Model Operationalization

With Governance and Model Risk Management

Shikhar Kwatra
Data & AI Architect | Master Inventor (200+ Patents)
Youngest AoT Member | ML Operationalization Leader
IBM Data and AI Expert Lab and Learning
skwatra@us.ibm.com
https://www.linkedin.com/in/shikharkwatra/

Agenda

- Introduction to ML Operationalization
- ML Operationalization - Process, Persona, Environments and Frameworks/Platforms
- ML Operationalization in Action with Governance and Model Risk Management – Demonstration
- Q&A/discussion

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

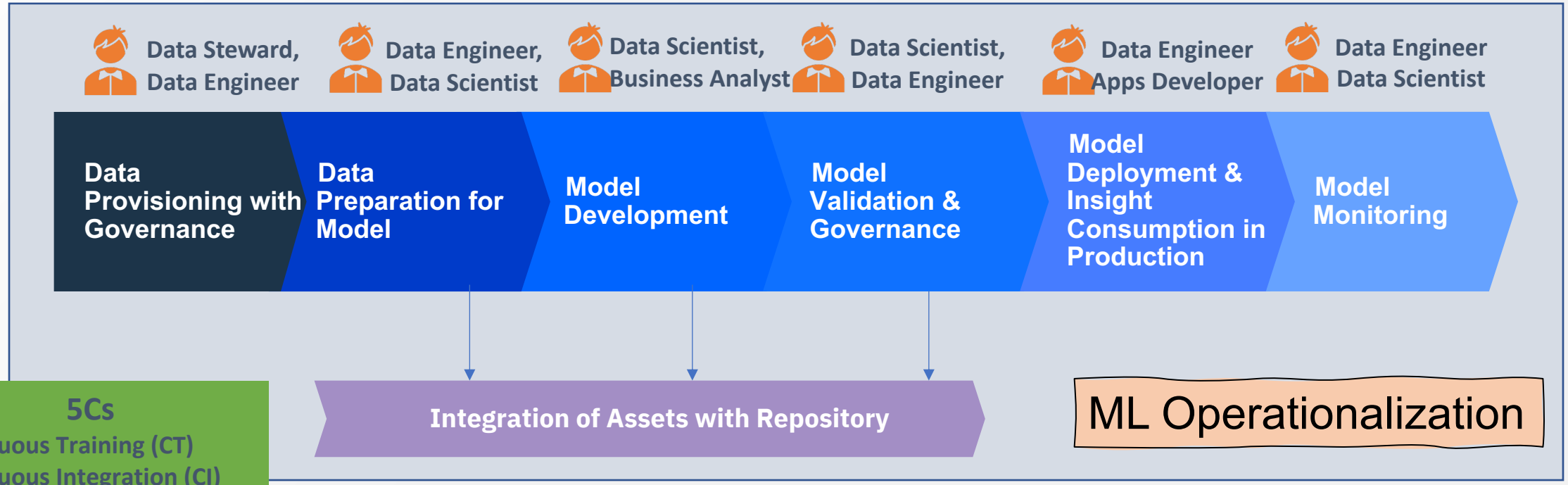
"Creating an ML model is just a starting point. To bring the technology into production service, you need to solve various real-world issues such as building a **data pipeline for continuous training**, **automated validation** of the model, **version control** of the model, creating a **scalable serving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



