Session: Build, Track, Deploy and Run a Model with Watson Studio

Goals:

- Understand IBM Watson Studio and its key features
- Learn how to use Watson Studio to
  - Train a model to predict which applicants qualify for mortgages
  - Set up tracking for the model to document the model history and generate an explanation for its performance
  - Deploy a model to a deployment space
  - Make a prediction request to see if an applicant is qualified for a mortgage

Required services:

- Watson Studio
- Watson Knowledge Catalog
- Machine Learning
- Cloud Object Storage

Watson Studio Overview

Watson Studio empowers data scientists, developers and analysts to **build, run and manage AI models, and optimize decisions** anywhere on IBM Cloud Pak for Data.

- Unite teams, automate AI lifecycles and speed time to value on an open multi-cloud architecture.
- Bring together open source frameworks like PyTorch, TensorFlow and scikit-learn with IBM and its ecosystem tools for code-based and visual data science. Work with Jupyter notebooks, JupyterLab and CLIs — or in languages such as Python, R and Scala.
- Flexible options to run
  - On IBM Cloud Pak for Data: on your own hybrid cloud using a unified data and AI platform
  - On IBM Cloud Pak for Data as a Service: fully managed cloud environment on IBM Cloud, AWS, Azure, or Google cloud
  - On IBM Cloud Pak for Data system: a hybrid cloud on-premises platform-in-a-box enterprise data and AI solution

Watson Studio provides the **environment and tools** for teams to collaboratively work on data to solve business problems. Each member can choose the tools needed to analyze and visualize data, to cleanse and shape data, to ingest streaming data, or to create and train machine learning models.

Features of Watson Studio

- **AutoAI for faster experimentation**

  Automatically build model pipelines. Prepare data and select model types. Generate and rank model pipelines.
- **Advanced data refinery**

  Cleanse and shape data with a graphical flow editor. Apply interactive templates to code operations, functions and logical operators.
- **Open source notebook support**

  Create a notebook file, use a sample notebook or bring your own notebook. Code and run a notebook.
- **Integrated visual tooling**

  Prepare data quickly and develop models visually with IBM SPSS Modeler in Watson Studio.
- **Model training and development**

  Build experiments quickly and enhance training by optimizing pipelines and identifying the right combination of data.
- **Extensive open source frameworks**

  Bring your model of choice to production. Track and retrain models using production feedback.

Features of Watson Studio

- **Embedded decision optimization**

  Combine predictive and prescriptive models. Use predictions to optimize decisions. Create and edit models in Python, in OPL (Optimization Programming Language) or with natural language.
- **Model management and monitoring**

  Monitor quality, fairness and drift metrics. Select and configure deployment for model insights. Customize model monitors and metrics.
- **Model risk management**

  Compare and evaluate models. Evaluate and select models with new data. Examine the key model metrics side-by-side.

Watson Studio - Projects

The architecture of Watson Studio is centered around the project. A project is a workspace where team members organize resources and work with data.

A project includes these resource types.

- **Collaborators** are the people on the team who work with the data. Data scientist tasks include analyzing data and building models. Data engineer tasks include preparing and integrating data.
- **Data assets** point to the data that is either in uploaded files or accessed through connections to data sources.
- **Operational assets** are the objects you create, such as scripts and models, to run code on data.
- **Tools** are the software you use to derive insights from data.

Projects fully integrate with the **catalogs** and **deployment spaces**:

- Catalogs are provided by the Watson Knowledge Catalog service
- Deployment spaces are provided by the Machine Learning service.

Watson Knowledge Catalog

Watson Knowledge Catalog, a core service of Cloud Pak for Data as a Service, includes **a secure enterprise catalog management platform** that provides high-quality data assets that are easy to find. The platform is supported by a data governance framework for users to ensure that data access is compliant with business rules and standards by preventing access to sensitive information by unauthorized users.

Watson Knowledge Catalog enables users to

- develop a knowledge core by curating data assets and enriching them with governance artifacts that describe their properties and meaning.
- use categories to control who can create and use governance artifacts for what purpose.
- create data protection rules that protect data across the platform.
- provide a self-service way to find and share assets across enterprise with catalogs

Data scientists find assets in catalogs and then copy the assets into projects where they analyze data and build models with Watson Studio and Machine Learning tools.

Machine Learning

Machine Learning provides a full range of tools and services to build, train, and deploy Machine Learning models. Users choose the tool with the level of automation or autonomy that matches individual needs.

These tools are available with the Machine Learning service:

- **AutoAI experiment builder** for automatically processing structured data to generate model-candidate pipelines. The best-performing pipelines can be saved as a machine learning model and deployed for scoring.
- **Notebooks** provide an interactive programming environment for working with data, testing models, and rapid prototyping
- **Deep Learning experiments** automates running hundreds of training runs while tracking and storing results.
- **Federated Learning** to train models using remote, disconnected data sources. The federated sources can contribute to building an accurate model without compromising security.
- Tools to **view and manage model deployments**.