BADM 4830 / BAIM 4200 Advanced Business Analytics

Module 4 - MLOps – What it is , Why MLOps

Ivan Portilla
Jesus.Portilla@Colorado.edu
Portilla@gmail.com
github.com/jiportilla/giveback

IBM

Objectives of This Module

Upon completion of this module, you will understand:

MLOps - What & Why

- Definition & People of MLOps - https://ml-ops.org/
- Key MLOps Features
  - Model Development
  - Monitoring
  - Productionalization & Deployment
  - Iteration & Lifecycle
  - Governance
- Lab: Intro to MLOps

Fall 2021
BADM 4830 / BAIM 4200 Advanced Business Analytics

- This course will give students the language, knowledge, and actionable methods to work alongside technical and non-technical members of your team to create AI solutions.

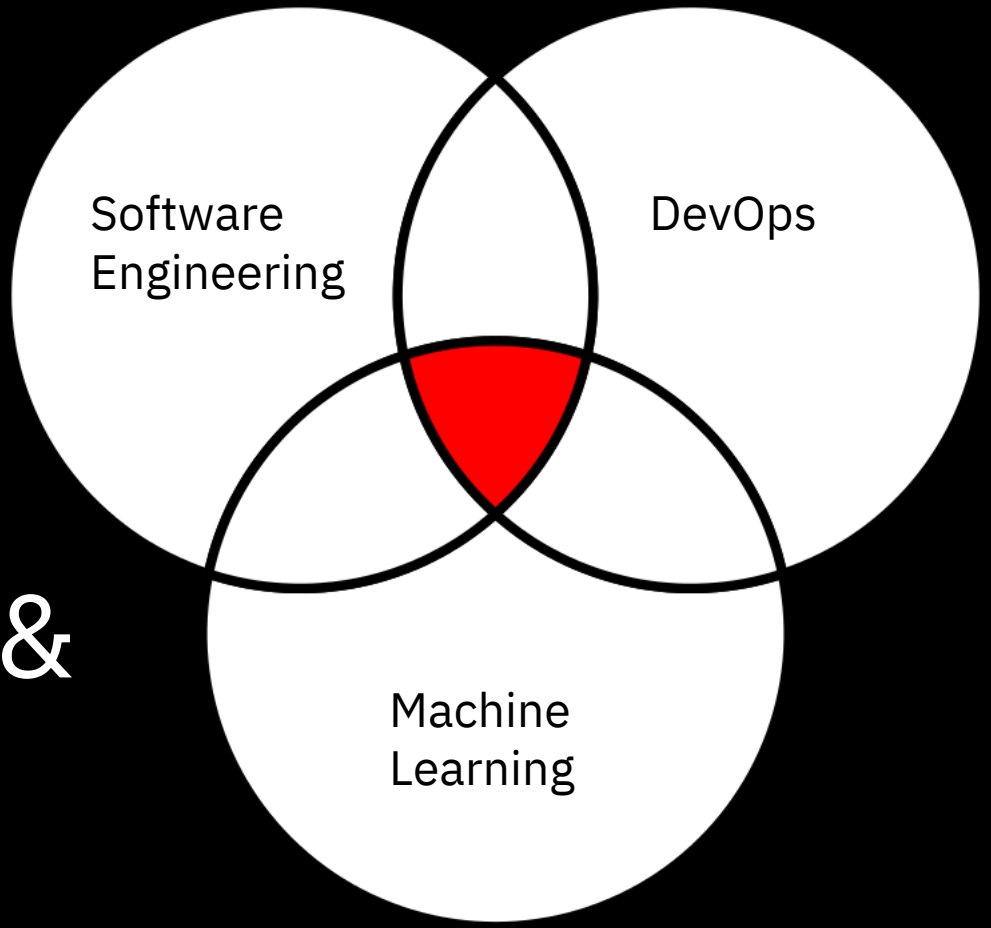
- Students will explore what it means to design artificial intelligence systems as a team, guided by a clear intent and a focus on people. This course will give you the framework and tools you need to recognize responsible AI design, align your team, and work with data sources to start building AI solutions.

- Students will learn the tools, technology, and practices that enable cross-functional AI teams to efficiently deploy, monitor, retrain, and govern models in production systems.