5Cs

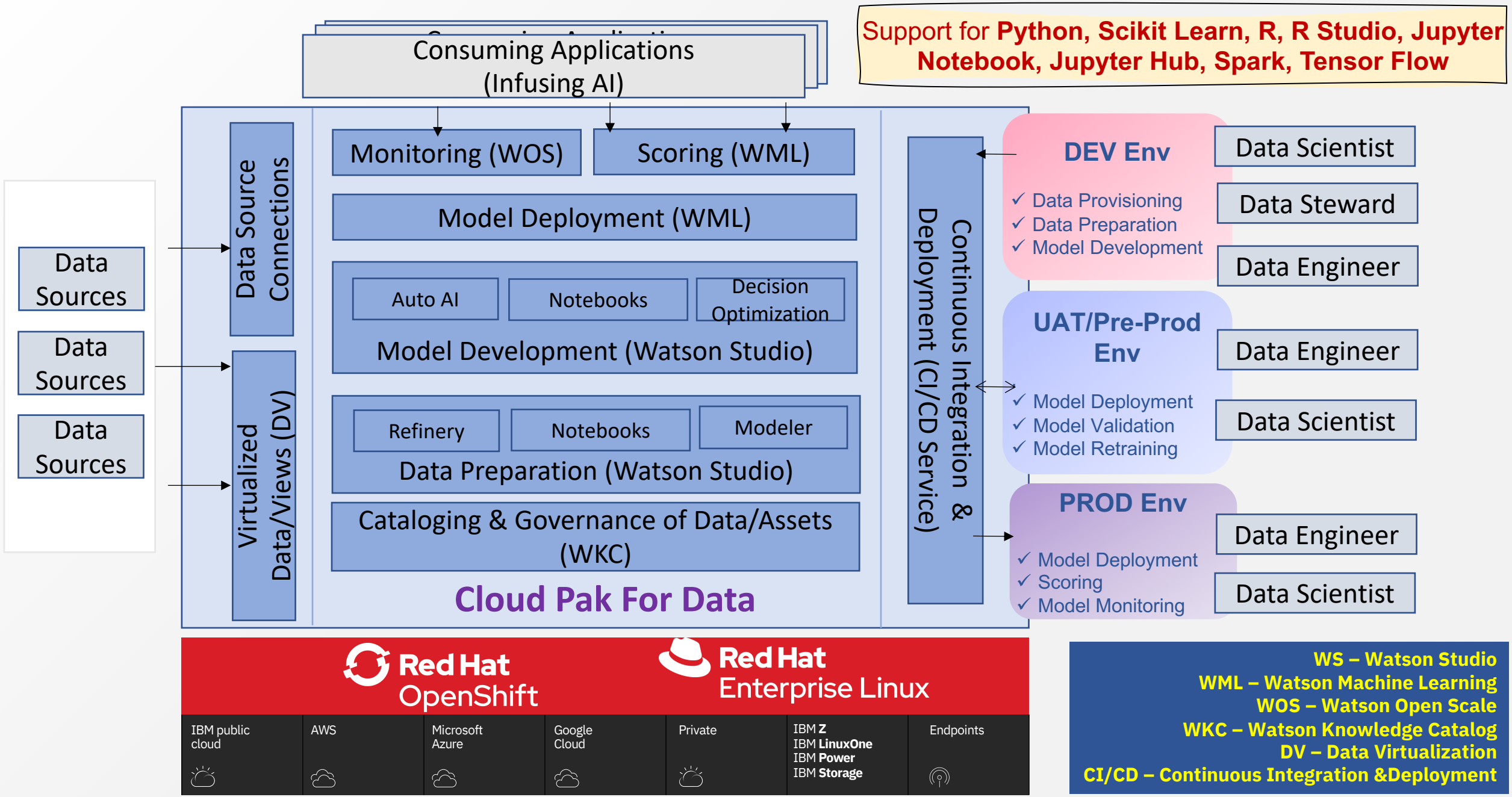
- Continuous Training (CT)
- Continuous Integration (CI)
- Continuous Validation (CV)
- Continuous Deployment (CD)
- Continuous Monitoring (CM)

