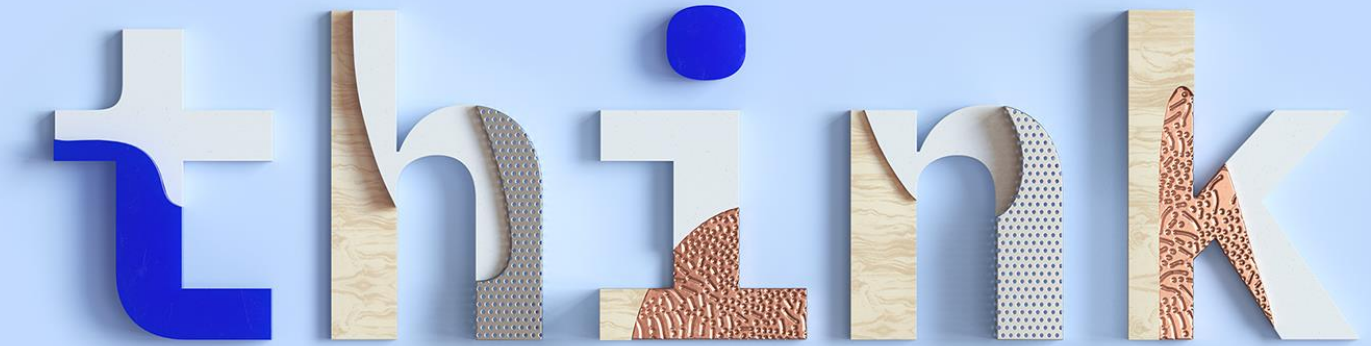
Bring Intelligence to Where Critical Transactions Run: IBM Machine Learning Technical Deep Dive

—

think 2018

Session 2702



Kewei Wei,
weikewei@c n.ibm.com
IBM Machine Learning Lead Architect

Maggie Lin,
meichi@us.ibm.com
IBM Machine Learning Development Lead

IBM

Machine learning is everywhere, influencing nearly everything we do…

Learn without being explicitly programmed

- Identifies patterns in historical data
- Builds behavioral models from patterns
- Makes recommendations

Waze personalized driving experience

Netflix personalized movie recommendations

7 out of 10 financial customers would take recommendations from a robo advisor

Agile Organizations

Convert insight into opportunity

- Personalizing every interaction
- Detecting fraud before a transaction completes
- Extending demographic reach
- Improving employee productivity
- Disrupting the competition and the disrupters

Look for innovative ways to improve insights & reduce costs

- Real-time actions are a game changer
- Transactional data is critical to real-time predictive insight
- Minimizing data movement drives cost efficiency, improves governance and security and reduces decision latency risk

IBM MACHINE LEARNING

Platform agnostic functionality with the same look and feel across deployment options

Integrates the most popular open source tools such as Spark, Python, R, Jupyter & Scala.

Machine Learning on IBM Cloud

PayGo consumption with as-a-service delivery, up & running in seconds

Integrated with IBM Spark-as-a-Service for compute, IBM Object Store for data, as well as other platform assets

Immediate cloud collaboration via RStudio and Jupyter notebooks

Machine Learning in IBM DSX

Scalable DSX cluster deployed on your private infrastructure

Can be deployed with Hortonworks Data Platform on-premises

LDAP integration for user management

Meaningful collaboration via notebooks, IDEs, community, and social features

Machine Learning for z/OS

On-premises, IBM Z, scalable Spark cluster deployment on private cloud

Cost effective strategy for lowest latency and highest degrees of resiliency and security

DSX is built into the ML for z/OS infrastructure

Close Integration with Z

Machine Learning for z/OS Brings Advanced Analytics and Compute to Your Data

IBM Z Point of View:

- Move compute to the data (Data Gravity)
- Industry leading encryption and security
- Resiliency on par with transactional apps
- Combine insight from any platform
- Leverage existing people, processes and infrastructure

Enabling:

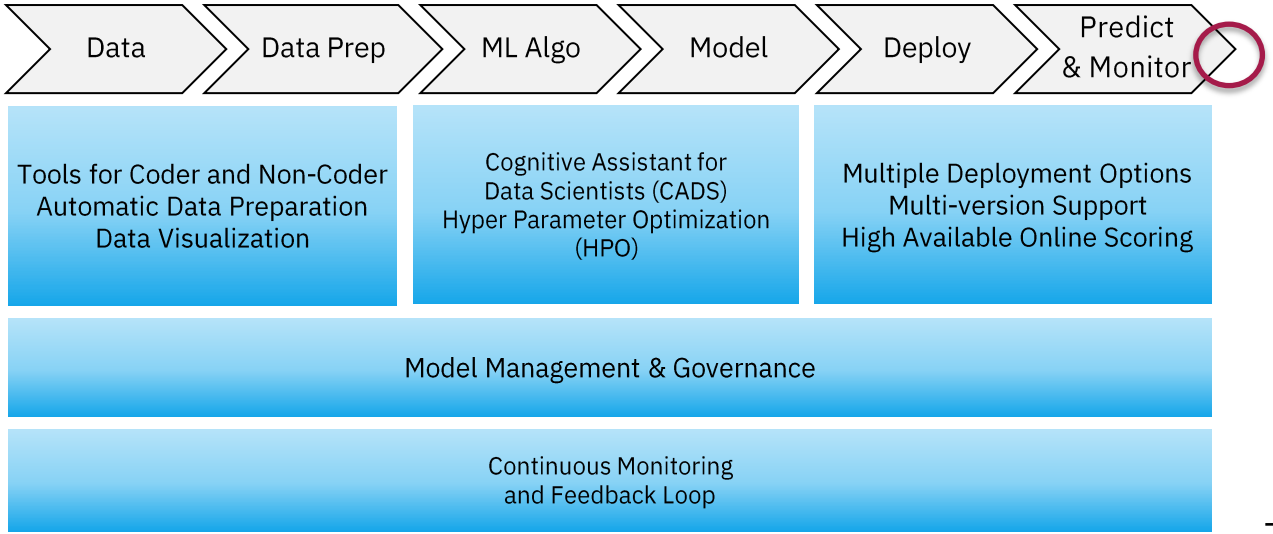
- Personalize every interaction
- Convert insight into opportunity and opportunity into revenue
- Automatically identify and minimize risk
- Disrupt the competition and the disrupters
- Drive down costs

Machine Learning for z/OS Provides Full Lifecycle Management for Faster ROI

Provides an end-to-end framework for ML in an enterprise environment

Leverage open source frameworks algorithms for ease of integration and accessibility of skills

Drastically simplifies workflow and assists with each step of the process from UI, data prep, model selection, parameter selection, deployment and most importantly monitoring for continuous feedback



