Model Operationalization

With Governance and Model Risk Management

Open Data Science Conference (East) 2020

Sourav Mazumder
The Open Group Distinguished Data Scientist
ML Operationalization Leader
IBM Data and AI Expert Lab and Learning
smazumder@us.ibm.com
https://www.linkedin.com/in/souravmazumder/

Shikhar Kwatra
IBM Master Inventor
ML Operationalization Architect
IBM Data and AI Expert Lab and Learning
skwatra@us.ibm.com
https://www.linkedin.com/in/shikharkwatra/

Join us in https://github.com/ibm-cloud-architecture/refarch-ml-ops

Agenda

- Introduction to ML Operationalization
- ML Operationalization - Process, Persona, Environments and Frameworks/Platforms
- ML Operationalization in Action with Governance and Model Risk Management - Demonstration
- Hands On (Self Paced)

What is ML Operationalization ?

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)

- Continuous Training
- Automated Validation and Deployment
- Insight Infusion at Scale
- Ensuring Transparency
- Removing Bias
- Business KPI Mapping
- Data and Model Governance
- Model Risk Management

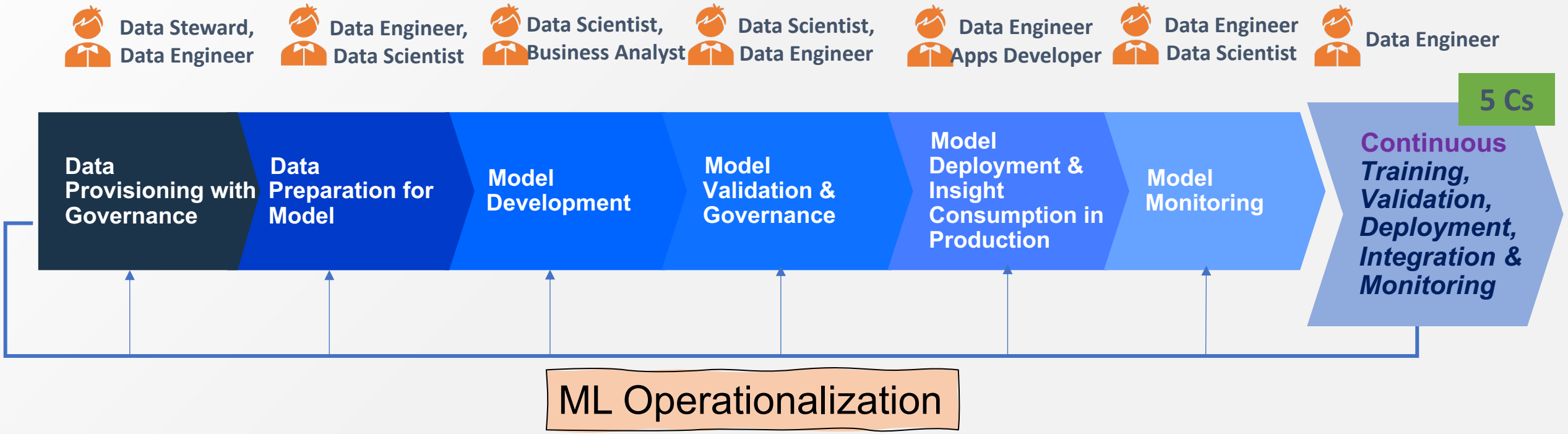
"Creating an ML model is just a starting point. To bring the technology into production service, you need to solve various real-world issues such as building a **data pipeline for continuous training**, **automated validation** of the model, **version control** of the model, creating a **scalable serving infrastructure**, and ongoing operation of the ML infrastructure with **monitoring and alerting.**"

Forrester

ML Ops can be daunting with different challenges faced by different organizations

