AIM301-300 LEVEL

Implement MLOps practices on AWS

Dr Sokratis Kartakis (he/him)
Snr MLOps Specialist Solutions Architect, EMEA
Amazon Web Services

Ajish Palakadan (he/him)
Practice Lead, Modern Data Platform
Reply

Agenda

What is MLOps
People, Processes, Technology

Why MLOps
Business Benefits and KPIs

MLOps People & Processes
Team Structure and Personas

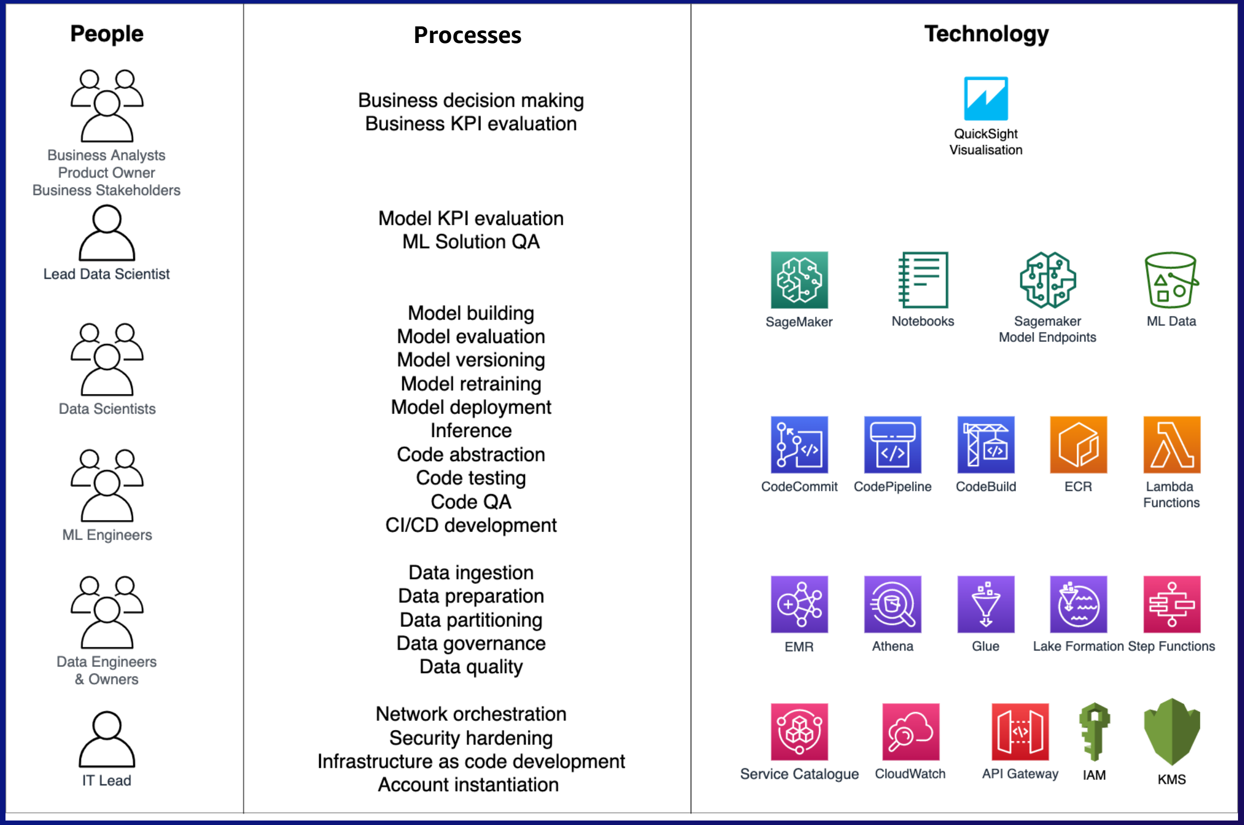
MLOps Foundation Roadmap - Technology
MLOps Maturity: Initial, Repeatable, Reliable, Scalable

MLOps Journey with Data Reply
Partnership, MLOps Accelerator and Customer Success Stories

Customer Challenges, ML Lifecycle and Personas

CONSIDERATIONS & CHALLENGES LEAD TO ML AND OPERATIONS (MLOPS)

- Culture
- Lack of cross-functional teams
- Priorities & needs (personas)
- Organizational structure
- Skillsets
- Unique aspects of ML lifecycle

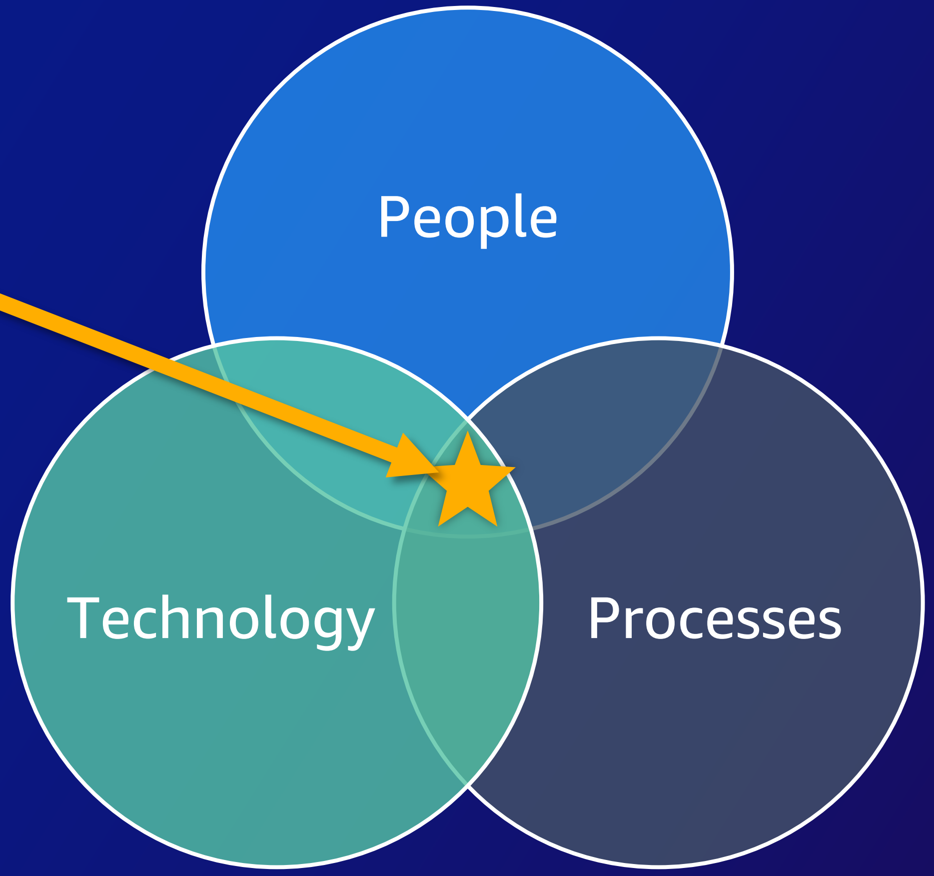


What is MLOps?

MLOps
Machine Learning
& Operations

The combination of people, processes, and technology to productionize ML solutions efficiently.

MLOps Definition



Why MLOps?
Expected Business Value

MLOps Foundation Expected Outcomes

STANDARDIZE OPERATIONS AND INFRASTRUCTURE FOR YOUR DATA SCIENCE

	Business Goal	Technical Metric	Before MLOps	MLOps Expected Outcomes	Business Value
1	Be more efficient in delivery	Time to value (from idea to production)	up to 12 months	< 3 months	Improve Speed-to-Value by 4x

MLOps Foundation Expected Outcomes

STANDARDIZE OPERATIONS AND INFRASTRUCTURE FOR YOUR DATA SCIENCE

	Business Goal	Technical Metric	Before MLOps	MLOps Expected Outcomes	Business Value
1	Be more efficient in delivery	Time to value (from idea to production)	up to 12 months	< 3 months	Improve Speed-to-Value by 4x
2	Simplify route-to-live	Time to productionize existing ML use cases	3-6 months	< 2 weeks	Reduce FTE overhead in average 8x

MLOps Foundation Expected Outcomes

STANDARDIZE OPERATIONS AND INFRASTRUCTURE FOR YOUR DATA SCIENCE

	Business Goal	Technical Metric	Before MLOps	MLOps Expected Outcomes	Business Value
1	Be more efficient in delivery	Time to value (from idea to production)	up to 12 months	< 3 months	Improve Speed-to-Value by 4x
2	Simplify route-to-live	Time to productionize existing ML use cases	3-6 months	< 2 weeks	Reduce FTE overhead in average 8x
3	Standardize infrastructure, data, & code	% Template driven development	n/a	> 85%	Focus on innovation increasing re-usability by 85%

MLOps Foundation Expected Outcomes

STANDARDIZE OPERATIONS AND INFRASTRUCTURE FOR YOUR DATA SCIENCE

	Business Goal	Technical Metric	Before MLOps	MLOps Expected Outcomes	Business Value
1	Be more efficient in delivery	Time to value (from idea to production)	up to 12 months	< 3 months	Improve Speed-to-Value by 4x
2	Simplify route-to-live	Time to productionize existing ML use cases	3-6 months	< 2 weeks	Reduce FTE overhead in average 8x
3	Standardize infrastructure, data, & code	% Template driven development	n/a	> 85%	Focus on innovation increasing re-usability by 85%
4	Standardize onboarding of new teams and ML use cases	Time to instantiate a new MLOps infrastructure & ML projects	40 days	< 1 hours	Accelerate ML adoption across all business areas

MLOps Foundation Expected Outcomes

STANDARDIZE OPERATIONS AND INFRASTRUCTURE FOR YOUR DATA SCIENCE

	Business Goal	Technical Metric	Before MLOps	MLOps Expected Outcomes	Business Value
1	Be more efficient in delivery	Time to value (from idea to production)	up to 12 months	< 3 months	Improve Speed-to-Value by 4x
2	Simplify route-to-live	Time to productionize existing ML use cases	3-6 months	< 2 weeks	Reduce FTE overhead in average 8x
3	Standardize infrastructure, data, & code	% Template driven development	n/a	> 85%	Focus on innovation increasing re-usability by 85%
4	Standardize onboarding of new teams and ML use cases	Time to instantiate a new MLOps infrastructure & ML projects	40 days	< 1 hours	Accelerate ML adoption across all business areas
5	Ensure high security standards	Execute the ML solutions without internet access in a private cloud	n/a	No internet	Your data is safe in your private cloud

MLOps Foundation Expected Outcomes

STANDARDIZE OPERATIONS AND INFRASTRUCTURE FOR YOUR DATA SCIENCE

	Business Goal	Technical Metric	Before MLOps	MLOps Expected Outcomes	Business Value
1	Be more efficient in delivery	Time to value (from idea to production)	up to 12 months	< 3 months	Improve Speed-to-Value by 4x
2	Simplify route-to-live	Time to productionize existing ML use cases	3-6 months	< 2 weeks	Reduce FTE overhead in average 8x
3	Standardize infrastructure, data, & code	% Template driven development	n/a	> 85%	Focus on innovation increasing re-usability by 85%
4	Standardize onboarding of new teams and ML use cases	Time to instantiate a new MLOps infrastructure & ML projects	40 days	< 1 hours	Accelerate ML adoption across all business areas
5	Ensure high security standards	Execute the ML solutions without internet access in a private cloud	n/a	No internet	Your data is safe in your private cloud

MLOps Foundation Expected Outcomes

STANDARDIZE OPERATIONS AND INFRASTRUCTURE FOR YOUR DATA SCIENCE

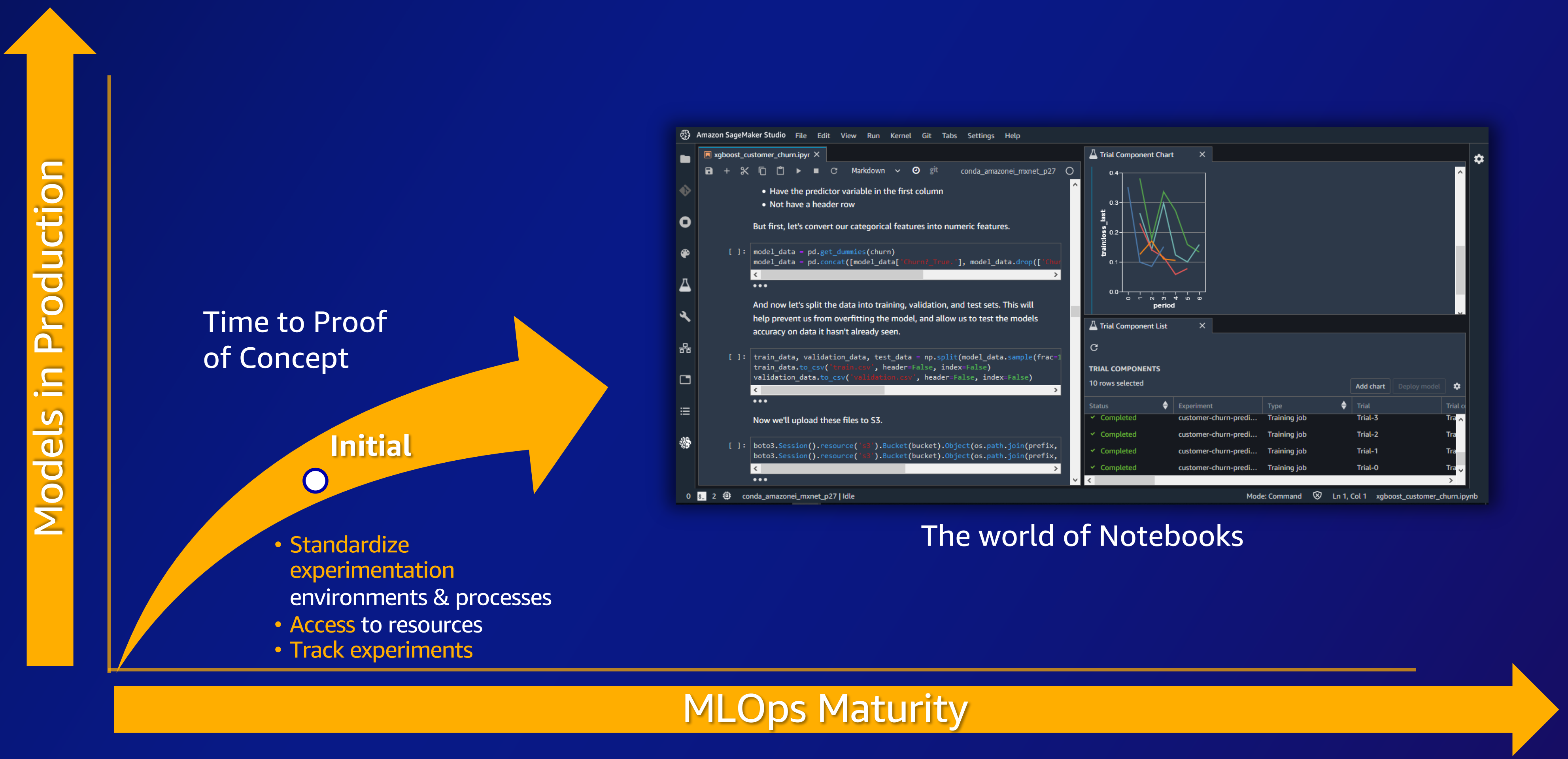
	Business Goal	Technical Metric	Before MLOps	MLOps Expected Outcomes	Business Value
1	Be more efficient in delivery	Time to value (from idea to production)	up to 12 months	< 3 months	Improve Speed-to-Value by 4x
2	Simplify route-to-live	Time to productionize existing ML use cases	3-6 months	< 2 weeks	Reduce FTE overhead in average 8x
3	Standardize infrastructure, data, & code	% Template driven development	n/a	> 85%	Focus on innovation increasing re-usability by 85%
4	Standardize onboarding of new teams and ML use cases	Time to instantiate a new MLOps infrastructure & ML projects	40 days	< 1 hours	Accelerate ML adoption across all business areas
5	Ensure high security standards	Execute the ML solutions without internet access in a private cloud	n/a	No internet	Your data is safe in your private cloud
	Reduce platform, people and operation costs				