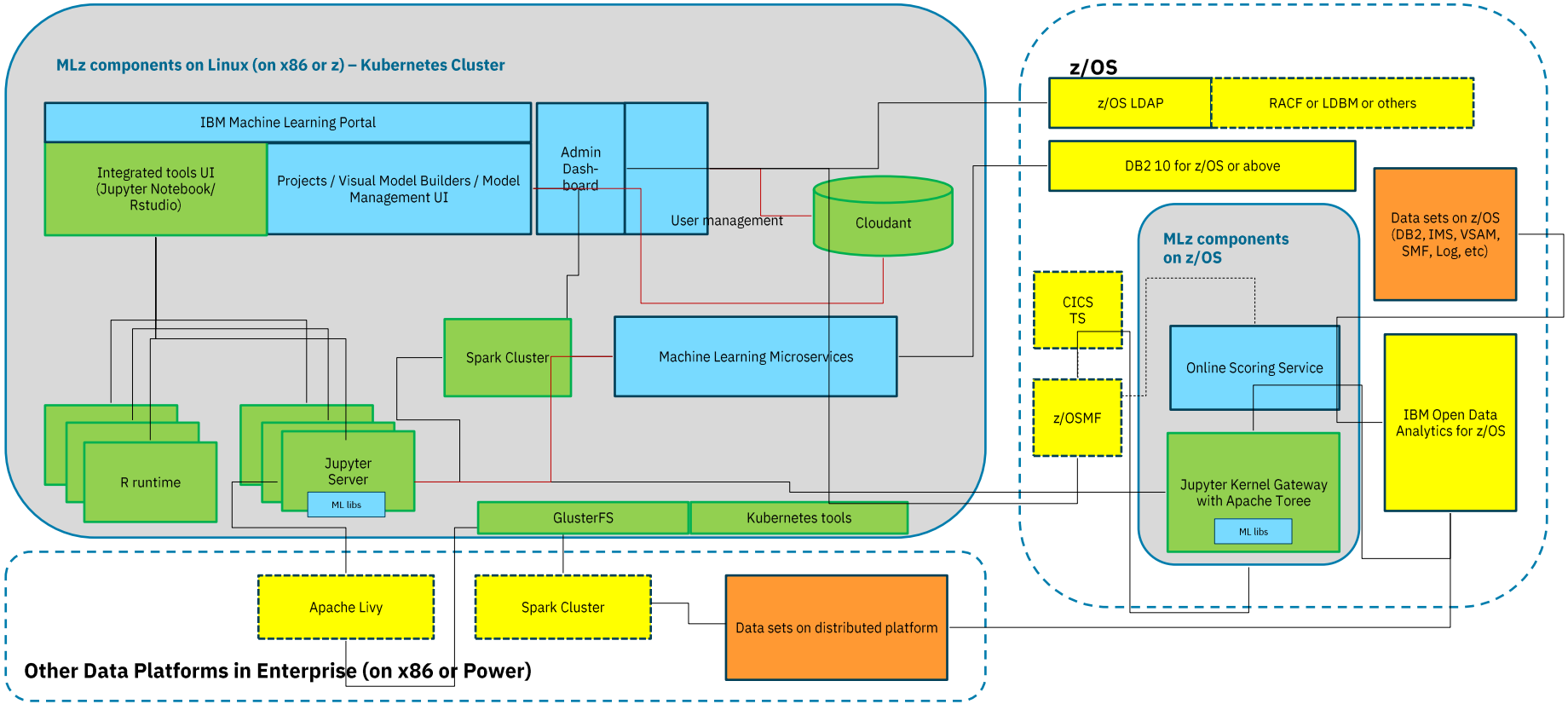
Data Gravity

Leverage current data for improved insights

Data-in-place for efficiency and more frequent model refresh

Optimized deployment at point of transaction

IBM Machine Learning for z/OS Architecture



IBM Machine Learning Roadmap in 2018

Integration

Moving compute to data - **Db2 Analytics Accelerator**

Best combination of rule-based decision with Machine Learning - **Optimization Decision Manager**

Compliance

Audit-able machine learning

- Audit traces for entire model life cycle
- Policy based trace for scoring request

Freedom

More choices for machine learning on z/OS

- Natural language processing packages
- Machine Learning libraries for R and xgboost

Simplification

Simplifying installation and upgrade for z/OS components of IBM Machine Learning

For Data Scientists

ML accelerators from IBM research

Cognitive Assists for Data Scientist (CADS) tools help select the optimal algorithm and parameters

Auto-Data Preparation tool helps transform data automatically

Data exploitation in-place

Data scientists can exploit Z data securely through Open Data Analytics for z/OS

MLz provides options to use any machine learning runtime close to data, structured and/or non-structured

Collaboration & Isolation

Multiple data scientists with different roles collaborate in one project

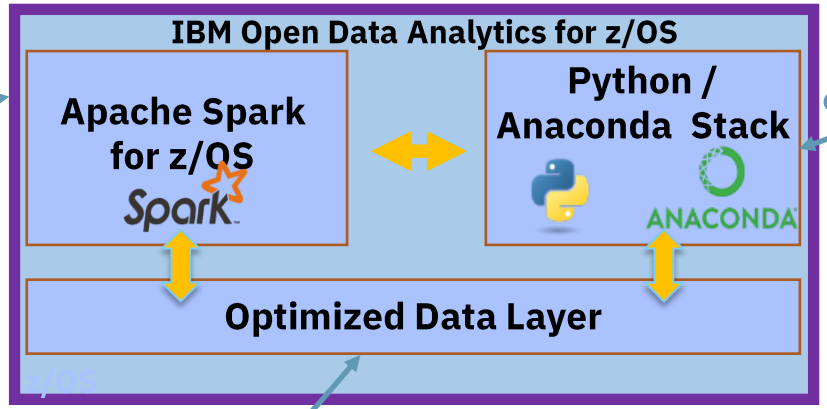
Project workspace is isolated and can have its dedicated resources and packages

Choices for all data scientists

MLz integrates tools for all types of data scientists – regardless of whether they are coders or non-coders, which programming language they learn, which ML libraries they prefer

IBM Open Data Analytics for z/OS: three major components

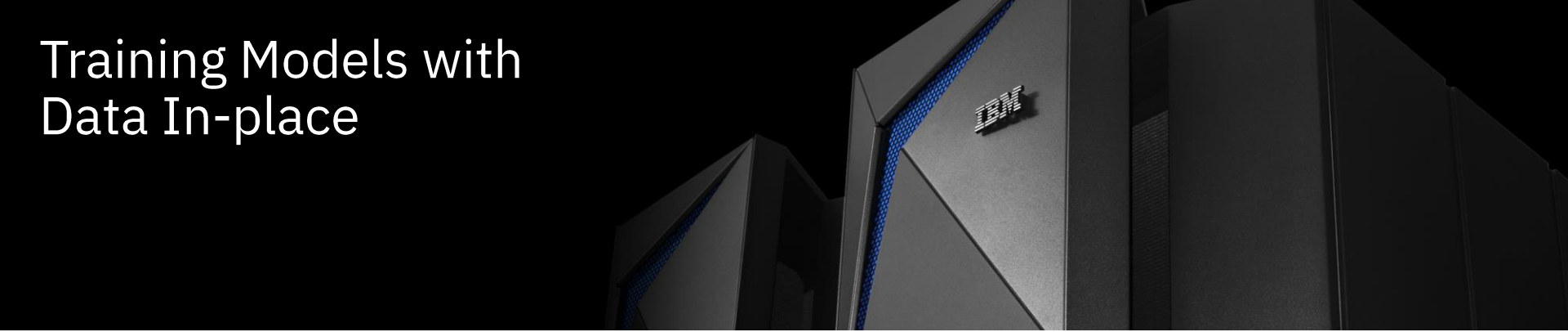
- In-memory analytics framework and runtime
- Supports applications in Java, Scala, Python
- Enables various styles of analytics: SparkSQL, Streaming, Machine Learning libraries, and Graph style processing
- Original availability March 2016
- New: Version 2.1.1 / 2.2.0 of Apache Spark
- New: Workload Management Integration
- New: Integrated end-to-end security