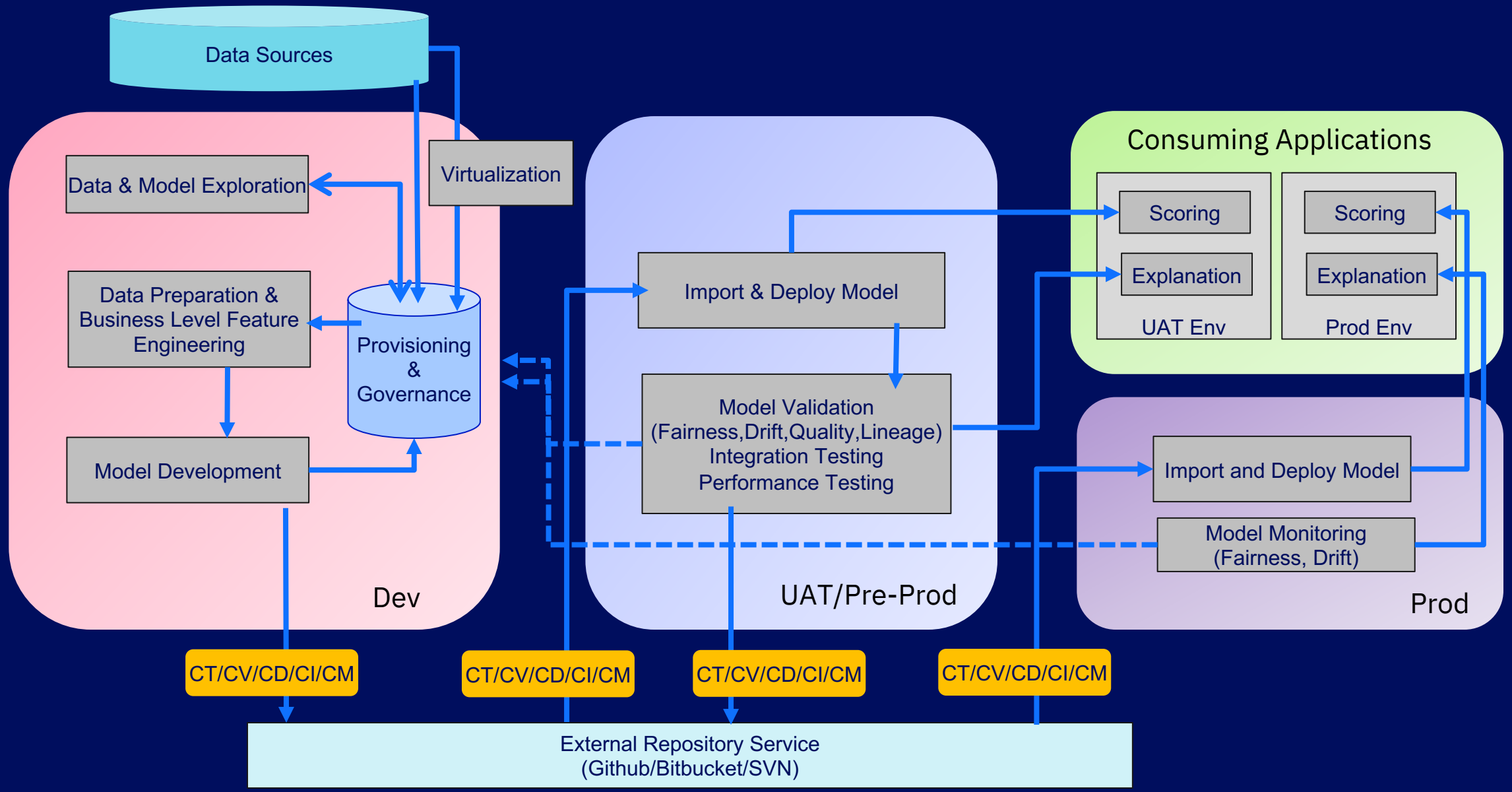
- Need to reduce the time between Model conception to use in Production
- Need of onboard large number of Data Scientists, track the datasets used for developing Models, how the Models are providing Business Value
- Enabling the Model to serve 10s of Millions of requests in a day and Monitor those requests
- Need to institutionalize collaborative approach involving multiple teams to deliver Models without Bias and ability to trace back Models' Lineage
- Explaining auditors why Model is predicting in certain way
- Need to ramp up no core Data Scientist in Data Science (Citizen Data Scientist)
- Need to get Explanation for every case predicted by a Predictive Model
- Need to have an optimized Infrastructure to support large number of Data Scientists and the Model

ML Operationalization – High Level Steps and Personas



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

ML Operationalization spread across Dev, UAT & Prod Environments



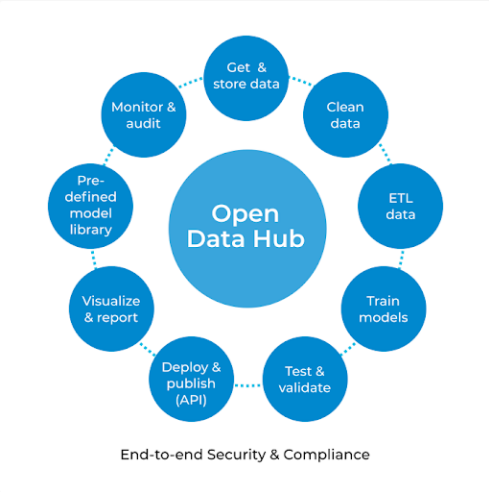
What to look for in ML Frameworks ?