Customer references building MLOps foundation and business benefits:

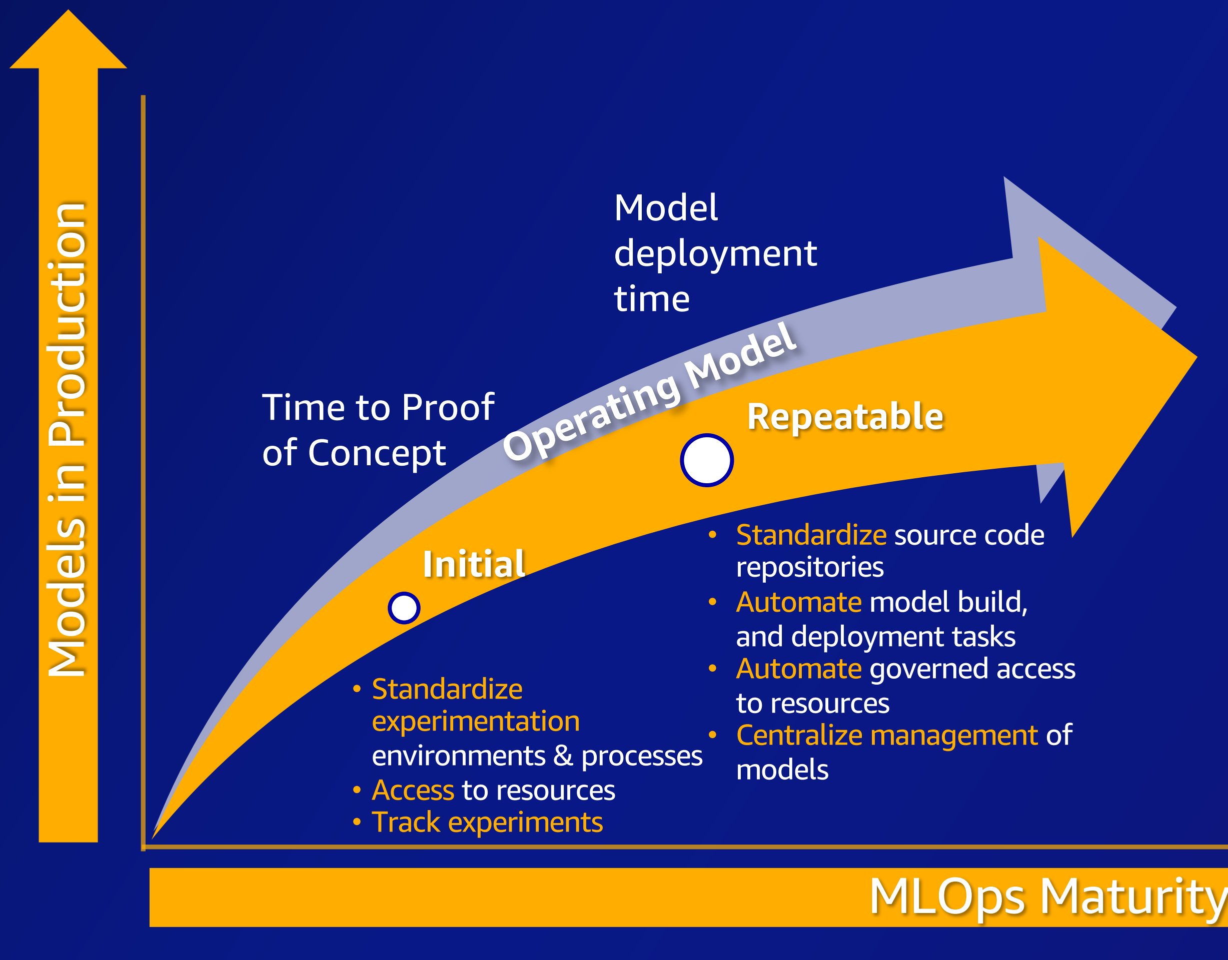
- NatWest: https://aws.amazon.com/solutions/case-studies/natwest-group-case-study
- BP: https://aws.amazon.com/solutions/case-studies/bp-machine-learning-case-study

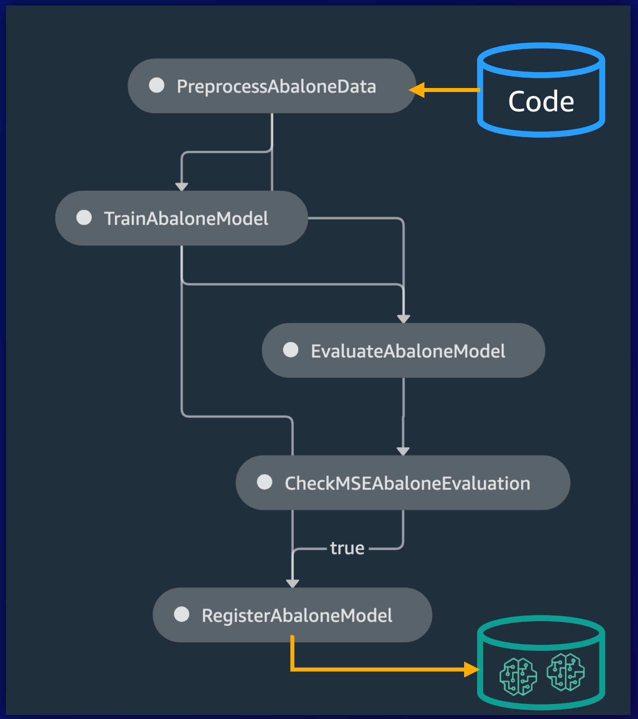
How to Mature on MLOps?

MLOps Maturity Model



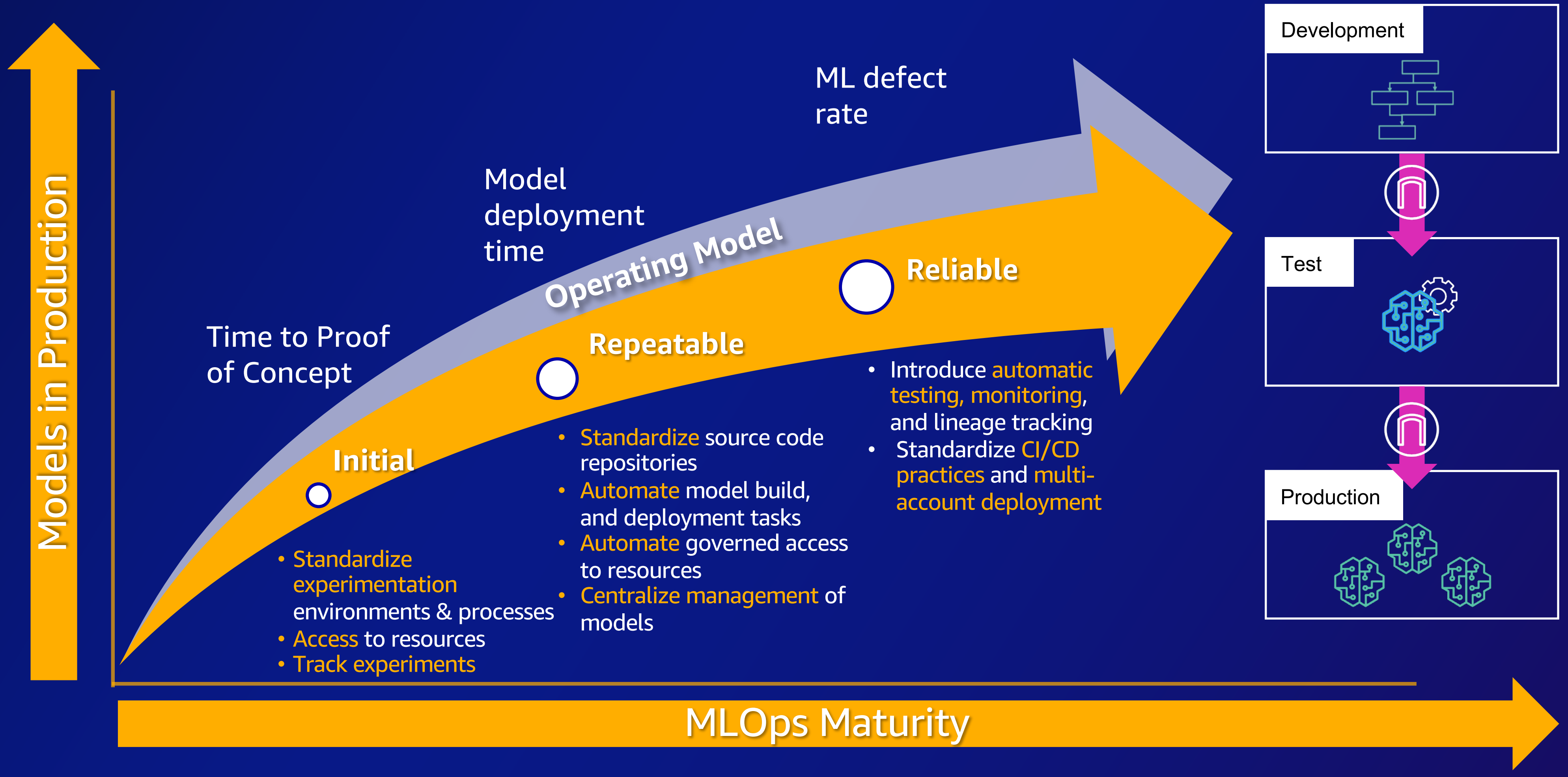
MLOps Maturity Model



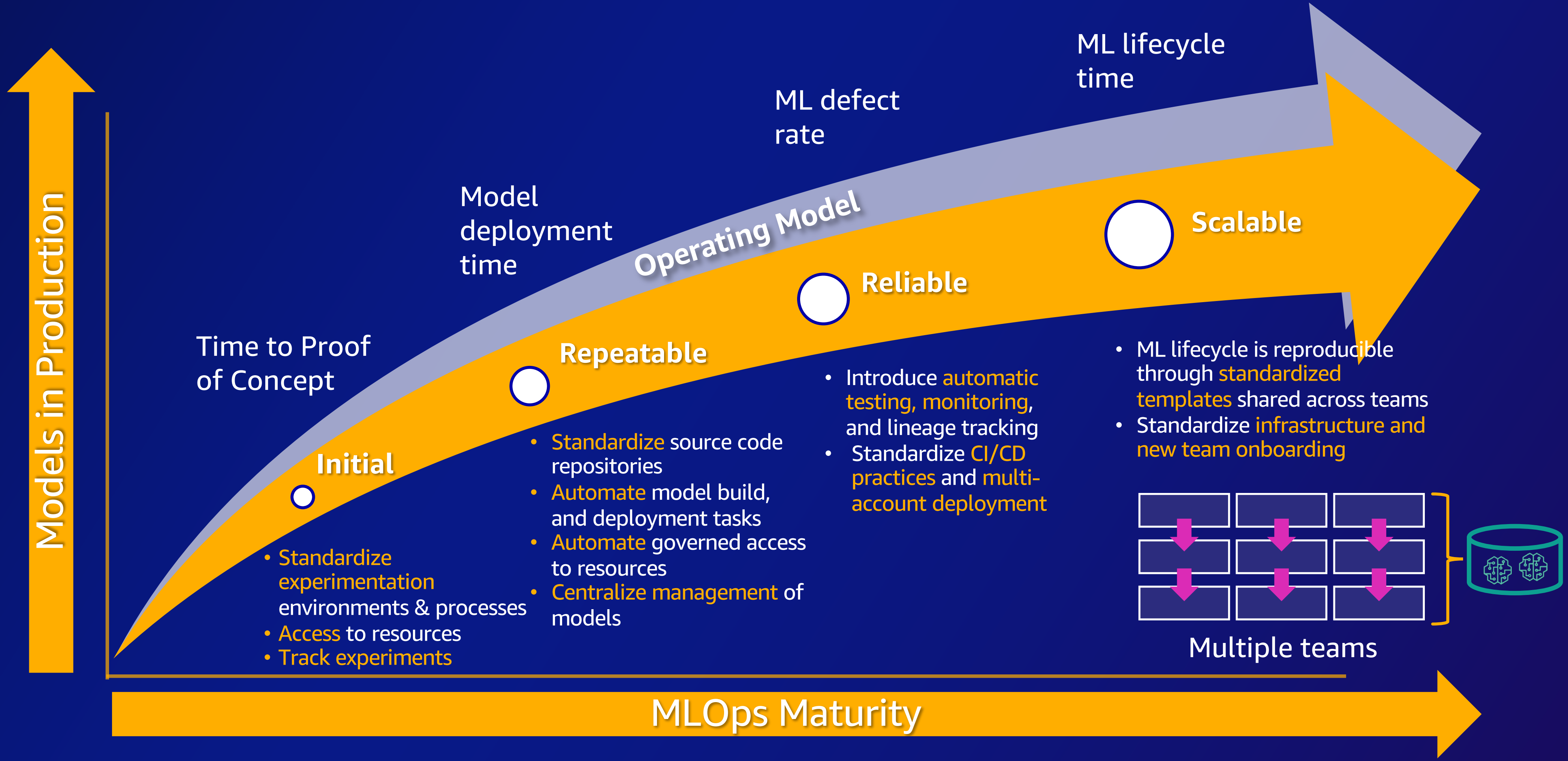


The world of automation and ML pipelines

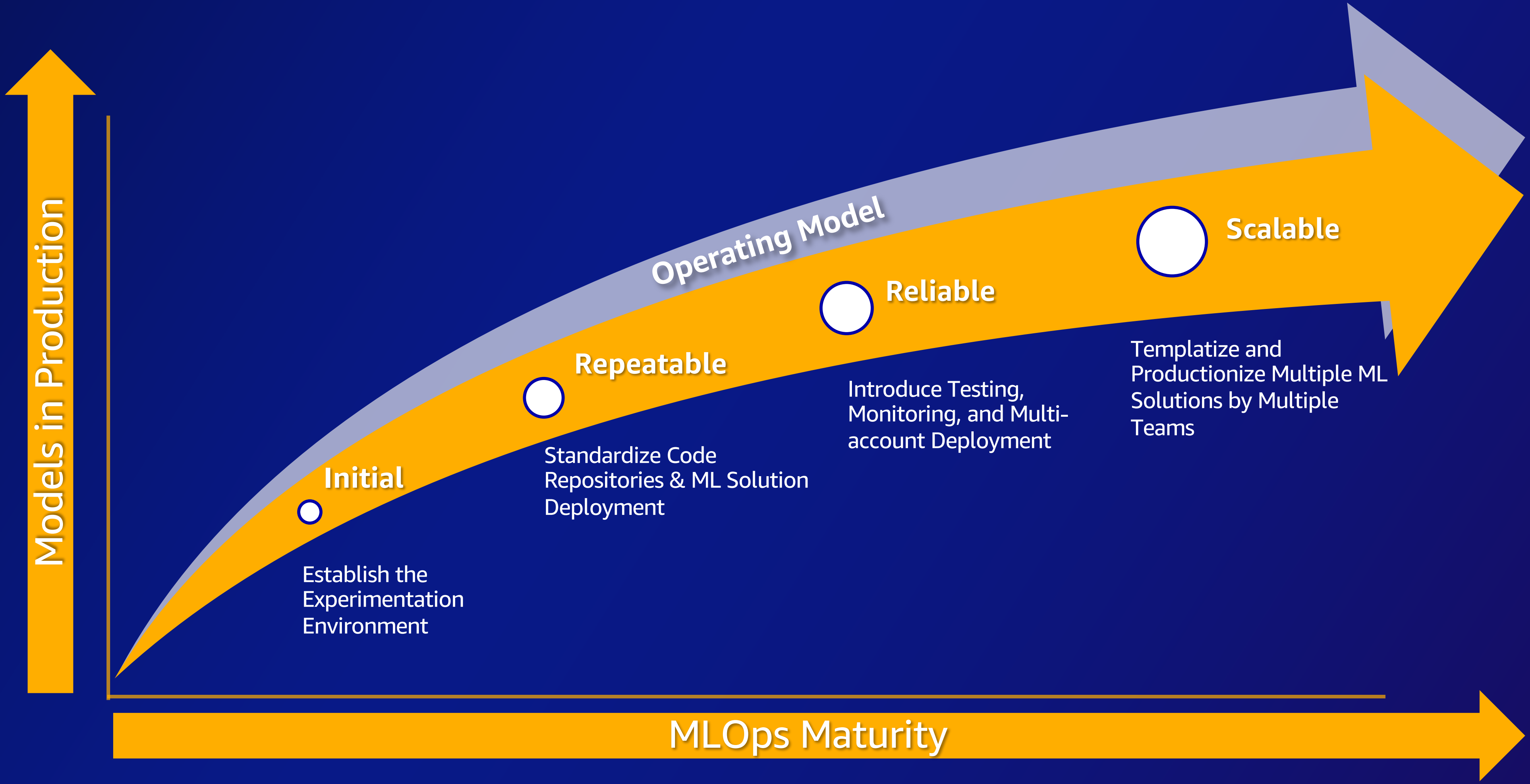
MLOps Maturity Model



MLOps Maturity Model



MLOps Maturity Model



MLOps People & Processes

MLOps Key Personas and Roles

Advance Analytics Team
Data Lake

Data Engineer
Prepare & Ingest data building ETL pipelines

Data Owners
Manage data sharing and provide access

Data Science Team
Experimentation & MLOps

Data Scientist
Create the best ML models to solve business problems

ML Engineer
Collaborate with DS to productionize ML

Platform Team
Secure Cloud/Data/ML Platform

MLOps Engineer/Admin
Standardize CI/CD, user/service role, model consumption, testing and deployment methodology