- Python 3.6.1 support
- Anaconda: core analytic libraries
- Enables access to the z/OS channel on the Anaconda Cloud for future package installs
- Python use in Jupyter Notebooks
- Python use from Apache Spark z/OS
- Integration with WLM and SAF capabilities

- Integrated , efficient access to data in place for wide variety of data on z/OS and off platform
- Access to Db2, IMS, IMS raw read (new), VSAM, PS, PDSE, ADABAS, IDMS (new), Virtual Tape, SMF, Syslog, Oracle Enterprise, Teradata, HDFS…
- Performance enhancements for memory-to-memory access of Spark / Python
- New: Integration with IBM Application Discovery and Delivery Intelligence
- New: Integration with business applications via persisted result store on z/OS
- New: Autogen of optimized data access Scala code
- New: Data Privacy Features

Training Models with Data In-place



Philosophy for model creation and deployment

– **Train models at where data resides, deploy models at where transactions run**

Machine Learning for z/OS supports

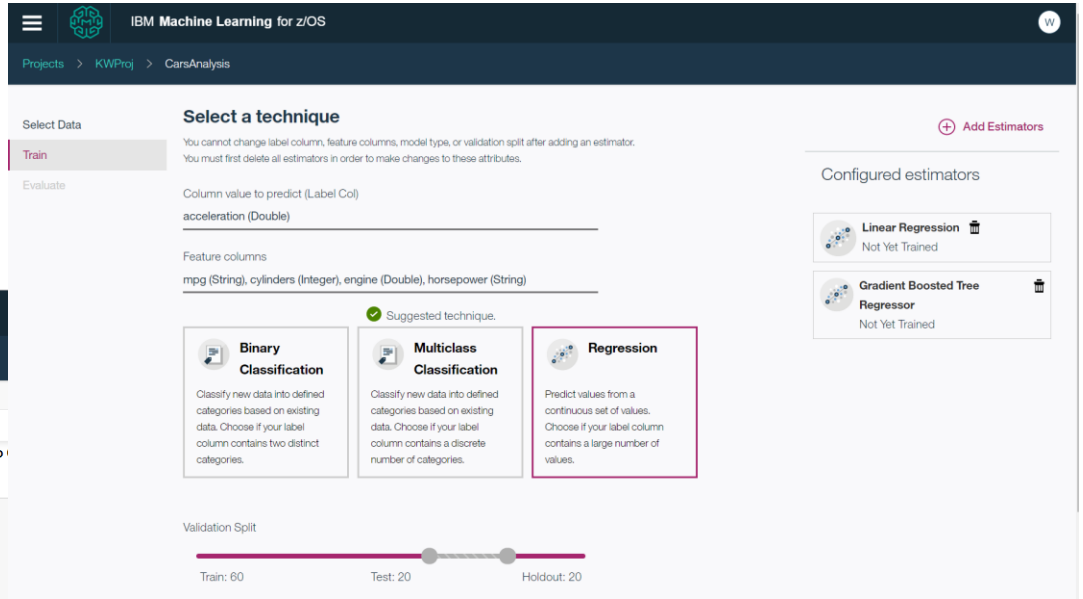
– **Train models on IBM Z**
  - With **IBM Open Data Analytics for z/OS (IzODA)**
    - If the majority of the training data is from z/OS
    - If the training data on z/OS cannot be offloaded for security or latency reason

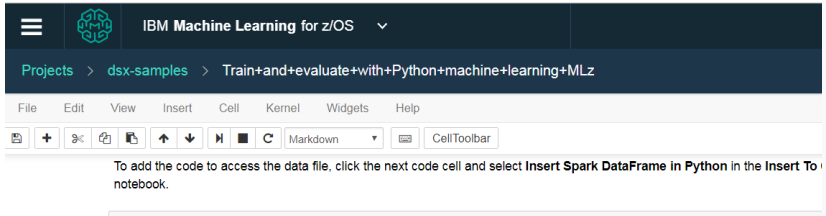
– **Train models off IBM Z**

- With **build-in Spark cluster**
    - If the majority of the training data is outside of z/OS, and
    - the training data set is not large <1G
    - Can add spark work resource through “**add compute node**” feature

– With **external Spark cluster**
    - If the majority of the training data is outside of z/OS

Model development tools for both coder and non-coder





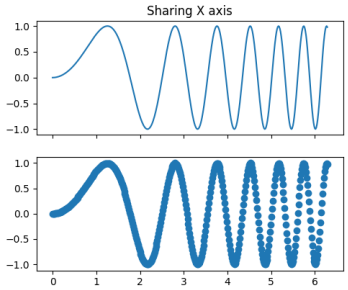
In [3]:
```
import os
from pyspark.sql import SQLContext
from pyspark import SparkContext
# initialize the SparkContext
sc = SparkContext.getOrCreate()
# Add asset from file system
filepath = dsx_core_utils.get_local_dataset(pc, 'cars.csv')
carsDataRaw = SQLContext(sc).read.csv(filepath, header='true', inferSchema = 'true')
```

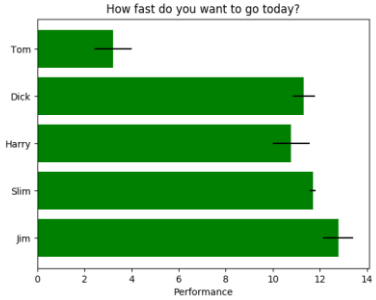
Note: Make sure the df variable in the following cell is the same as the generated code from insertToCode.

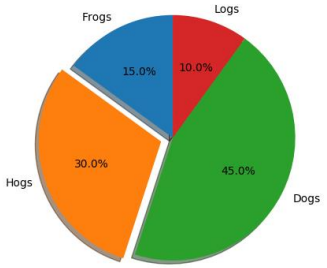
Due to missing data in `mpg` and `horsepower` columns, they will be excluded from the dataset for model training.