Re-cap

1. MLOps definition
2. Data Project Roles
3. AutoAI experiment

MlOps is the **convergence** of Software Engineering, Machine learning & DevOps

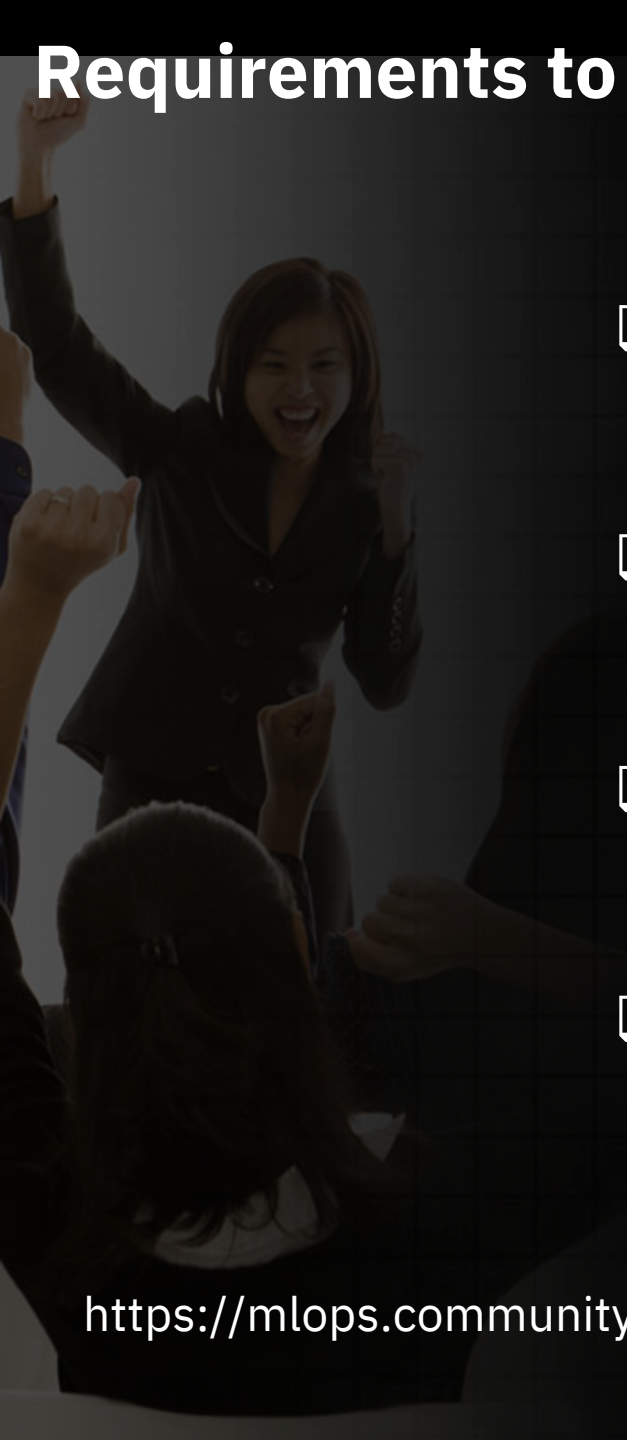


Requirements to Achieve MLOps

- ❑ Reproducible
- ❑ Accountable
- ❑ Collaborative
- ❑ Continuous

https://mlops.community//

Requirements to Achieve MLOps



- ❑ Reproducible
- ❑ Accountable
- ❑ Collaborative
- ❑ Continuous

Must be able to re-train a 9-month-old model to within few %

Must be able to track back from model in Production to its provenance

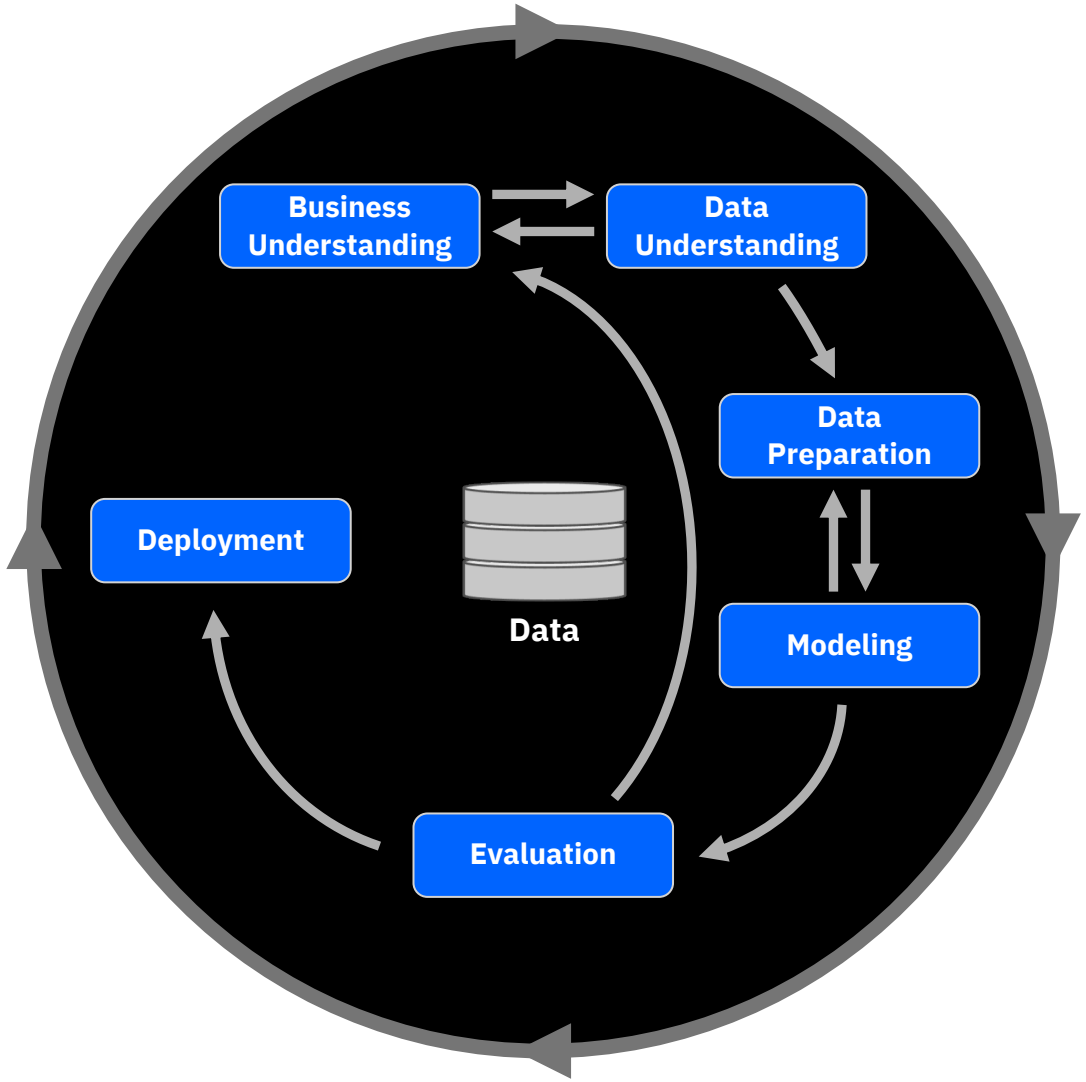
Must be able to do asynchronous collaboration

Must be able to deploy automatically & monitor statistically

https://mlops.community//

Solution Development Method Approach

CRossIndustry Standard Process for Data Mining (CRISP-DM)



Seven steps to successful Data Mining/Predictive Analytics

1. Define the business challenge in a precise statement
2. Define the data model and data requirements
3. Source data from all available repositories
4. Evaluate the data quality
5. Select the machine learning algorithm
6. Interpret the results and iterate to improve model
7. Deploy the model into your business

AI Project Roles

Drives governance policy effectiveness while tracking how data is used and its value to the company

Data Steward

Builds data pipelines that power dashboards and data platforms while ensuring high quality

Data Engineer

Prepares data to tease out the insights they're looking for, without IT involvement

