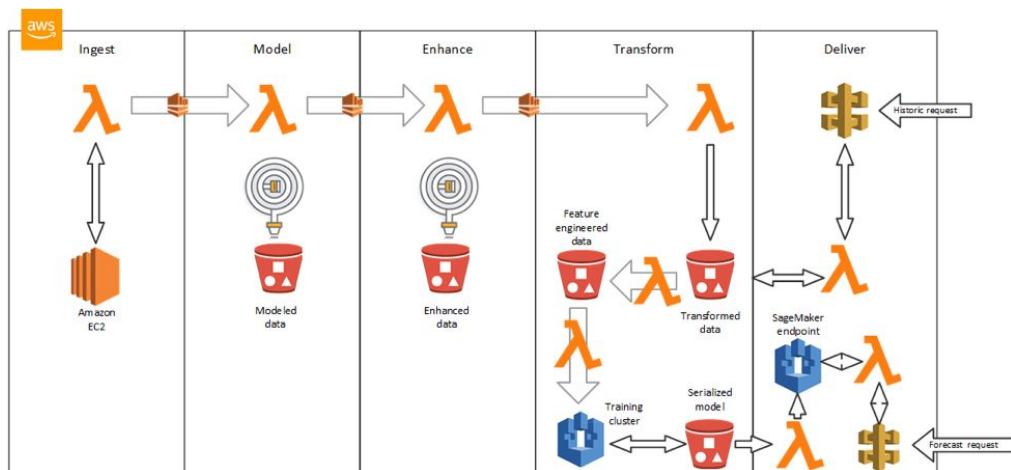
MODEL



Reuse, Repeat, Recycle



13 sample architectures to kickstart your Google Cloud journey



Machine Learning Lens AWS Well - Architected Framework

Science != Engineering





Process

ML Dev Process

Discovery

PoC

Development

Deployment

- Work with subject matter experts and stakeholders
- Understand challenge
- Define end goal (KPIs, baseline performance)
- Sketch MVP implementation
- Summarise for sign off



ML Dev Process

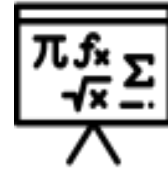
Discovery

PoC

Development

Deployment

- Exploratory data analysis
- Identification of 'no go' scenarios (blockers for the project)
- Identification of other project risks
- Bare bones implementation
- Evidence that the solution is viable and will solve the problem (sign off again)



databricks

ML Dev Process

Discovery

PoC

Development

Deployment

- Dev sprints
- Git strategy
- Unit and Integration tests
- Model version control
- Baseline performance tests/results



