Where Insight and Innovation Converge

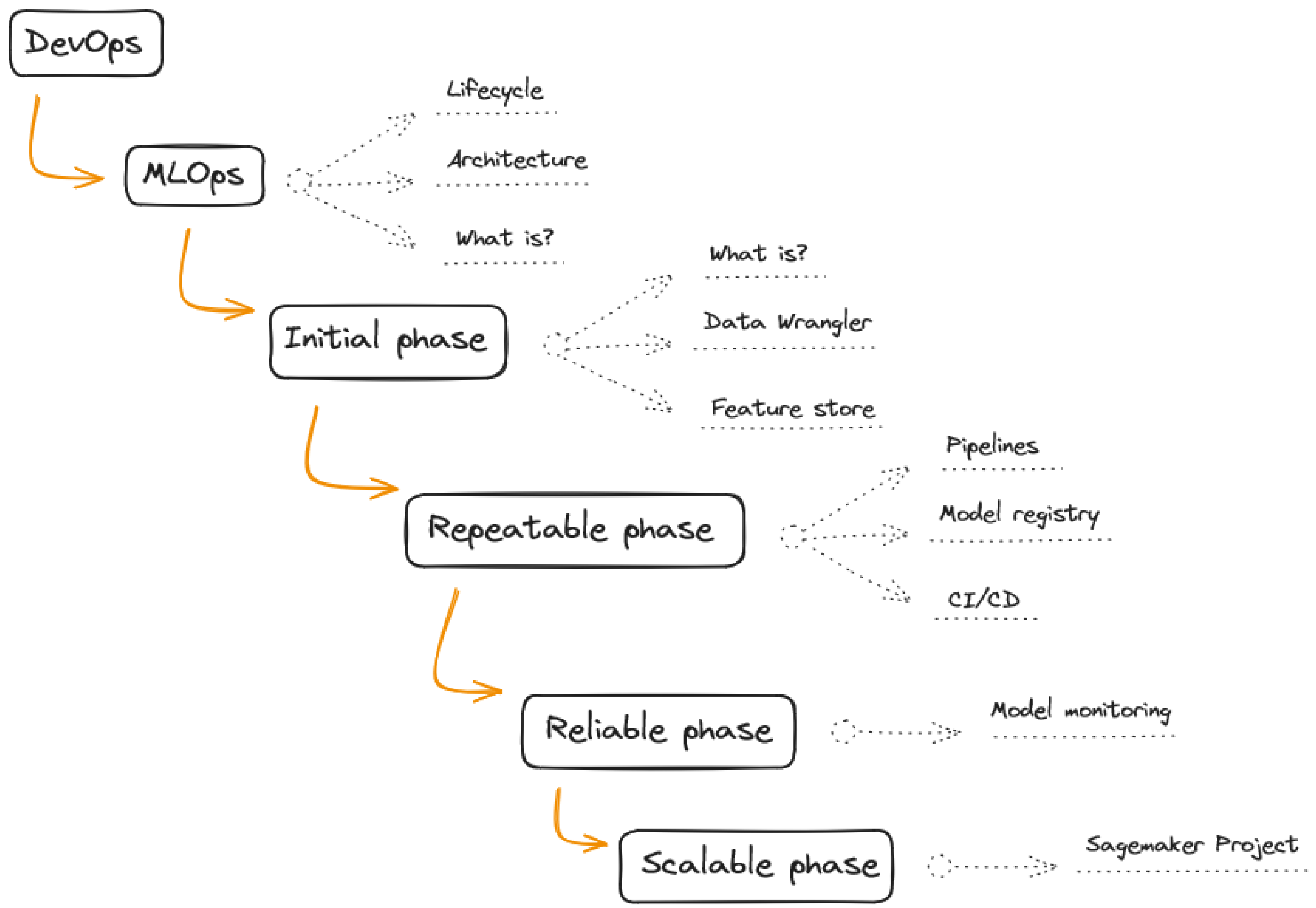
MLOps with Amazon SageMaker

Simone Cardis

ML Engineer

scardis@sorint.tek

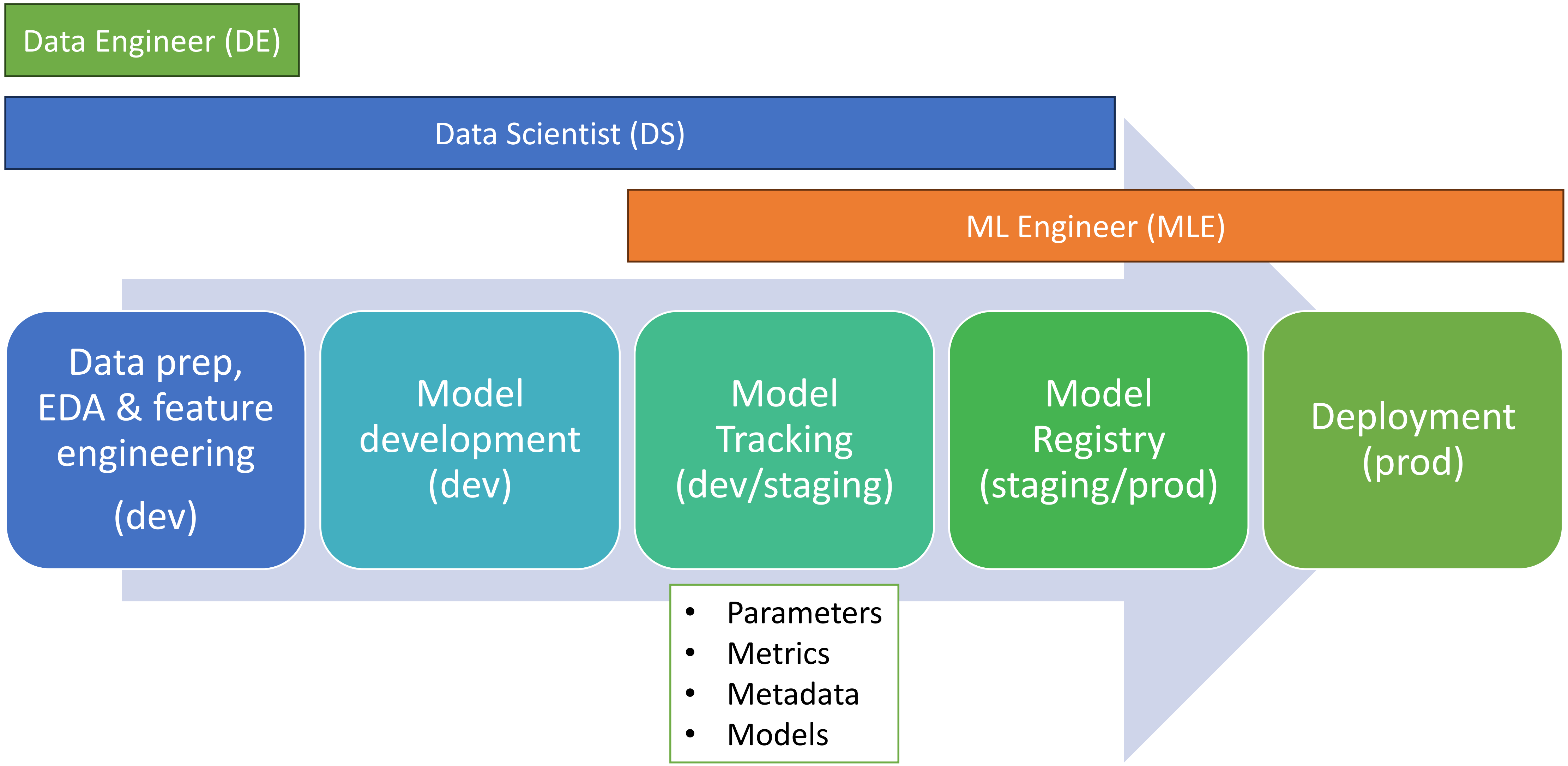
MLOps


DevOps
MLOps
Lifecycle
Architecture
What is?
Initial phase
What is?
Data Wrangler
Feature store
Repeatable phase
Pipelines
Model registry
CI/CD
Reliable phase
Model monitoring
Scalable phase
Sagemaker Project

MLOps

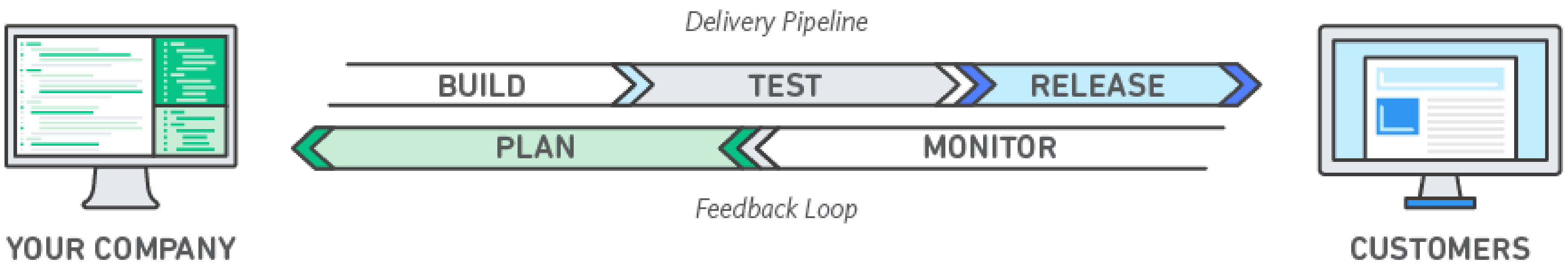
Introduction

ML lifecycle and roles



What is DevOps?