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

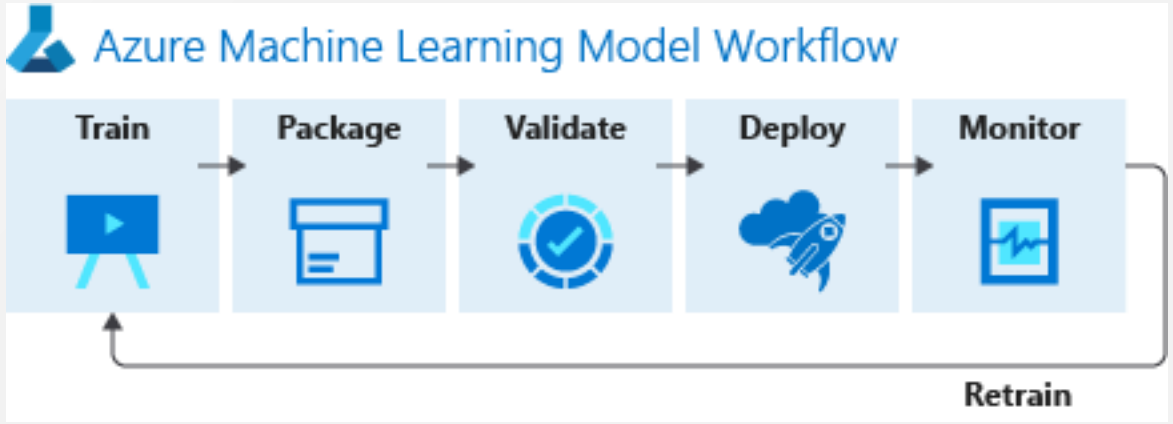
Kubeflow