Security
Assess data, user, and service access creating policies and guardrails

Architects/ SysOps Engineer
Standardize account infrastructure, connectivity, user roles implementation

Business
Viz Dashboards, ML Adoption, & ROI

Business Stakeholder
Product Owners
Define business problem, business KPIs, and make business decisions

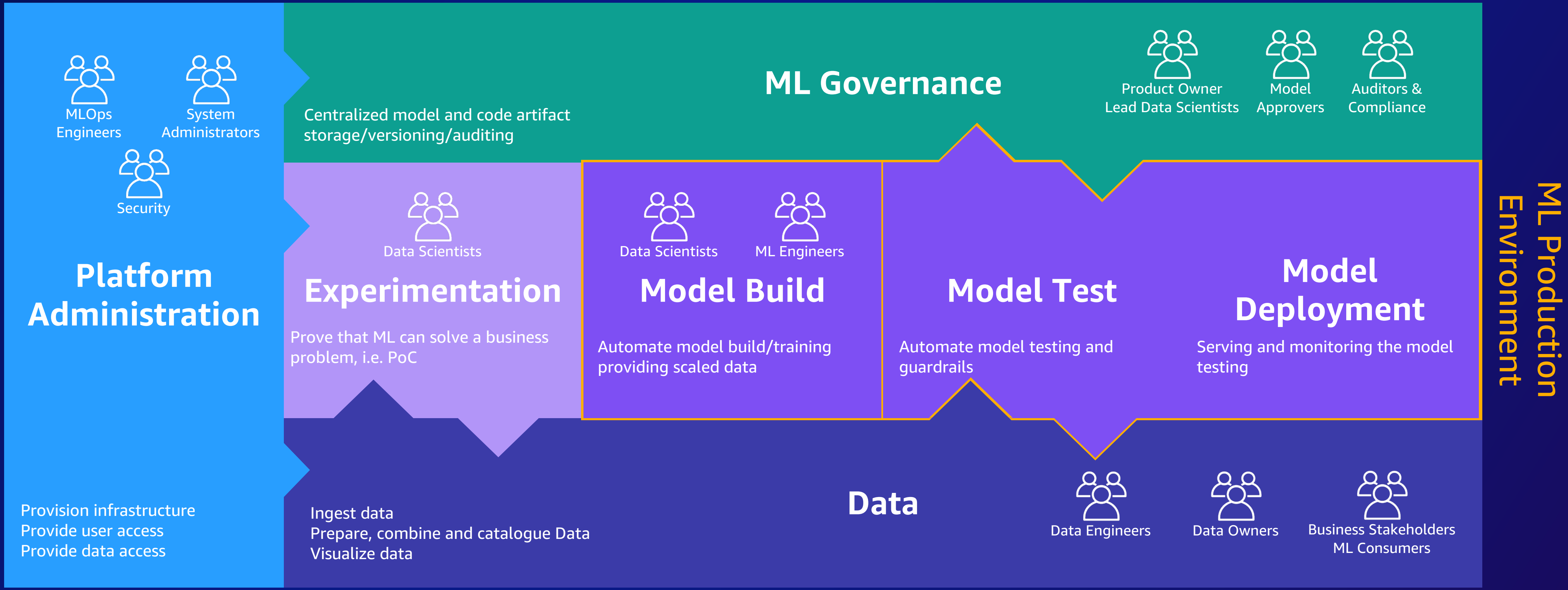
Business Stakeholder
Data & ML Consumers
Consumers of ML results from other BUs, driving business decision making