A set of **processes** and **tools** for **optimize development** and **deployment** integrations.



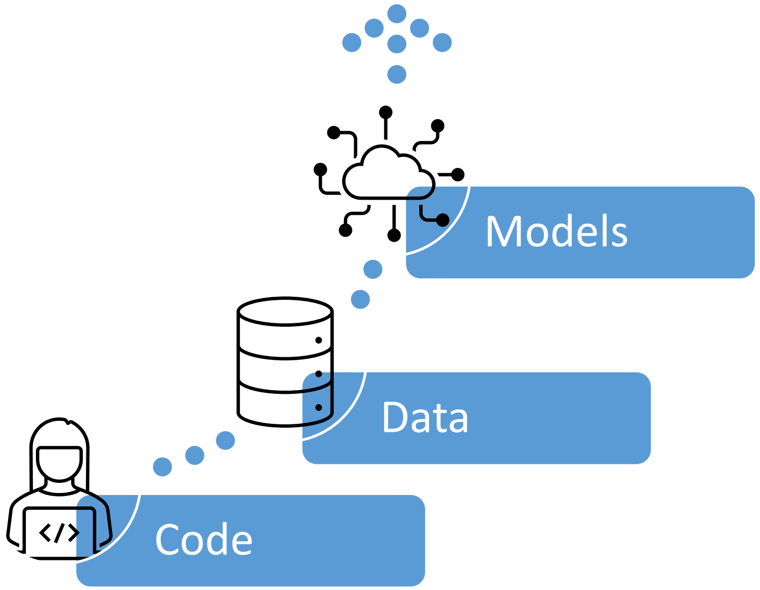
https://aws.amazon.com/devops/what-is-devops/

Benefits of DEVOps

Scale & Repeatability	Infrastructure as a code is an example of efficient manner to automate deployment at scale
Security	DevOps tools help in managing resources and control them preserving compliance
Reliability	With CI/CD practices deploy quickly with increased quality. Using testing and logging best practices it’s possible to monitor and check every function before deployment
Speed	Devops best practices like continuous delivery and microservices enable more speed and productivity
Rapid Delivery	With CI/CD (continuous integration and continuous delivery) products are improved and deployed in shorter time

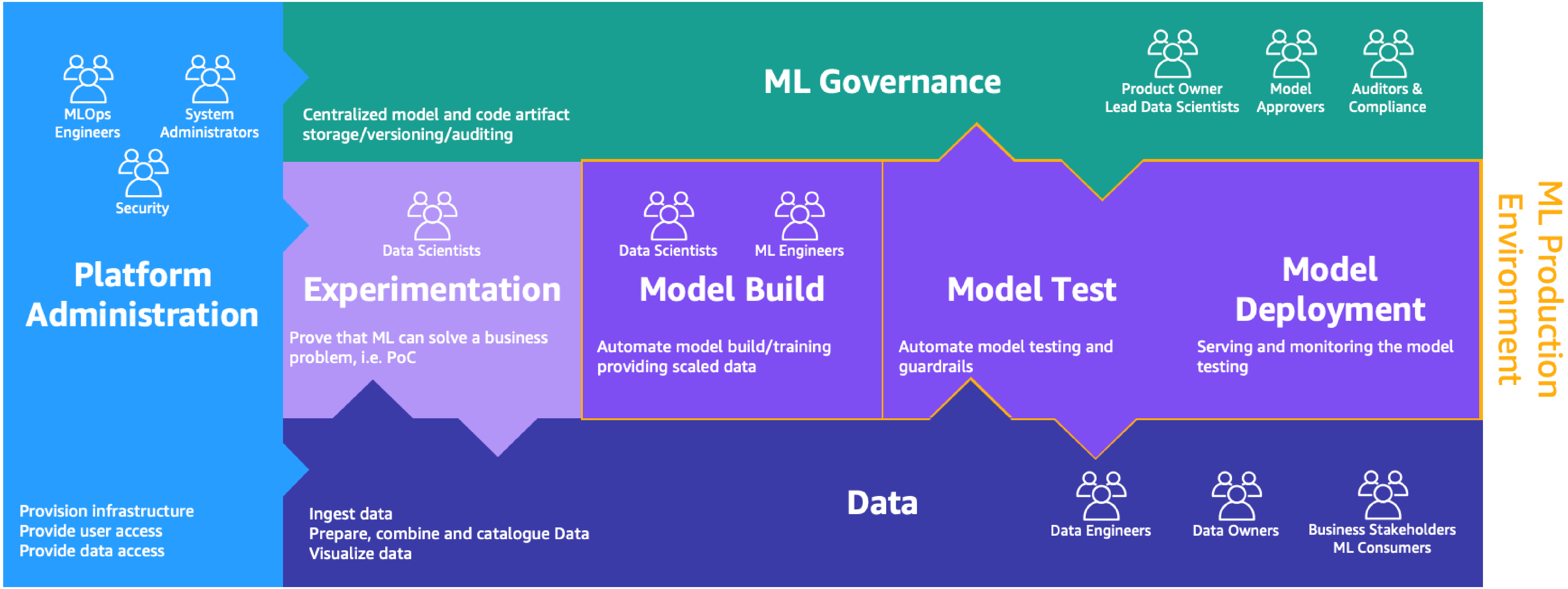
What is ML and Operations (MLOps)?