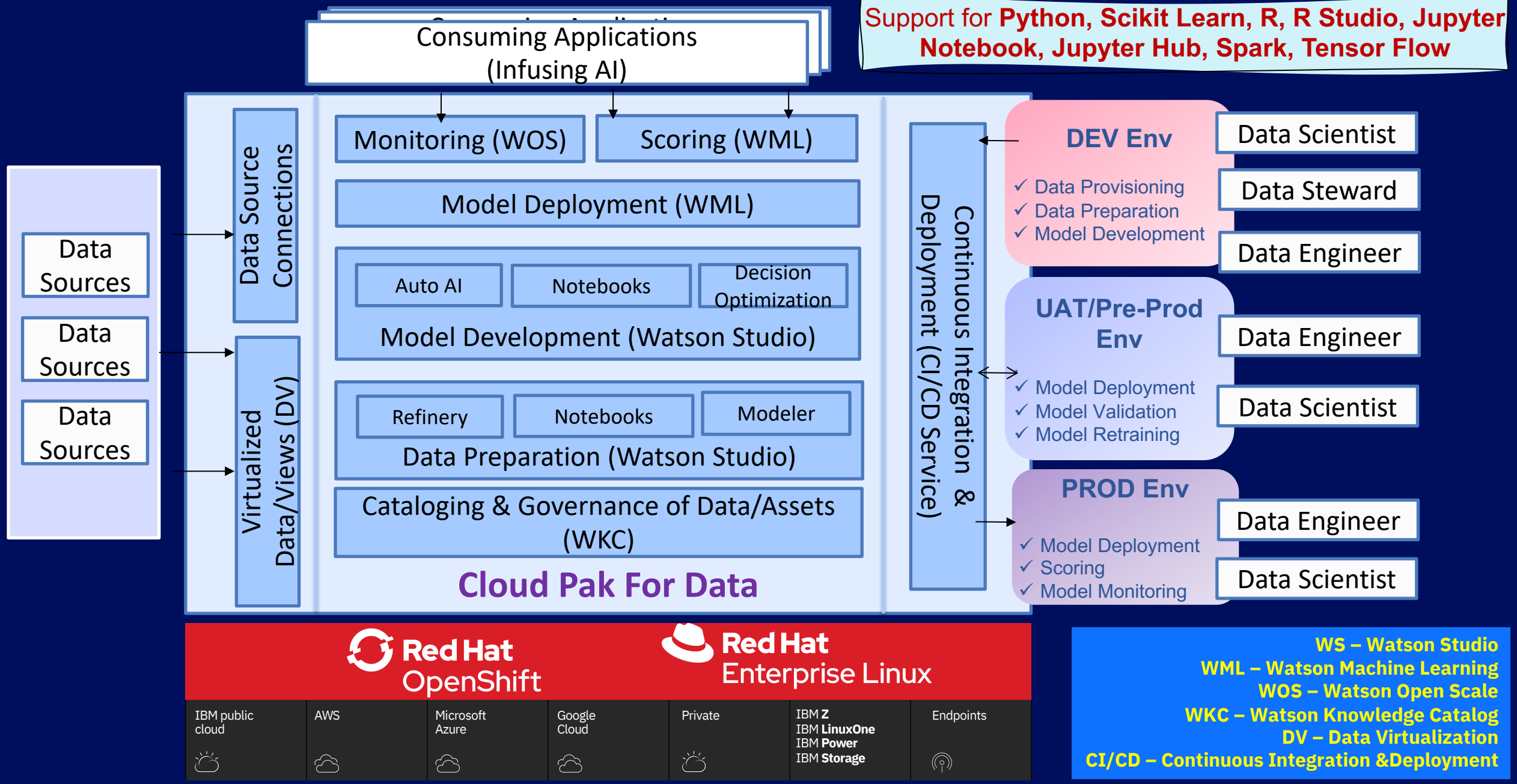
mlflow

Apache Airflow

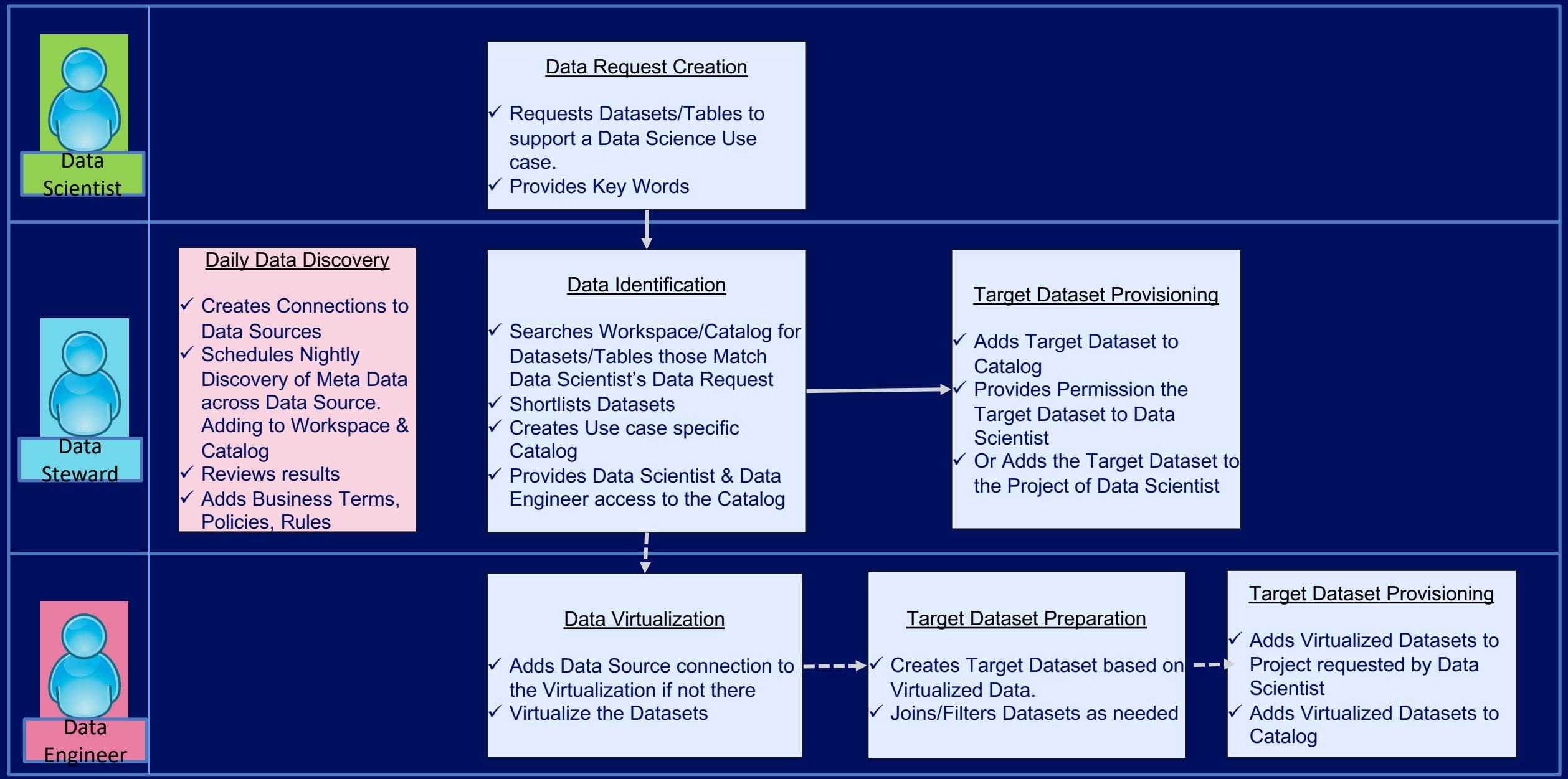
