

MLOps on AWS: a hands-on tutorial

Workshop · March 26, 2022
In-person event · STCC,
Lausanne, Switzerland

Workshop Organizers





<https://github.com/xtreamsrl/amld22-mlops-on-aws>

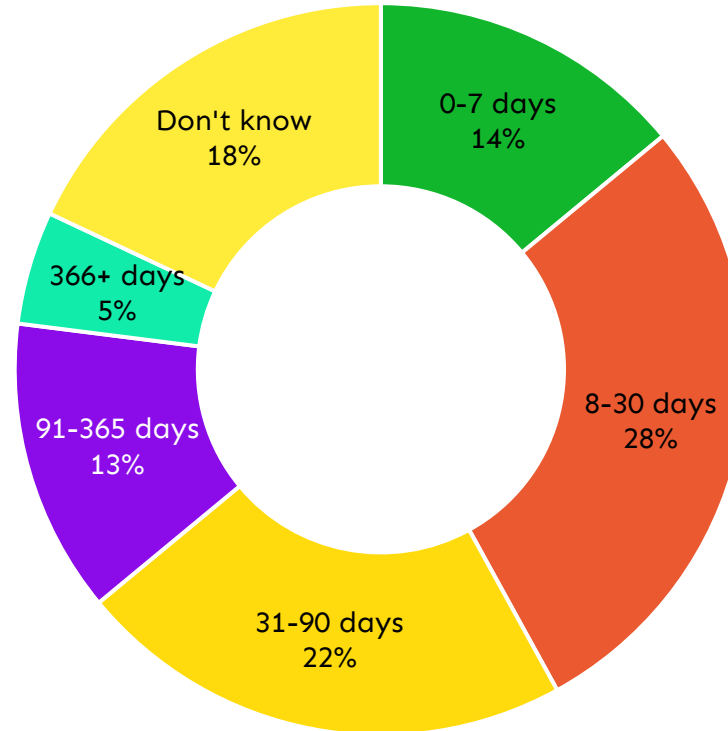




**Why
MLOps?**