Data Scientist

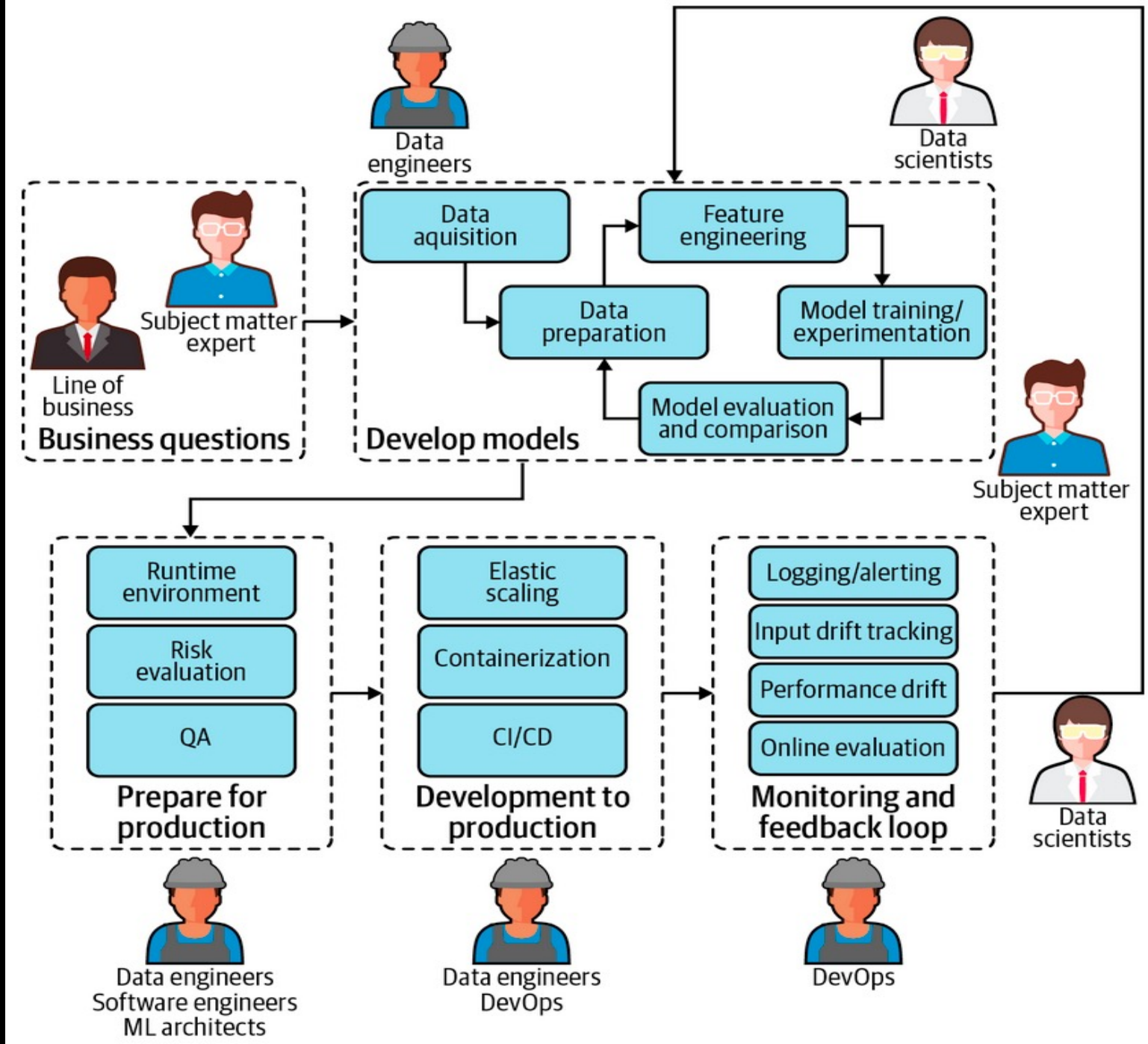
Business Analyst

Works with data to apply insights to the business strategy

App Developer

Makes insights immediately actionable and adds intelligence to apps in straightforward manner

Enterprise ML Lifecycle



Use case

1. Operationalized AI stages
2. Best practices

Stages in Operationalizing AI

Continuous integration / continuous delivery pipeline

Scope AI project(s)

Explore, prioritize, select use cases – feasibility vs impact

Detail selected use case(s) – KPIs, data, workflow, success criteria

Technical design: tools, infrastructure, data details, approach


Business


Data Scientist


IT/ Ops


Data provider

DataOps

All: Define sources & needs

Provider: Set up catalog for discovery, lineage, curation, access rules

Steward: Create data policies & access

Consumer: Find, understand, add, explore, review, share


Data provider


Data steward

Build AI assets

Connect data

Analyze data

Prepare data

Build models

Train models

Visualize & evaluate models

Test for Bias, Fairness


Data engineer


Data scientist

ML Ops – Deploy

Review, 3rd party oversight, unit tests, validation, approve deployable version

Deploy model to ML runtime engine

Monitor & evaluate model execution

Manage against thresholds

Validation reports


Software engineer


Data scientist


IT/ Ops

Manage and Trust

Configure monitoring
Configure integration with other systems

Monitor and manage:

1. Quality
2. Performance
3. Custom metrics
4. Bias/fairness
5. Accuracy & Data Drift

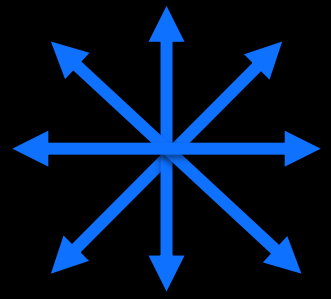
Explain on demand


Risk & Compliance

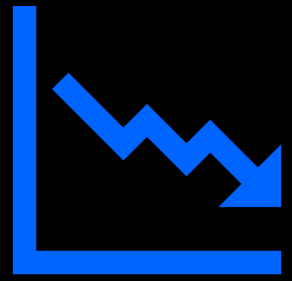

Business

Trustworthiness presents new challenges to operationalizing AI

Bias



Quality



Drift

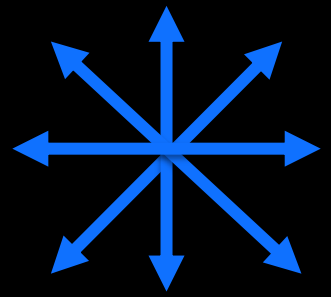
Explainability

Trustworthiness presents new challenges to operationalizing AI

Bias

Training data and AI models may be biased.

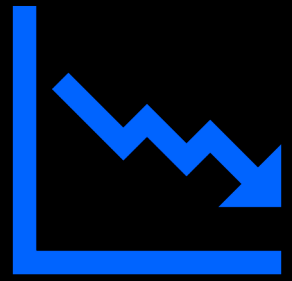
Are privileged groups at a systematic advantage compared to unprivileged groups?



Quality

Models need to perform well across the AI/ML lifecycle.

Are relevant performance metrics being monitored over time?



Drift

Changes in input data cause model to make inaccurate decisions.

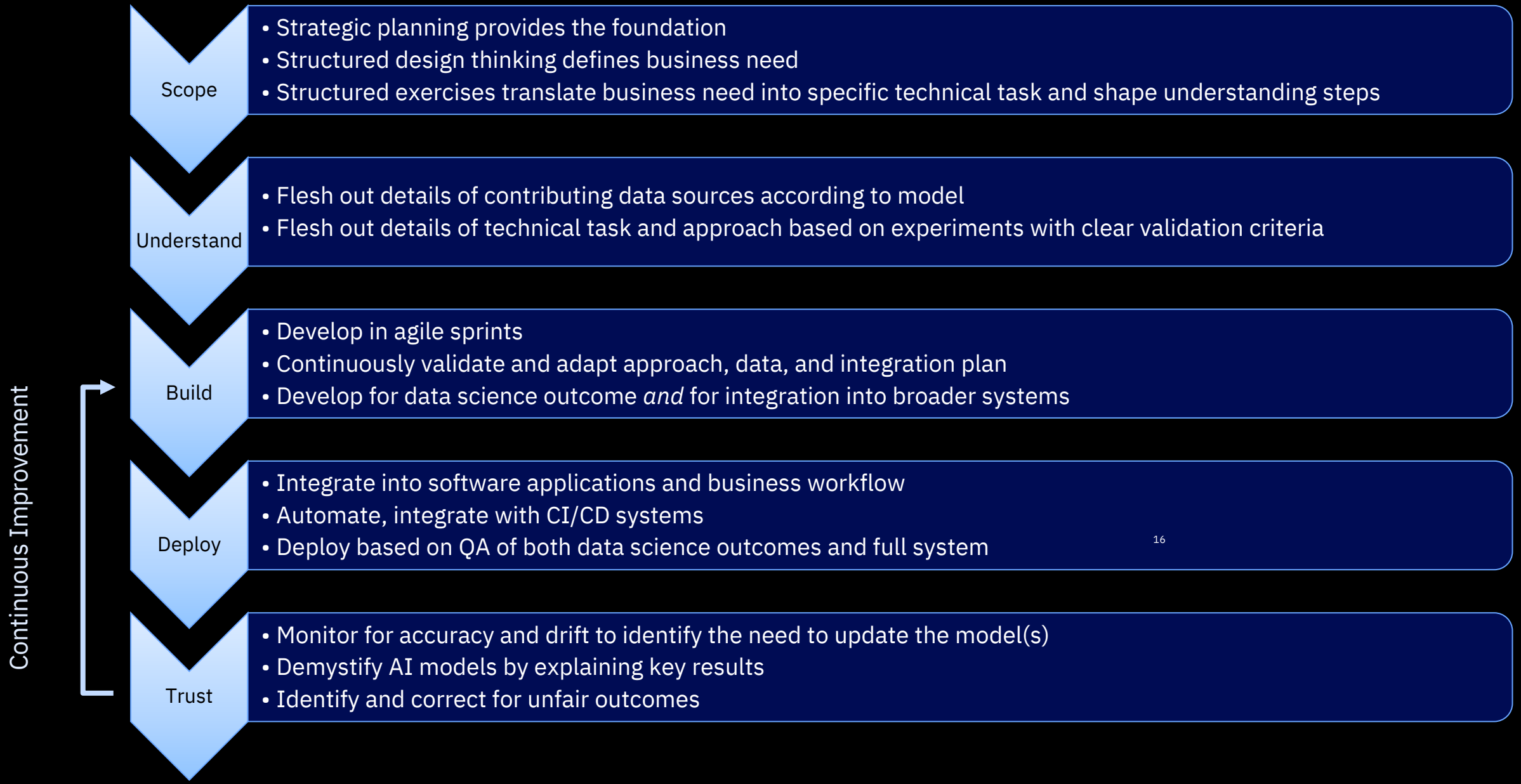
Training data may not include data ranges or combinations seen in real life