PyTorch



databricks

Spark
MLlib

pytest

mlflow



git



TensorFlow

ML Dev Process

Discovery

PoC

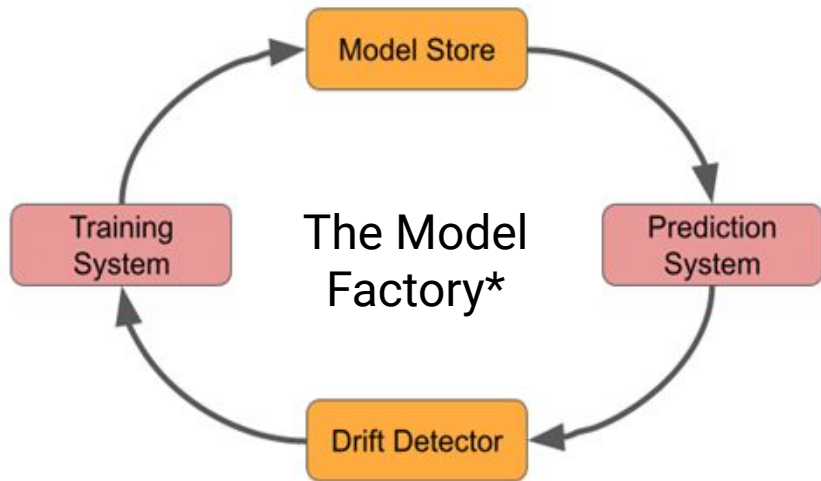
Development

Deployment

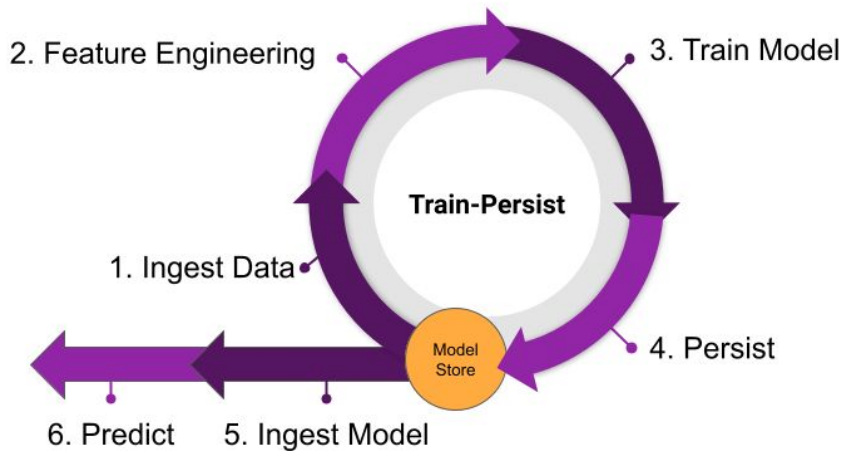
- Final packaging
- Deployment and update strategy
- Infrastructure instantiation
- Documentation completion
- Performance monitoring
- CI/CD pipeline definition



MLEng → MLOps

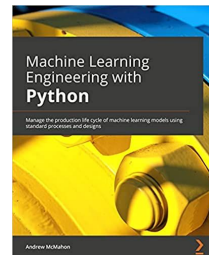


The Train-Persist Process*



*ML Engineering In Python, Andrew McMahon (Packt Publishing), 2021

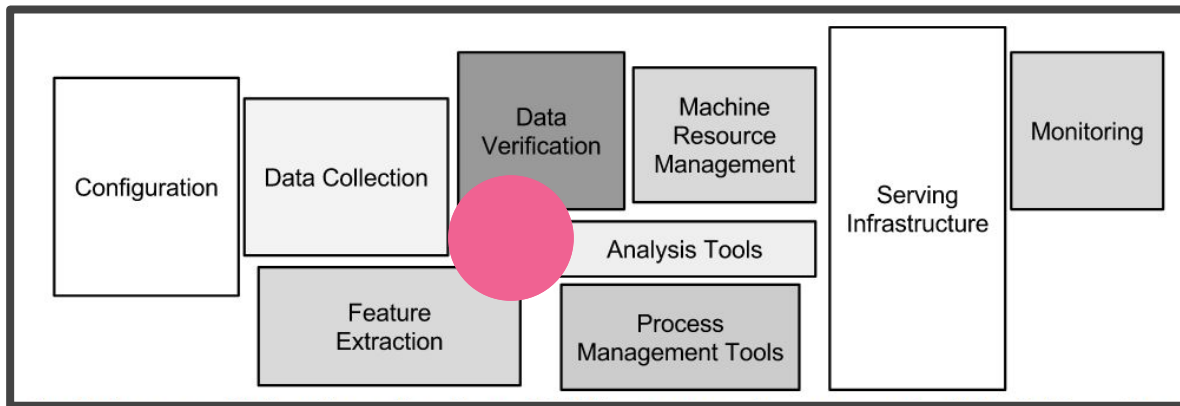
<https://www.amazon.co.uk/Machine-Learning-Engineering-Python-production/dp/1801079250>





Product

Products != Models != Analyses



ML Engineering

[Hidden Technical Debt in Machine Learning Systems, Google](#)

Products work: Test, test, test ...

Products get delivered: Don't reinvent the wheel ...

Products come in a variety of shapes: Build for flexibility ...

Products != Models != Analyses





Thank you!

Where can you find me?



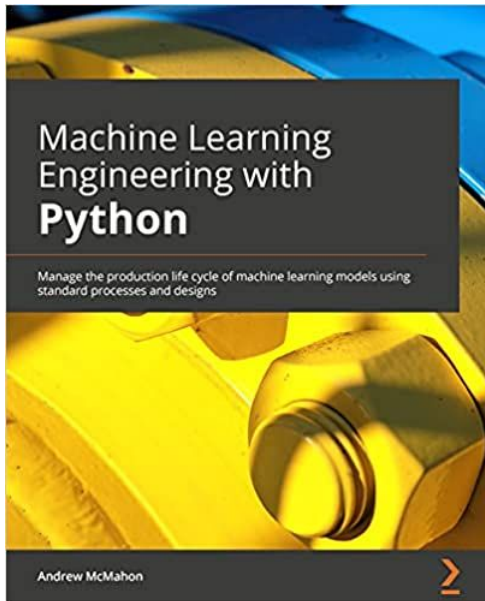
@electricweegie



AndyMc629



Andy McMahon



electricweegie.com



Appendix