From Software Vendors



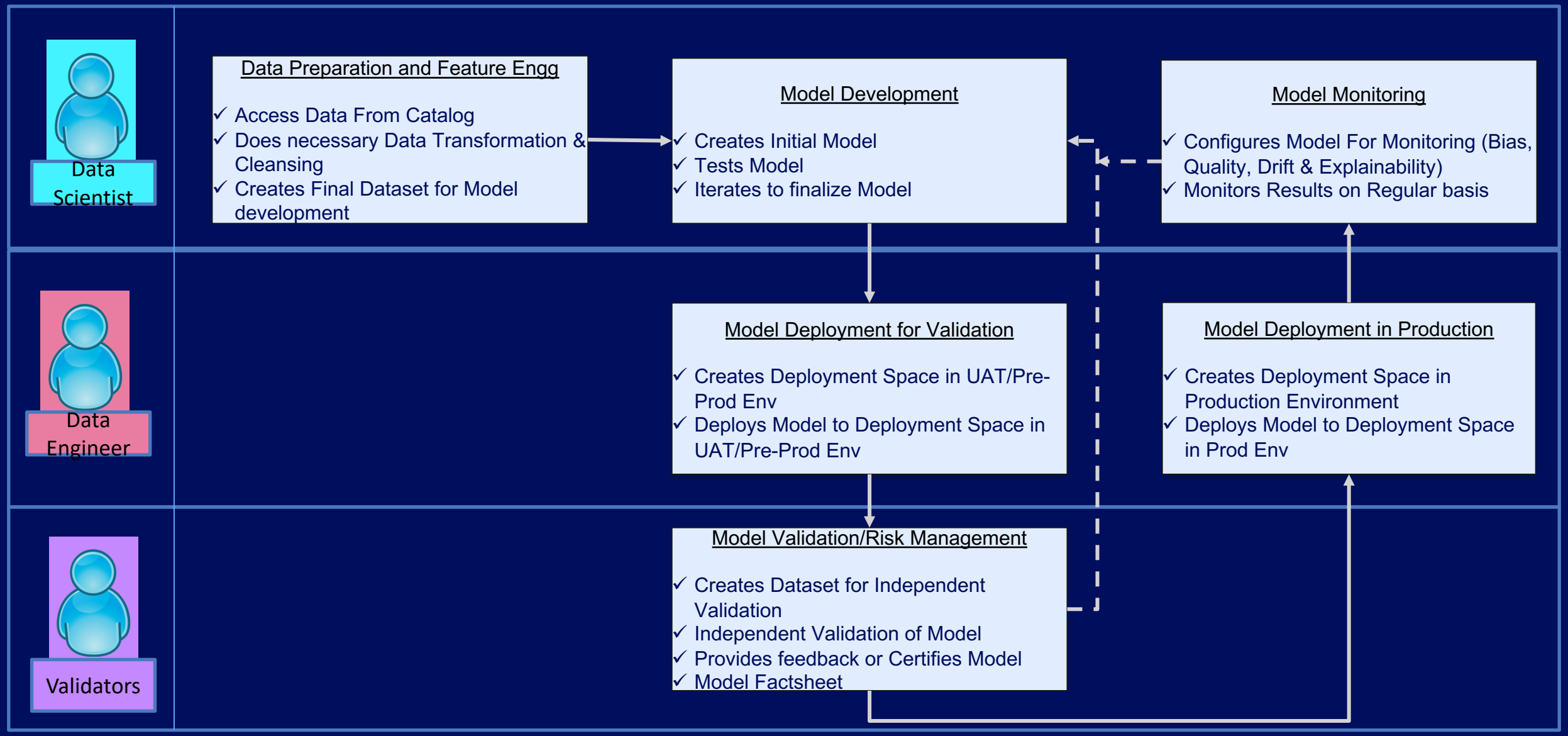
ML Operationalization with IBM Cloud Pak For Data