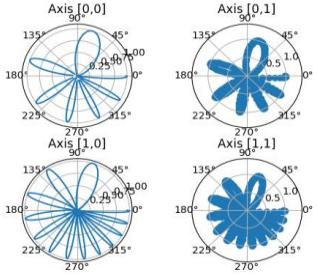
In [4]:
```
carsModData = carsDataRaw.drop("mpg").drop("horsepower")
carsModData.show(5)
```

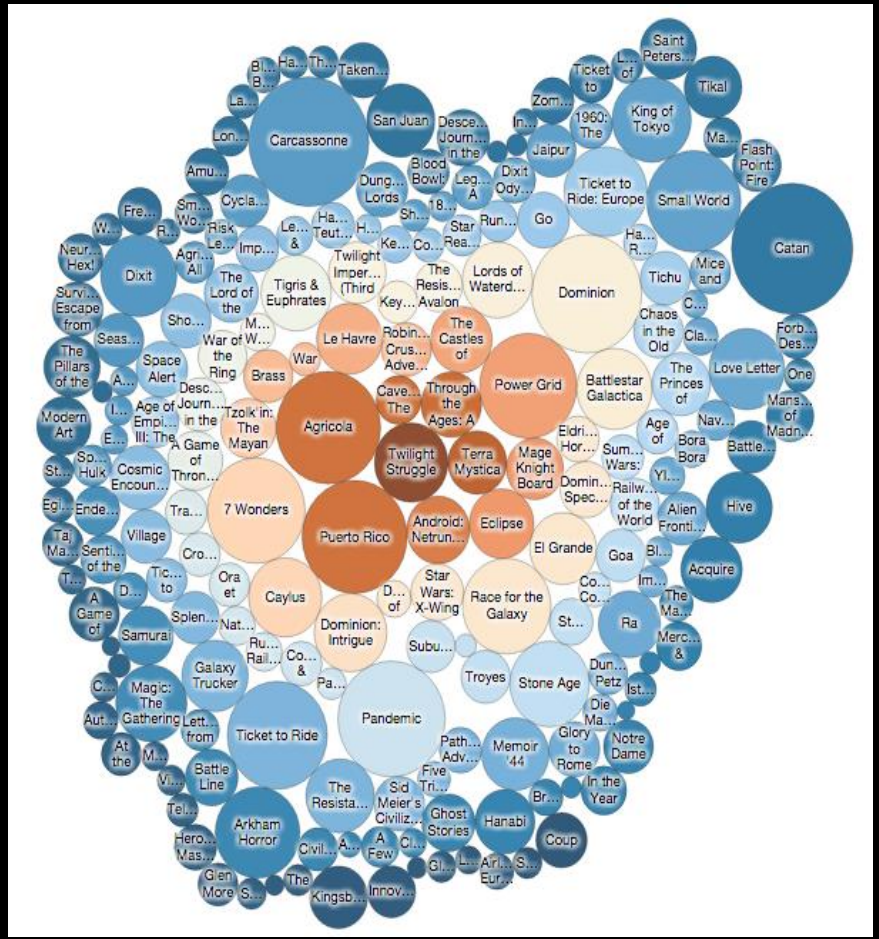
Built-in Visualization Libraries







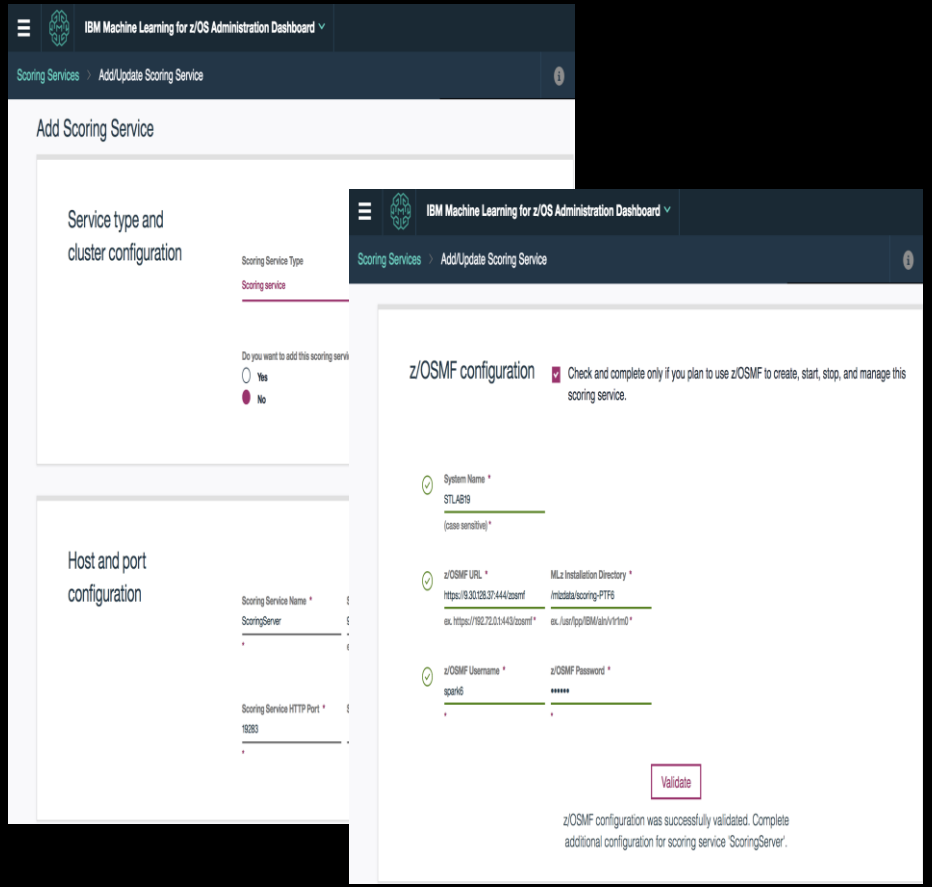




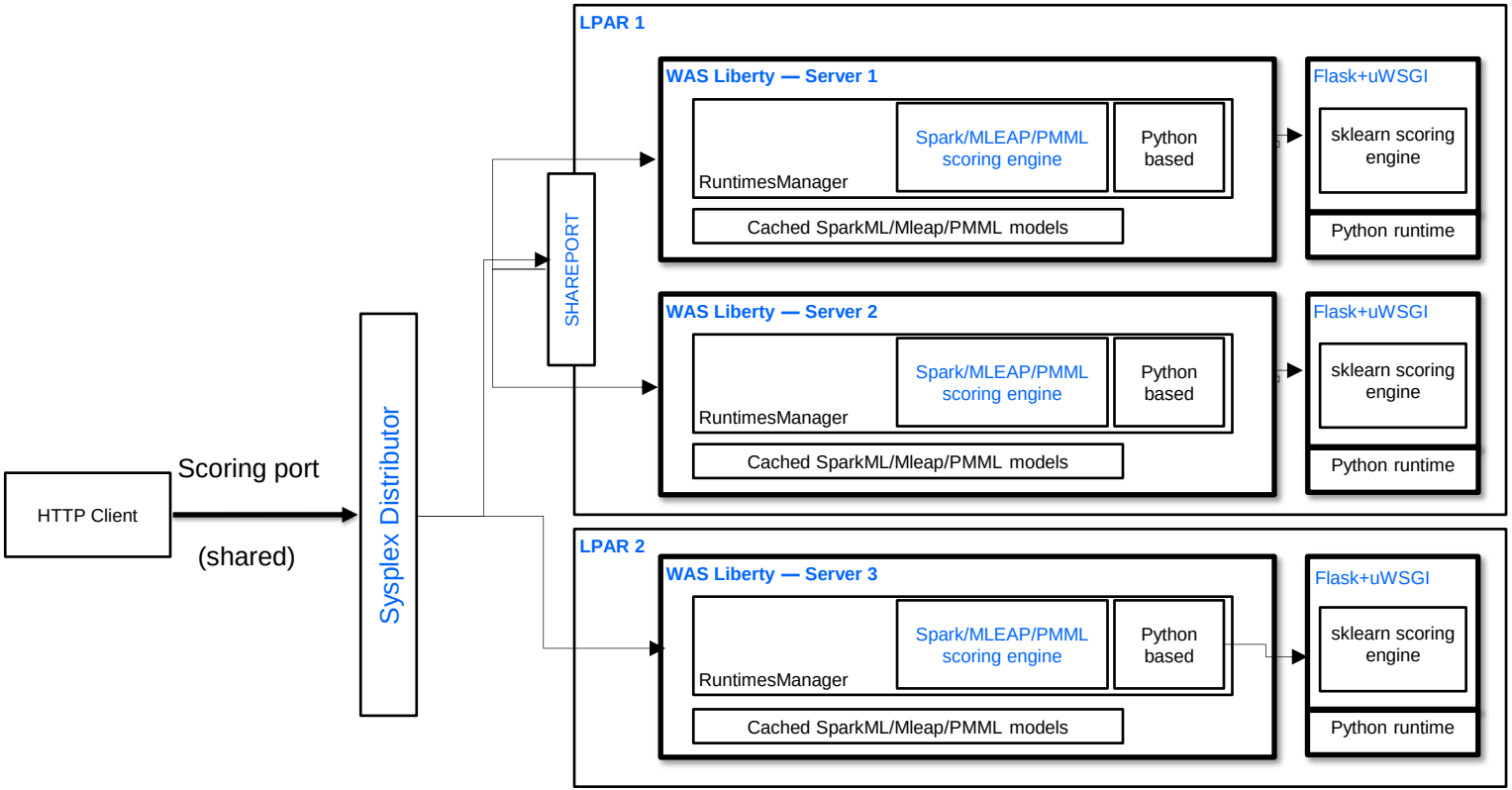
Manage Scoring Service through MLz UI

Scoring service configuration

- Define the scoring service through MLz UI administration dashboard
  - Continue to support HTTP and HTTPS access to the scoring server