Risk & Compliance
Approve & Review Models

Auditors/Risk & Compliance
Review models, data sources, code artifacts

MLOps Foundation People & Processes

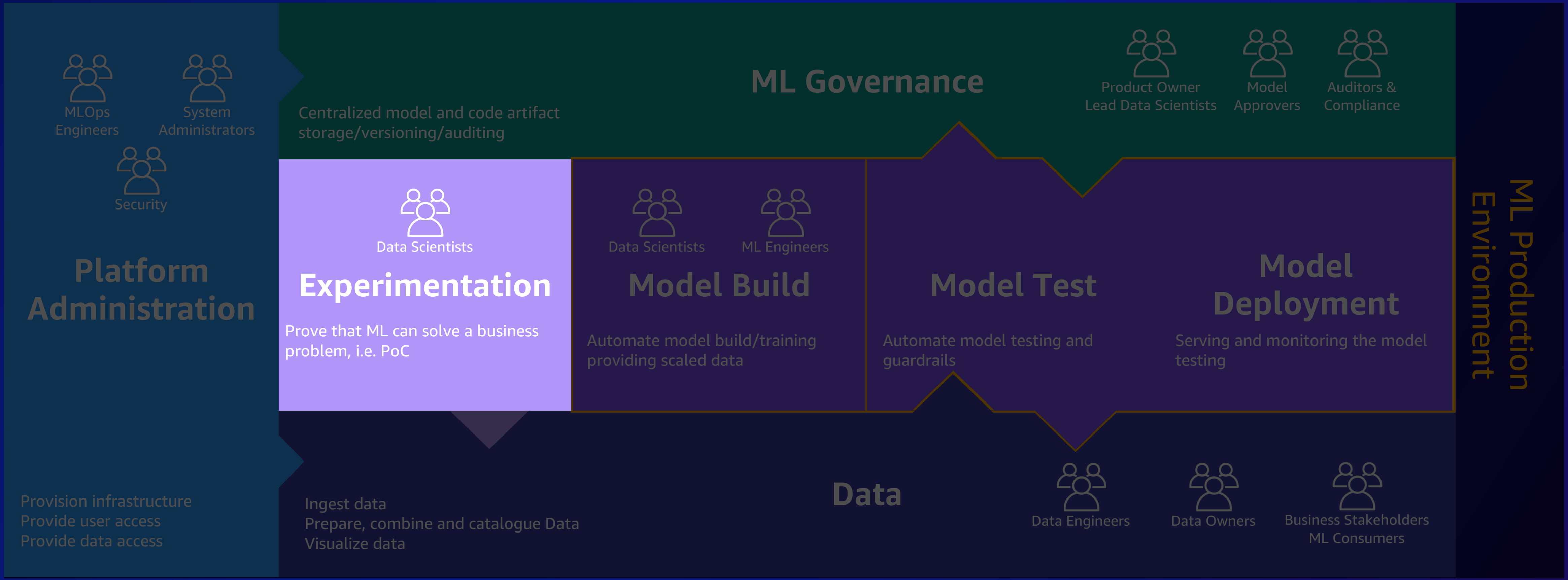
SEPARATION OF CONCERNS IS KEY FOR SUCCESS



MLOps Foundation Roadmap
Technology

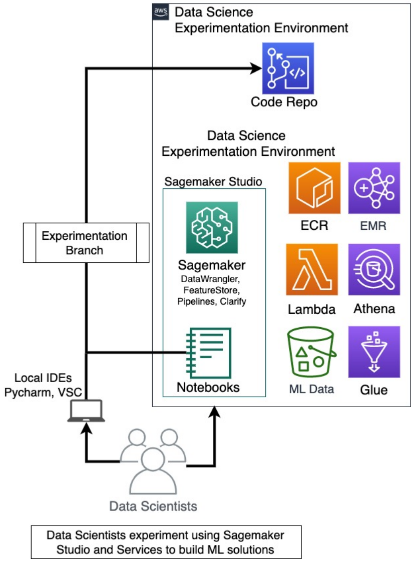
MLOps Foundation People & Processes

INITIAL PHASE



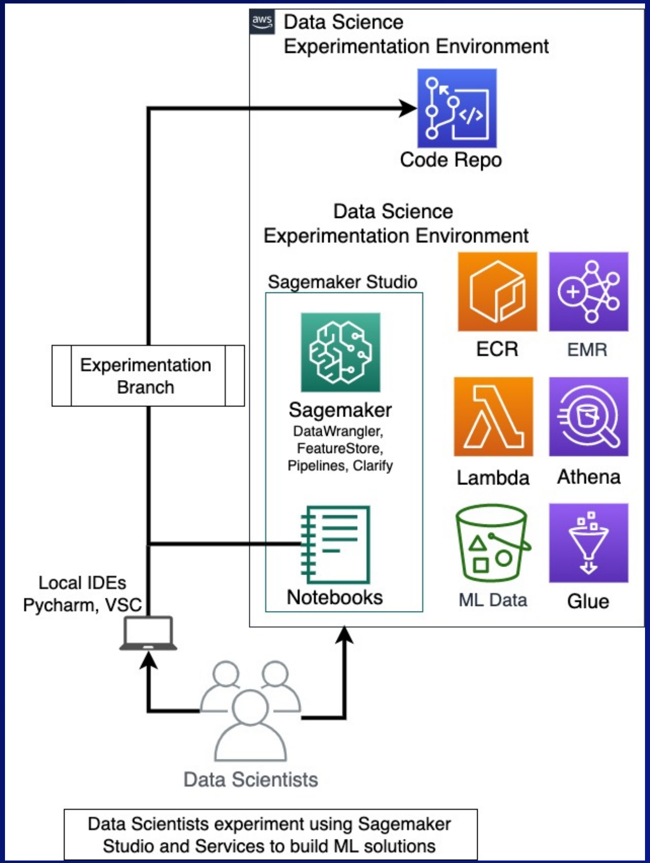
MLOPs Initial Phase

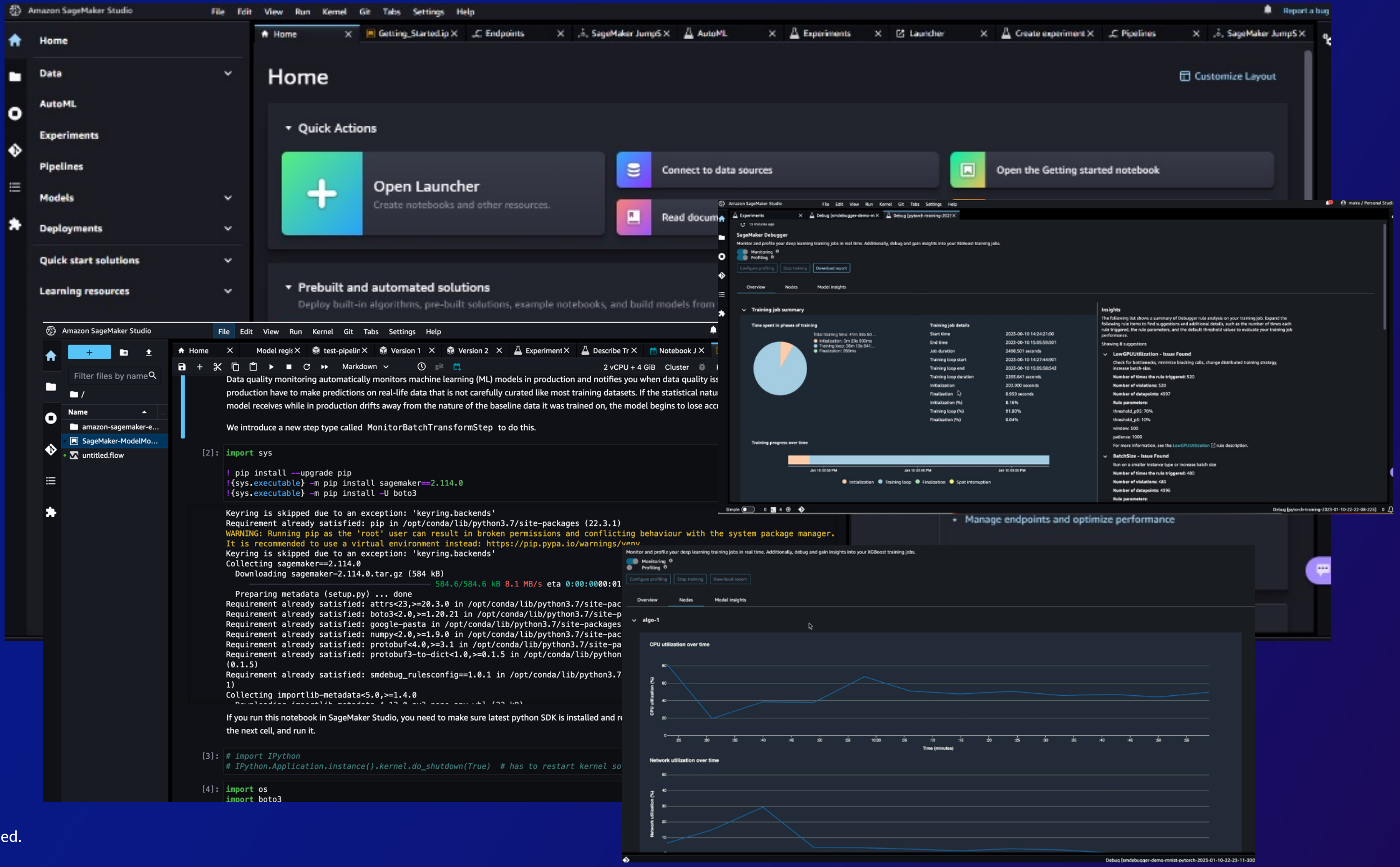
ML EXPERIMENTATION ON AWS USING AMAZON SAGEMAKER STUDIO NOTEBOOKS



Amazon SageMaker Studio

ML EXPERIMENTATION ON AWS USING AMAZON SAGEMAKER STUDIO NOTEBOOKS



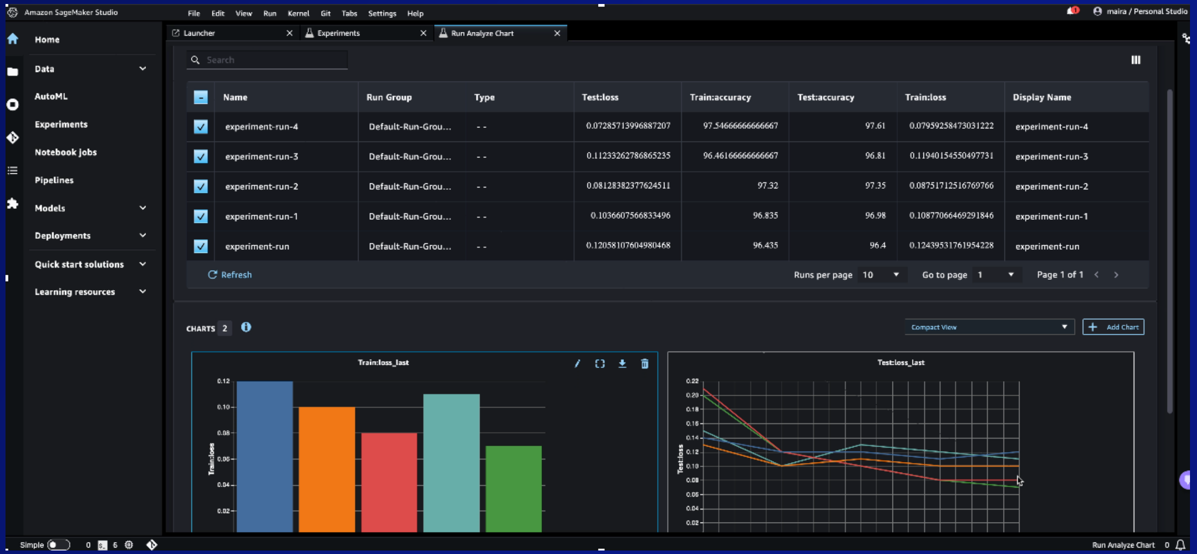


Amazon SageMaker Experiments

ORGANIZE, TRACK, AND COMPARE MACHINE LEARNING EXPERIMENTS

```python
from sagemaker.experiments.run import Run
from sagemaker.session import Session
# create an experiment, start a new run and log a parameter
experiment_name = "experiment_name"
run_name = "run_name"
with Run(
    experiment_name=experiment_name,
    run_name=run_name,
    sagemaker_session=Session()
) as run:
    run.log_parameters({
        "normalization_mean": 0.1307,
        "normalization_std": 0.3081,
     })
```

```python
from sagemaker.experiments.run import load_run
from sagemaker.session import Session
# load the run from context
with load_run(
    sagemaker_session=Session()
) as run:
    run.log_metric("precision", 0.9, step=step)
    run.log_confusion_matrix(
        target,
        pred,
        "Confusion-Matrix-Test-Data"
    )
```



Tracking at scale

Track parameters and metrics across experiments and users

Custom organization

Organize experiments by teams, goals, and hypotheses

Visualization

Easily visualize experiments and compare

Metrics and logging

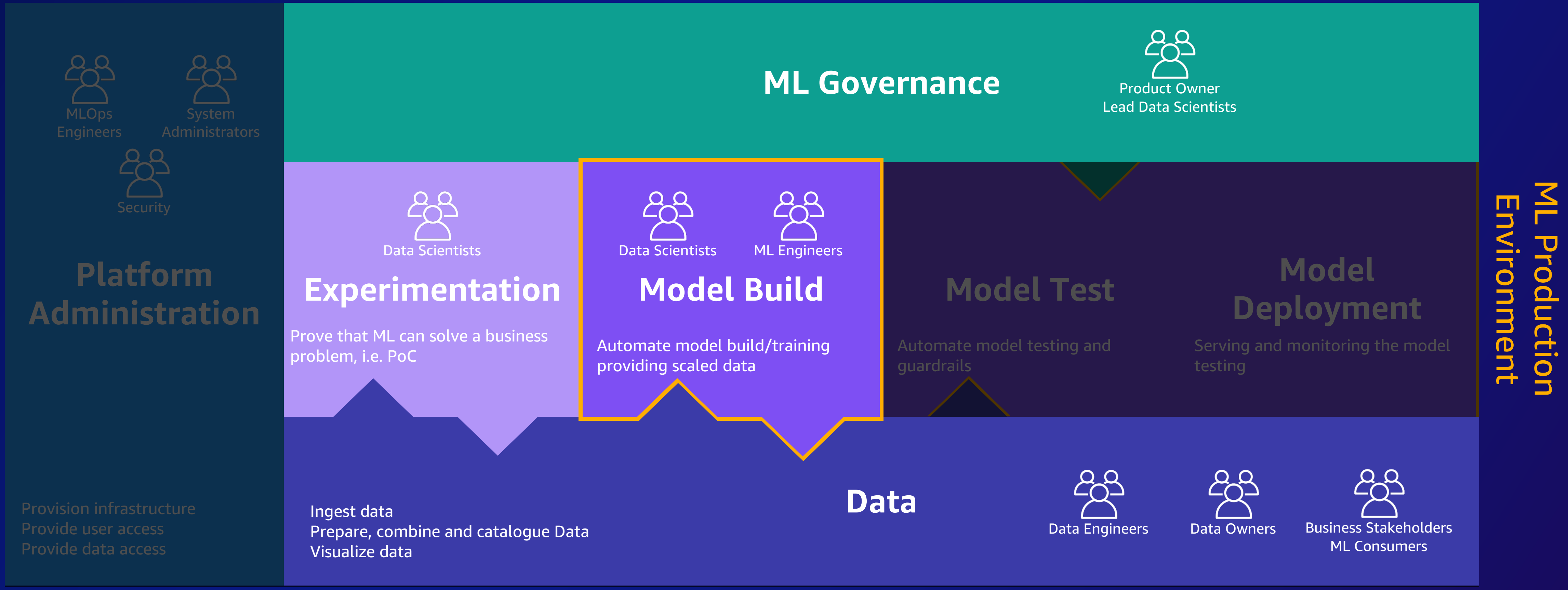
Log custom metrics using the Python SDK and APIs

Fast iteration

Quickly go back and forth, and maintain high-quality

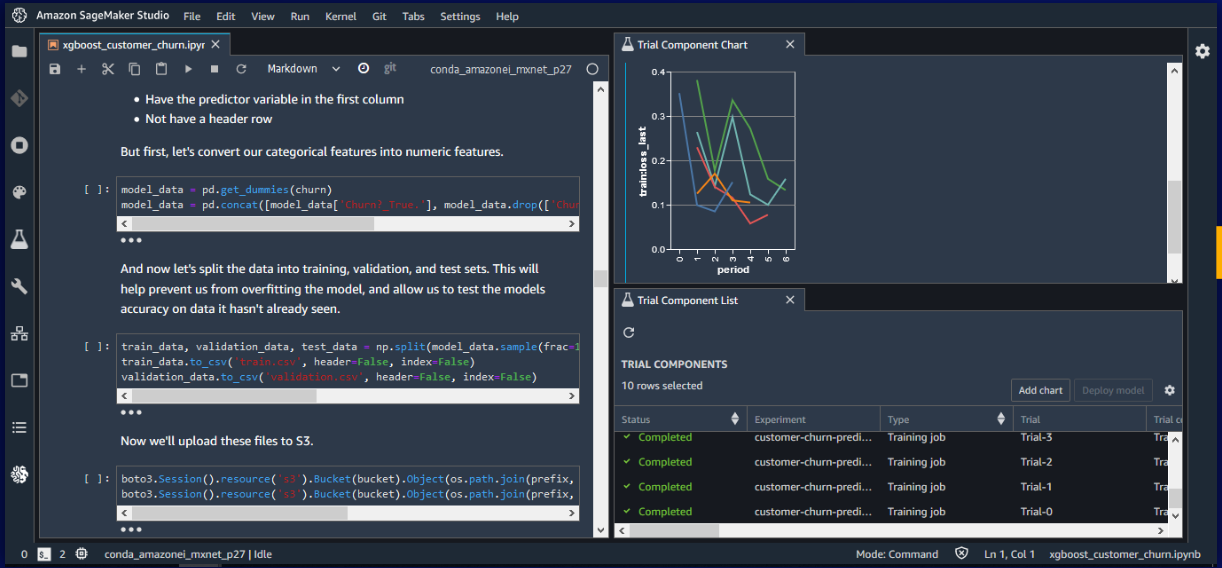
MLOps Foundation People & Processes

REPEATABLE PHASE

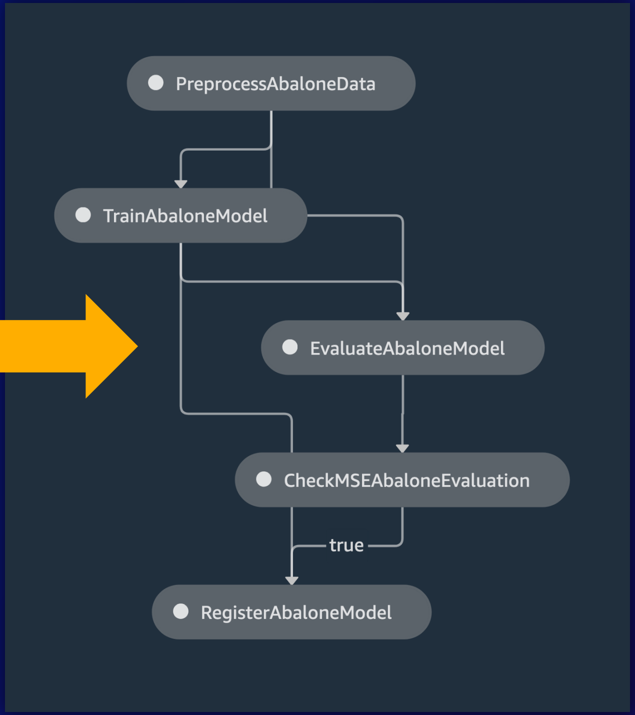


ML Solution Lifecycle Automation

CREATING AUTOMATED WORKFLOWS ON AWS – ML PIPELINES



Move from the world of Notebooks…

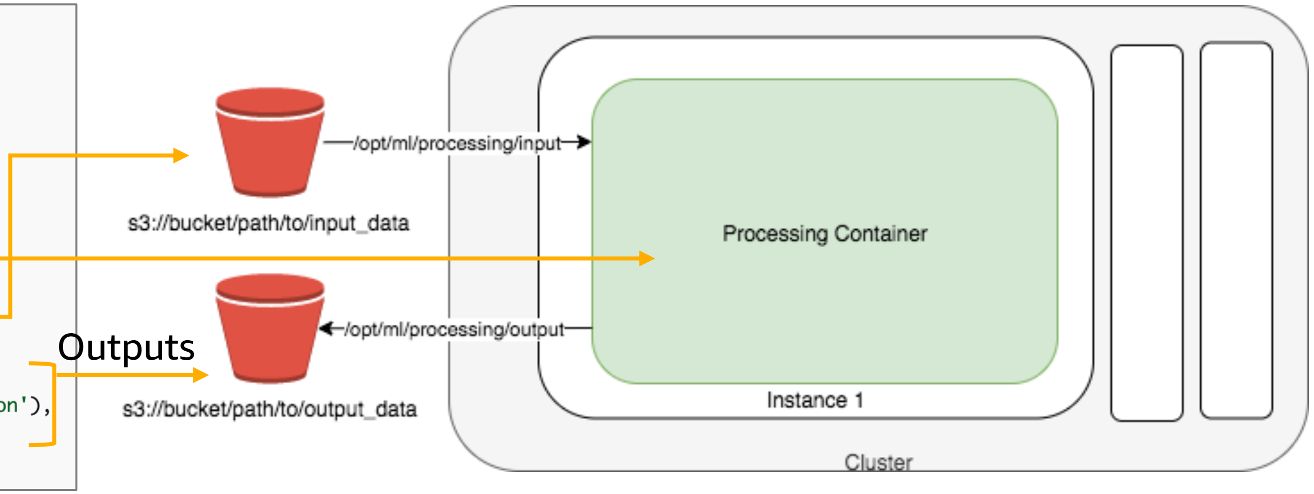


…to the world of automation

Amazon SageMaker Processing

PREPROCESS DATA & OFFLOAD DATA SCIENTISTS ON MANAGING CONTAINERS

```python
from sagemaker.sklearn.processing import SKLearnProcessor
from sagemaker.processing import ProcessingInput, ProcessingOutput

sklearn_processor = SKLearnProcessor(framework_version='0.20.0',
                                     role=role,
                                     instance_type='ml.m5.xlarge',
                                     instance_count=1)
sklearn_processor.run(code='preprocessing.py',
                      inputs=[ProcessingInput(
                        source='s3://path/to/my/input-data.csv',
                        destination='/opt/ml/processing/input')],
                      outputs=[ProcessingOutput(source='/opt/ml/processing/output/train'),
                               ProcessingOutput(source='/opt/ml/processing/output/validation'),
                               ProcessingOutput(source='/opt/ml/processing/output/test')]
                     )
```



Supported pre-built processors: scikit learn, PySpark, HUGGING FACE, mxnet, TensorFlow, XGBoost + Custom Containers

Framework Processors

Amazon SageMaker Pipelines

AUTOMATED WORKFLOWS TO PRE-PROCESS, TRAIN, EVALUATION, & REGISTER MODELS

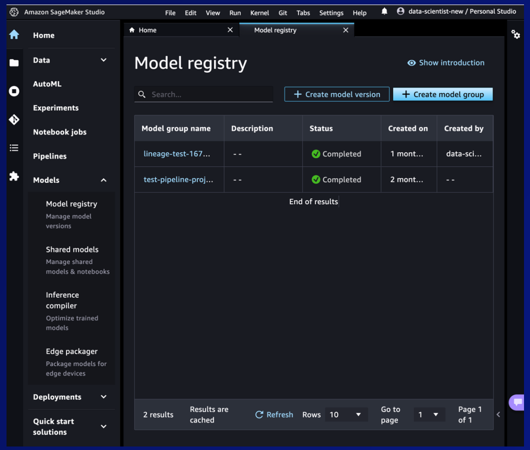

```python
from sagemaker.sklearn.processing import SKLearnProcessor

sklearn_processor = SKLearnProcessor(framework_version='0.20.0',
                                     role=<role>,

from sagemaker.workflow.pipeline import Pipeline

  pipeline_name = f"AbalonePipeline"
  pipeline = Pipeline(
      name=pipeline_name,
      parameters=[
          processing_instance_type,
          processing_instance_count,
          training_instance_type,
          model_approval_status,
          input_data,
          batch_data,
      ],
      steps=[step_process, step_train, step_eval, step_cond],
  )
        ProcessingOutput(output_name="test", source="/opt/ml/processing/test")
    ],
    code="abalone/preprocessing.py"
)
```

Step Types

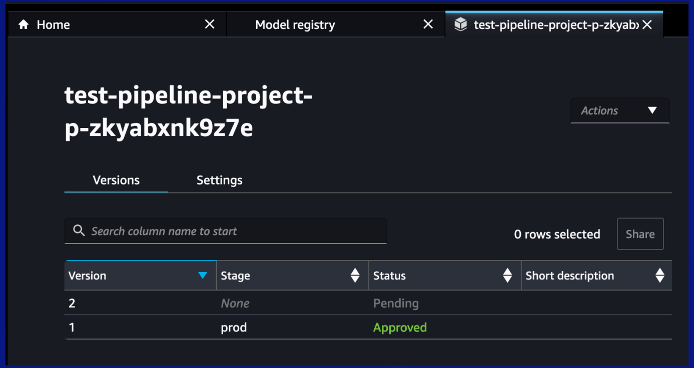
- Processing
- Training
- Tuning
- CreateModel
- RegisterModel
- Transform
- Condition
- Callback
- Lambda Function
- ClarifyCheck
- QualityCheck
- Amazon EMR
- Fail