A set of **processes** and **automation** for **managing** **ML artifacts**

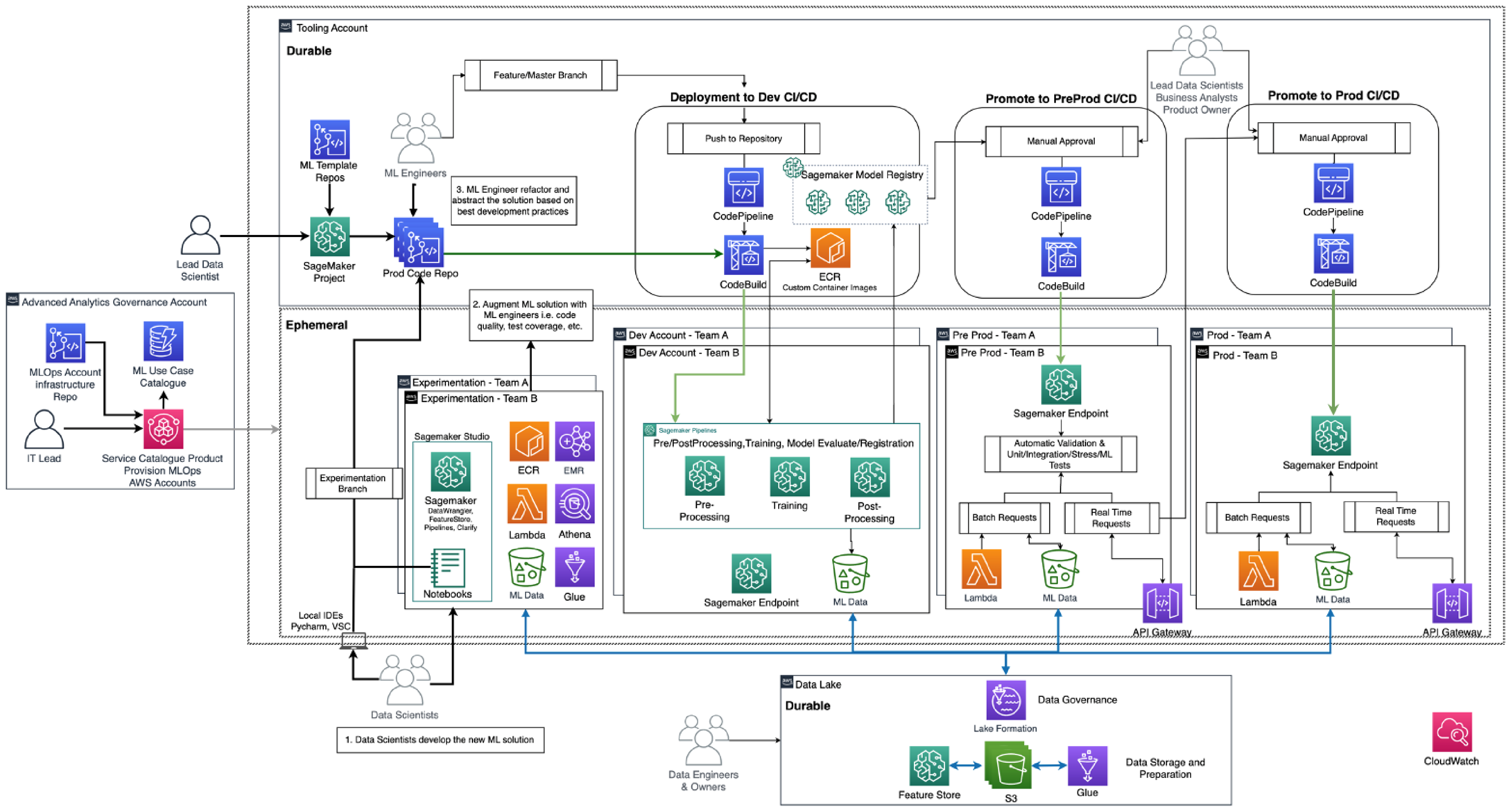


through each stage of their lifecycle

ML Lifecycle and personas



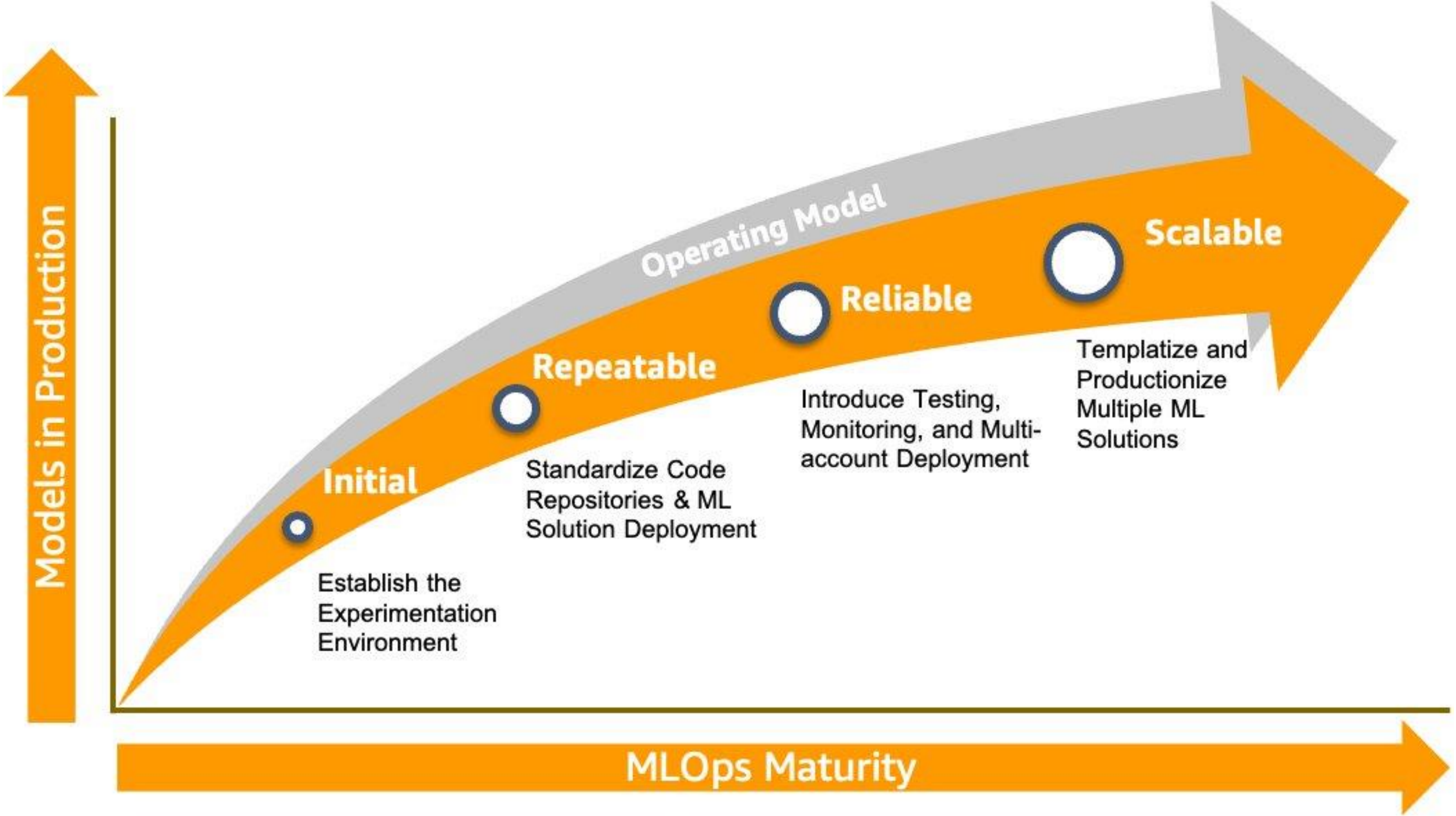
AWS MLOps Reference Architecture



Benefits of MLOps

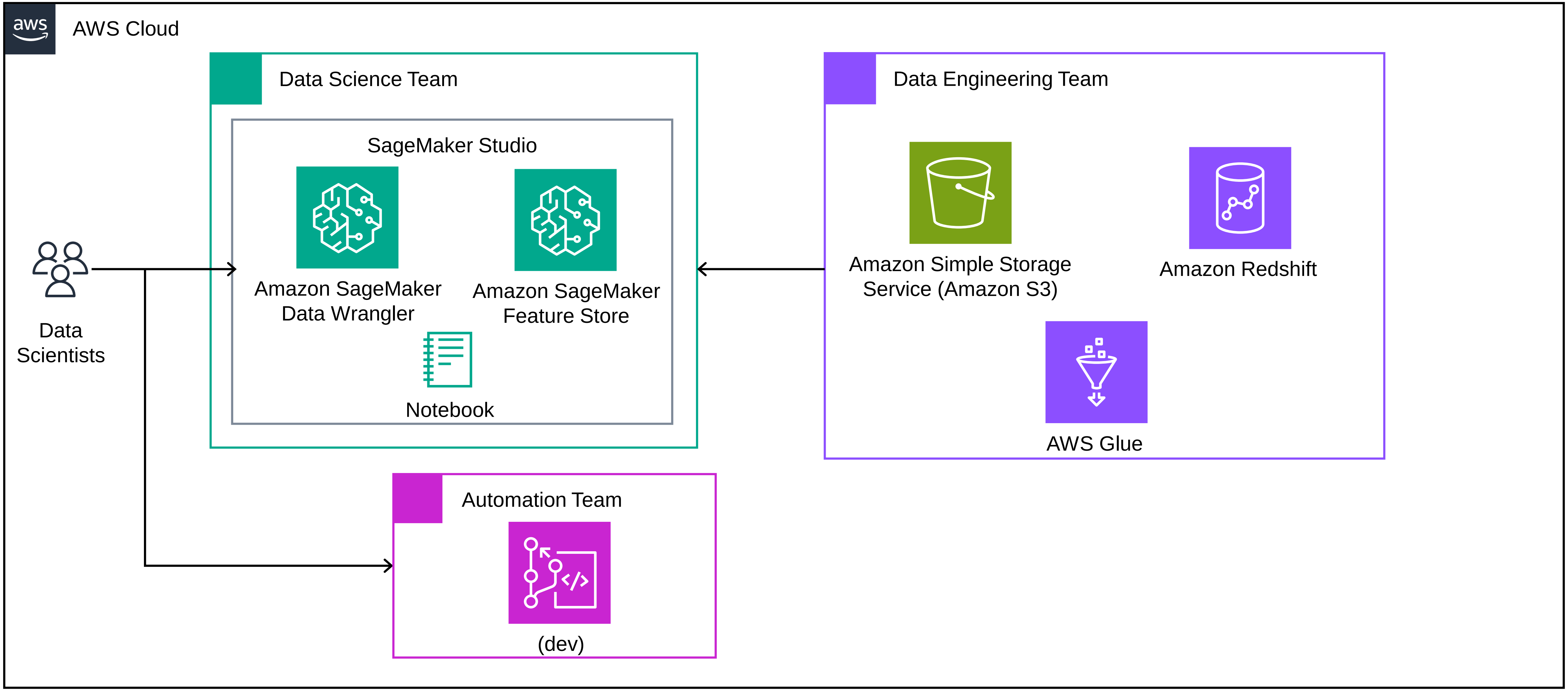
Productivity	Self-service environments with access to curated data sets data scientists waste no time with missing or invalid data
Repeatability	Automating to ensure a repeatable process Including model training, evaluation, versioning, and deployment
Reliability	With CI/CD practices deploy quickly with increased quality
Auditability	Versioning all inputs and outputs, including experiments, source data and trained model, to demonstrate exactly how the model was built
Data and model quality	Track changes to data statistical properties and model quality over time

