Introduction to Watson Studio

Power of data. Simplicity of design. Speed of innovation.

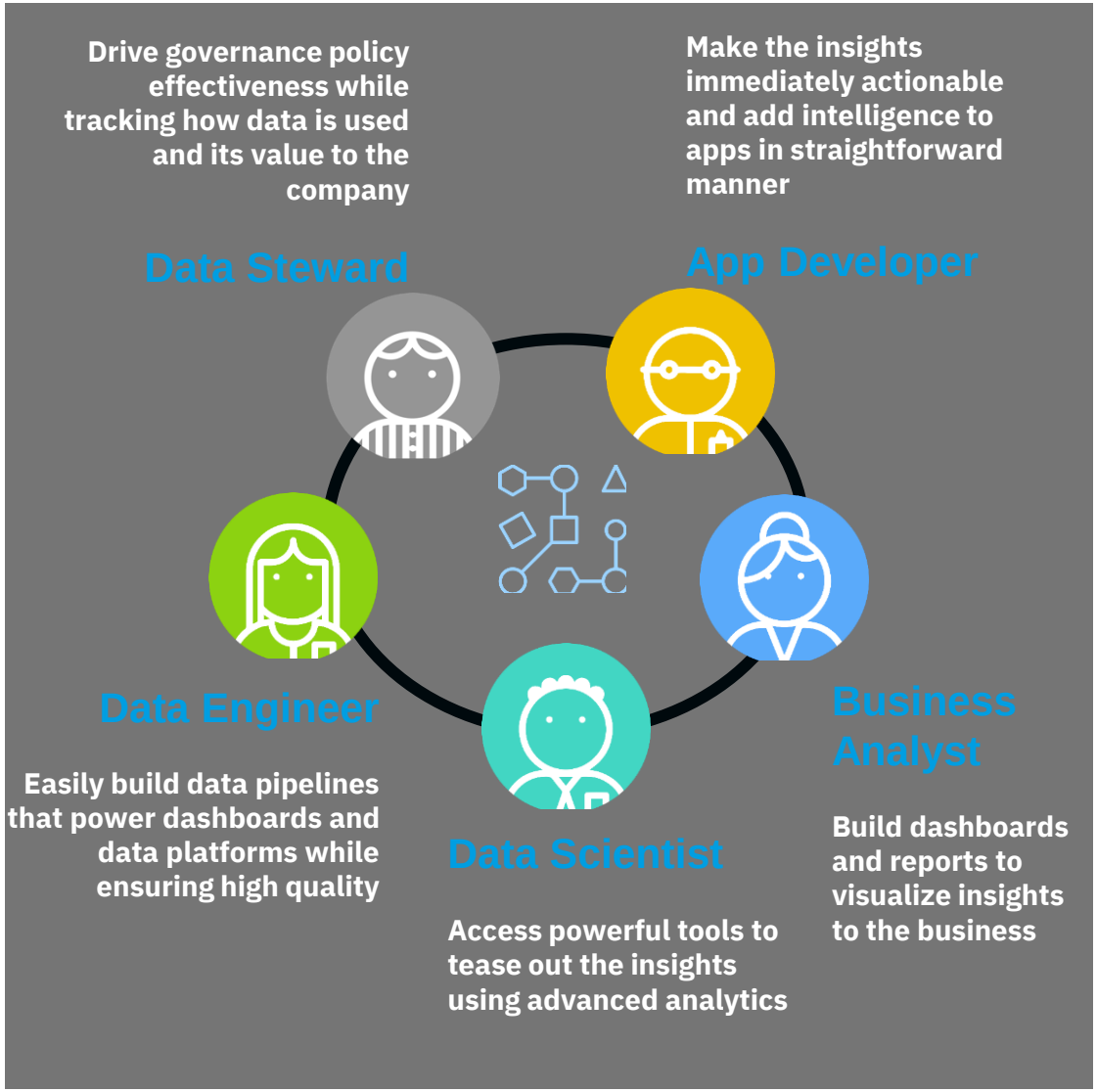
Bernie Beekman
Joel Patterson
Michael Cronk

IBM takes an Enterprise Approach to Data Science

- Freedom of Choice
  - Choose programming languages, open source libraries, IBM value-add capabilities
  - Code/Click
  - Machine Learning/Deep Learning/Decision Optimization.
- Operationalize Machine Learning
  - Manage complete ML lifecycle – Build, Deploy, Manage, Scale, Monitor, Re-train
- Hybrid ML
  - Build where you want, deploy where you want
- Governance
  - Ensure that right people get access to the right data
- Automation, Automation, Automation

IBM Watson Studio Platform

An integrated platform of tools, services, data, and metadata that help companies or agencies accelerate their shift to be data-driven organizations.