  - Users continue need to configure the scoring server config file (e.g. scoring.cfg) and start/stop scoring server at the backend z/OS system
- A new capability in MLz v 1.1.0.5 (Feb/2018 release), user can manage scoring service completely through MLz UI
  - With this capability, the scoring server can be created/configured/started/stopped/removed directly through MLz UI Admin Dashboard
    - ➢ Support for SPARK/MLeap/Python scoring server.



High Available Online Scoring Service Cluster



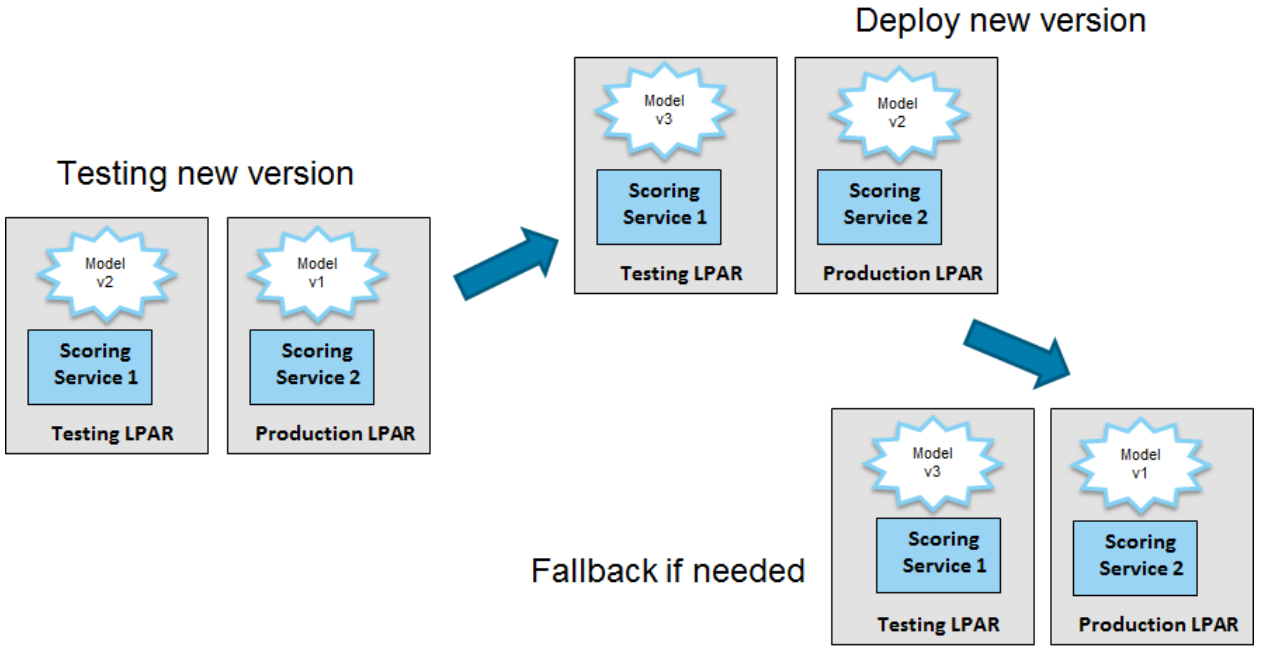
Each instance has its unique IP + administrative port