Explainability

Traditional statistical models are simpler to interpret and explain.

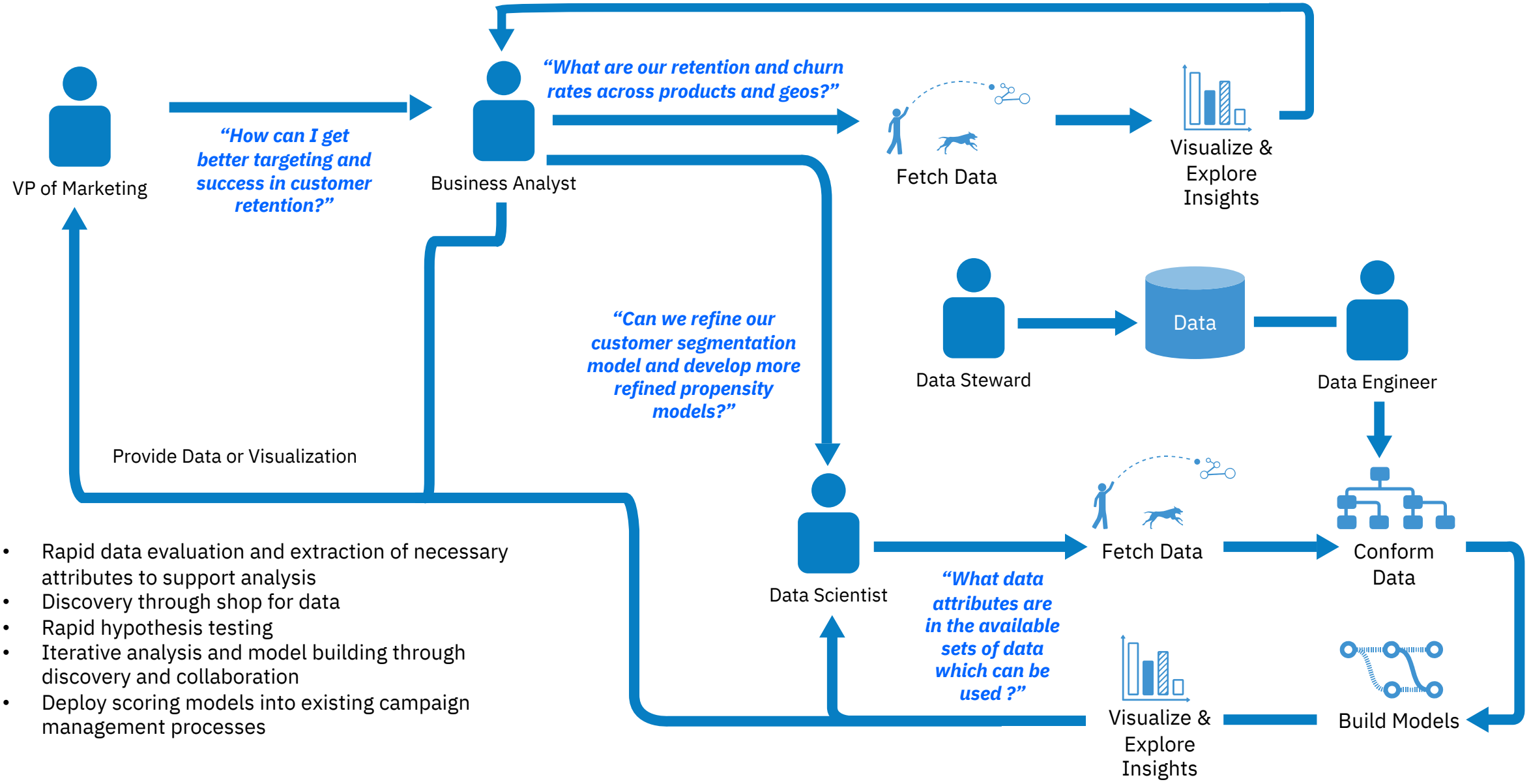
At what point would the outcome have been different?