Amazon SageMaker Model Registry

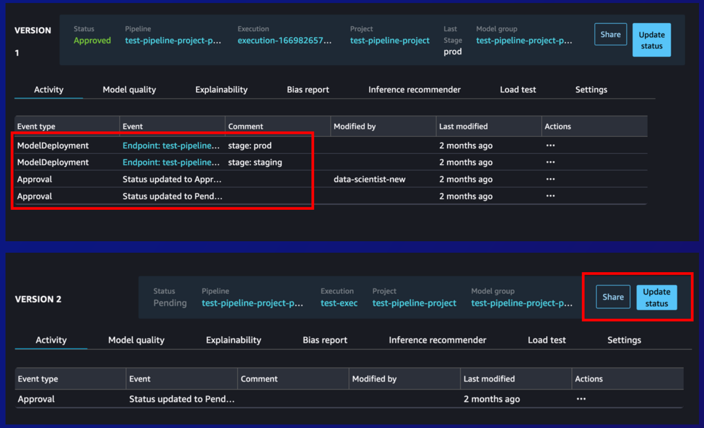
STORE, VERSION, AND TRIGGER MODEL PROMOTION



Create model groups in your model registry



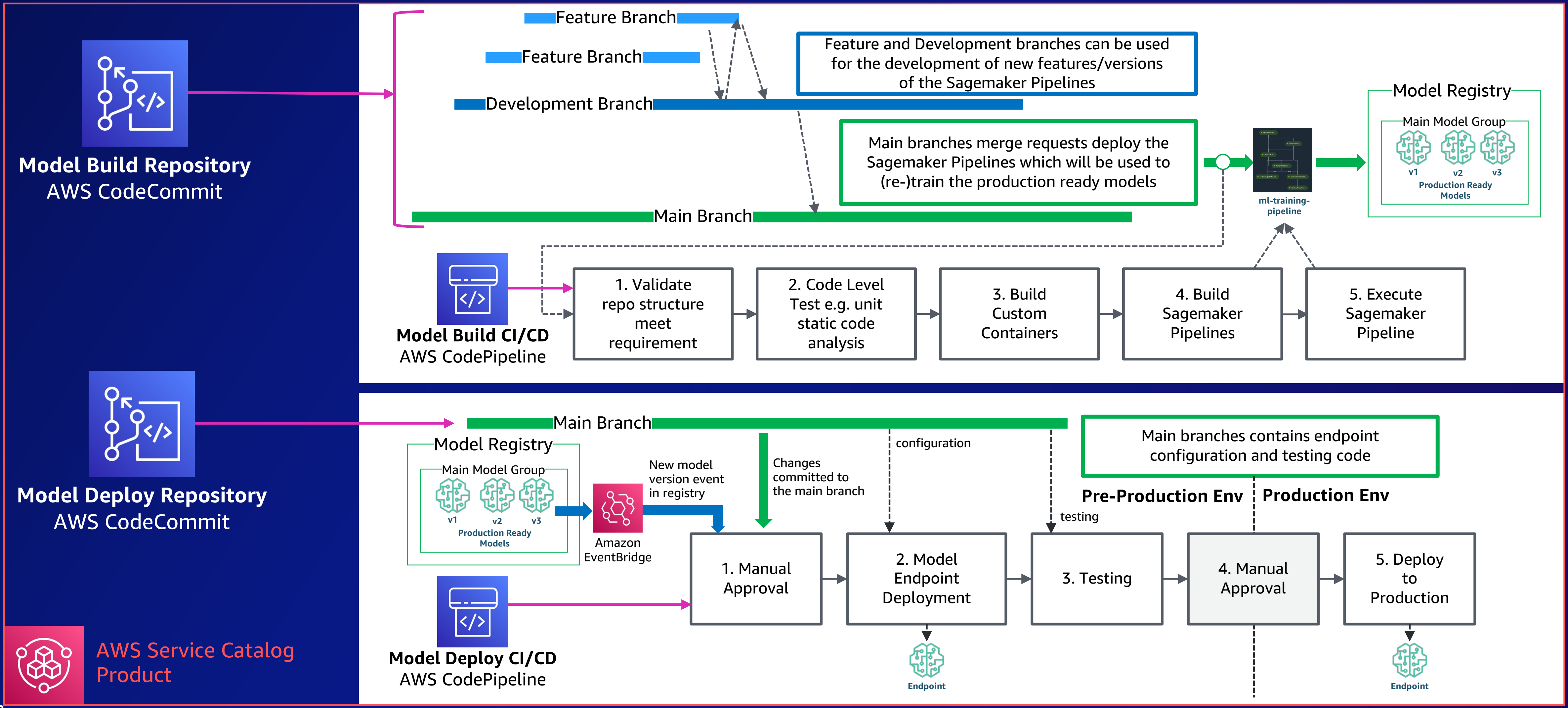
Benchmark and observe the model versions in your model group and promote versions of the model by changing their status