MLOps maturity model

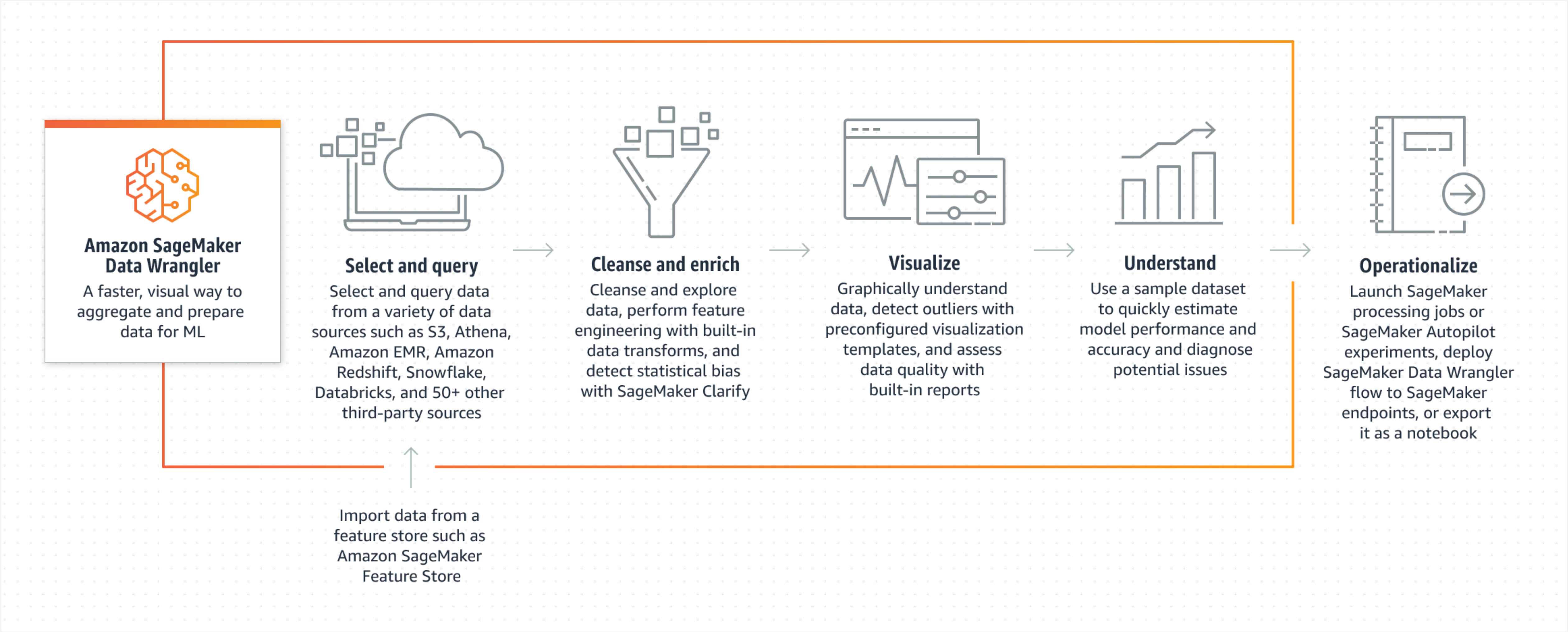


Initial Phase

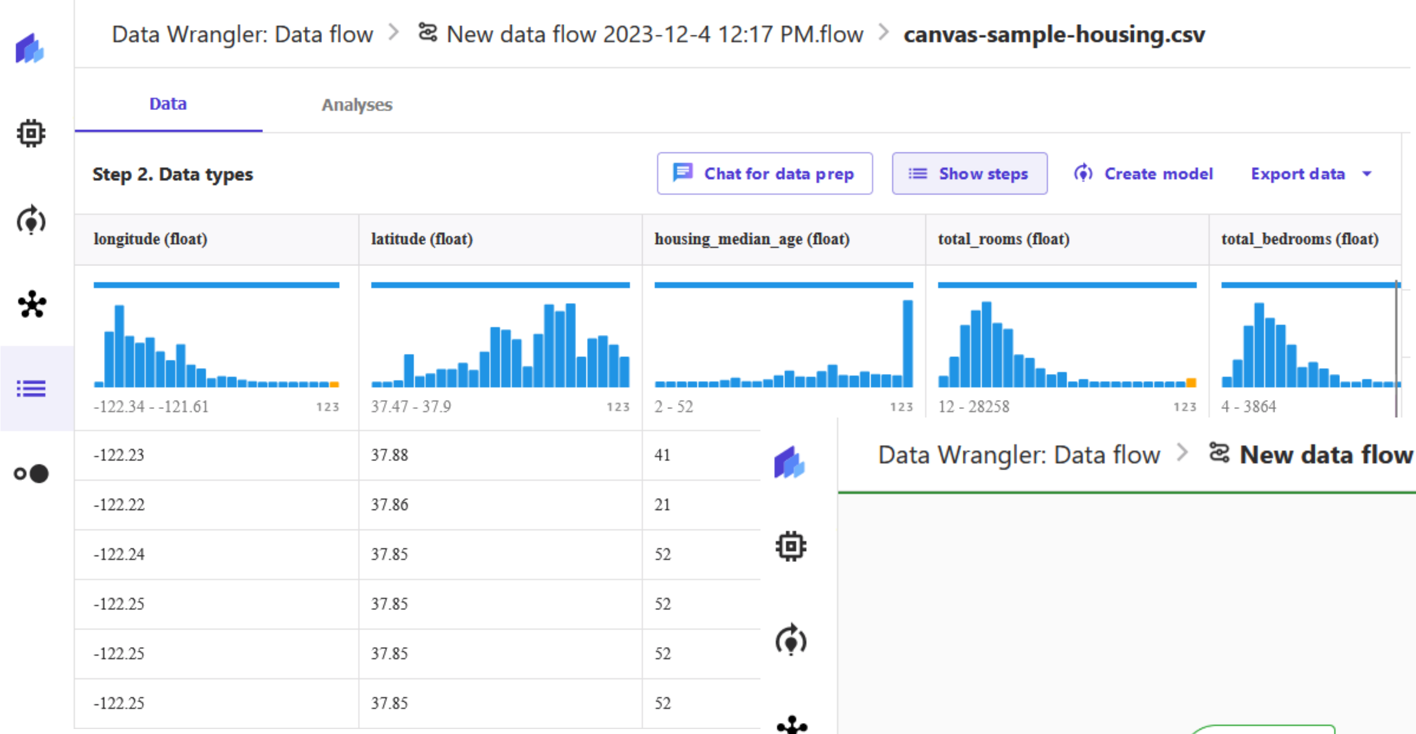
Initial Phase: Experimentation

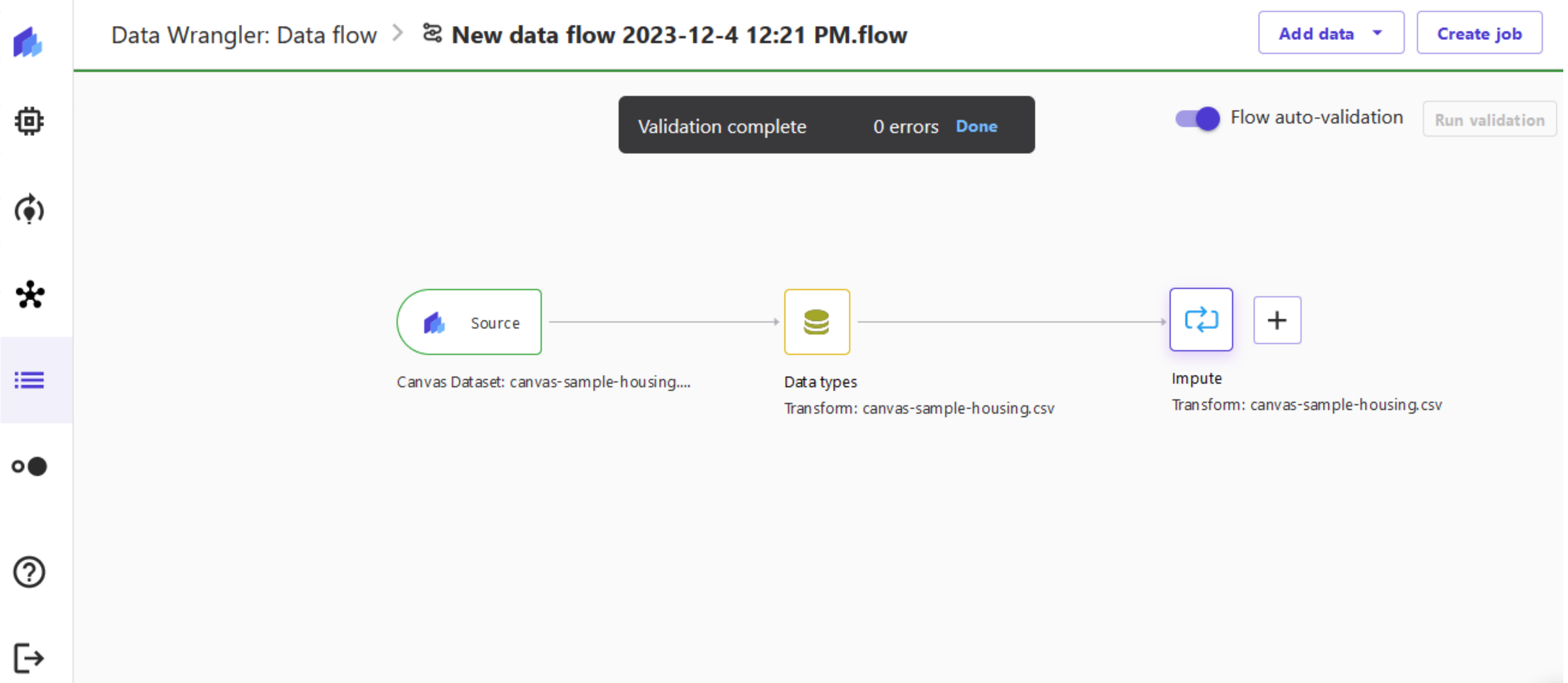


• Amazon SageMaker Data Wrangler



Data Wrangler Data flow





https://aws.amazon.com/blogs/machine-learning/accelerate-data-preparation-for-ml-with-comprehensive-data-preparation-capabilities-and-a-natural-language-interface-in-amazon-sagemaker-canvas/

Amazon SageMaker Feature Store

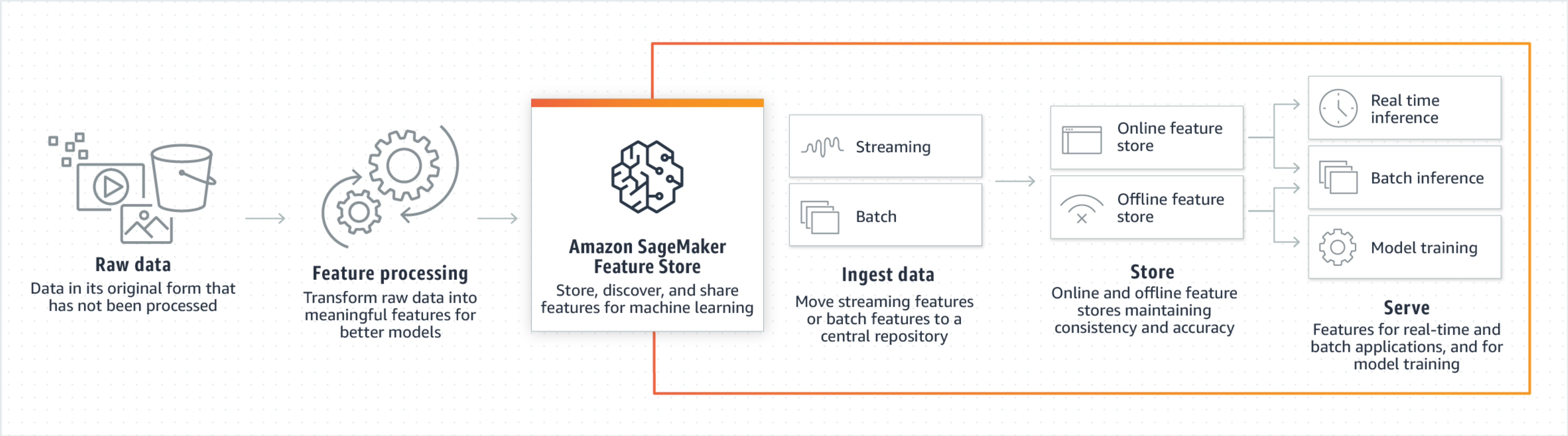
SORINT TEK