Process: Best Practices Throughout the Life Cycle



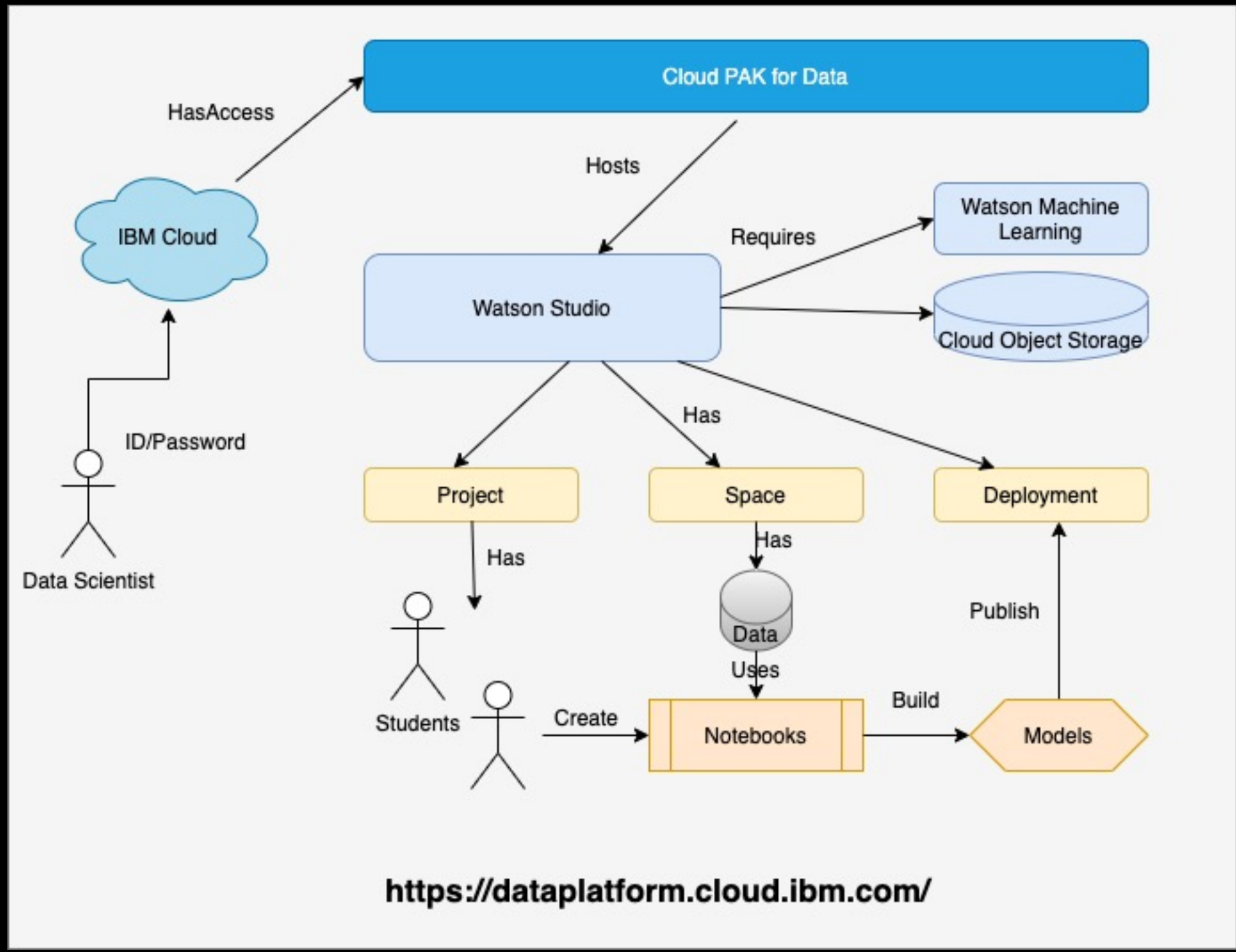
Demos - Lab

Sample business problem: Churn Analysis

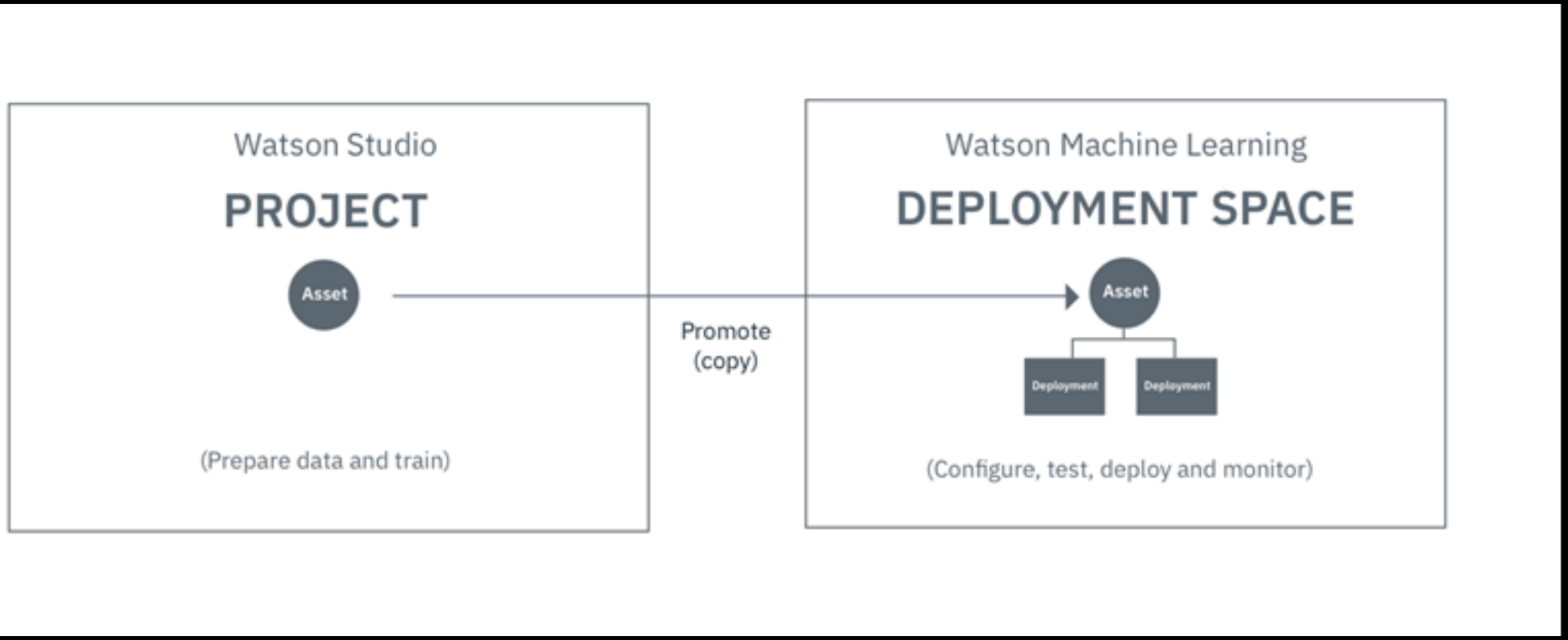


- Rapid data evaluation and extraction of necessary attributes to support analysis
- Discovery through shop for data
- Rapid hypothesis testing
- Iterative analysis and model building through discovery and collaboration
- Deploy scoring models into existing campaign management processes