Versioning support for Models and Model Deployment

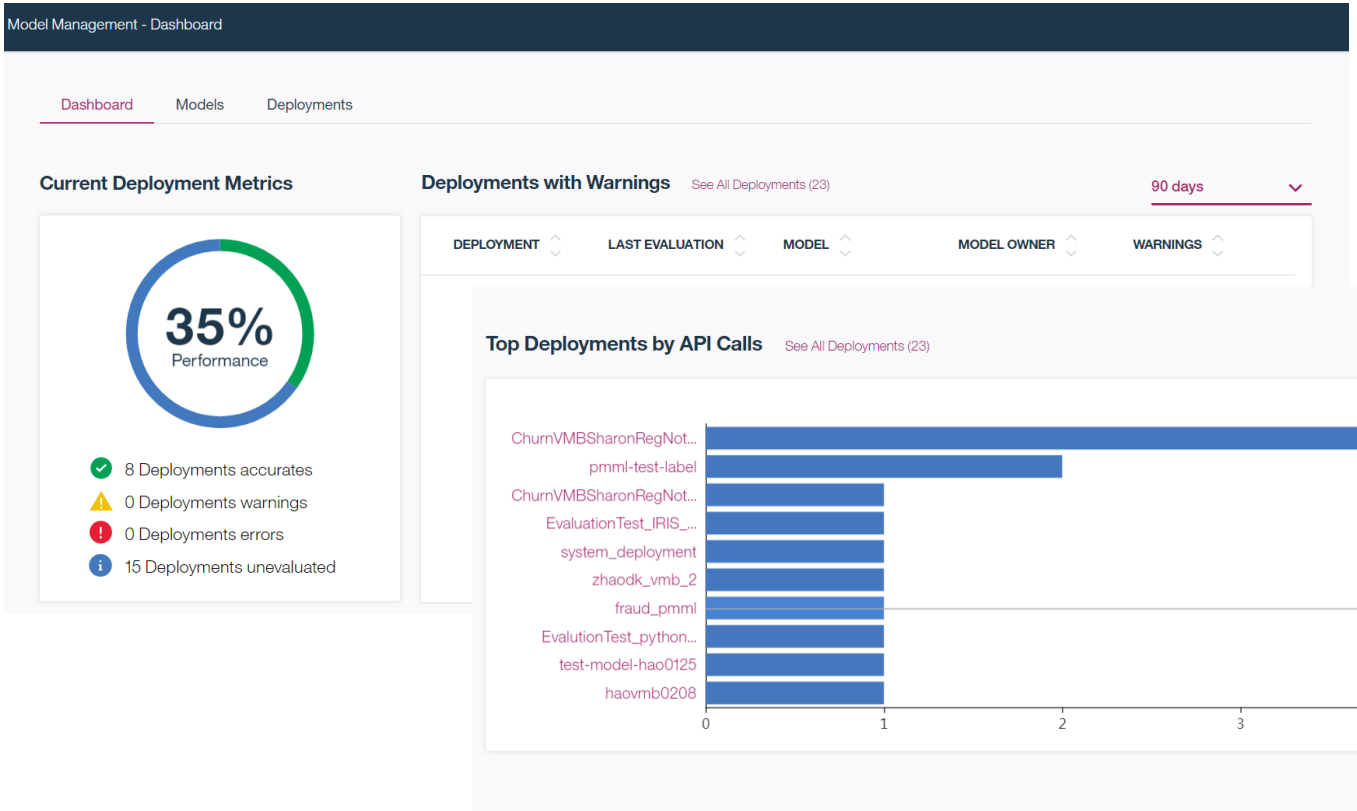
A model can have multiple versions

Different versions of the same model can be deployed to different online scoring services

Model version upgrade is transparent to applications



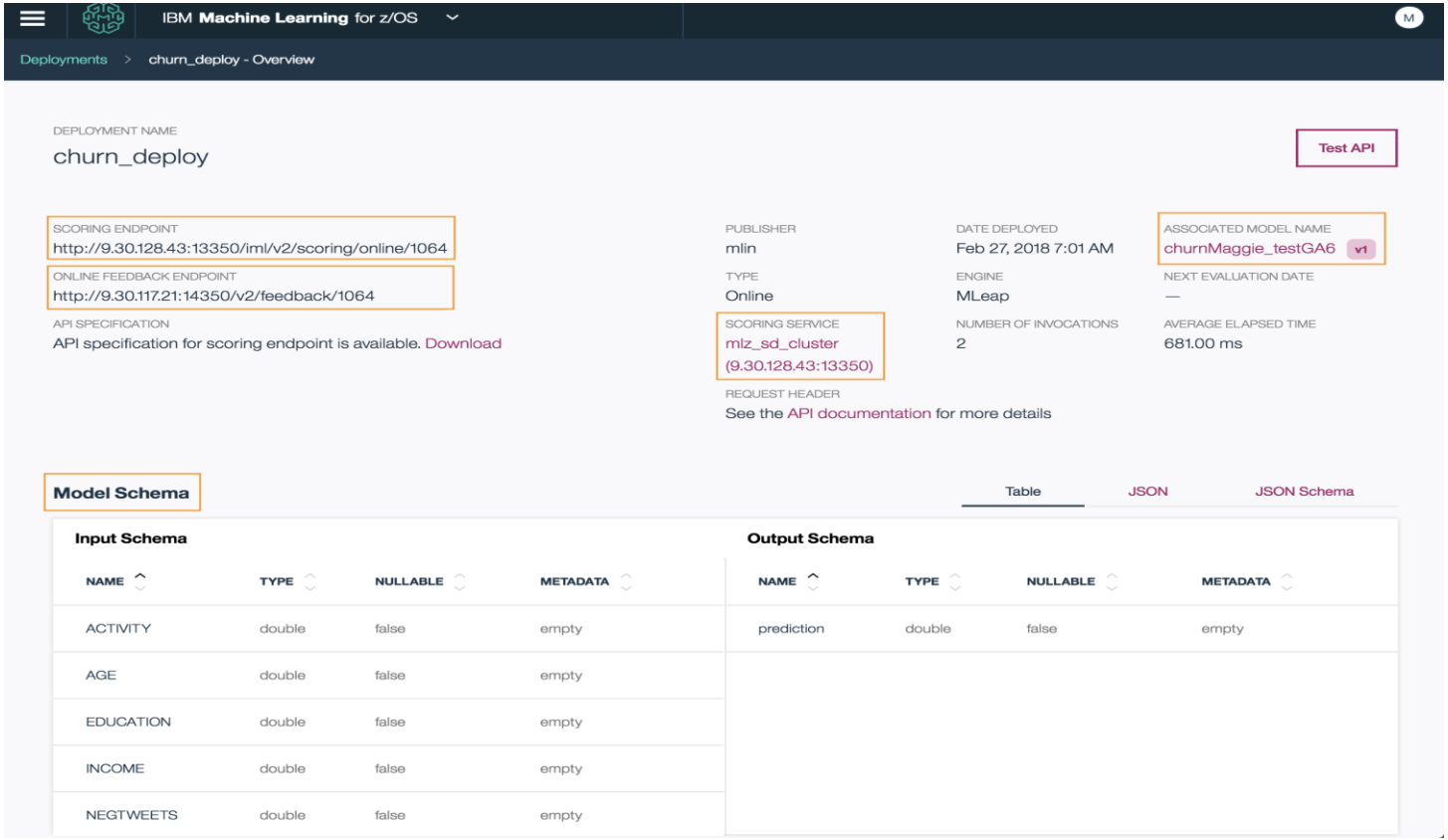
Model Management for Machine Learning Administrator