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

The Non Functional Capabilities to look for in ML Ops Platforms/Frameworks

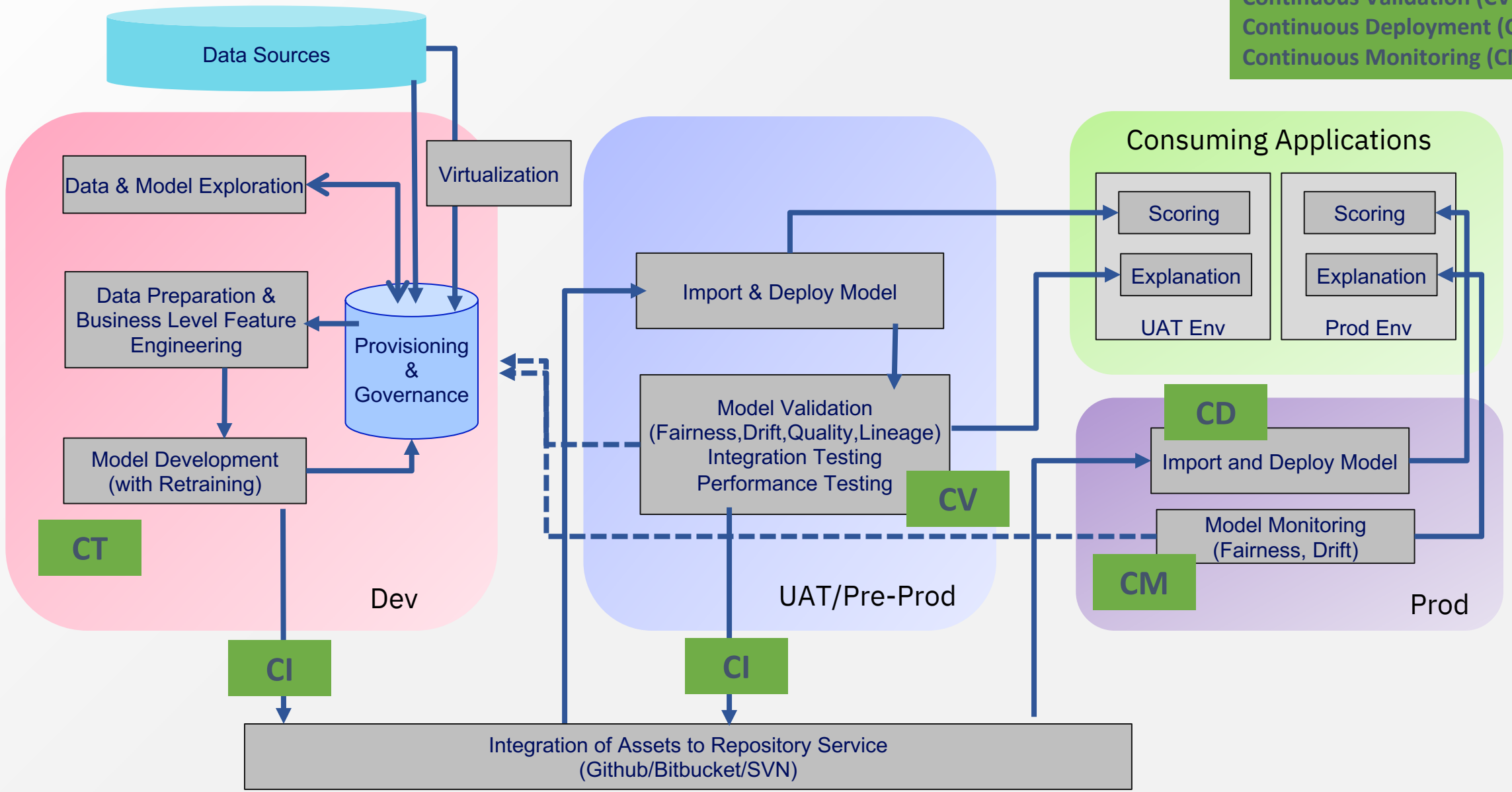
Features	Description
Flexibility/Customizability	How flexible is the platform in integrating and/or customizing new frameworks for AI model development.
Ease of Use	How easy is it to leverage these tools and proposed techniques from setup to application.
Integrations	How well does the platform integrate 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

ML Operationalization with IBM Cloud Pak For Data



ML Operationalization spread across Dev, UAT & Prod Environments

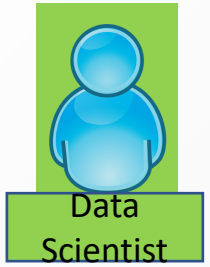
5Cs
- Continuous Training (CT)
- Continuous Integration (CI)
- Continuous Validation (CV)
- Continuous Deployment (CD)
- Continuous Monitoring (CM)



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

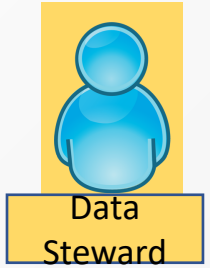
Watson Knowledge Catalog (WKC)