Single source of truth to store, **retrieve**, **remove**, **track**, **share**, **discover**, and control access to features

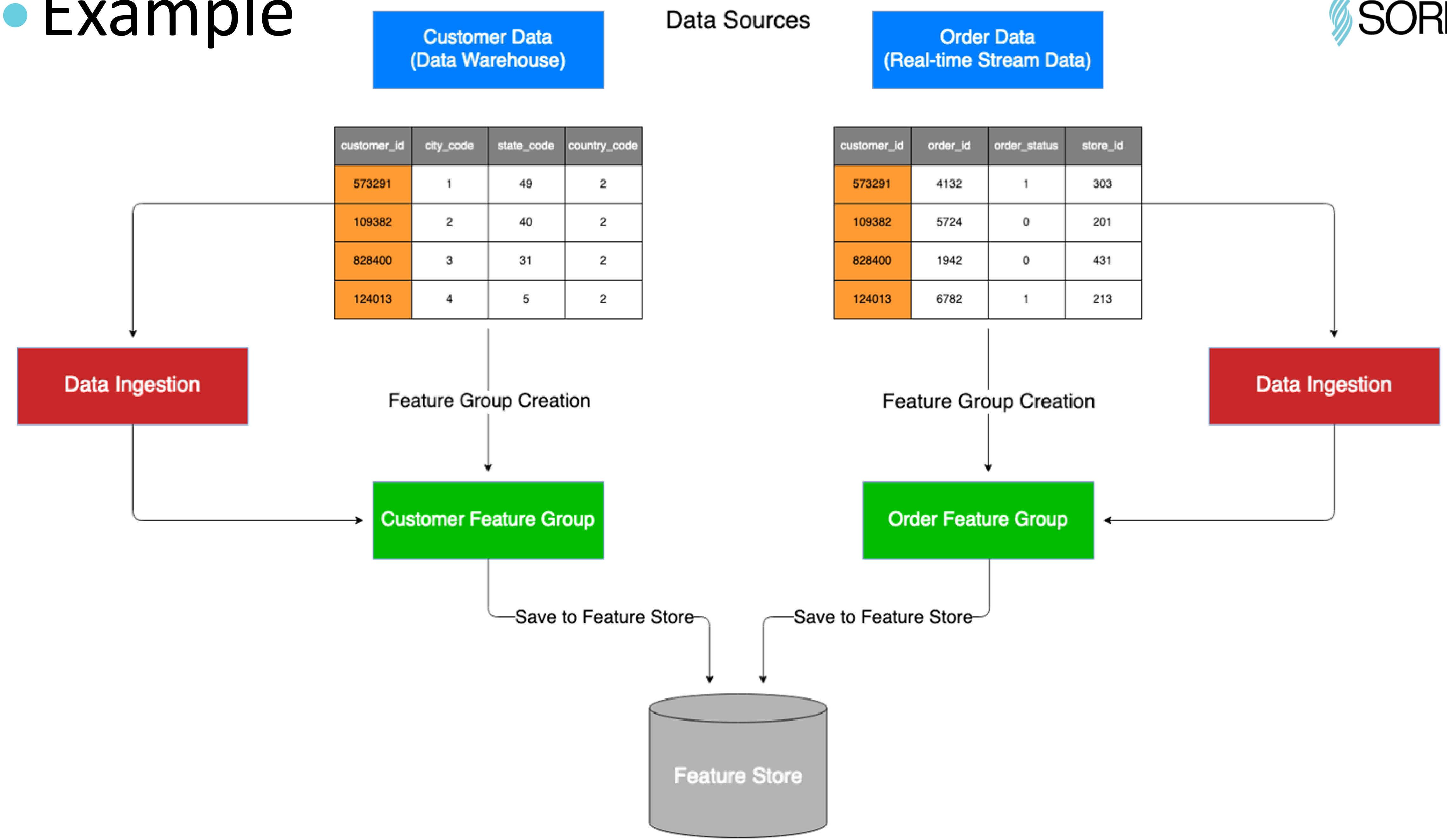
- **Secure**ly store and serve features for **real-time and batch applications**
- Accelerate model development by **sharing and reusing features**
- Provide **historical** data access to **recreate training datasets** at a given point in time in the past.
- **Reduce training-serving skew** (discrepancy between model training and inference serving), which can cause models to perform worse than expected in production.
- Enable **data encryption** and **access control**