Watson Studio Model Deployment



Watson Studio Model Deployment



Agenda

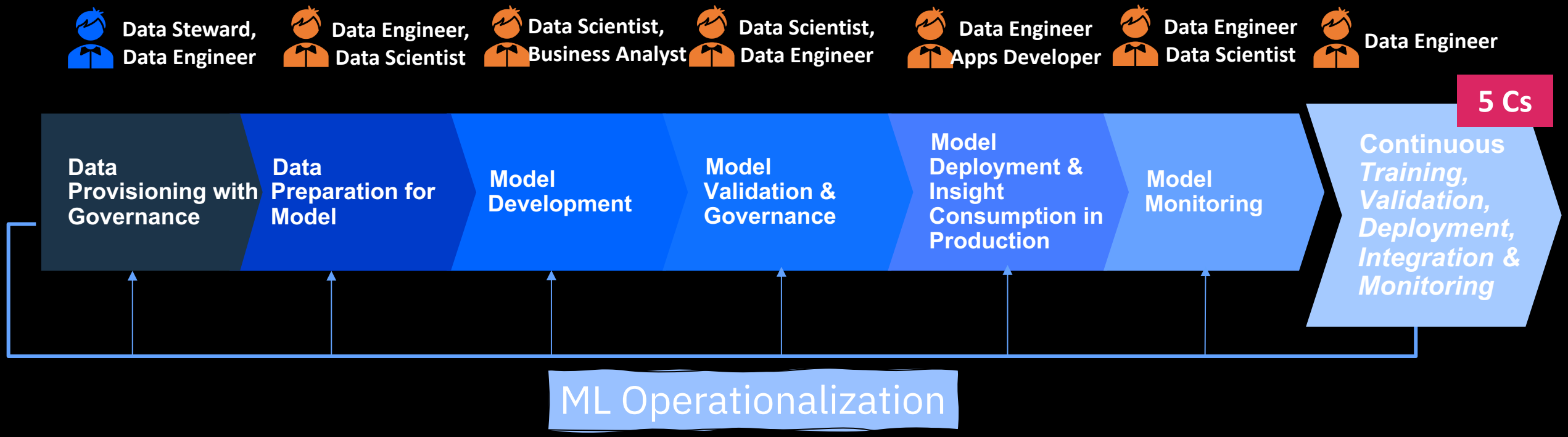
MLOps What & Why

1. ML Operationalization stages & personas
2. ML operationalization environments
3. MLOps Frameworks
4. MLOps in Action

ML Operationalization – High Level Steps and Personas

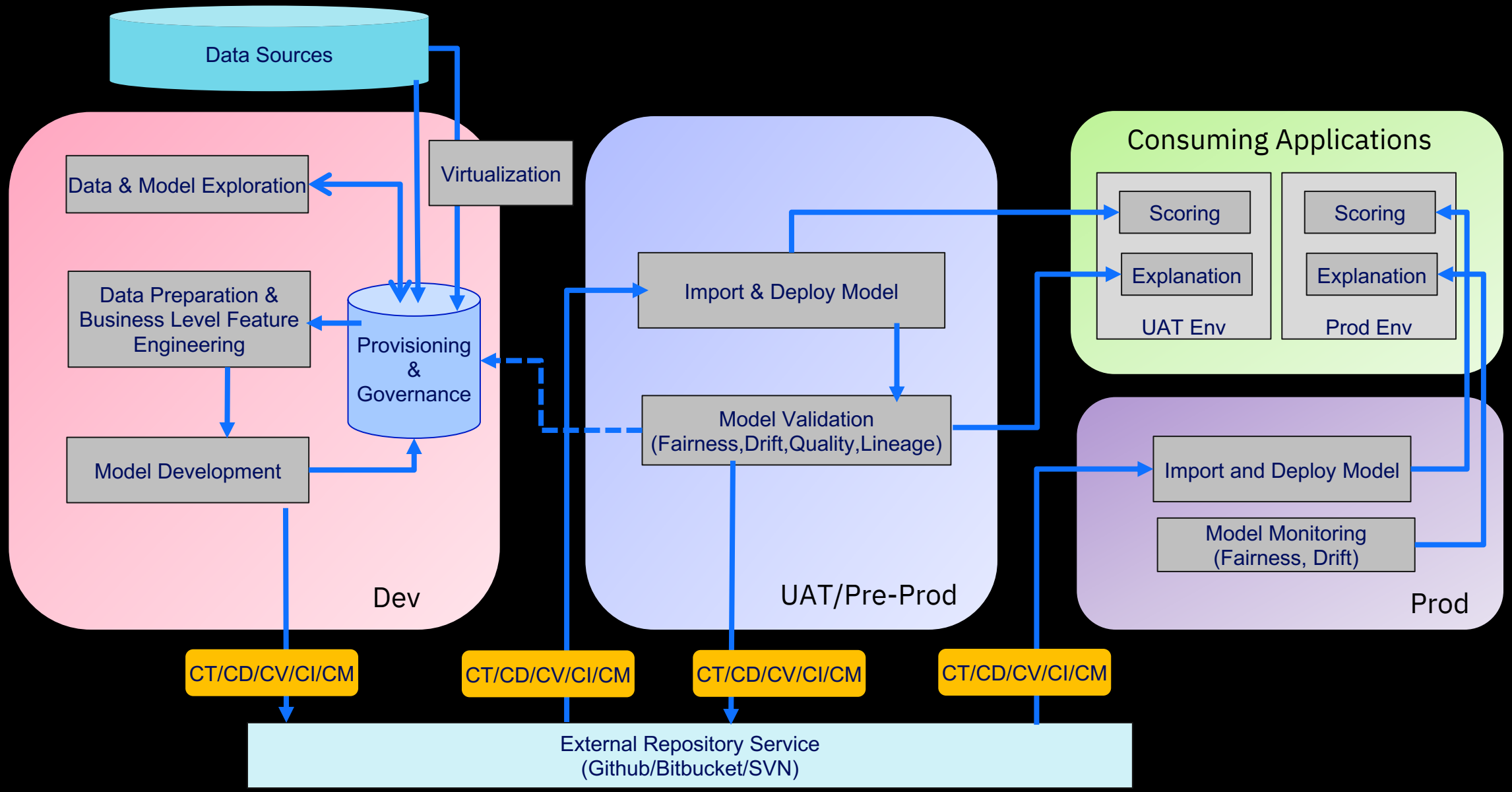
ML Operationalization refers to operationalization of Machine Learning Models for production use to realize business value out of those Models.

ML Operationalization overlays paradigm of DevOps on Model Lifecycle management process (CRISP-DM)



For Conceptual View of ML Ops please check - https://ibm.co/AI-Ops

ML Operationalization spread across Dev, UAT & Prod Environments



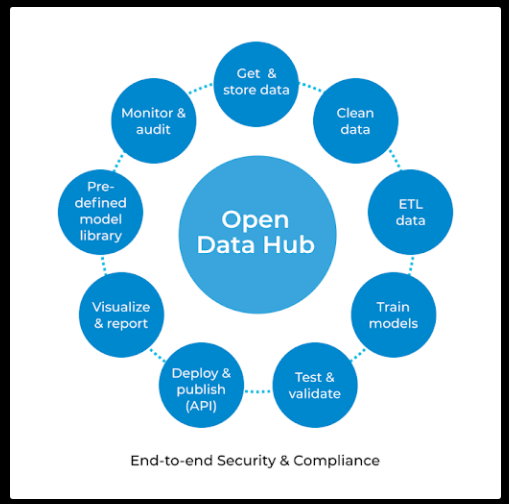
What to look for in ML Frameworks ?

Features	Description
Flexibility/Customizability	How flexible is the platform in integrating and/or customizing new frameworks for AI model development.
Ease of Use	How easy is it to leverage these tools and proposed techniques from setup to application.
Integrations	How well does the platform integrate with Git or other model versioning and source control tools, catalogs (for governance and discoverability) or various data sources.
Governance	How well does the solution support governance and discoverability of assets (data assets, models, notebooks, ...)
Platform	Support for various platforms (public cloud, on-prem, hybrid cloud), and compute types (CPU/GPU) for training and scoring (or inference) AI models
Monitoring	How well does the solution support monitoring AI models for performance / explainability / fairness
Scalability	How scalable is the platform in supporting various Data & AI users in different roles to explore, develop, and deploy AI models.
Openness	How well does the platform support open-source technologies which has become a key differentiator for platform providers.
Security	How well does the platform support enterprise-grade security access to the platform in terms of authorization and authentication
Support for 5 Cs	Support for Continuous Training, Validation, Deployment, Integration and Monitoring