A single view to manage models and deployments

Deployment details for the model

API support for Application Developer

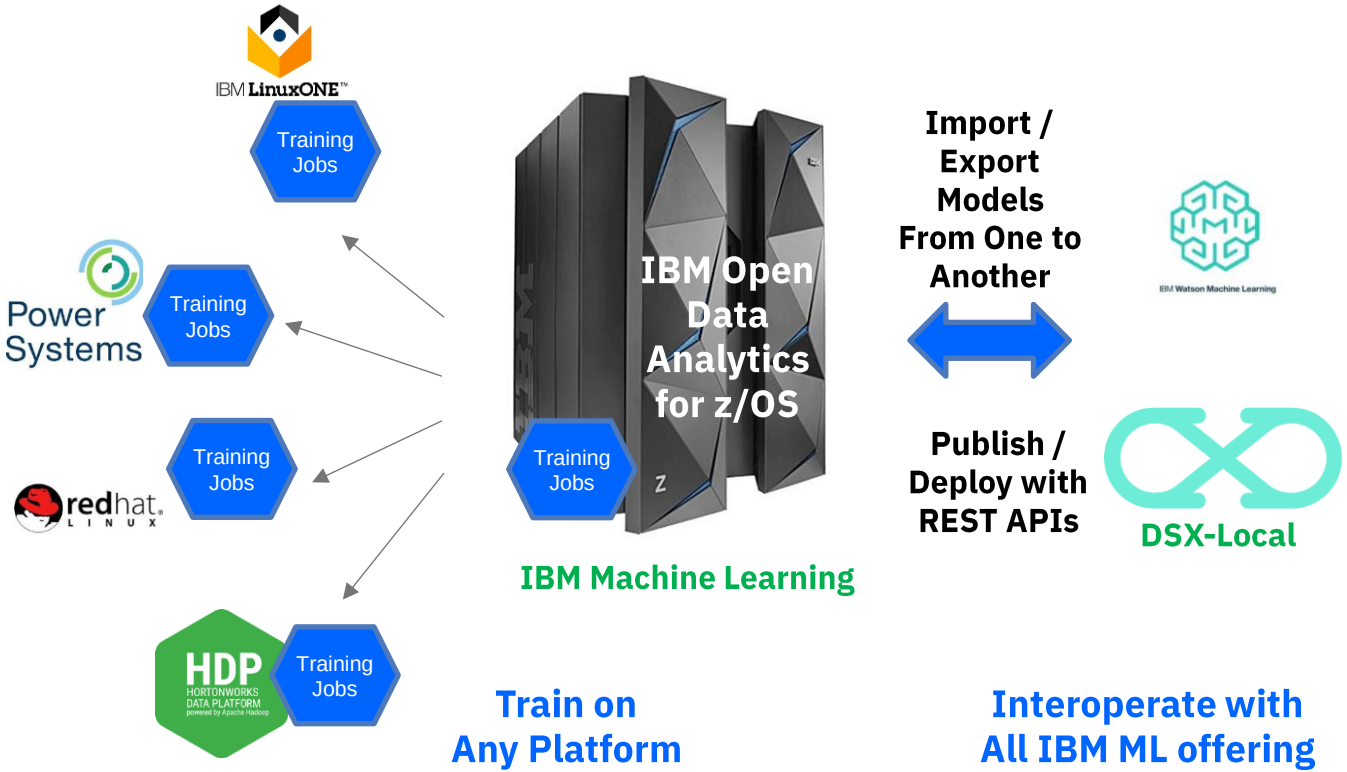


Flexible Deployment Options

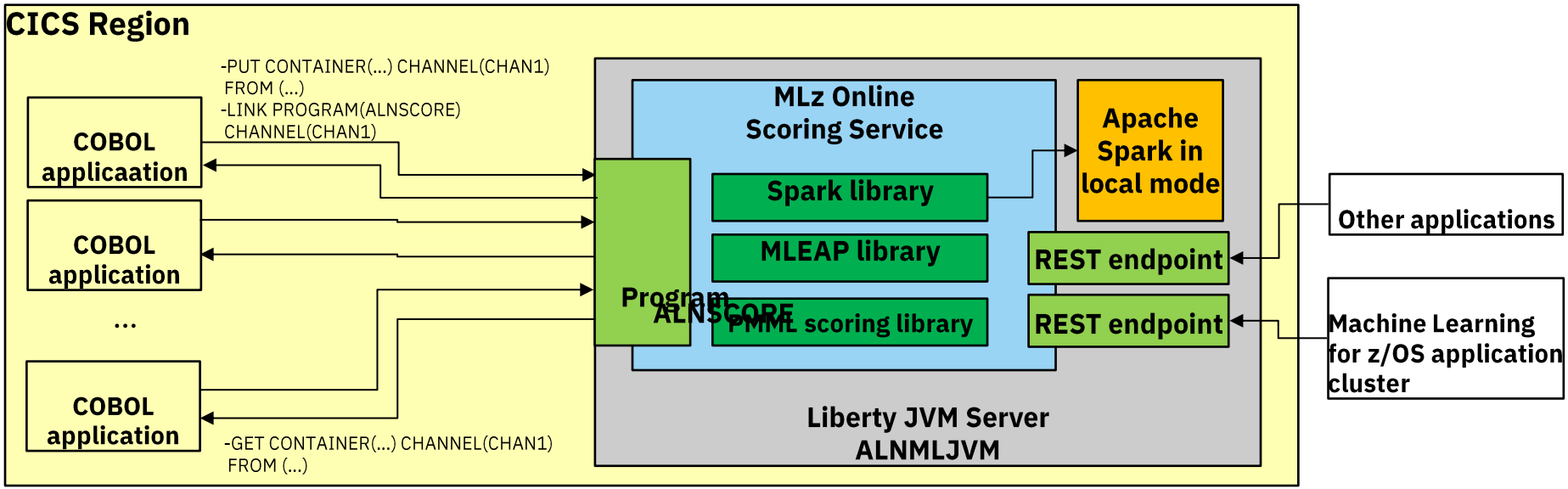
Train models on any platform, deploy on any platform