Data Virtualization (DV)



Data Request Creation — WKC

- ✓ Requests Datasets/Tables to support a Data Science Use case.
- ✓ Provides Key Words



Daily Data Discovery — WKC

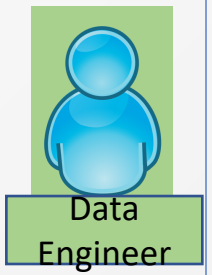
- ✓ Creates Connections to Data Sources
- ✓ Schedules Nightly Discovery of Meta Data across Data Source. Adding to Workspace & Catalog
- ✓ Reviews results
- ✓ Adds Business Terms, Policies, Rules

Data Identification — WKC

- ✓ Searches Workspace/Catalog for Datasets/Tables those Match Data Scientist's Data Request
- ✓ Shortlists Datasets
- ✓ Creates Use case specific Catalog
- ✓ Provides Data Scientist & Data Engineer access to the Catalog

Target Dataset Provisioning

- ✓ Adds Target Dataset to Catalog
- ✓ Provides Permission the Target Dataset to Data Scientist
- ✓ Or Adds the Target Dataset to the Project of Data Scientist



Data Virtualization — DV

- ✓ Adds Data Source connection to the Virtualization if not there
- ✓ Virtualize the Datasets

Target Dataset Preparation — DV

- ✓ Creates Target Dataset based on Virtualized Data.
- ✓ Joins/Filters Datasets as needed

Target Dataset Provisioning — DV

- ✓ Adds Virtualized Datasets to Project requested by Data Scientist
- ✓ Adds Virtualized Datasets to Catalog

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

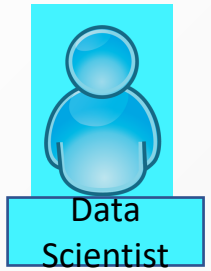
Watson Studio (WS)
Refinery, Notebook (Python, R, Spark, Open Source Libraries) , Auto AI, SPSS

Watson Machine Learning (WML)

Watson Open Scale (WOS)

3rd Party Model Development Tools (3rd Party Dev)

3rd Party Model Deployment Tools (3rd Party Dep)



WS (Refinery, Notebook, SPSS)

Data Preparation and Feature Engg

- ✓ Access Data From Catalog
- ✓ Does necessary Data Transformation & Cleansing
- ✓ Creates Final Dataset for Model development

3rd Party Dev

WS (AutoAI, Notebook, SPSS)

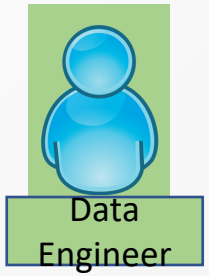
Model Development

- ✓ Creates Initial Model
- ✓ Tests Model
- ✓ Iterates to finalize Model

WOS

Model Monitoring

- ✓ Configures Model For Monitoring (Bias, Quality, Drift & Explainability)
- ✓ Monitors Results on Regular basis



WML

3rd Party Dep

Model Deployment for Validation

- ✓ Creates Deployment Space in UAT/Pre-Prod Env
- ✓ Deploys Model to Deployment Space in UAT/Pre-Prod Env

WML

3rd Party Dep

Model Deployment in Production

- ✓ Creates Deployment Space in Production Environment
- ✓ Deploys Model to Deployment Space in Prod Env