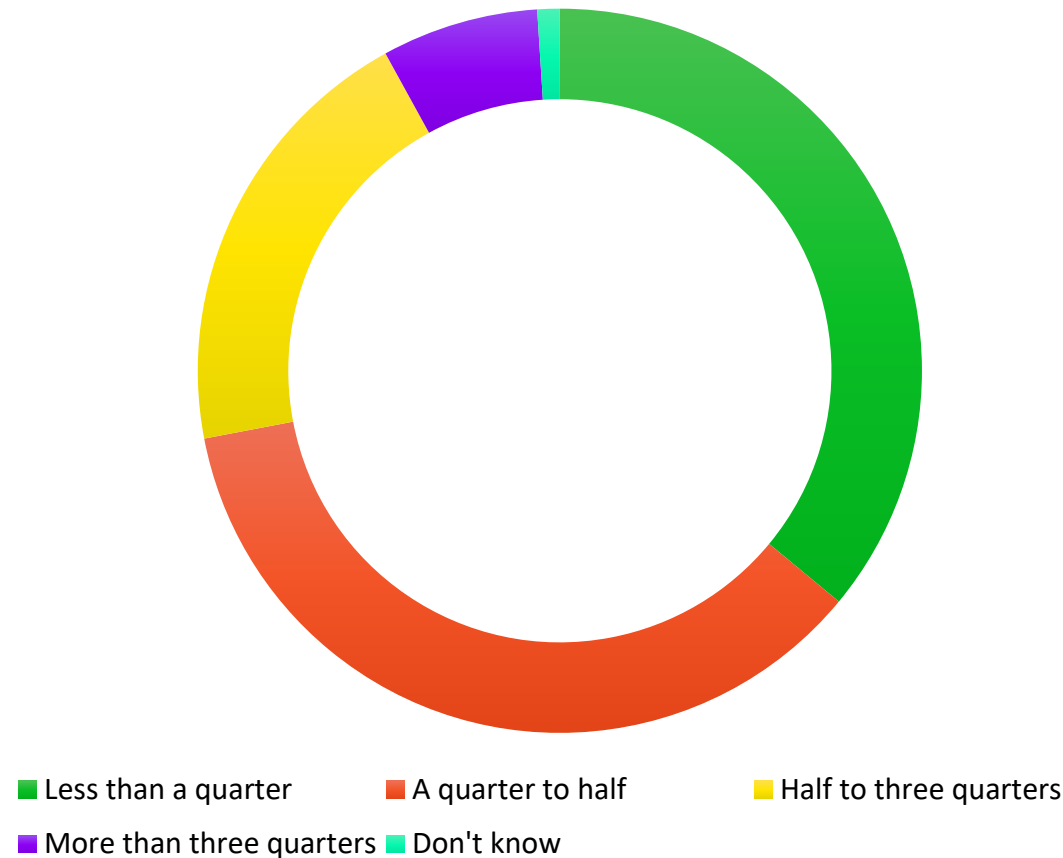
ML Development is slow

Time to ship a new model to production

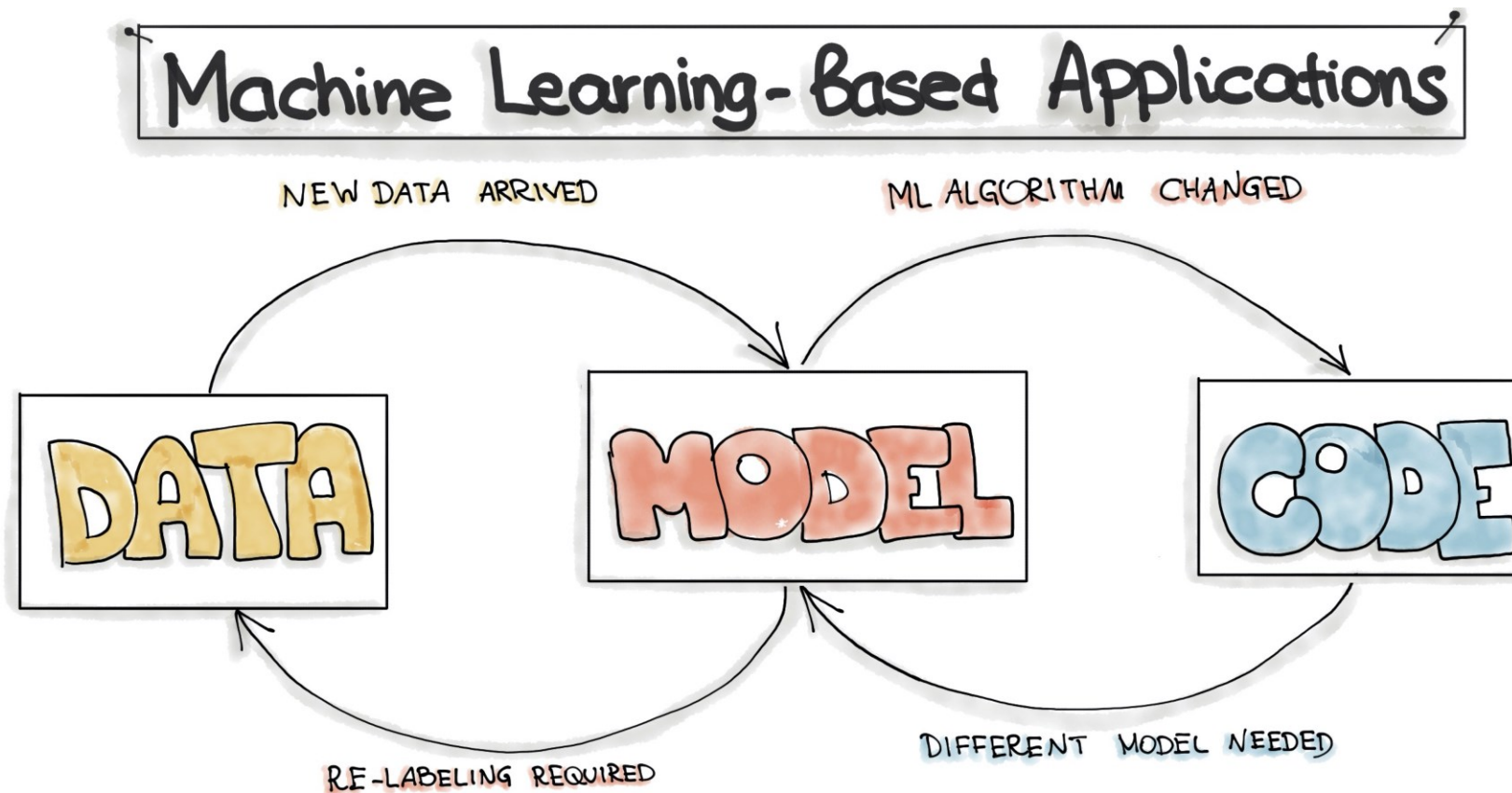


ML Deployment is time-consuming

Share of time spent by data scientists on model deployment