Train models in any IBM ML offerings (MLz, WML, DSX-Local), deploy on another

- Import / export
- Unified APIs



CICS-integrated Online Scoring Service

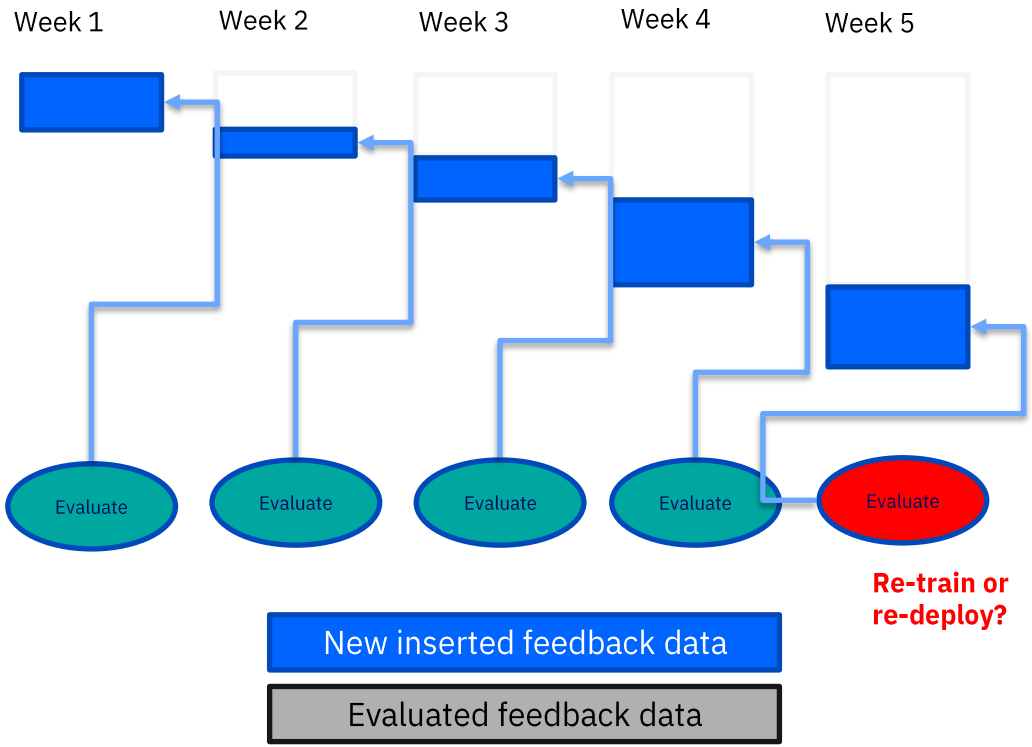


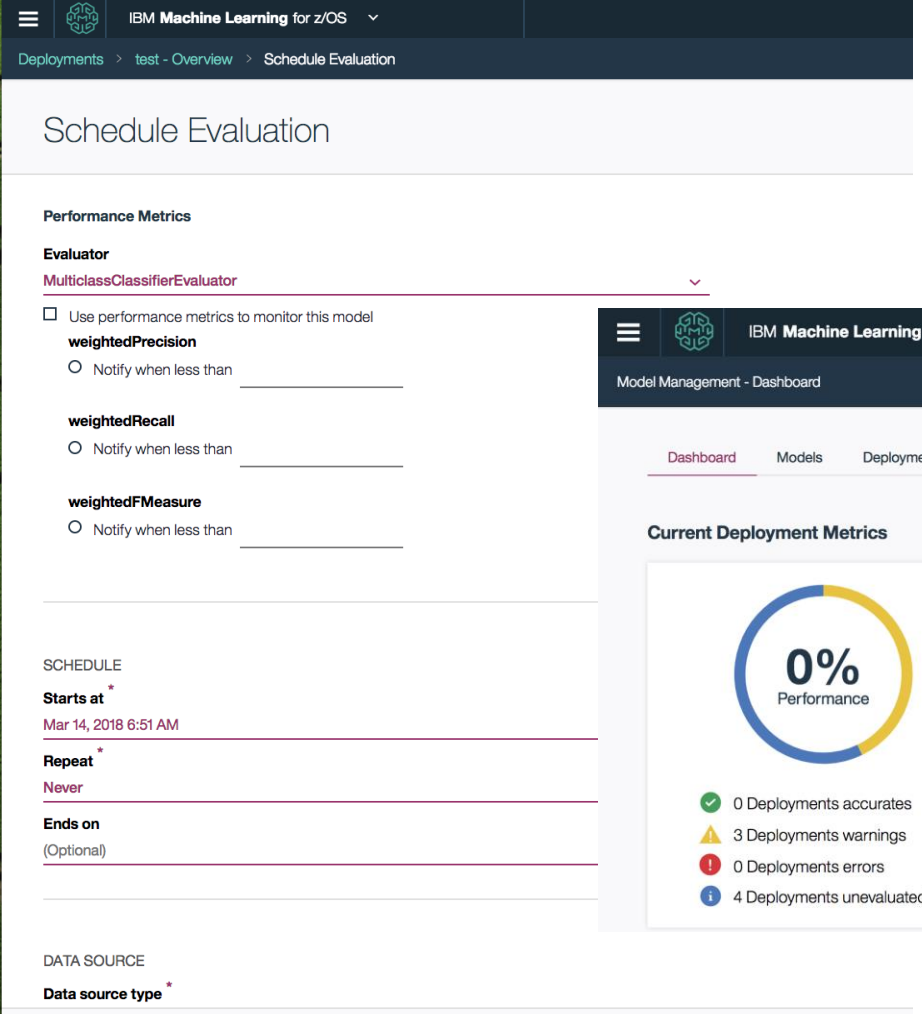
Continuous Learning System

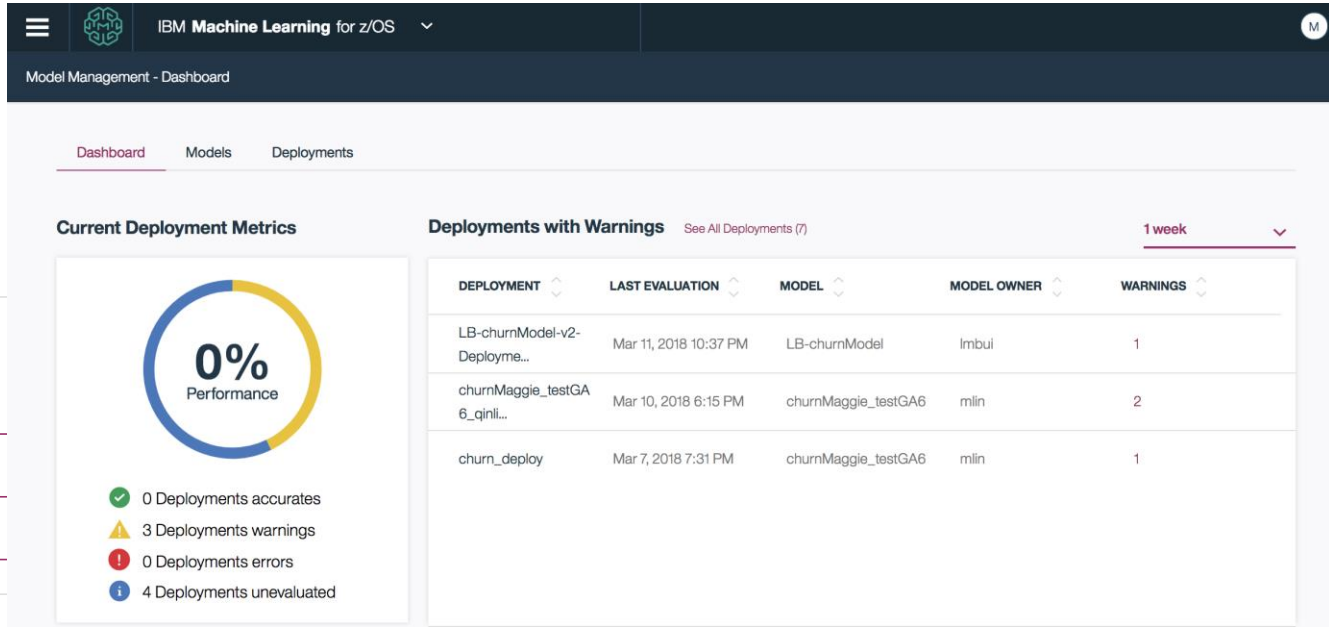
Feedback data is ingested to feedback dataset

An evaluation task monitors performance of a model with pre-defined threshold

Model is automatically re-trained (and optional re-deployed) to improve its own performance




IBM Machine Learning for z/OS
Deployments > test - Overview > Schedule Evaluation
Schedule Evaluation
Performance Metrics
Evaluator
MulticlassClassifierEvaluator
Use performance metrics to monitor this model
weightedPrecision
Notify when less than
weightedRecall
Notify when less than
weightedFMeasure
Notify when less than
SCHEDULE
Starts at *
Mar 14, 2018 6:51 AM
Repeat *
Never
Ends on
(Optional)
DATA SOURCE
Data source type *


IBM Machine Learning for z/OS
Model Management - Dashboard
Dashboard
Models
Deployments
Current Deployment Metrics
0%
Performance
0 Deployments accurates
3 Deployments warnings
0 Deployments errors
4 Deployments unevaluated
Deployments with Warnings
See All Deployments (7)
1 week
DEPLOYMENT
LAST EVALUATION
MODEL
MODEL OWNER
WARNINGS
LB-churnModel-v2-Deployme...
Mar 11, 2018 10:37 PM
LB-churnModel
lmbui
1
churnMaggie_testGA6_qinli...
Mar 10, 2018 6:15 PM
churnMaggie_testGA6
mlin
2
churn_deploy
Mar 7, 2018 7:31 PM
churnMaggie_testGA6
mlin
1

Redbooks
ibm.com/redbooks

Turning Data into Insight with IBM Machine Learning for z/OS

Some highlights from the content

- Value proposition of Machine Learning for z/OS
- Open Data Analytics for z/OS
- Installation and customization
- Upgrade and migration
- Administration and operation
- Model development and deployment
- Use cases for using Machine Learning for z/OS
  - Anti-fraud
  - ITOA
  - Banking churn
  - Investment advisory
  - Loan approval

IBM

Redbooks

Thank You