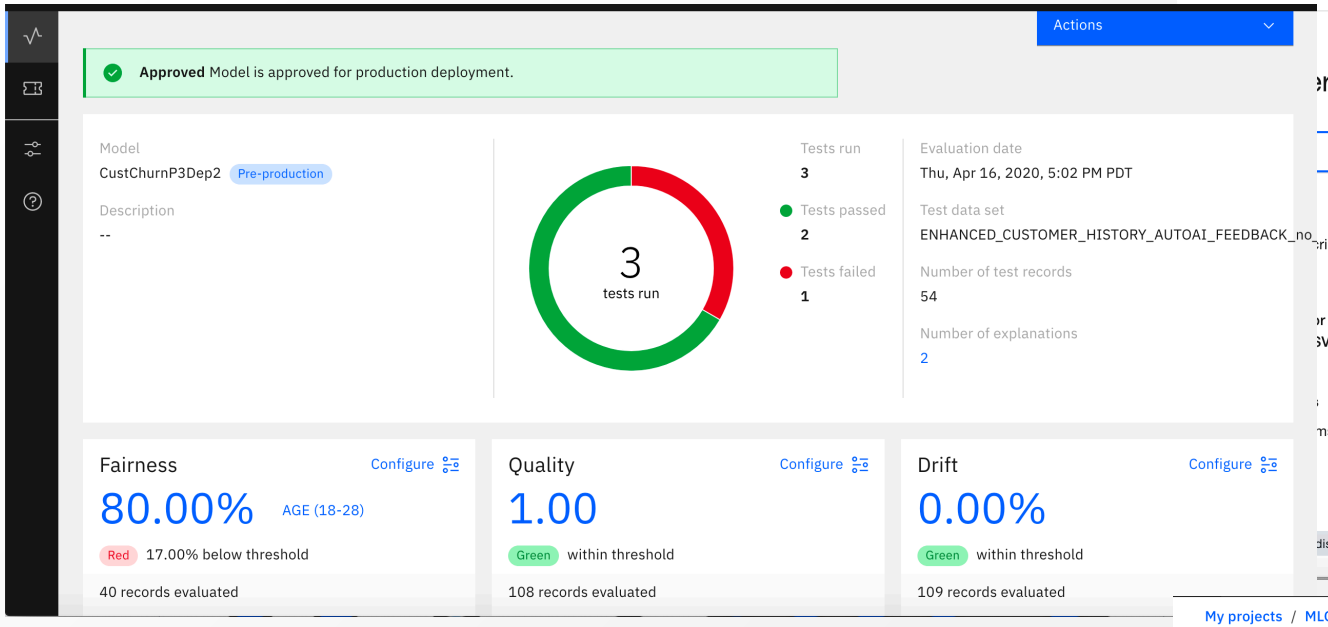
ML Ops In Action (1/2) - Data Provisioning and Governance