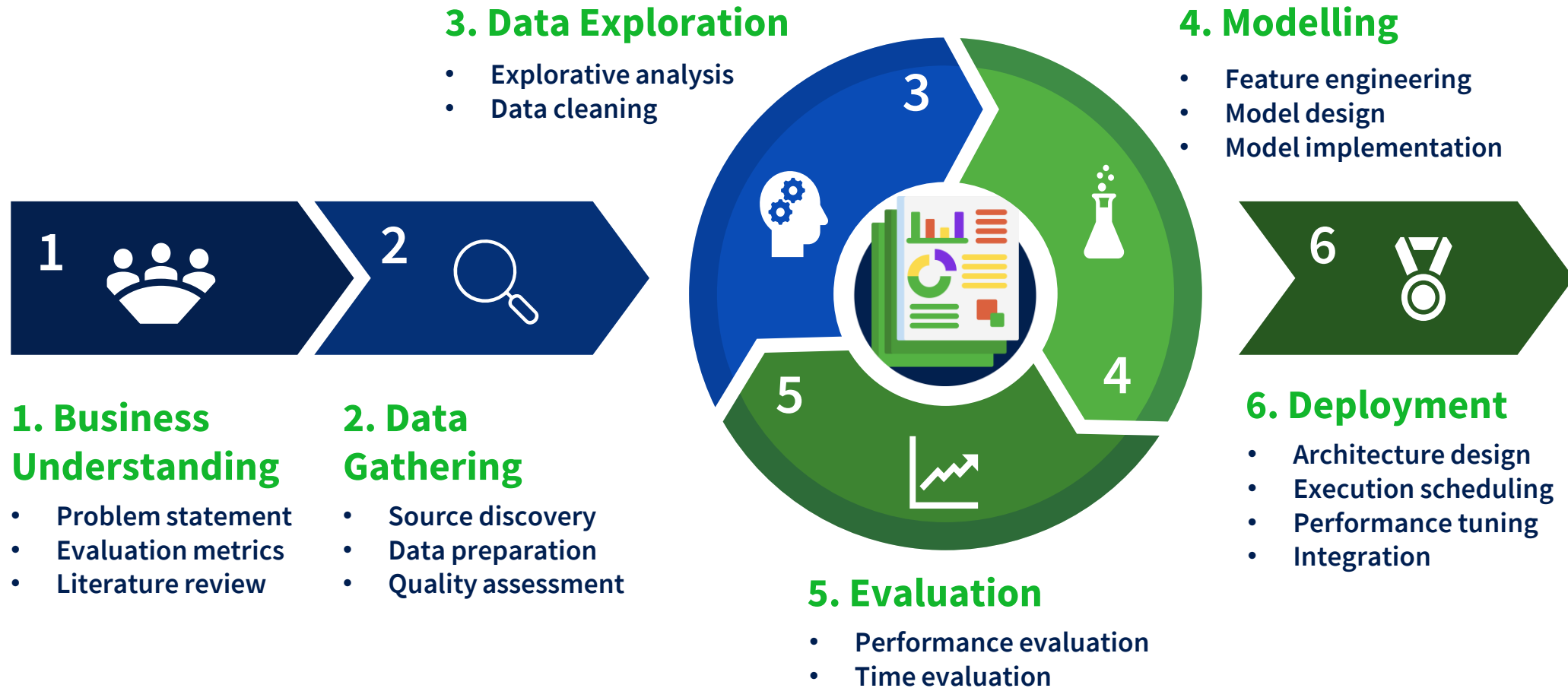
Updates cascade



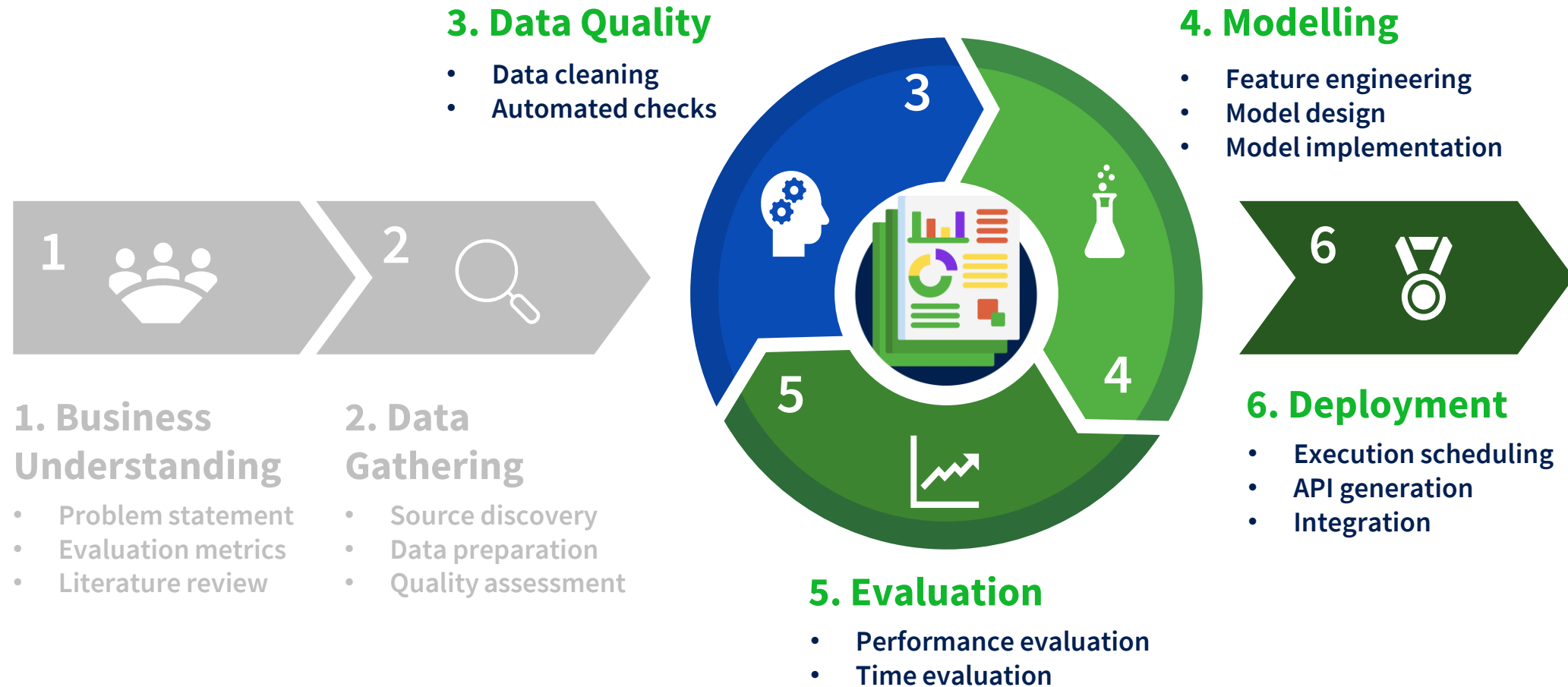


**What is
MLOps?**

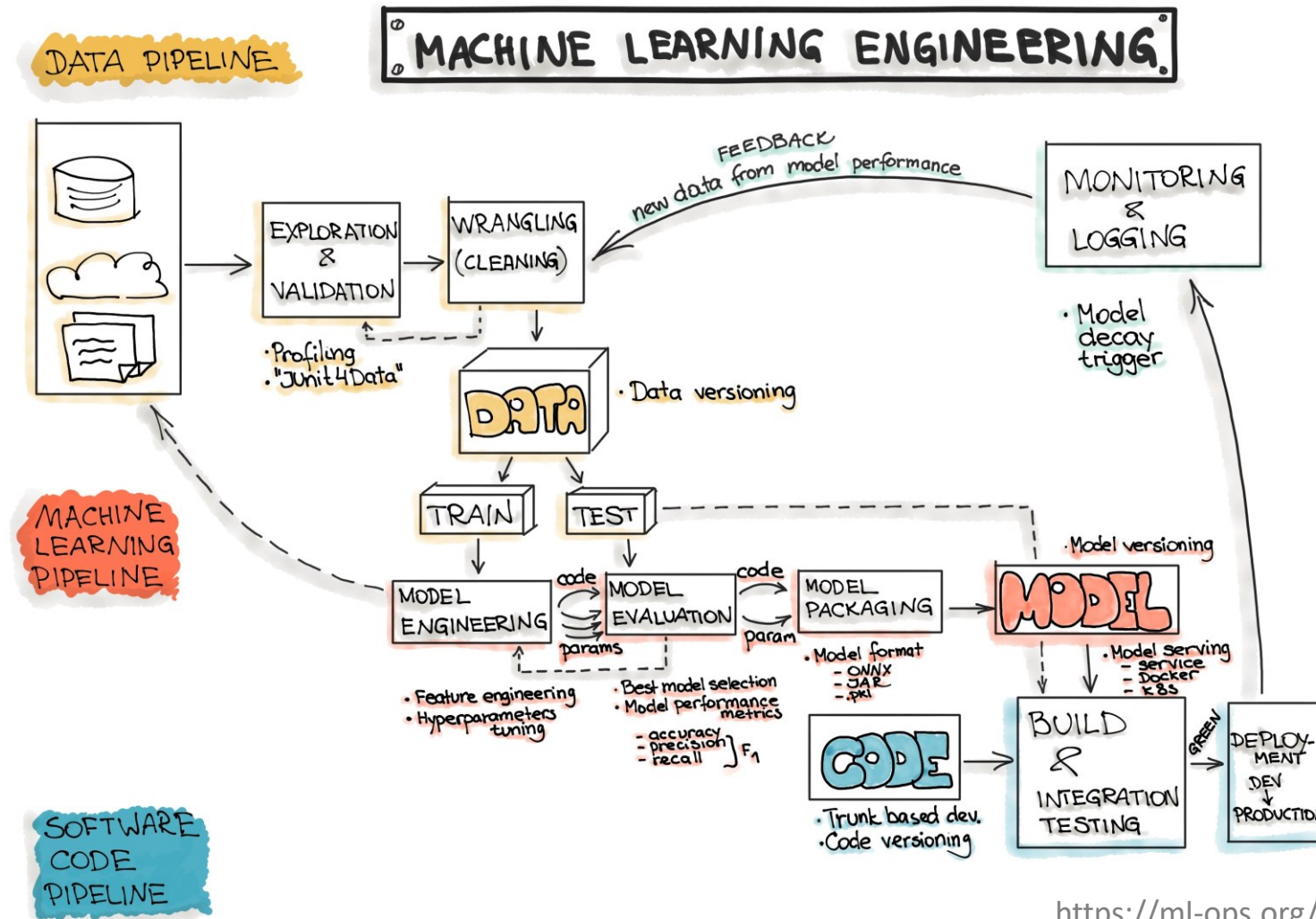
The Machine Learning process



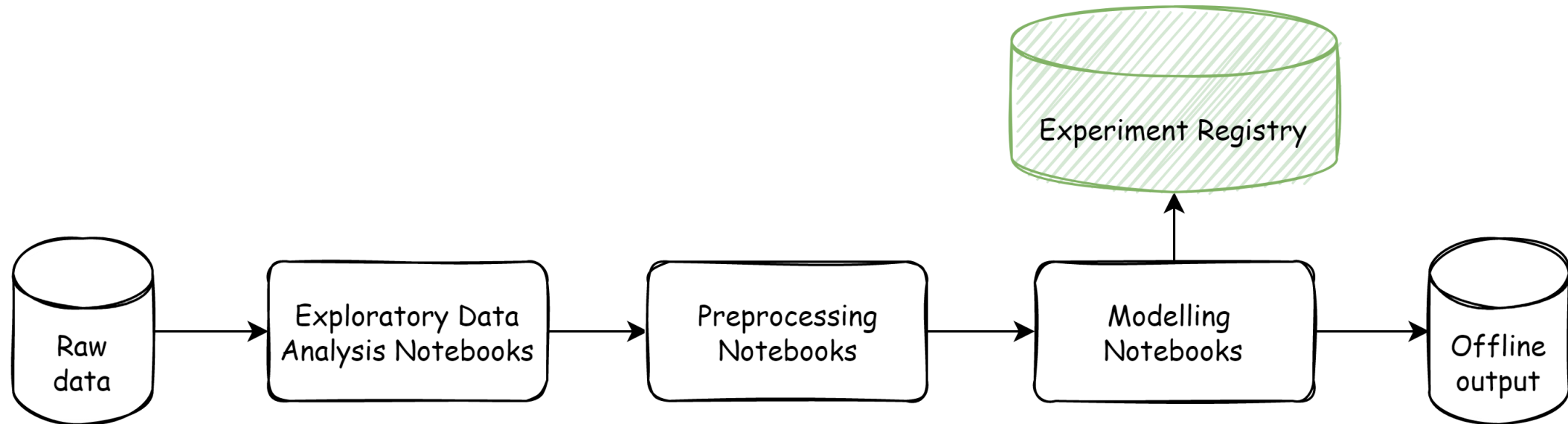
The Machine Learning process