Popular ML Ops Tools and Frameworks

Open Source Frameworks

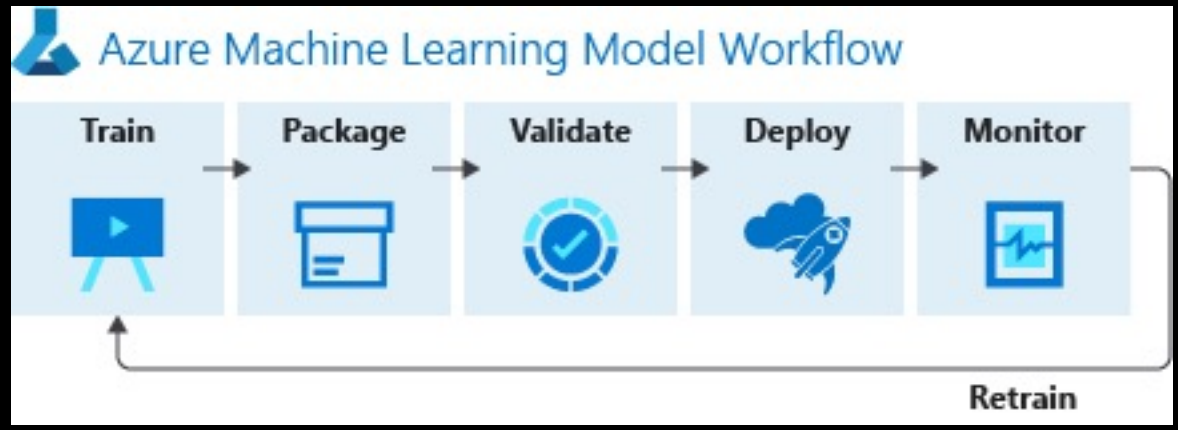
Kubeflow



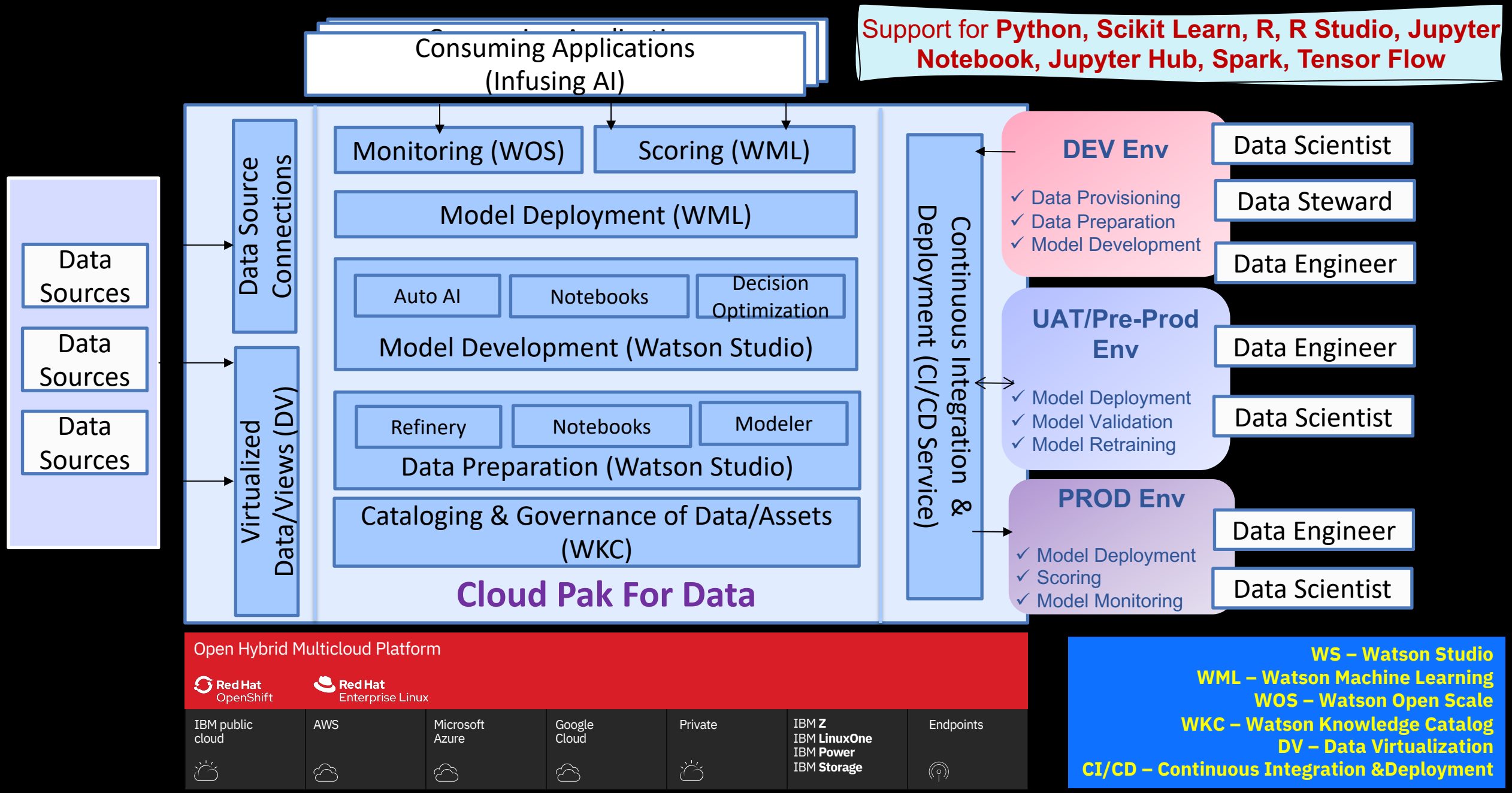
mlflow

Apache Airflow

From Software Vendors

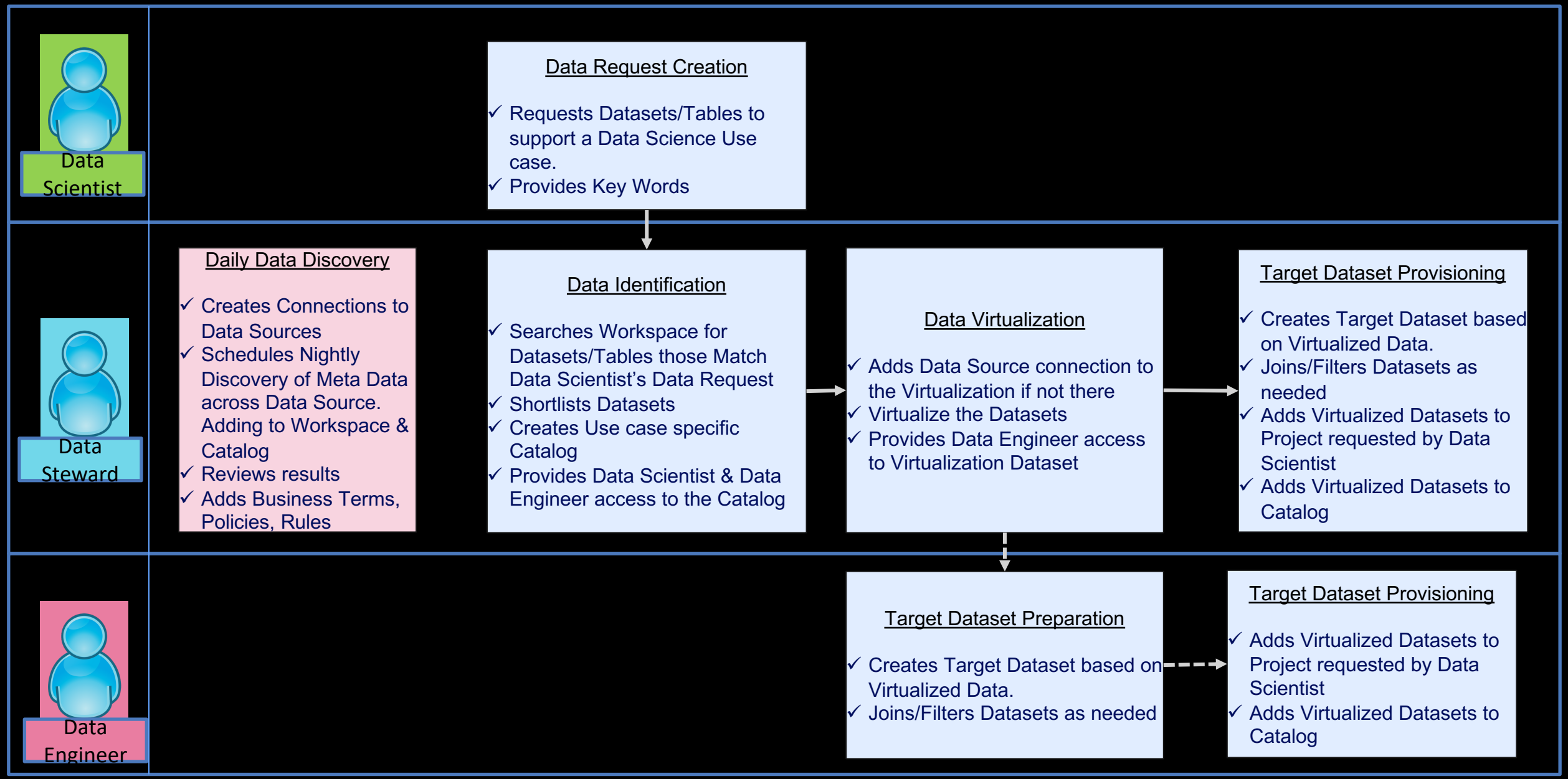


ML Operationalization with IBM Cloud Pak For Data

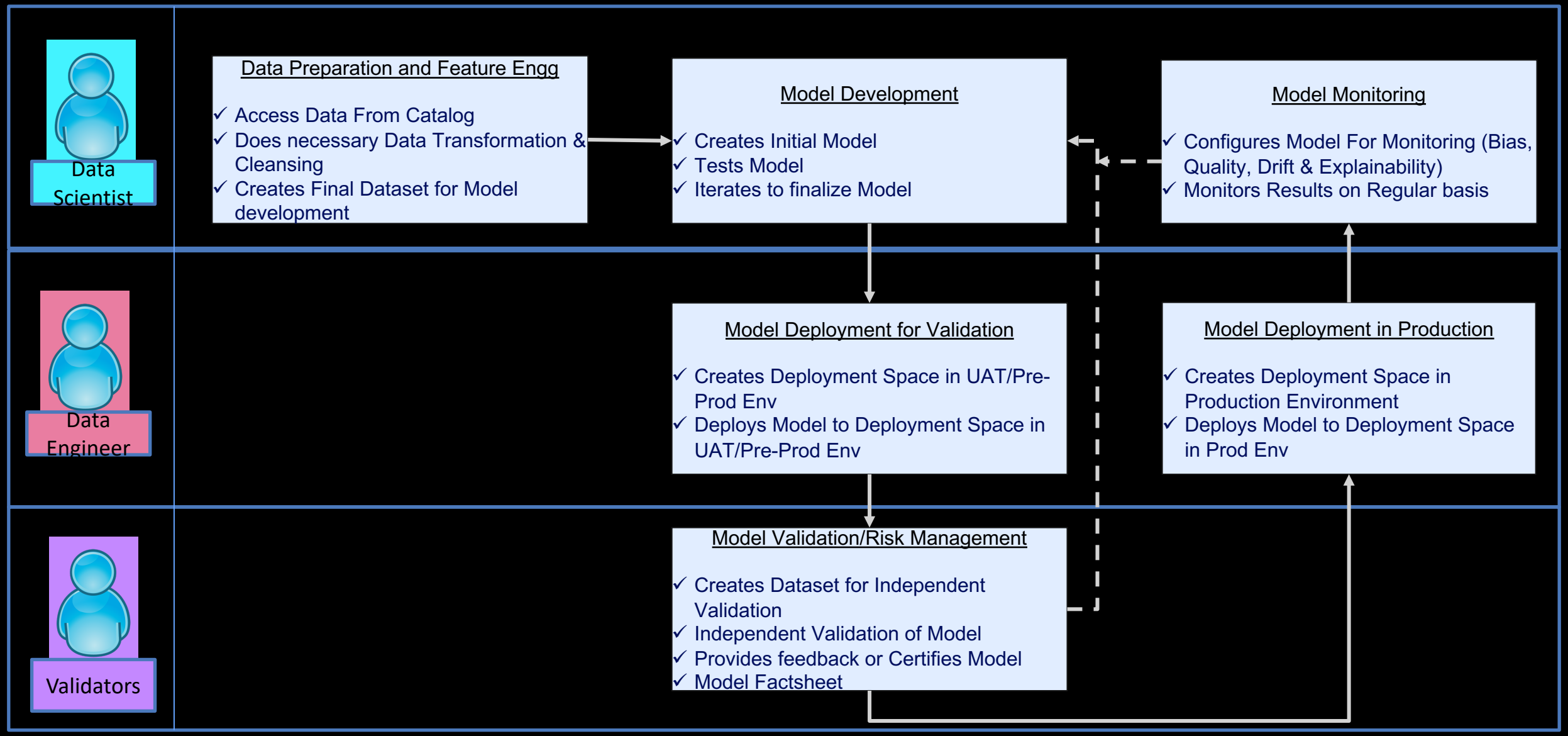


WS – Watson Studio
WML – Watson Machine Learning
WOS – Watson Open Scale