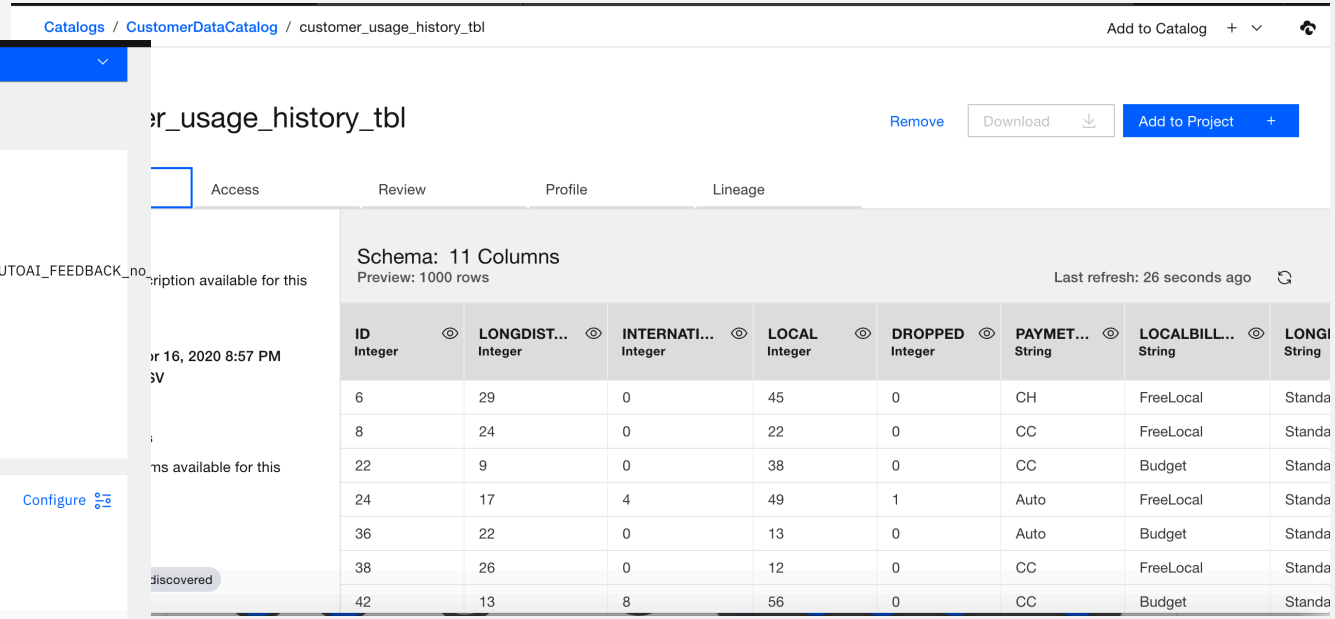


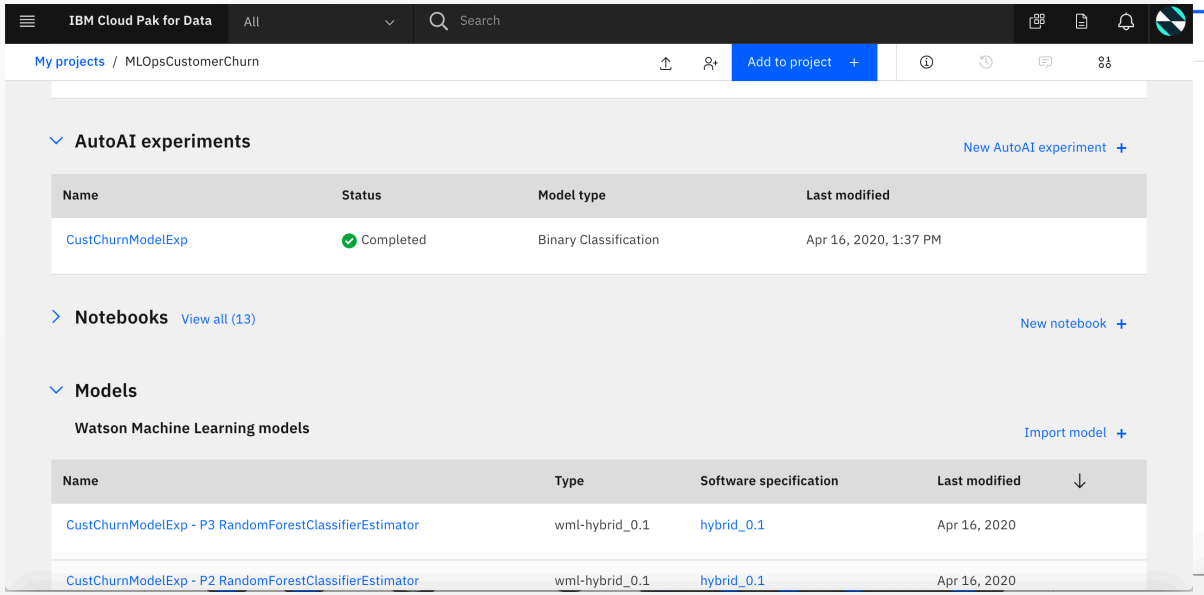
ML Ops In Action (2/2) - Model Development, Deployment & Monitoring