Amazon SageMaker Feature Store

- The **processing logic for data** is **authored only once**
  - ✓ Real-time inference for features stored in the **online store**
  - ✓ **Offline store** for model training and batch inference



Example



customer_id	city_code	state_code	country_code
573291	1	49	2
109382	2	40	2
828400	3	31	2
124013	4	5	2

customer_id	order_id	order_status	store_id
573291	4132	1	303
109382	5724	0	201
828400	1942	0	431
124013	6782	1	213

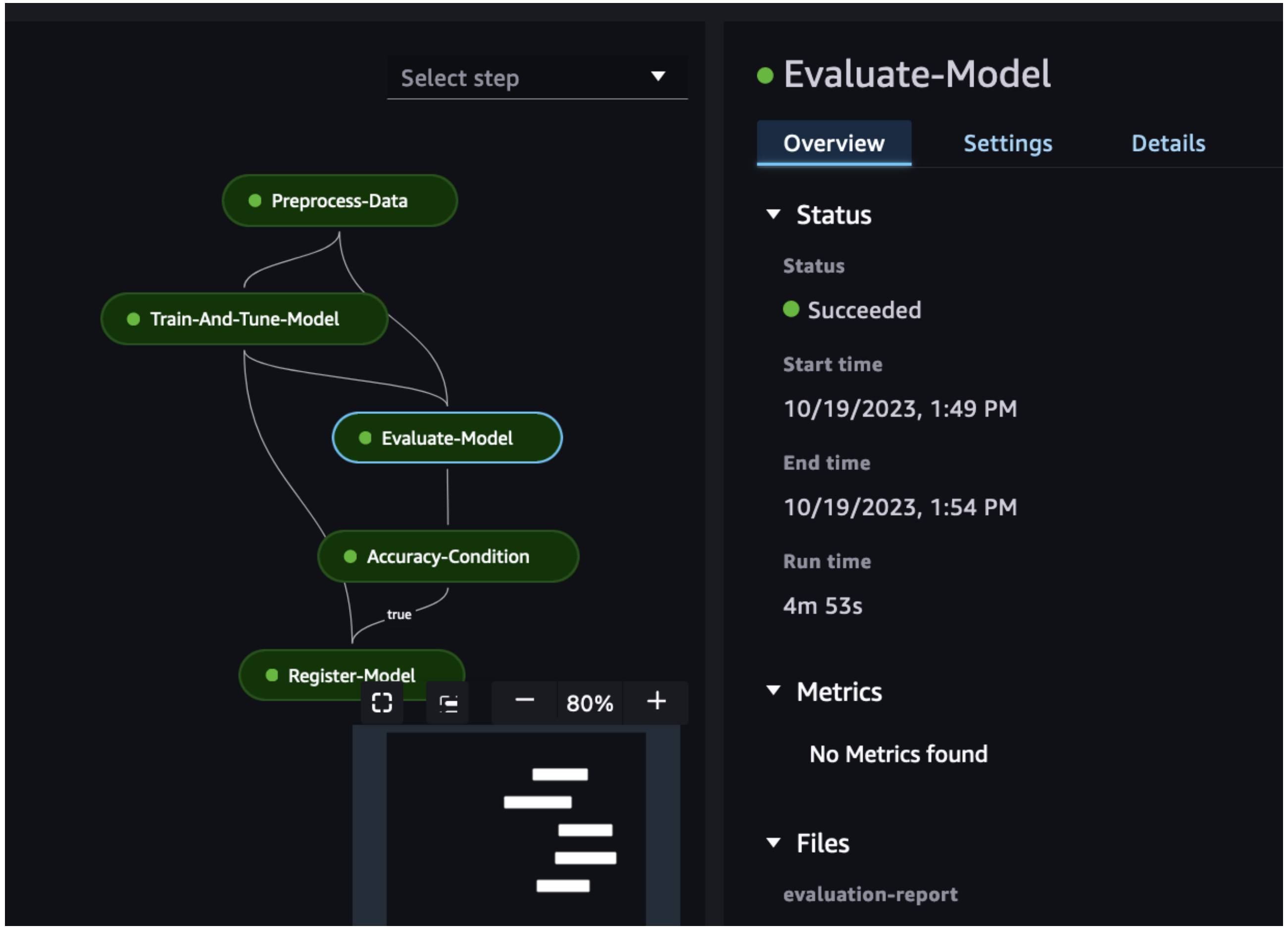
Repeatable Phase

SageMaker Pipelines

A series of interconnected steps (SageMaker processing jobs, training, HPO) that is defined by a Directed Acyclic Graph (DAG) using a Python SDK

Benefits:

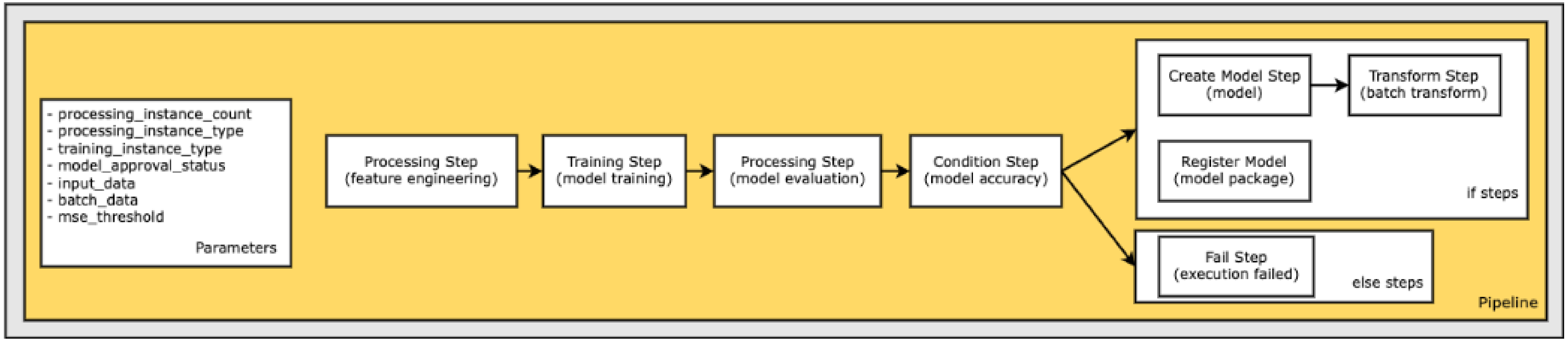
- Integrated in SageMaker Studio
- Full Pipeline Lineage



SageMaker Pipeline Example

```
from sagemaker.workflow.pipeline import Pipeline


pipeline_name = f"AbalonePipeline"
pipeline = Pipeline(
    name=pipeline_name,
    parameters=[
        processing_instance_count,
        instance_type,
        model_approval_status,
        input_data,
        batch_data,
        mse_threshold,
    ],
    steps=[step_process, step_train, step_eval, step_cond],
)
```



https://github.com/aws/amazon-sagemaker-examples/blob/main/sagemaker-pipelines/tabular/abalone_build_train_deploy/sagemaker-pipelines-preprocess-train-evaluate-batch-transform.ipynb