WKC – Watson Knowledge Catalog
DV – Data Virtualization
CI/CD – Continuous Integration &Deployment

ML Ops In Action (1/2) - Data Provisioning and Governance



ML Ops In Action (2/2) - Model Development, Deployment & Monitoring

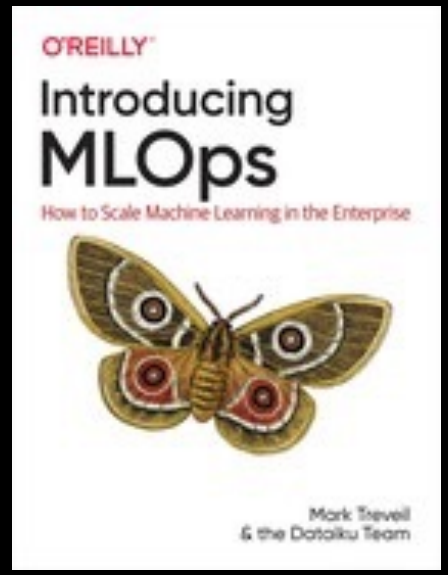


Q&A

Ivan Portilla

ivanp@us.ibm.com

@iportilla



https://medium.com/inside-machine-learning/ai-ops-managing-the-end-to-end-lifecycle-of-ai-3606a59591b0