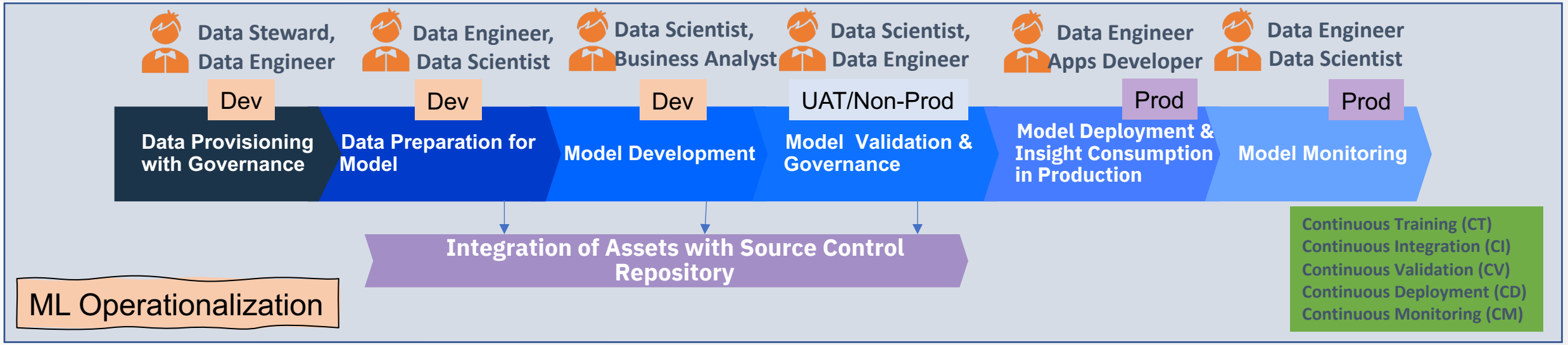
WOS

Model Validation/Risk Management

- ✓ Creates Dataset for Independent Validation
- ✓ Independent Validation of Model
- ✓ Provides feedback or Certifies Model
- ✓ Model Factsheet

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 Catalog | Watson Studio | Watson Machine Learning | Watson Open Scale | Data Virtualization