SageMaker Model Registry

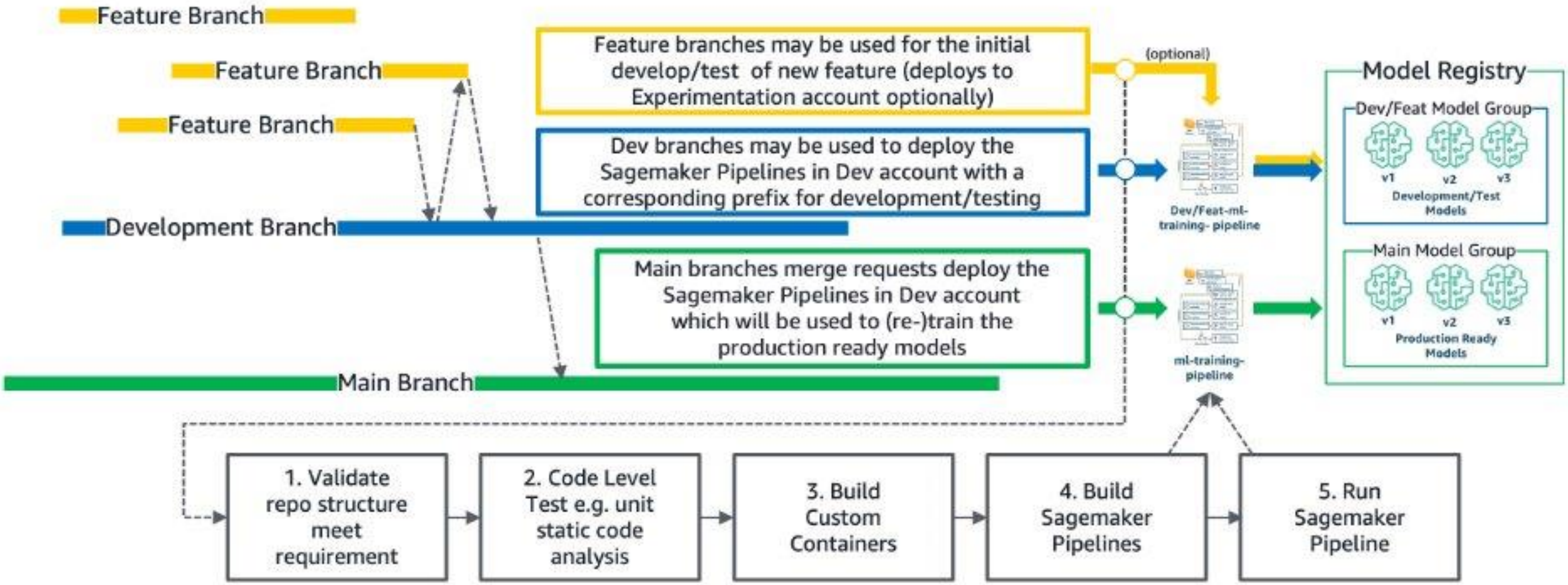
Features

- Manage model **versions**
- Associate **metadata**, e.g. training metrics
- Initiates CI/CD deployment for approved model versions

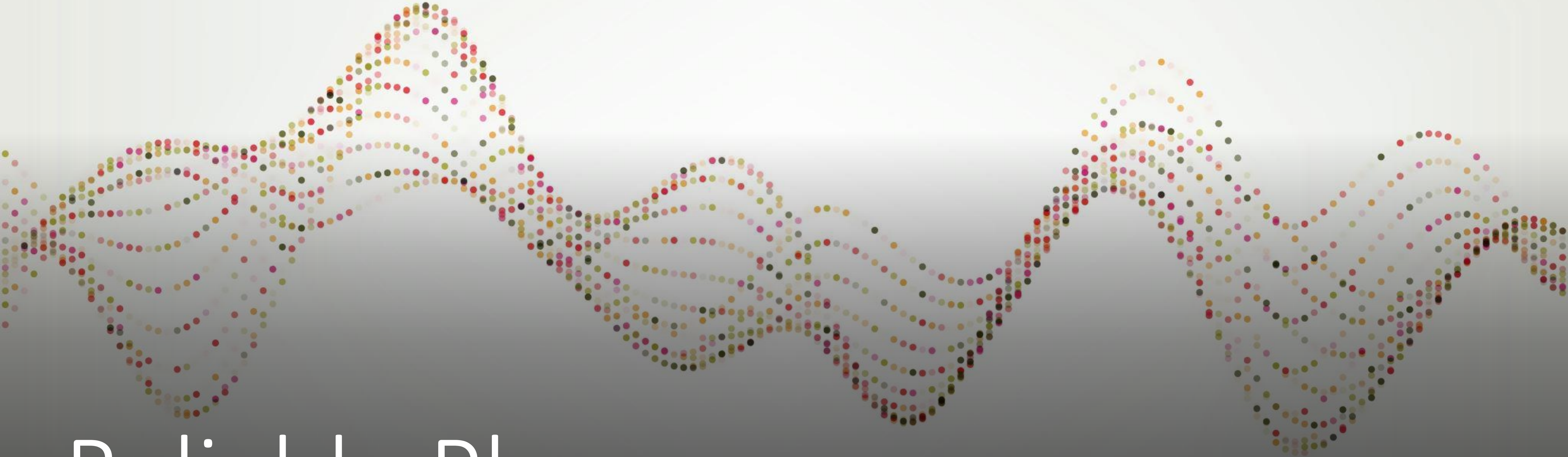
Typical workflow

- Create a **Model Group** that tracks all the models trained for a task
- A training pipeline run registers a model version in the Model Group
- The best model version is chosen and deployed for inference

• CI/CD integration



https://aws.amazon.com/blogs/machine-learning/improve-your-data-science-workflow-with-a-multi-branch-training-mlops-pipeline-using-aws/



Reliable Phase

SageMaker Model Monitoring

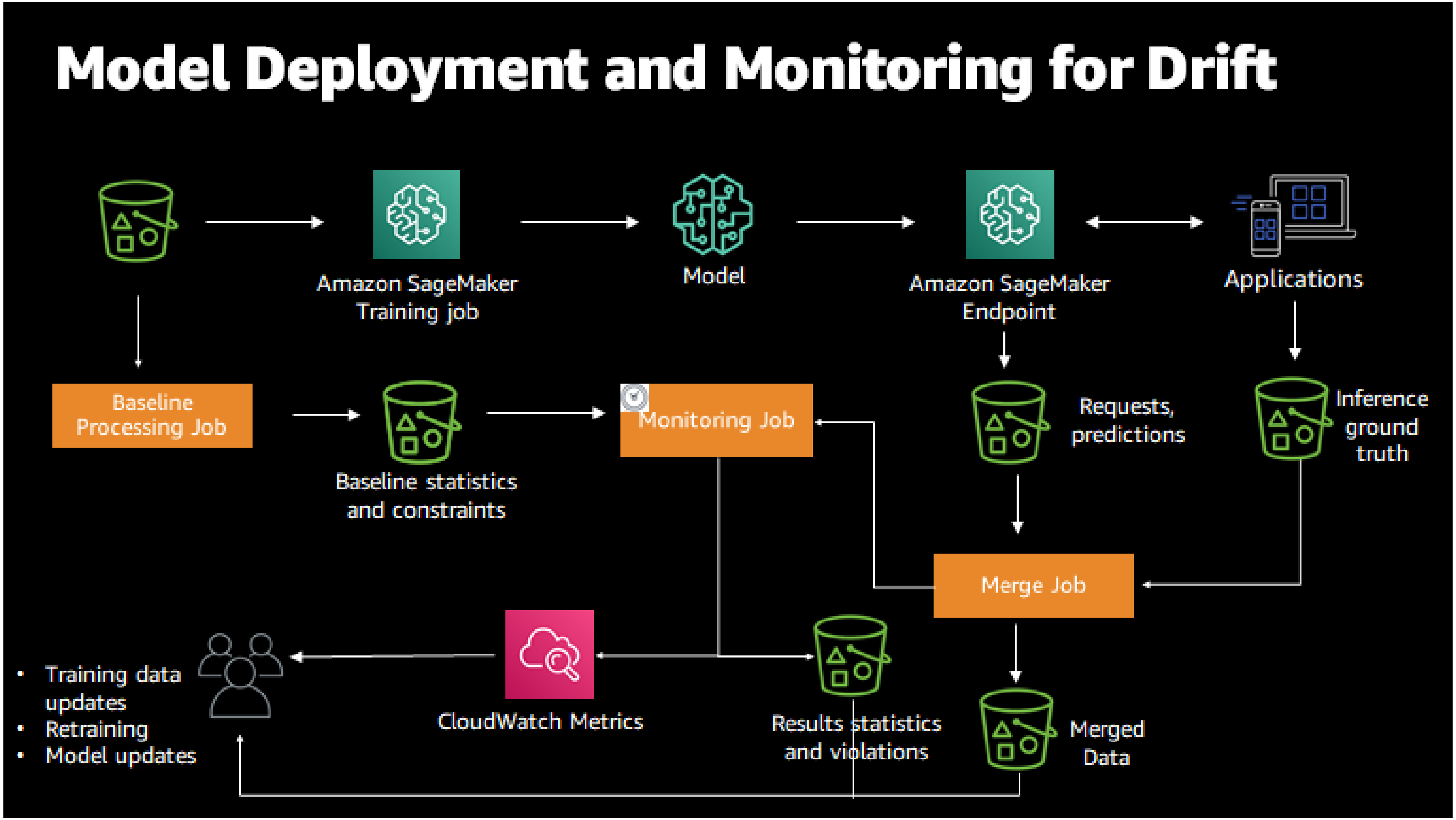
Types of monitoring

- **Data quality**
  monitor drift (change in statistics) of production data from baseline training data (use deequ)
- **Model quality**
  monitor drift in model quality metrics, such as accuracy
- **Feature attribution drift**
  monitor drift in feature attribution, i.e. indicate how much each feature in your model contributed to the predictions for each given instance in training and in production (live data)
- **Model bias**
  bias in model's predictions

Model Bias metrics example: DPPL

Metric	Description	Example	Interpretation
Difference in Positive Proportions in Predicted Labels (DPPL)	Measures the difference in the proportion of positive predictions between the favored facet (population) *a* and the disfavored facet (population) *d*.	Has there been an imbalance across demographic groups in the predicted positive outcomes that might indicate bias?	• Positive values indicate that the favored facet a has a higher proportion of predicted positive outcomes. • Values near zero indicate a more equal proportion of predicted positive outcomes between facets. • Negative values indicate the disfavored facet d has a higher proportion of predicted positive outcomes.