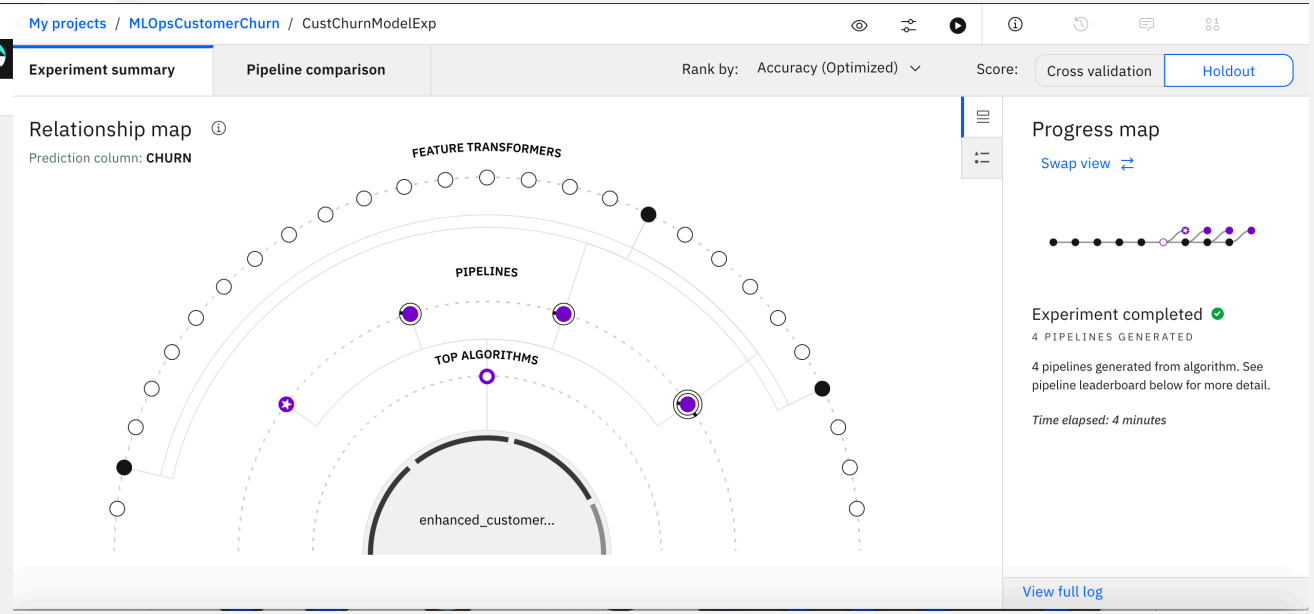


ML Ops - Demonstration



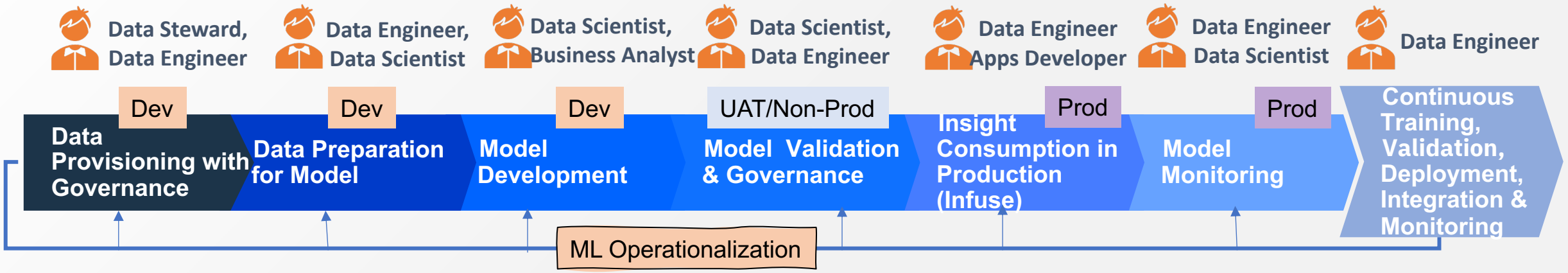






Get Started with ML Operationalization using ML Ops Starter Kit

End 2 End framework to help get started with ML Operationalization



Instructions | Code | Configurations | Data | Best Practices

1 Click Import

ML Ops Starter Kit

1 Click Configuration

Watson Knowledge Catalogue | Watson Studio | Watson Machine Learning | Watson Open Scale | Data Virtualization

Components/services

Dev/Non-Prod Edition | Standard Edition | Enterprise Edition | One Enterprise Edition | Cloud Pak For Data Experience (Community Version)

Cloud Pak For Data

Other ML Ops Frameworks

ML Ops – Self paced Hands On

- Get your free access in IBM Cloud Pak For Data Experience. https://www.ibm.com/cloud/paks/experiences/cloud-pak-for-data. You need to create an IBM account if you don't have one already.

- To start the Lab with the ML Ops Starter Kit –
  - Go to IBM Cloud Architecture Center- https://www.ibm.com/cloud/architecture/architectures/dataAIArchitecture/solutions
  - Click on the tile 'Drive business results from your machine learning Models' in Solution section
  - In the next page click on Get the Code – this will take you to Public Github of ML Ops Starter Kit
  - Go to Projects folder
  - Follow the ReadMe there to get started
  - Please note that due to the time limitation right now the Lab caters to part of the demonstration (Data Preparation, Model Development, Model Deployment, Infusing Model's Prediction and Model Monitoring) you saw today. Full lab would be coming soon in forthcoming version of Cloud Pak For Data (in May). You can get access to the same from the same github. **You are also welcome to contribute.**

Let us Discuss more about ML Ops

Q and A

Join us in https://github.com/ibm-cloud-architecture/refarch-ml-ops

Thank You