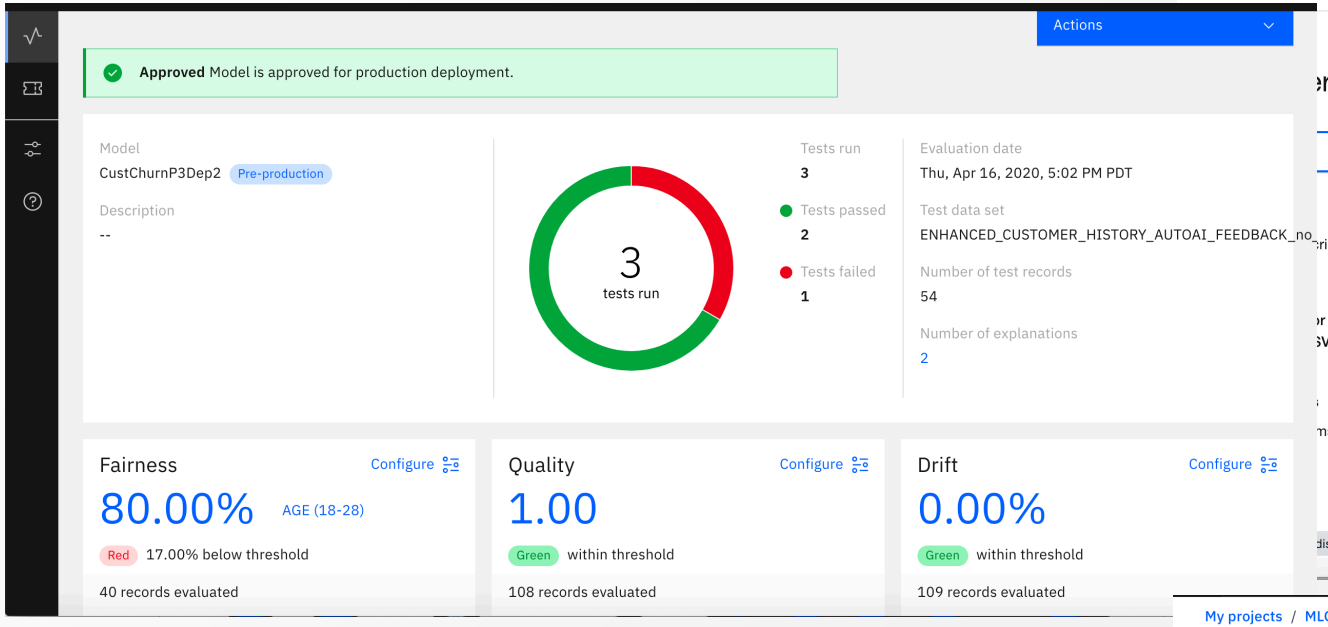
Components/services

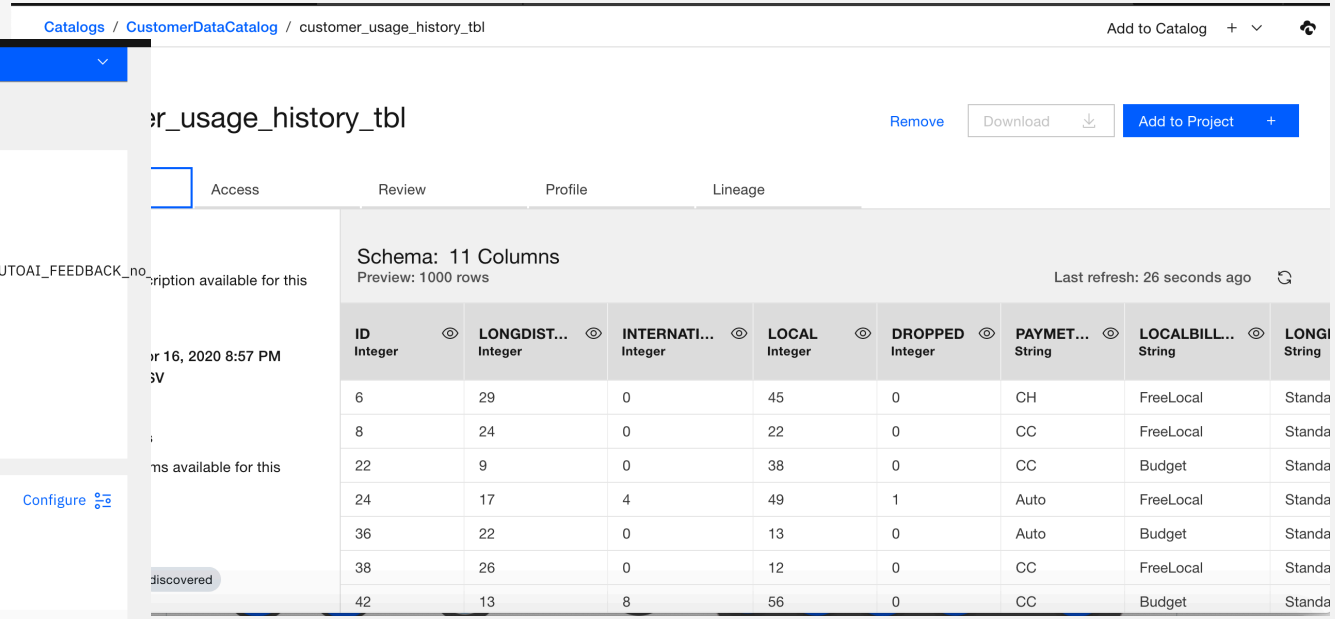
Dev/Non-Prod Edition | Standard Edition | Enterprise Edition | One Enterprise Edition | Cloud Pak For Data Experience (Community Version)