SageMaker Model Monitoring



Scalable Phase

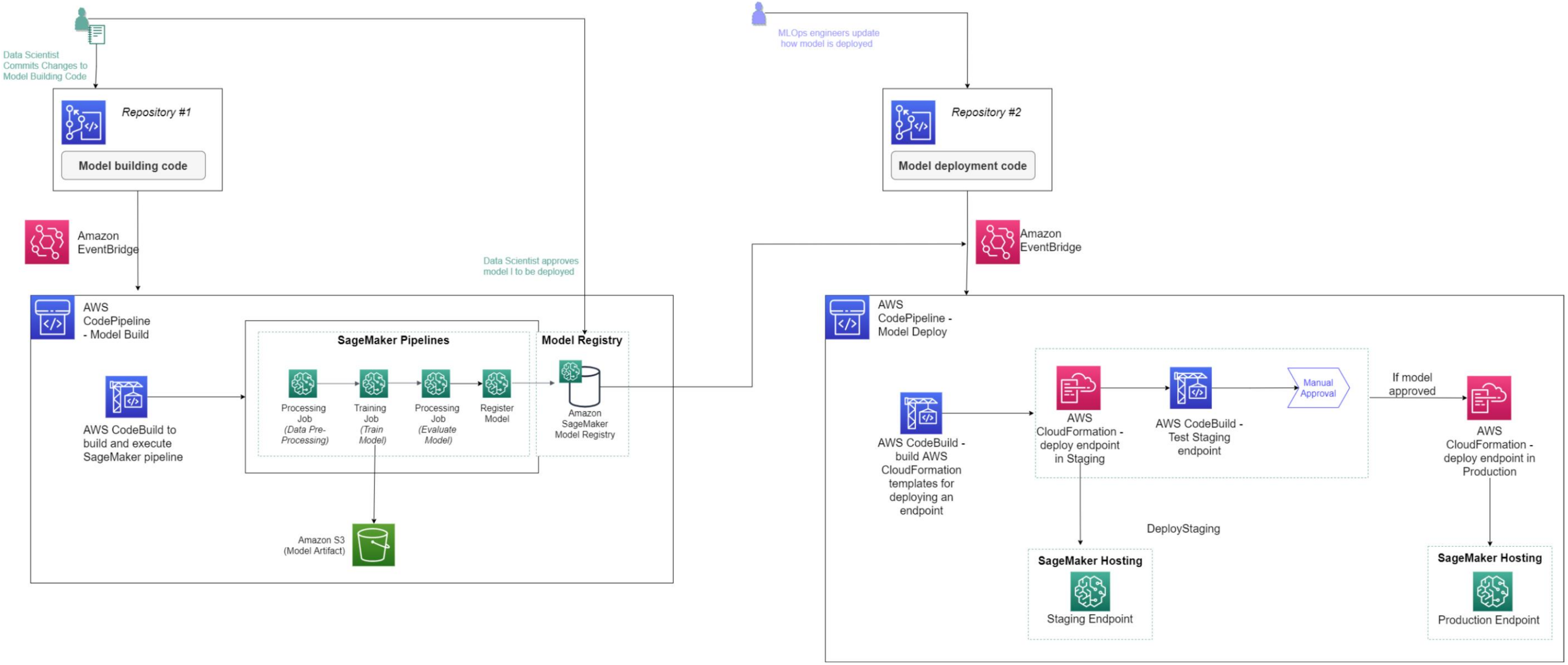
SageMaker Project

Project offers custom CloudFormation templates to define and control resources for ML workflows

- **Key Responsibilities**:
  - Configuring IAM roles and policies.
  - Enforcing resource tags.
  - Implementing encryption.
  - Decoupling resources across multiple accounts.
- **Benefits**
  - **For Organizations:**
    - Resource control and enhanced security
  - **For Data Scientists:**
    - Easy selection of templates to set up and pre-configure ML workflows.

SageMaker Project

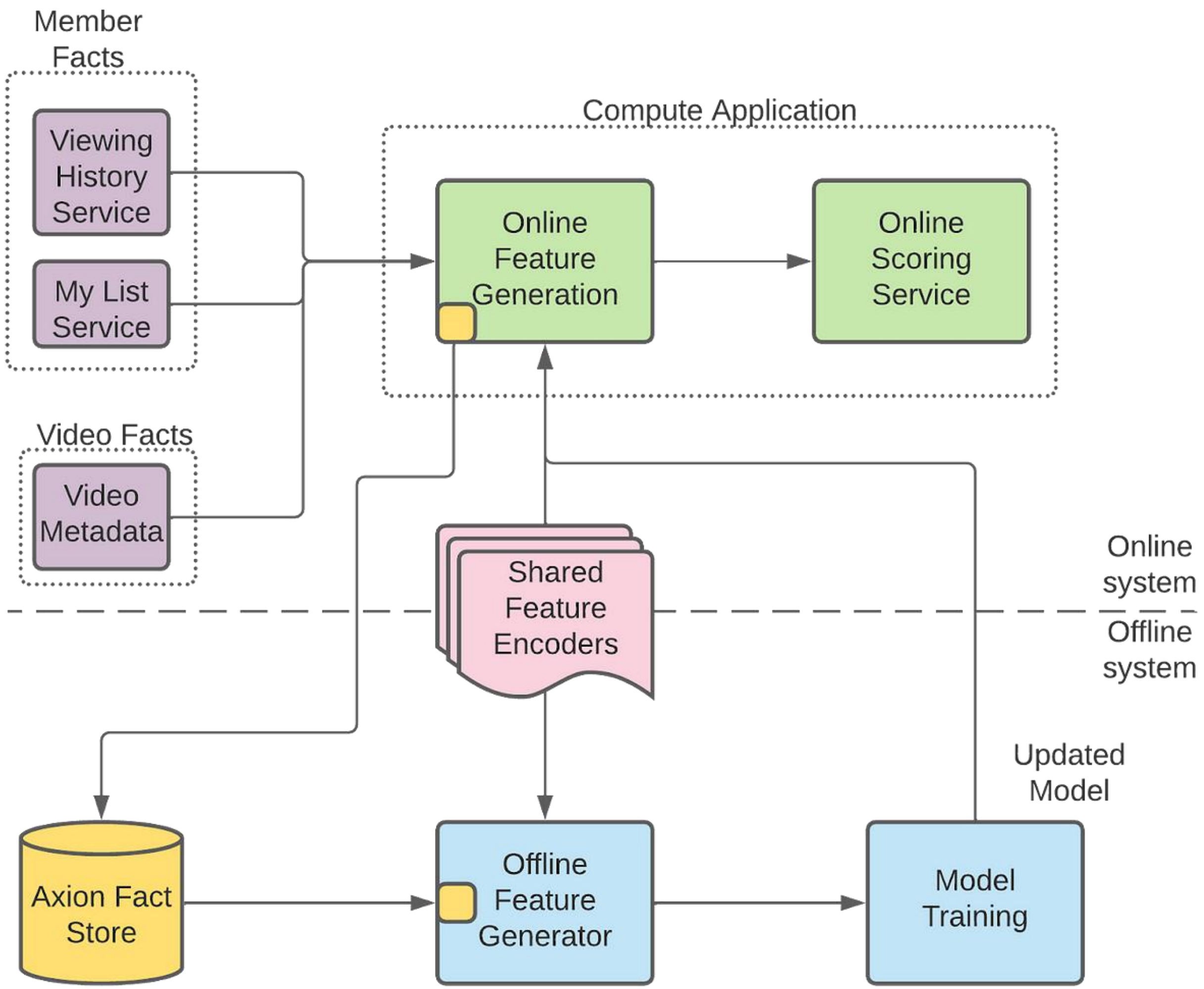
SORINTTEK



Netflix Fact Store

https://netflixtechblog.com/evolution-of-ml-fact-store-5941d3231762

"We make sure there is **no training/serving skew** by using the same data and the code for online and offline feature generation"



Grazie

scardis@latek.it

SORINT TEK