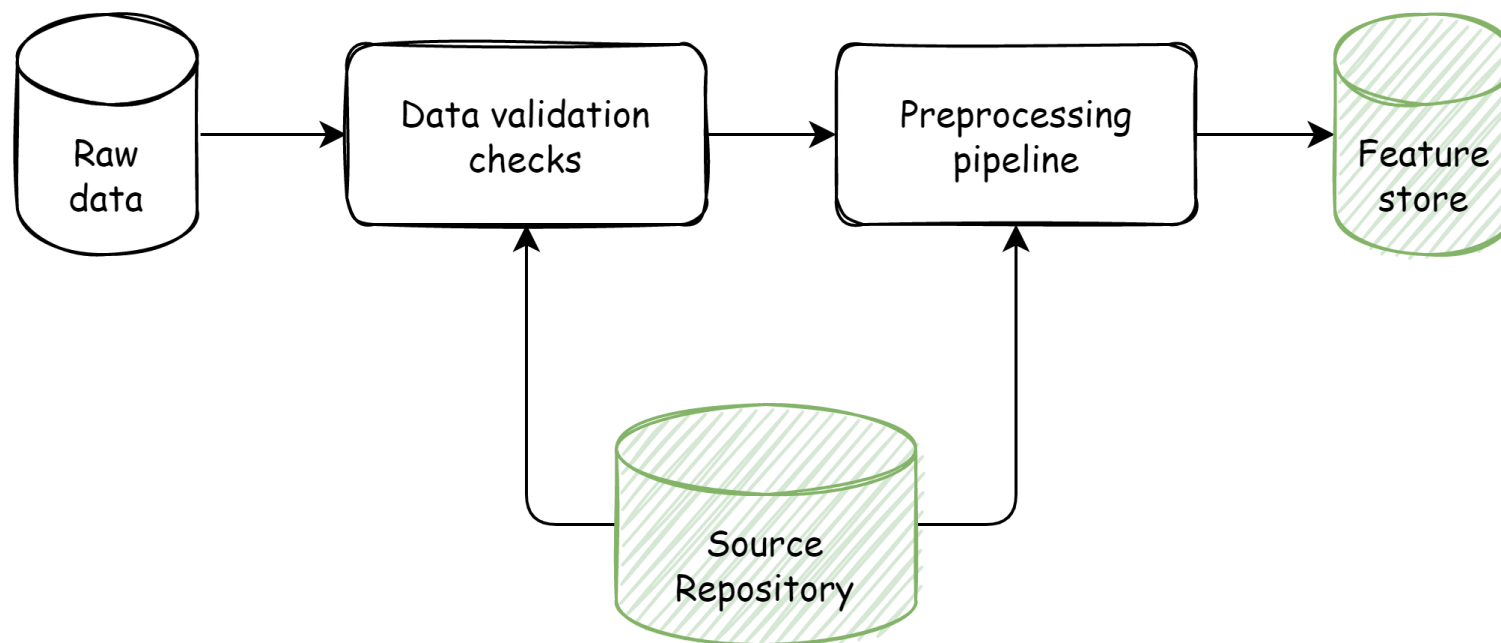
The Machine Learning process



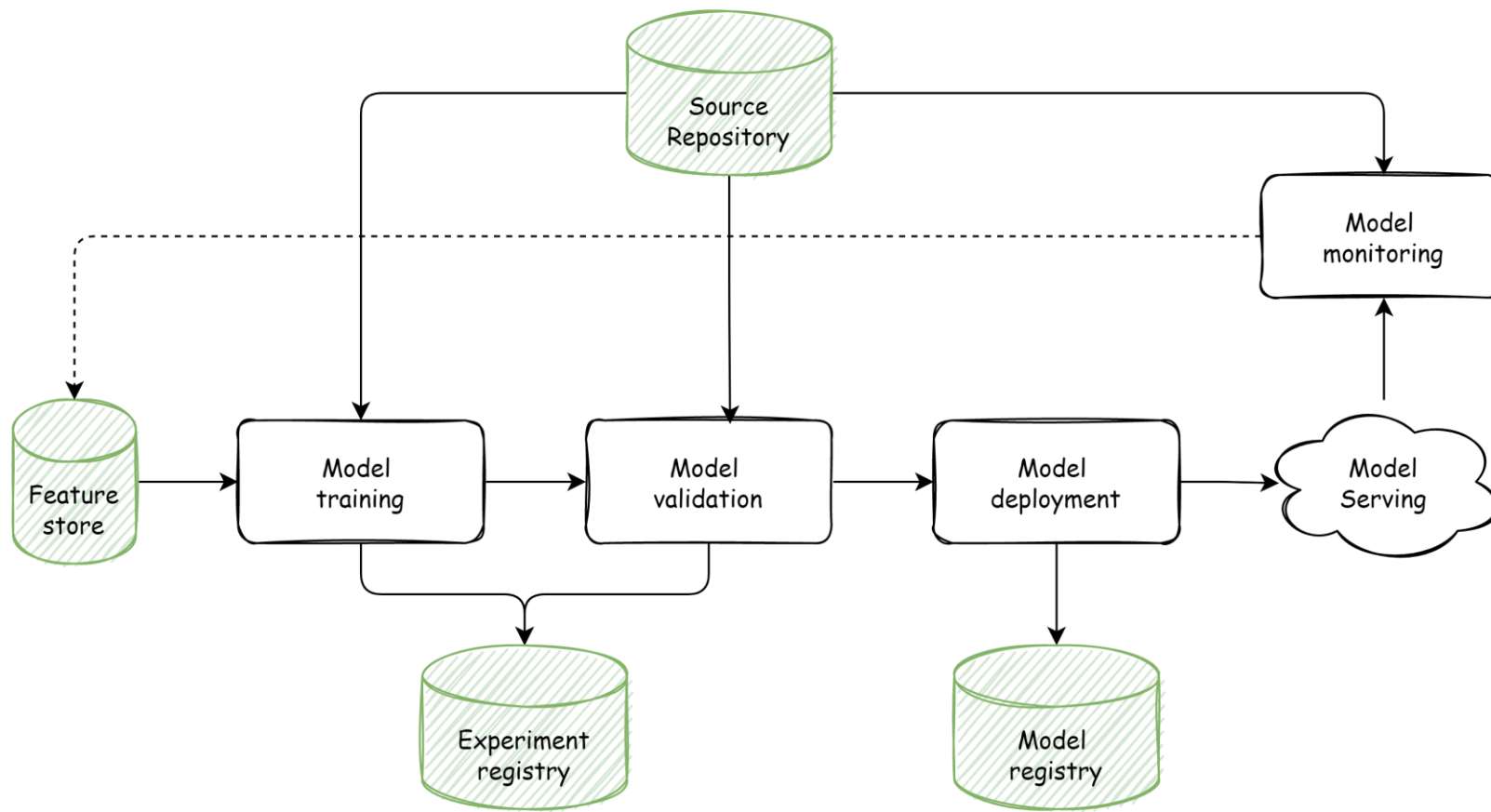
Offline development



Automated data pre-processing



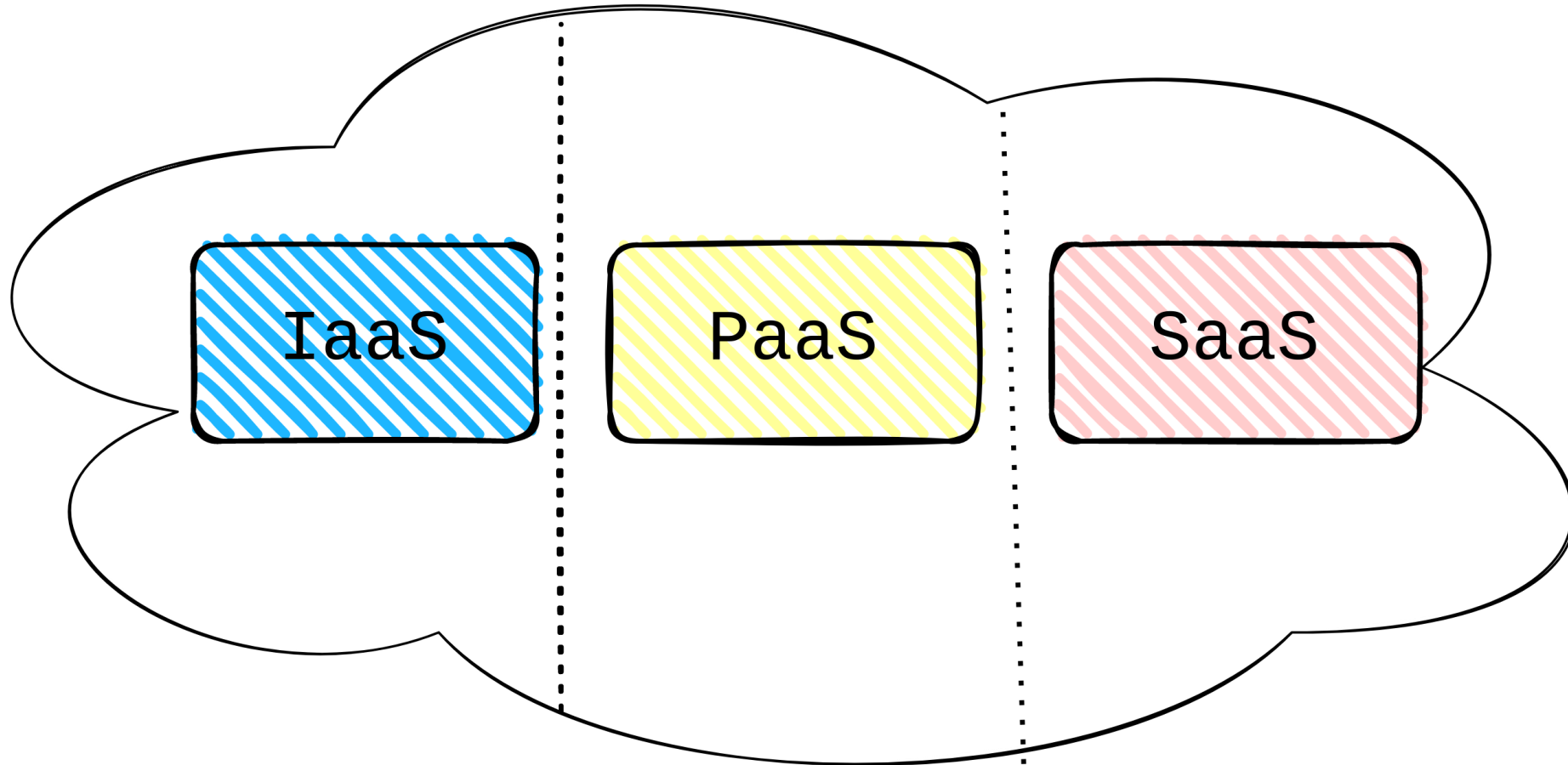
Automated model lifecycle