Track the activities, metrics, and settings per model version

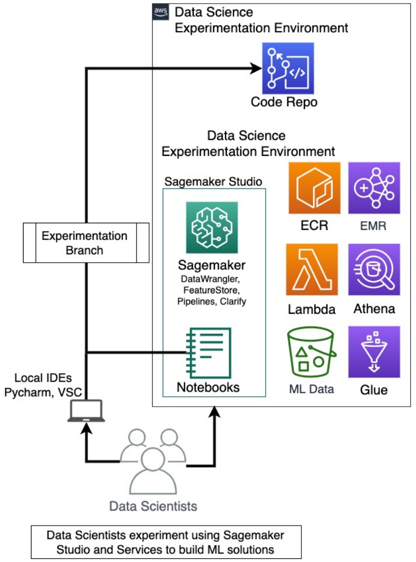
Amazon SageMaker Projects

EXAMPLE REPOSITORIES & CI/CD PIPELINES PER SAGEMAKER PROJECT



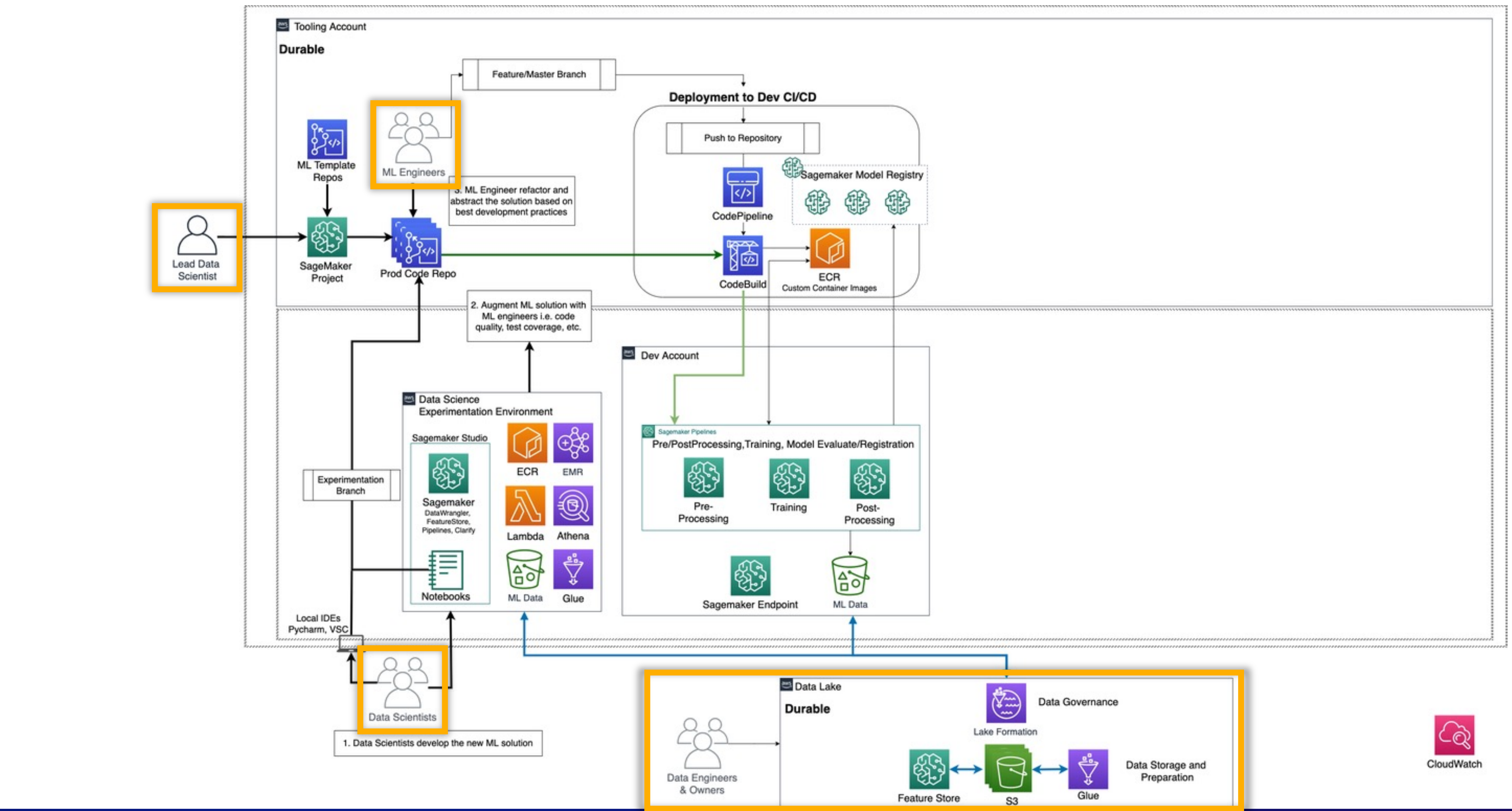
MLOPs Initial Phase

ML EXPERIMENTATION ON AWS USING AMAZON SAGEMAKER STUDIO NOTEBOOKS



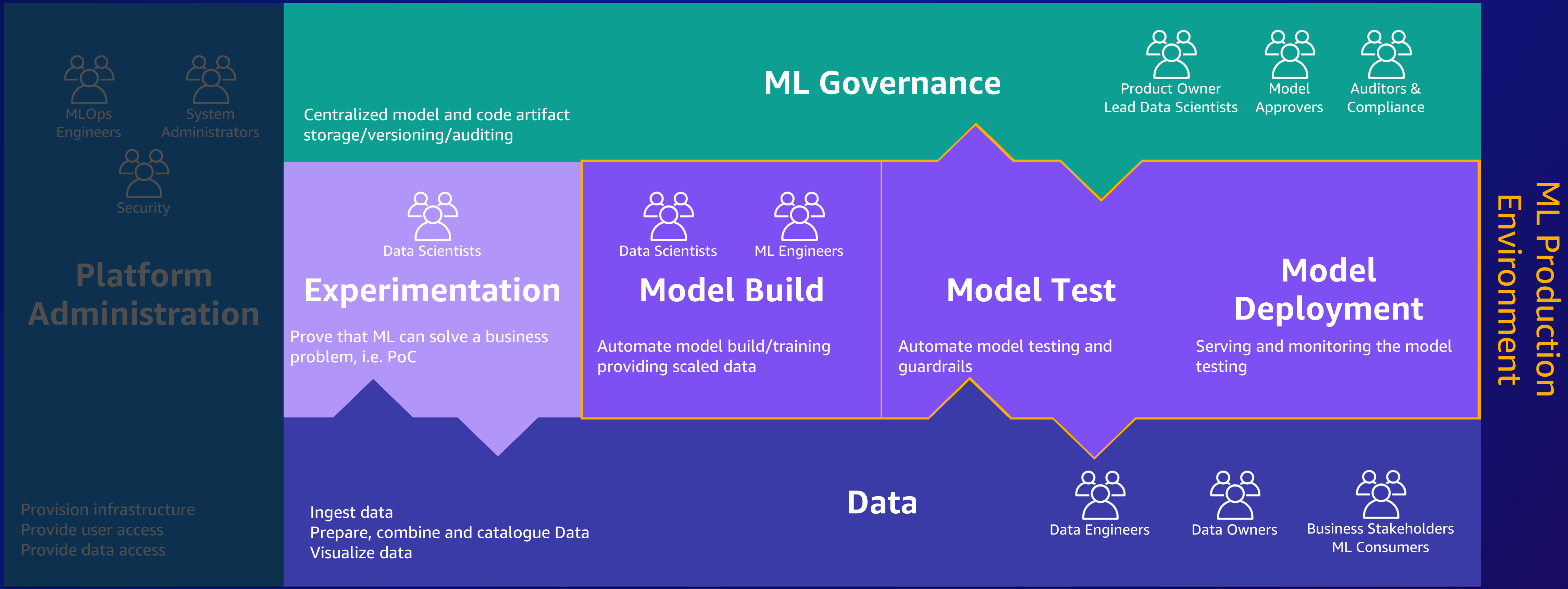
MLOPs Repeatable Phase

FROM RESEARCH NOTEBOOKS TO ML PIPELINES & AUTOMATION



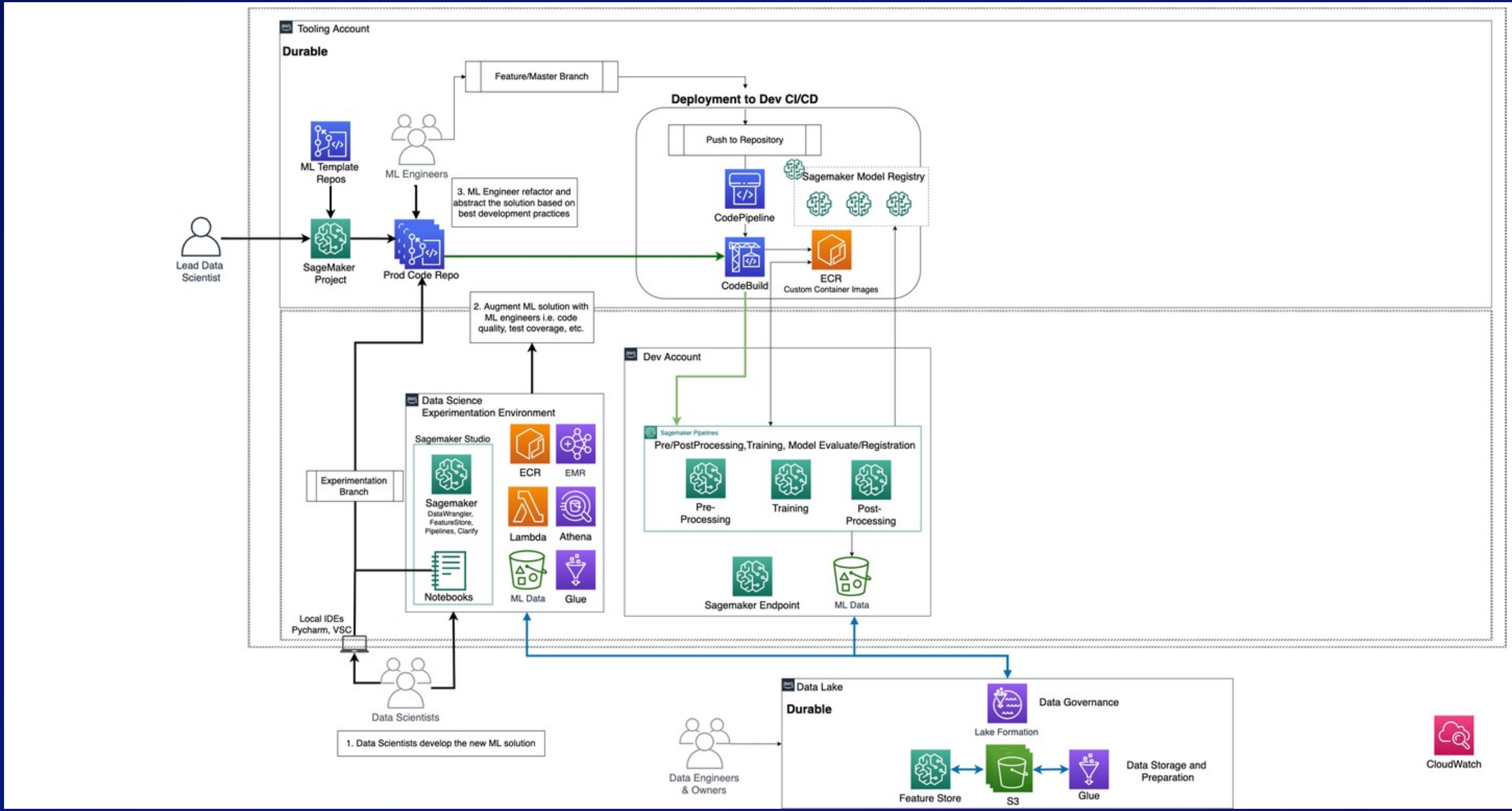
MLOps Foundation People & Processes

RELIABLE PHASE



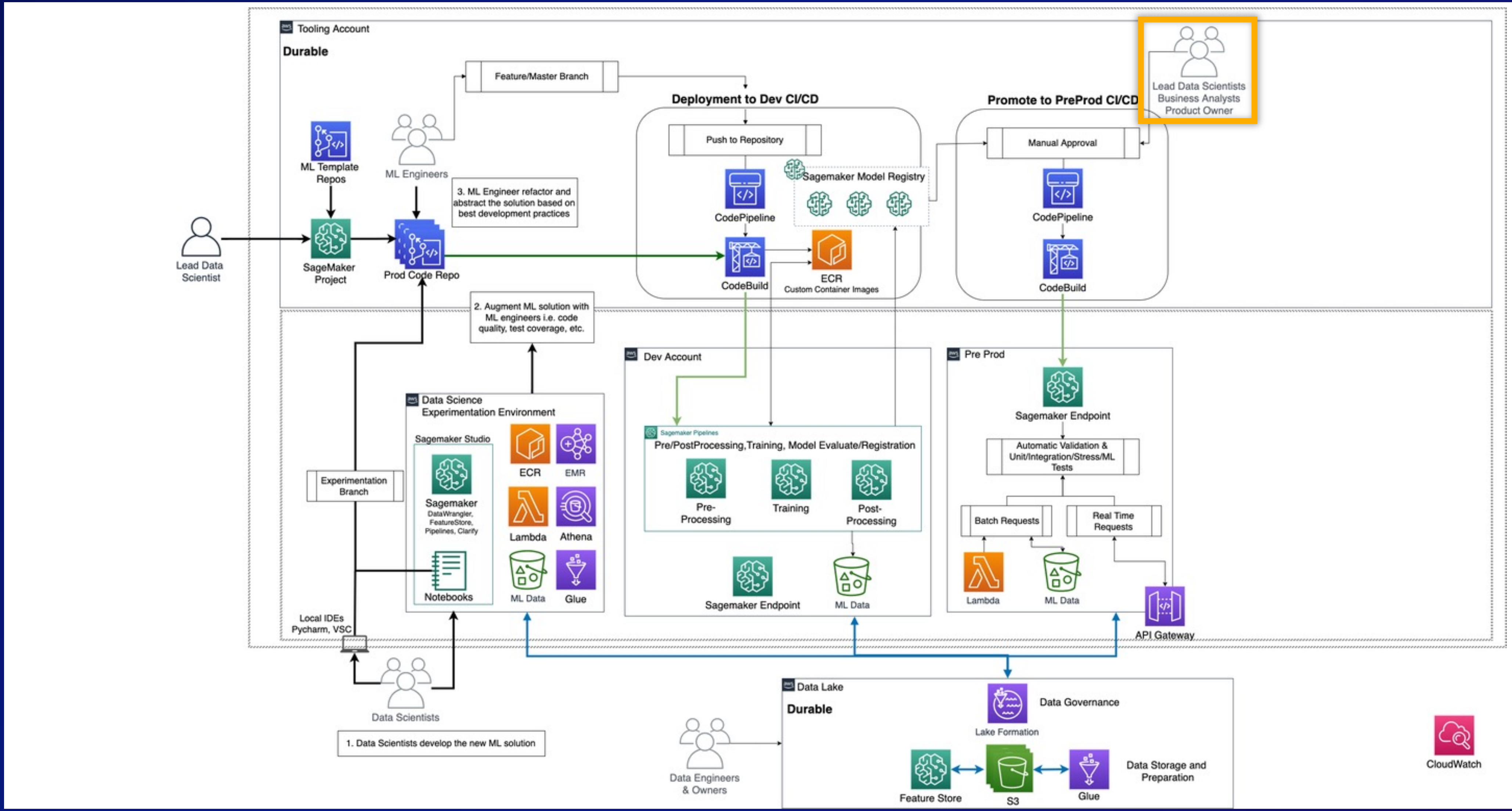
MLOPs Repeatable Phase

FROM RESEARCH NOTEBOOKS TO ML PIPELINES & AUTOMATION



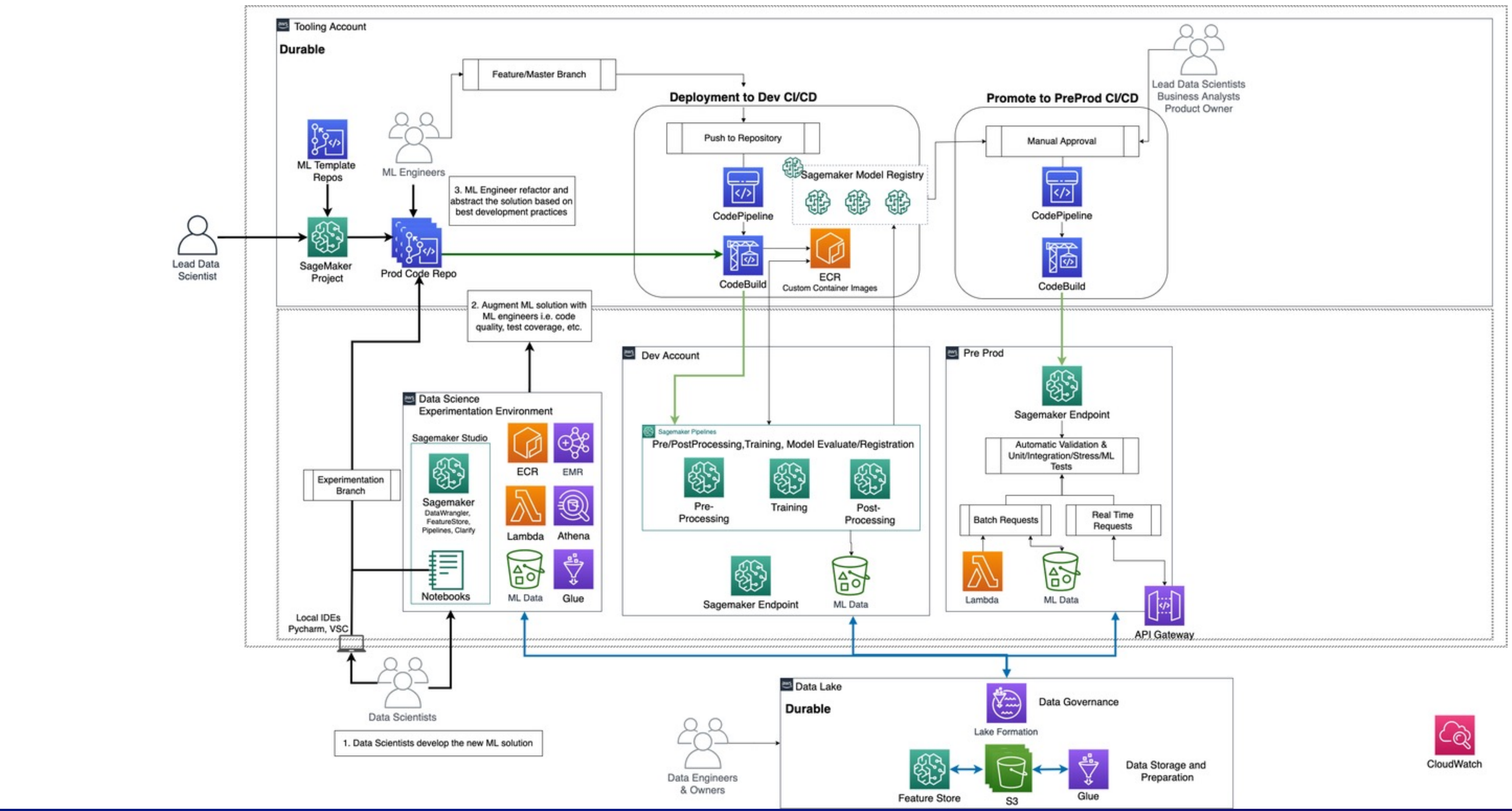
MLOPs Reliable Phase 1/2

UNDERSTAND THE NEEDS OF MLOPS & INTRODUCE AUTOMATED TESTING



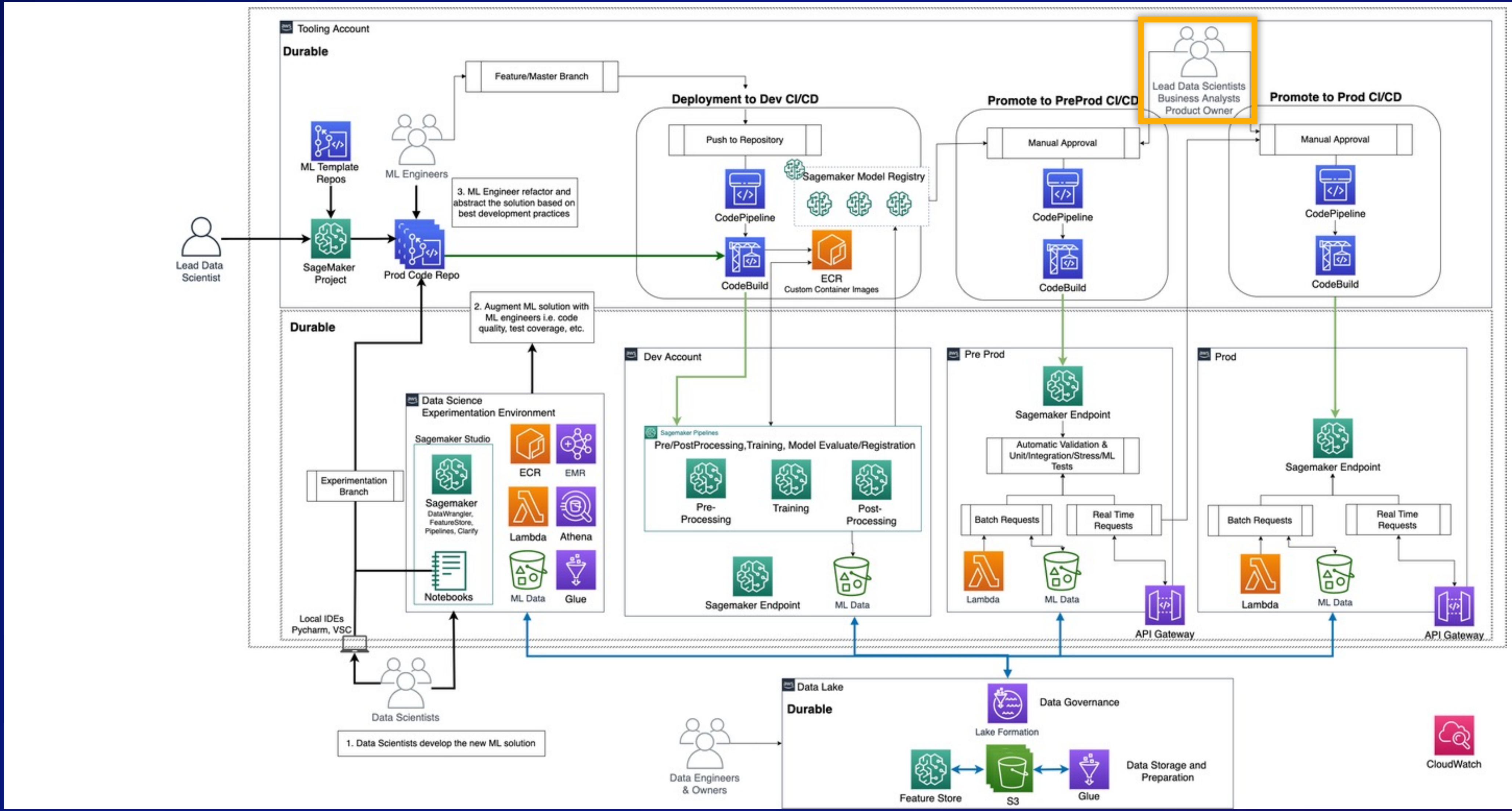
MLOPs Reliable Phase 1/2

UNDERSTAND THE NEEDS OF MLOPS & INTRODUCE AUTOMATED TESTING



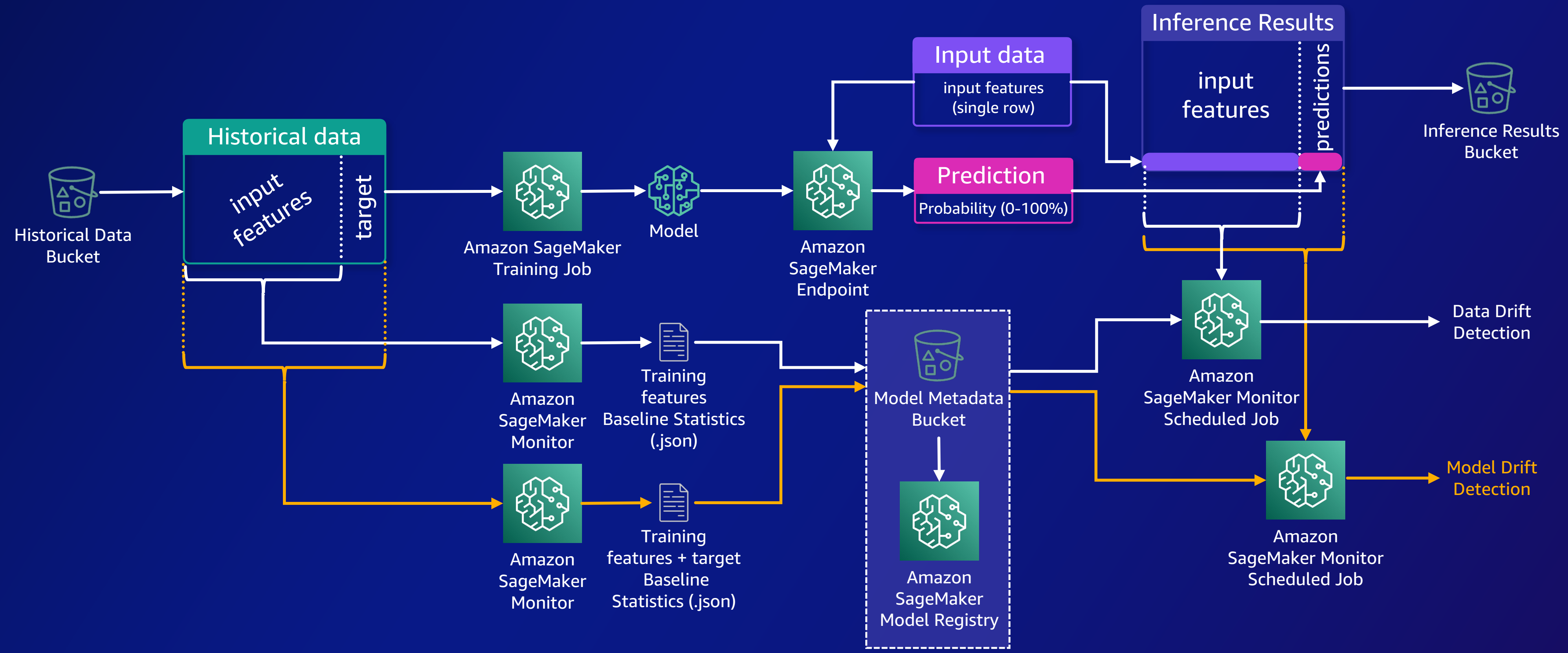
MLOPs Reliable Phase 2/2

BEST PRACTICE ON MLOPS, ROBUST AND SECURE MODELS



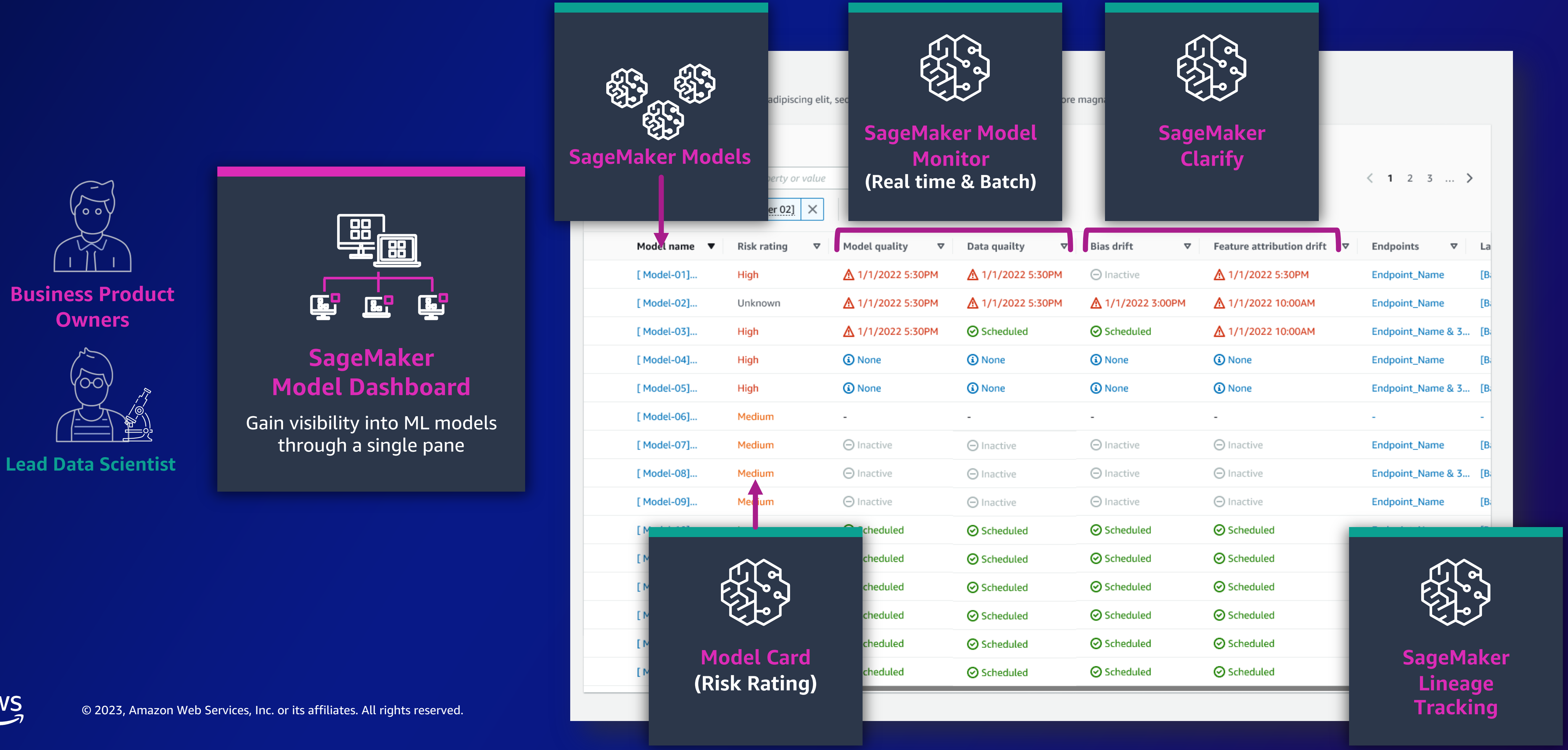
Amazon SageMaker Model Monitor

AUTOMATICALLY DETECT DATA AND MODEL QUALITY DRIFTS



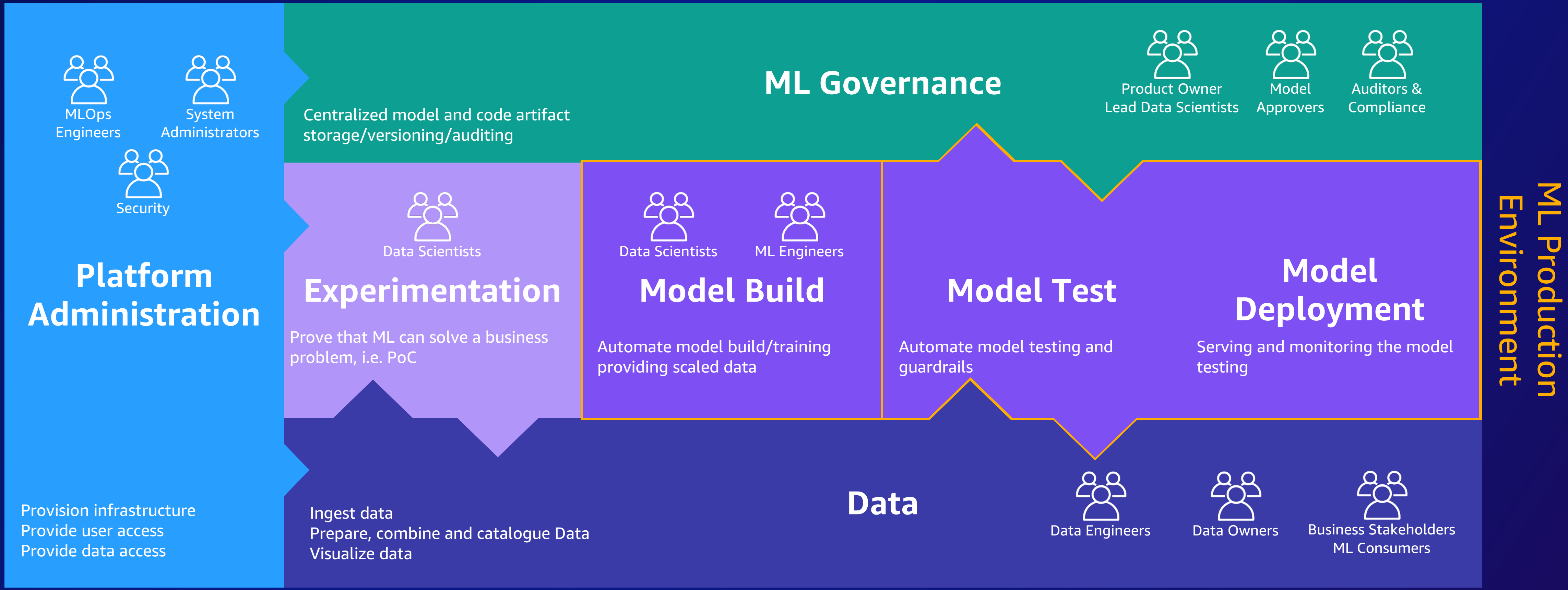
ML Governance – Model Dashboard

UNIFIED MONITORING ACROSS ALL YOUR MODELS IN PRODUCTION



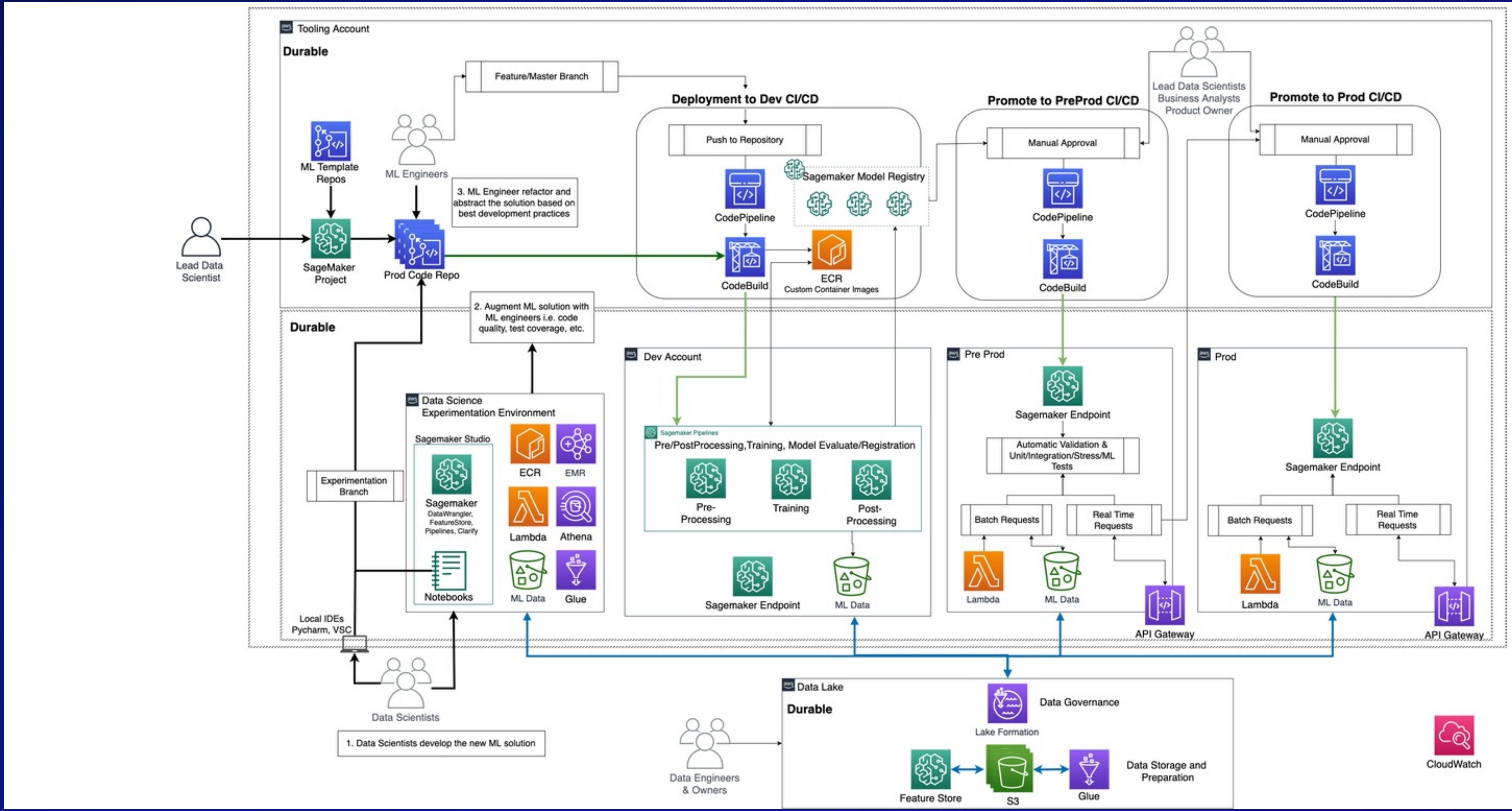
MLOps Foundation People & Processes

SCALABLE PHASE



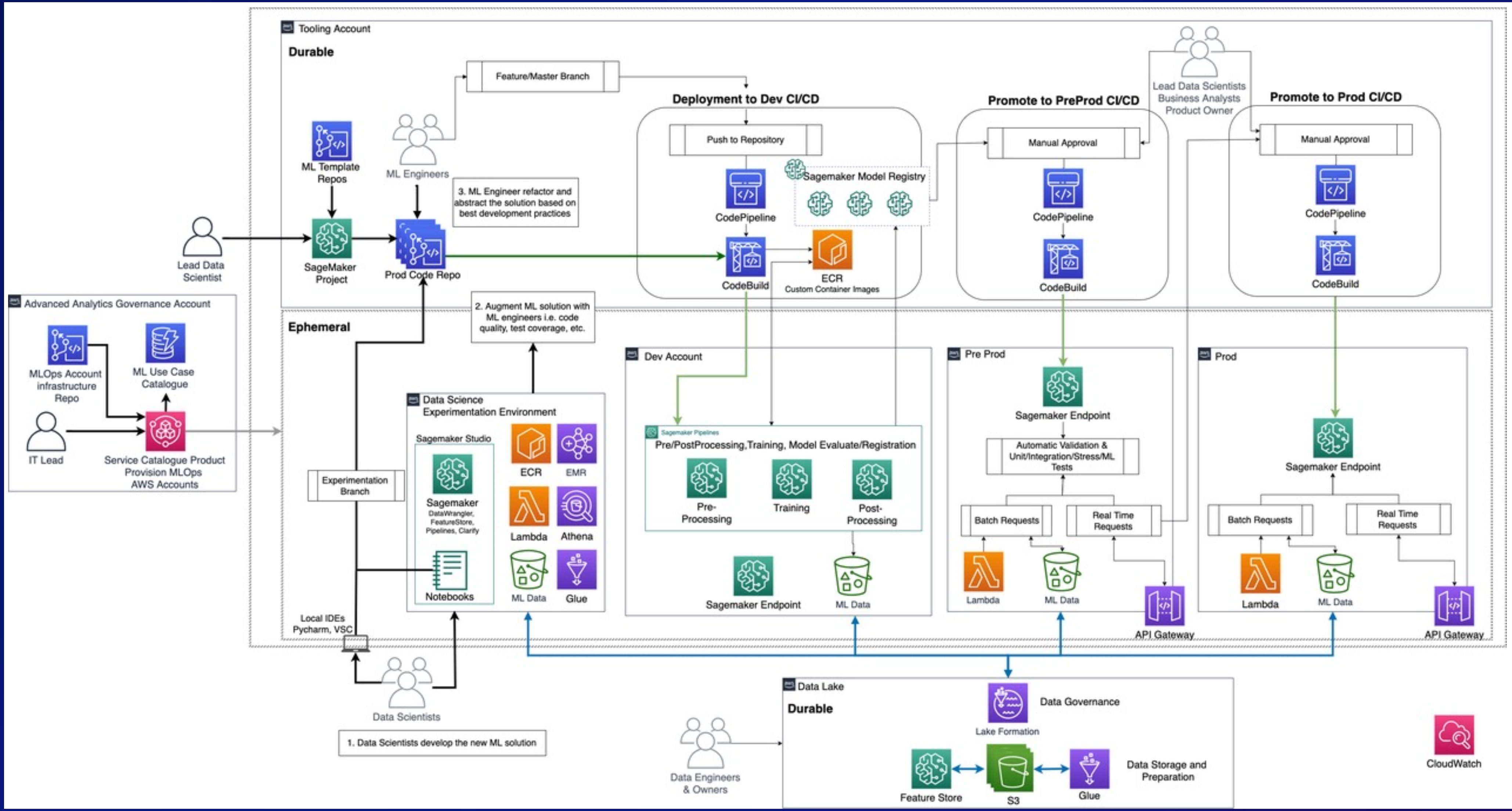
MLOPs Reliable Phase

BEST PRACTICE ON MLOPS, ROBUST AND SECURE MODELS



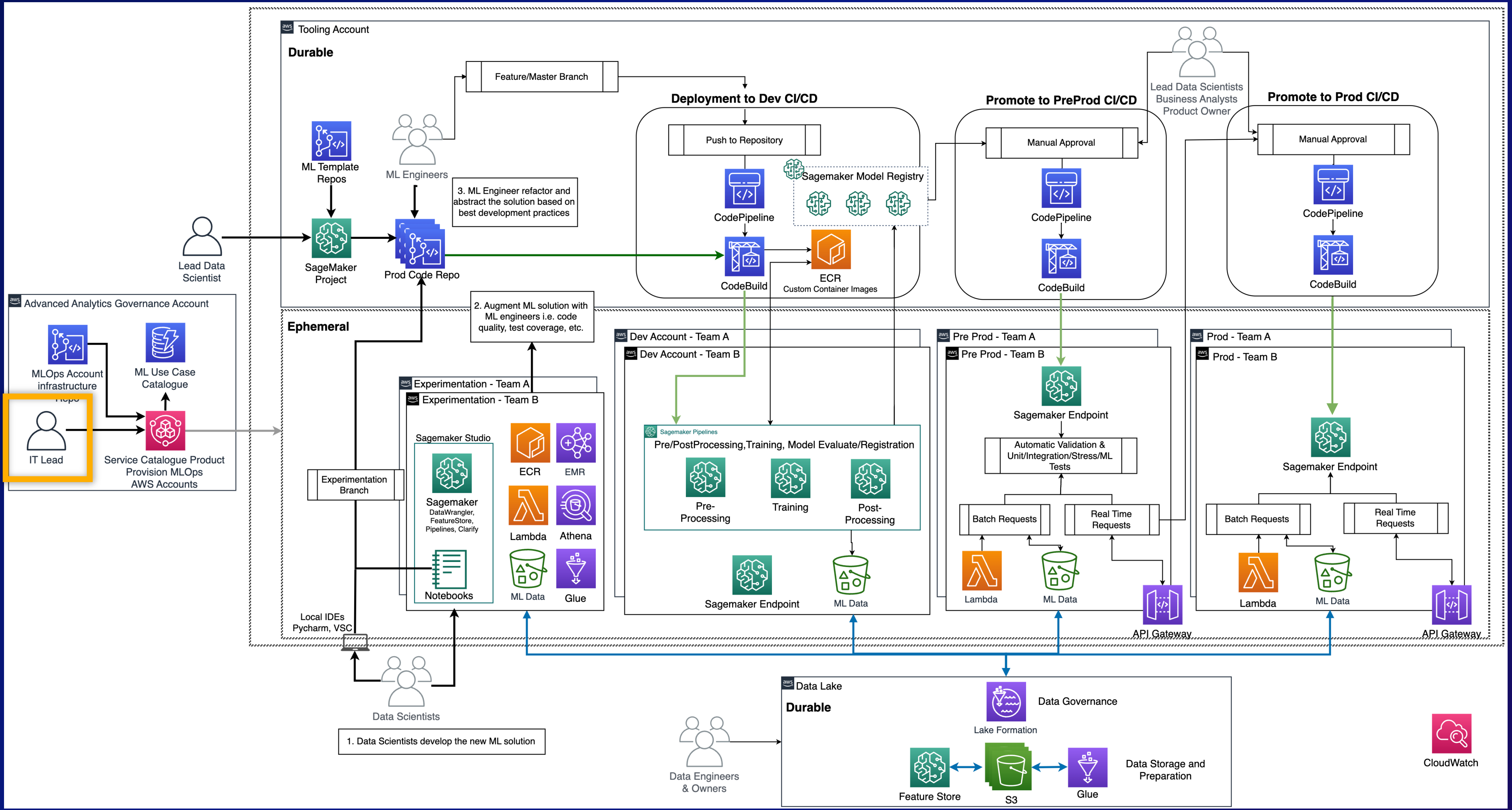
MLOPs Scalable Phase

MULTIPLE TEAMS AND ML USE CASES ADOPT MLOPS



MLOPs Scalable Phase

MULTIPLE TEAMS AND ML USE CASES ADOPT MLOPS



MLOPs Foundation Technology

MULTIPLE TEAMS AND ML USE CASES ADOPT MLOPS

Large architecture but data scientists need to be enabled only on ….

Amazon SageMaker

The Data Reply UK MLOps Journey

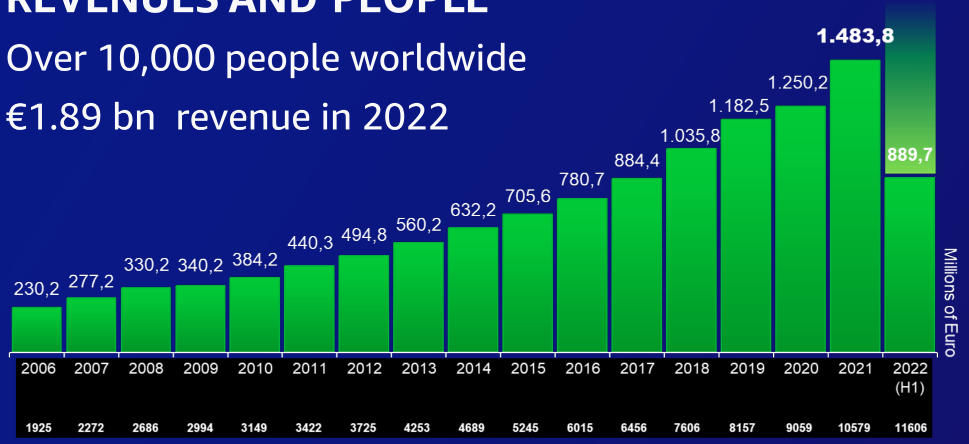
REPLY in A Nutshell

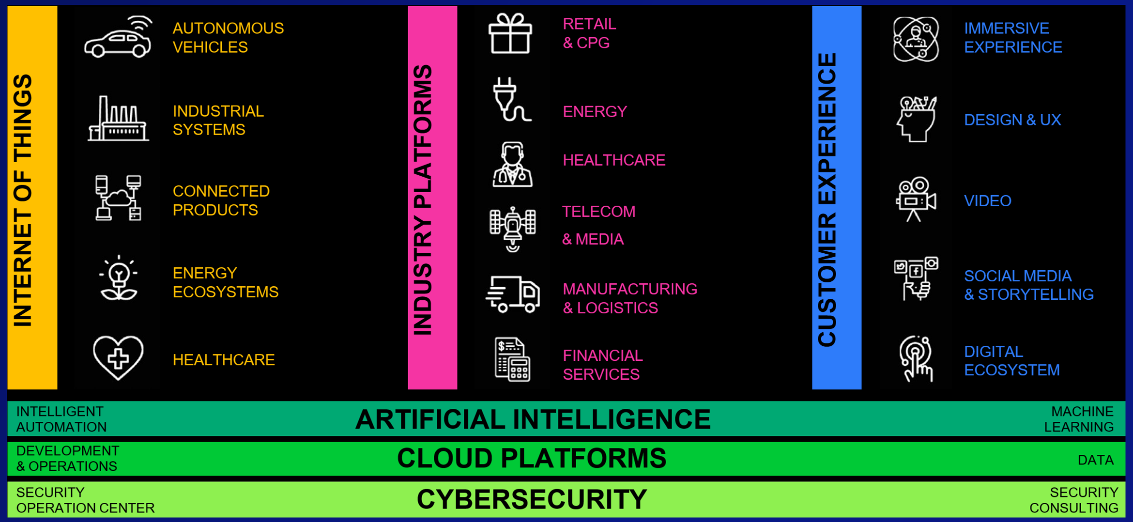
Founded in 1996, Reply is a company specialized in **System integration, Digital Services and Consulting** with a focus on solutions **design** and **implementation.**

REVENUES AND PEOPLE

Over 10,000 people worldwide

€1.89 bn revenue in 2022





REPLY DATA

Reply and AWS Partnership

A PASSION FOR TECHNOLOGY

Reply is among the companies with the most extensive number of competencies, including in **Data and Analytics and Machine Learning**, Service Delivery Validations, Certifications and programmes participation.

Data Reply is a **Launch partner of the AWS MLOps Competency** in 2021

aws PARTNER Premier Tier Services

13 AWS Competencies

16 AWS Service Delivery Validations

9 AWS Programmes

700+ AWS Certifications

REPLY DATA