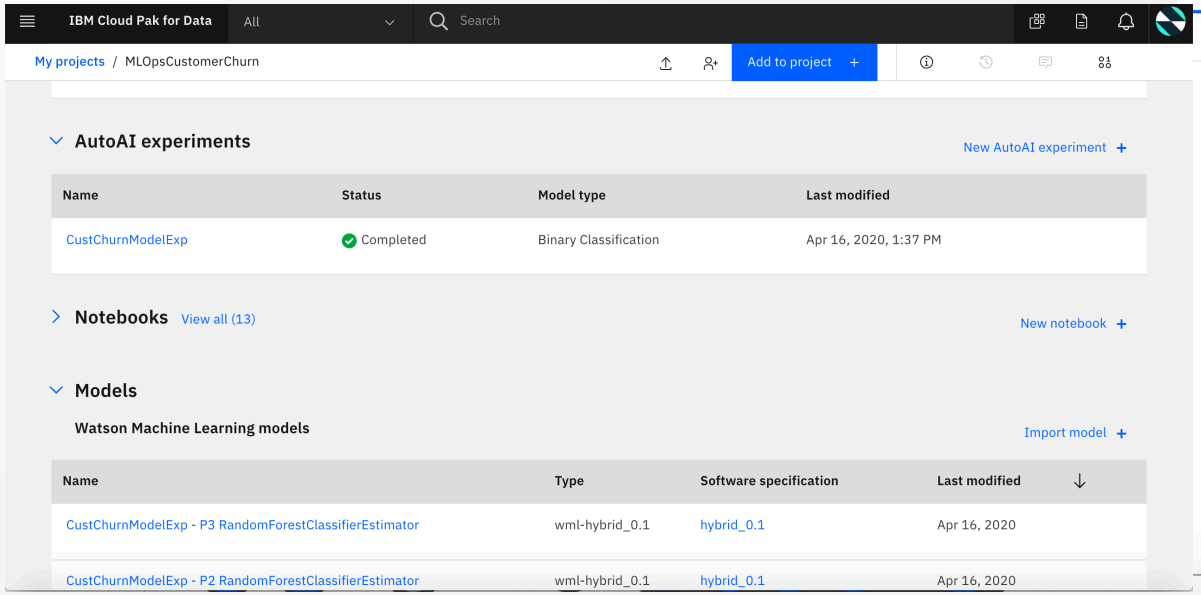
Cloud Pak For Data

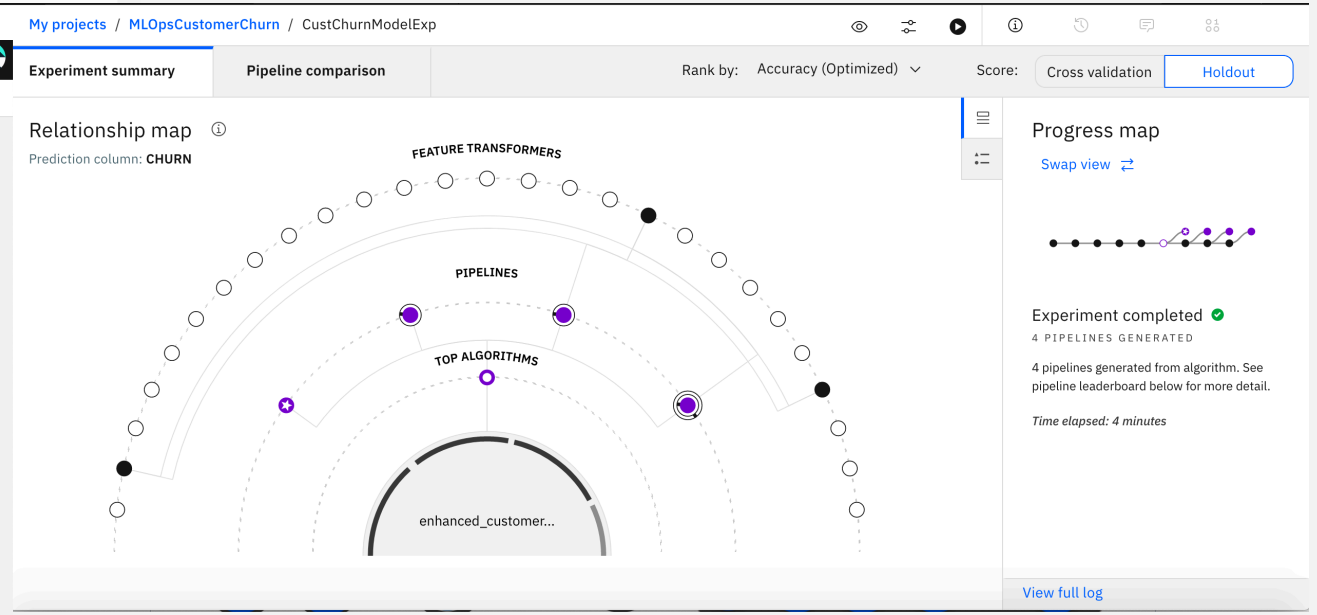
Other ML Ops Frameworks

ML Ops - Demonstration









Thank You