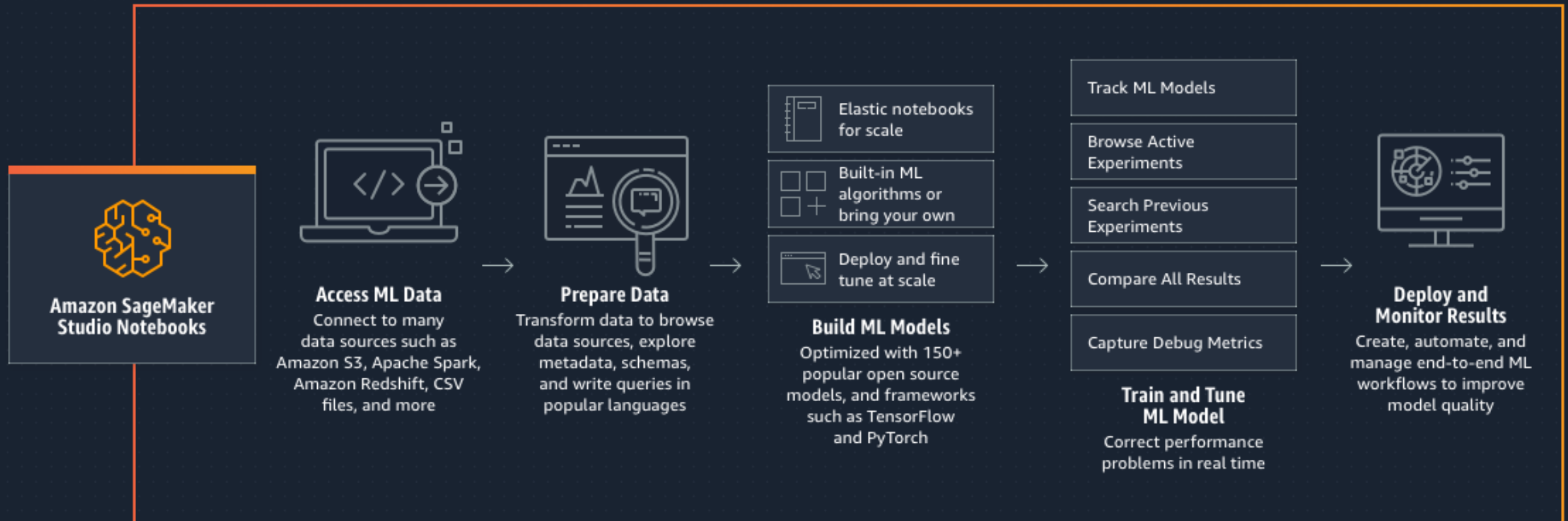


MLOps on AWS

AWS Cloud basics



Sagemaker Components

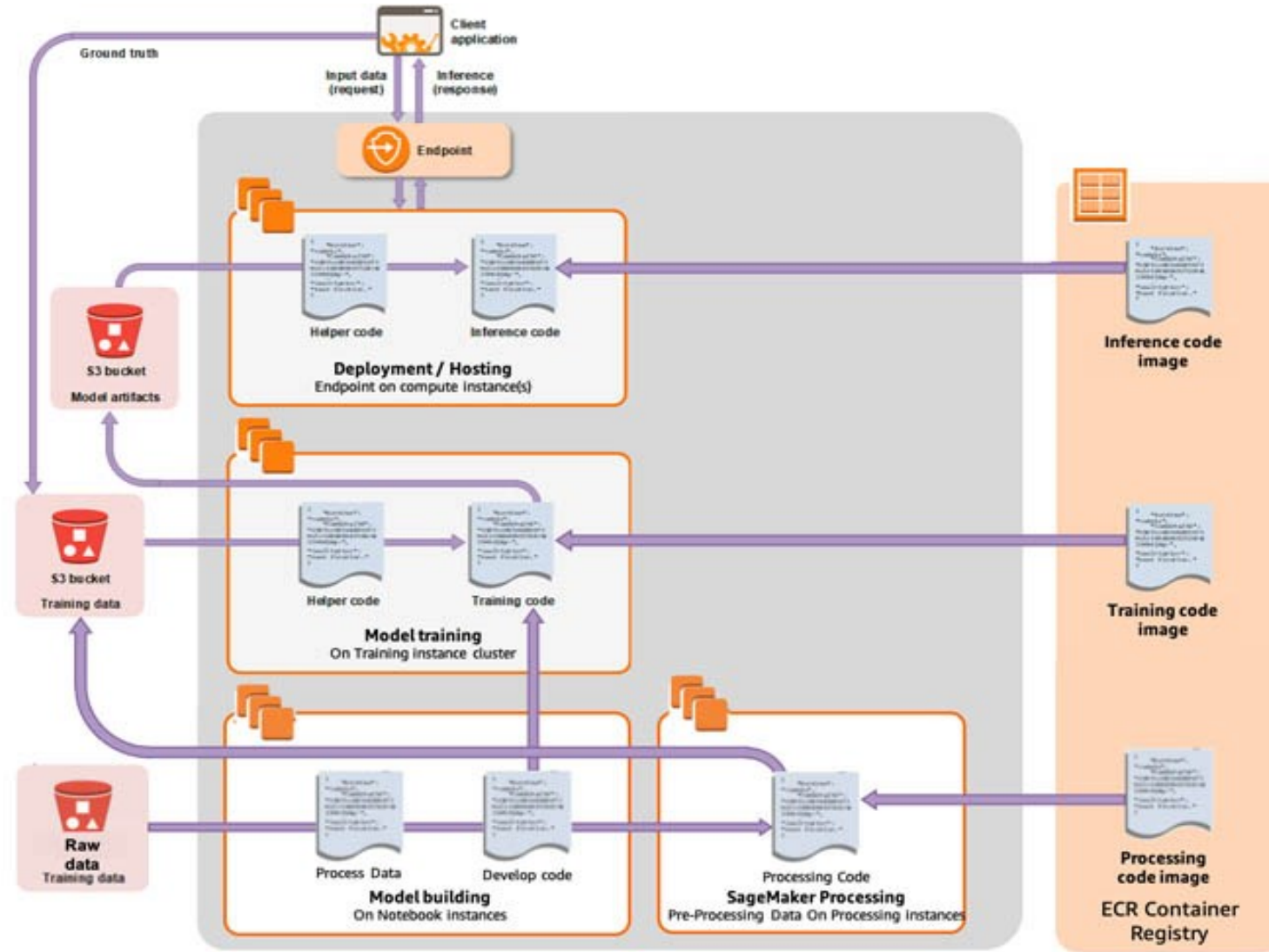


Sagemaker Training

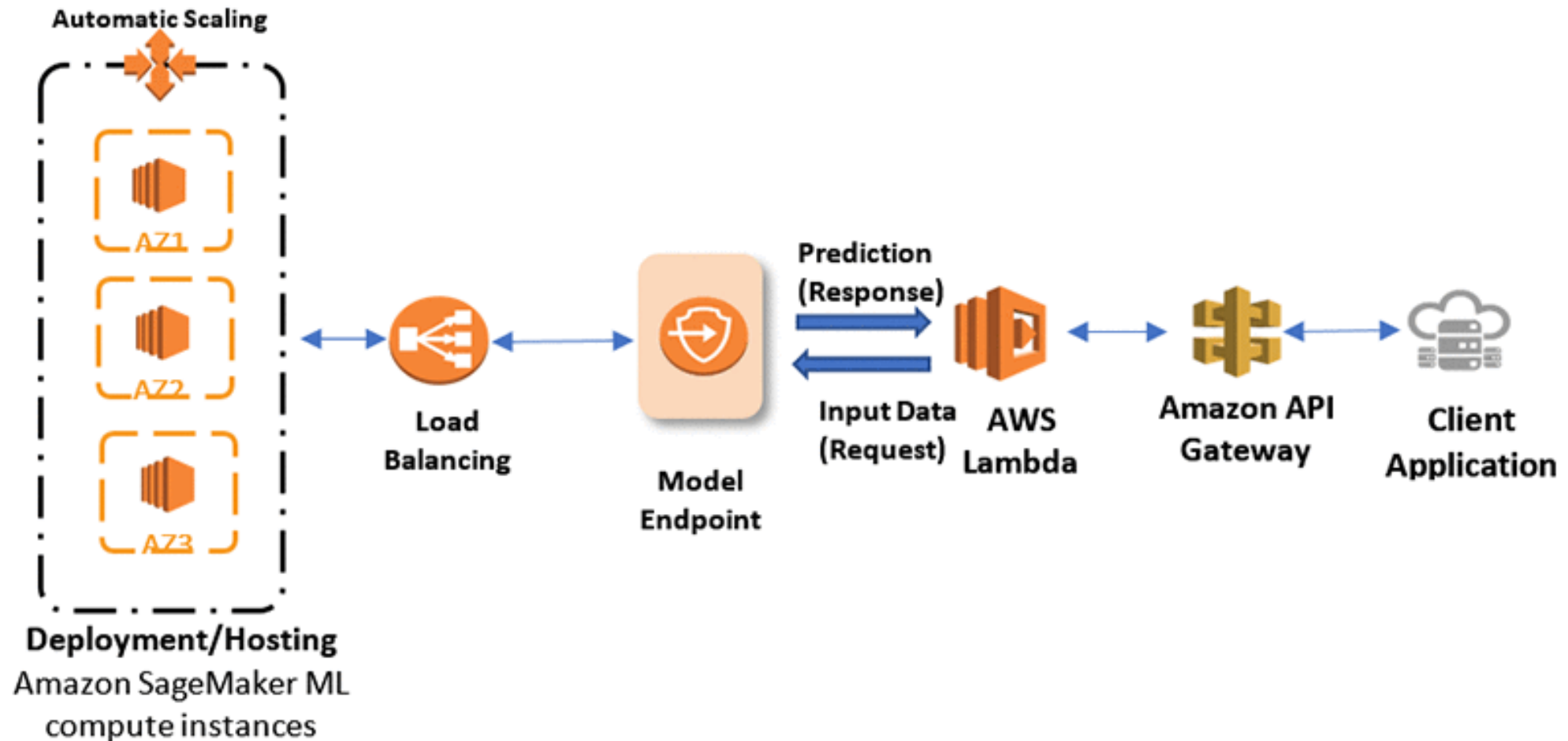
And then you deploy your model to production for real-time inference (or you use Batch Transform for batch inference*)

Then you train and tune your model

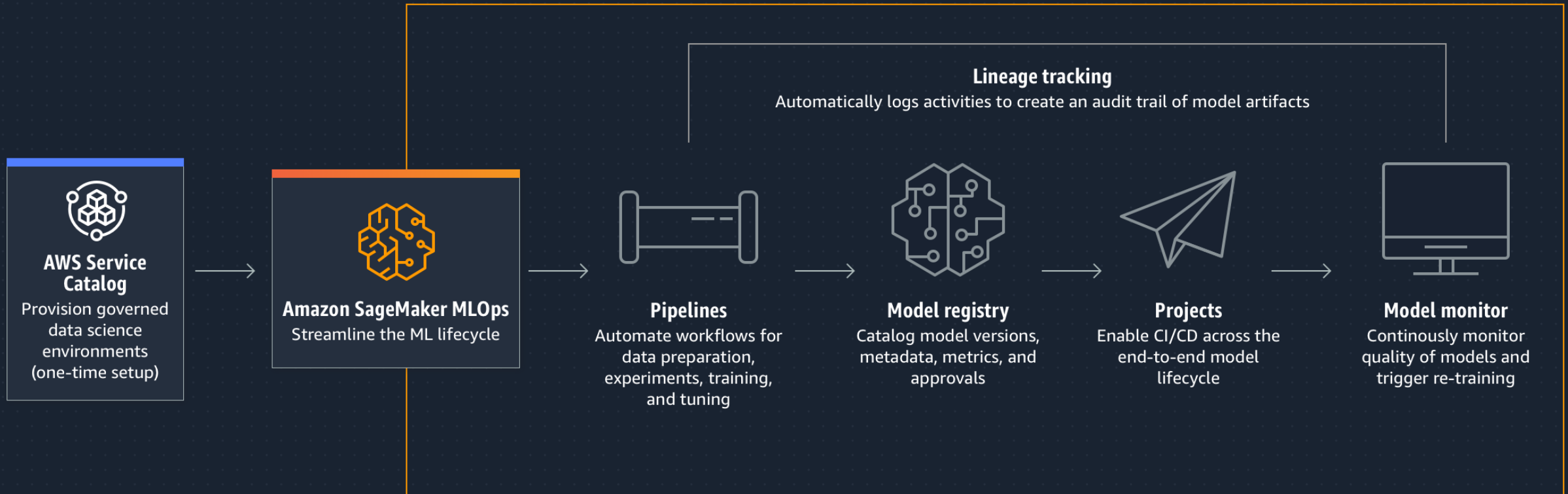
You start by building your model and processing raw data, developing your training data



Sagemaker Endpoint



AWS MLOps: Sagemaker Studio



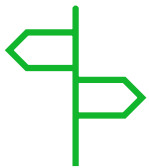
Before we begin...



This is a simplified version of the reality – we have taken some shortcuts: we will highlight which ones along the way.



We will use very simple models; we will talk briefly how to improve them.



We will not cover all the features of SageMaker and AWS, often there are multiple ways of achieving the same goal. We will comment on this whenever possible.



**Check the billing!!
Always.
All the time.**



**Jump
on
AWS**