Data Reply and AWS MLOPs Partnership

Data Reply is an AWS trusted partner in MLOps, through AWS MLOps partner enablement program

We developed an MLOps accelerator using native AWS services and successfully leveraged with a number of our customers

OUTCOMES

MLOps Platform Implementation

9-12 Months

3 Months

ML Route to Live

3 Months

<2 Weeks

Onboarding Data Science Teams

2 Months

<2 Weeks

MLOps Accelerator

MLOps Accelerator is a framework which aims to industrialise and operationalise ML lifecycle on AWS using best practices and Amazon SageMaker as the key principles.

EXPERIMENTATION

Amazon SageMaker:
- Studio
- Experiments

MODEL BUILD

Amazon SageMaker:
- Processing
- Pipelines
- Model Registry
- Project

MODEL TESTING & DEPLOYMENT

Amazon SageMaker:
- Endpoints
- Shadow Testing
- Model Monitor

MODEL GOVERNANCE

Amazon SageMaker:
- Model Cards
- Dashboards

PLATFORM

Amazon SageMaker:
- Custom Project Templates

AWS Service Catalog

DATA

Amazon SageMaker:
- Data Wrangler
- Feature Store

AWS CodePipeline | **GitLab** | **GitHub** | **CDK** | **Jenkins** | **Python**

Why Choose Our MLOps Accelerator

It allows customer to implement the MLOps capabilities up to 50% faster

Based on AWS native services, but can be integrated with your existing cloud platform

Modular and reusable across different use cases and environments

Adheres to AWS's best practices (AWS WAF ML Lens)

Validated & Trusted by AWS MLOps Solutions Architects

Case Study:
ML Modernisation & MLOps

INDICIA WORLDWIDE

Client:
Indicia Worldwide, a global marketing agency.

Goal:
To standardise and scale ML capabilities on AWS improving productivity, repeatability and time to market with acceptable ROI

Challenge:
- Gap in capabilities to productionise ML use cases
- Manual input to deploy models
- Slow speed of development
- Frustration for internal teams, lack of collaboration
- Lack of cloud operations for ML platform

Solution:
- Transitioned ML development from on-prem to Amazon SageMaker
- The MLOps journey began with the MLOps Assessment followed by MLOps solution implementation
- AWS Sagemaker services used to automates the build, test, and deploy models in production

Results:
- Successfully **productionised 15 use cases** on the cloud, with several more in the pipeline.
- Reduced 'time-to-market' by **60-70%**
- Data scientists can **focus on clients** , not infrastructure
- Happier, more productive staff

Customer Quotes

INDICIA WORLDWIDE

"**The platform** that Data Reply built for us **has massively reduced our time to value**. What used to take months, **we can now do in weeks**. It's not just about the models. It's about testing, deploying, and scaling, as well as integrating them into existing systems."

Graham Lannigan

Head of Data Platform at Indicia Worldwide

TUI

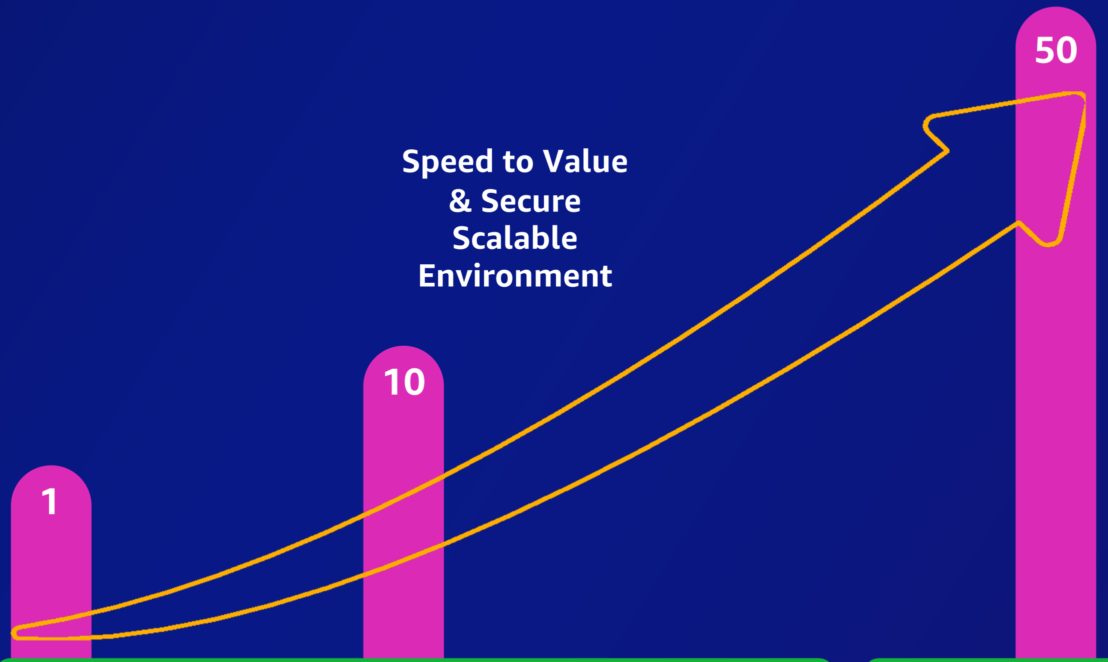
"Data Reply UK is helping us with streamlining and standardising our end-to-end ML processes. With the new MLOPs Framework leveraging Amazon CDK and Amazon SageMaker services, we are set to **productionise 10 ML models in the next 6 months** , scaling **to tens of models next year.** This will help with optimising the recommendation system on the TUI's web app and the website leading to an increased customer engagement and a seamless customer experience!"

Stefan Grossman

Customer Analytics Lead at TUI

Our MLOps Engagement Approach

FOCUSING ON BUSINESS VALUE



Speed to Value
& Secure
Scalable
Environment

Increased number
of ML use-cases
& Return
on Investment

ASSESS	ACCELERATE	ELEVATE
4 weeks	6 – 12 weeks	Project Dependable
Discovery & MLOPs Solution Roadmap	Multiple use cases leveraging MLOPs Accelerator	Scale further through the business

Initial → Repeatable → Reliable → Scalable

MLOps Discovery Workshop Offer

A HALF DAY WORKSHOP AT ZERO COST TO THE CUSTOMER

Workshop overview

Why attend?

A customer is looking to accelerate and scale their ML initiatives but has limited capability, skills and resources

Who should attend?

Head of Data Science/ML, Head of Data Platform, Engineering, DevOps

Outcomes

- Key ML development challenges and pain points identified
- Understanding MLOps value
- Initial use case(s) for MLOps is identified
- Stakeholders identified for the MLOps project mobilisation

Thought Leadership & Additional Resources

Data Reply is a Launch Partners Of The AWS MLOPs category Of The ML Competency.

aws
Partner Network

Our MLOps Whitepaper

APN Blog Post – Leveraging MLOps on AWS

aws partner network

Our offering is featured on AWS Marketplace

Our video Case Study with Indicia Worldwide is on APN TV

These clients have trusted Data Reply with MLOps

INDICIA WORLDWIDE

St James's Place

TUI

BMW

LAVAZZA

nexi

Scan for more information about the Data Reply MLOps Capability Assessment

MLOps Resources

MLOps Resources

MLOps on AWS by Amazon SageMaker
https://aws.amazon.com/sagemaker/mlops

Amazon SageMaker documentation
https://docs.aws.amazon.com/sagemaker/index.html

MLOps foundation roadmap for enterprises
https://aws.amazon.com/blogs/machine-learning/mlops-foundation-roadmap-for-enterprises-with-amazon-sagemaker

Secure multi-account MLOPs platform source code examples
https://github.com/aws-samples/sagemaker-custom-project-templates/tree/main/mlops-multi-account-cdk

Thank you!

Please complete the session survey in the mobile app

Dr Sokratis Kartakis
/in/sokratis-kartakis
@sokratis.kartakis

Ajish Palakadan
/in/ajish-palakadan