Watson Studio Deployment Options

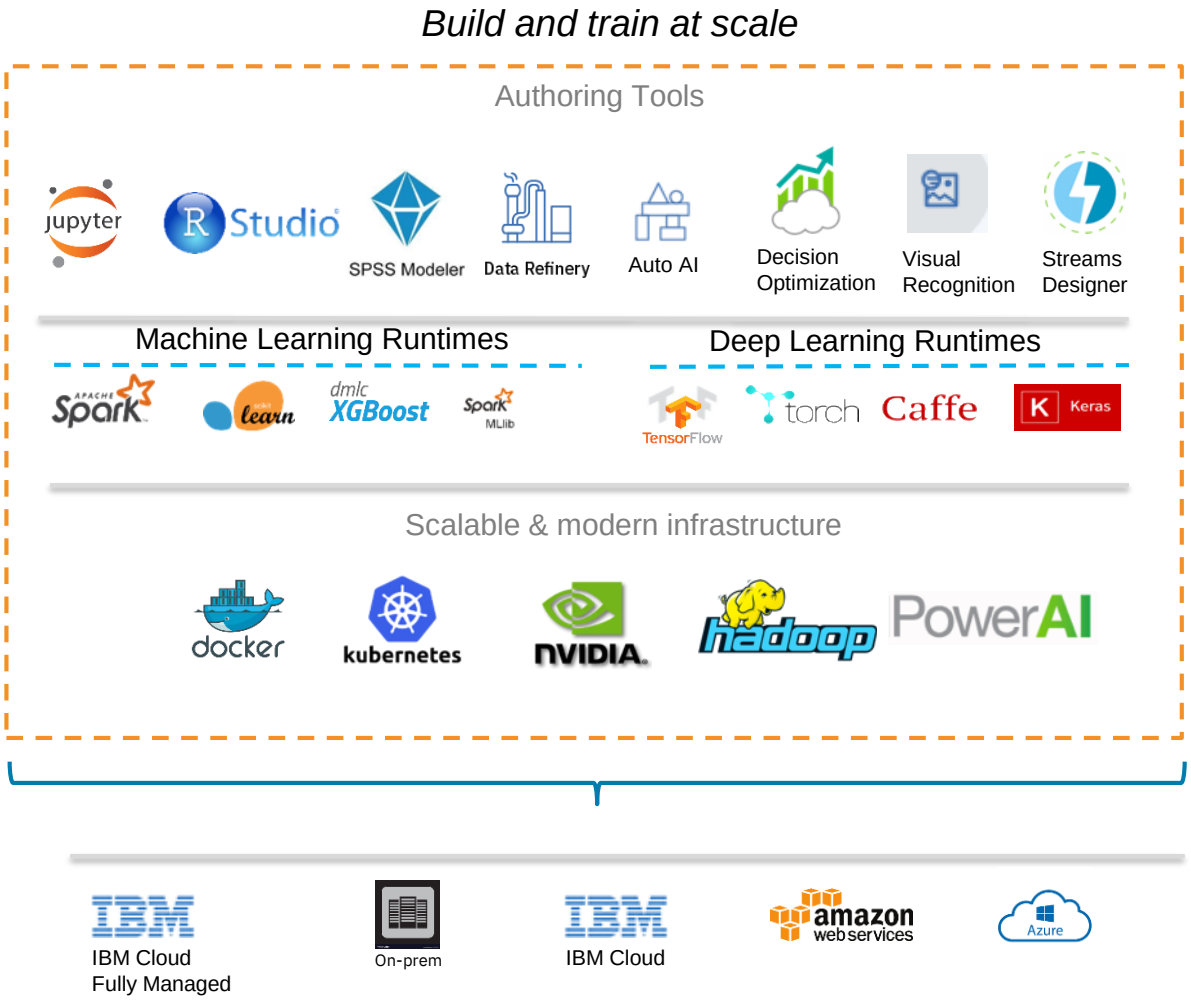
- Watson Studio on IBM Cloud
  - Managed offering provided by IBM
- Watson Studio Local
  - On-premise – Private Cloud
  - IBM Cloud, AWS, Azure
- Watson Studio Desktop
- IBM Cloud Private for Data
  - Watson Studio Local

Watson Studio Tools

- Using best of breed - Open source & IBM tools
- Code (R, Python or Scala) and no-code/visual modeling tools

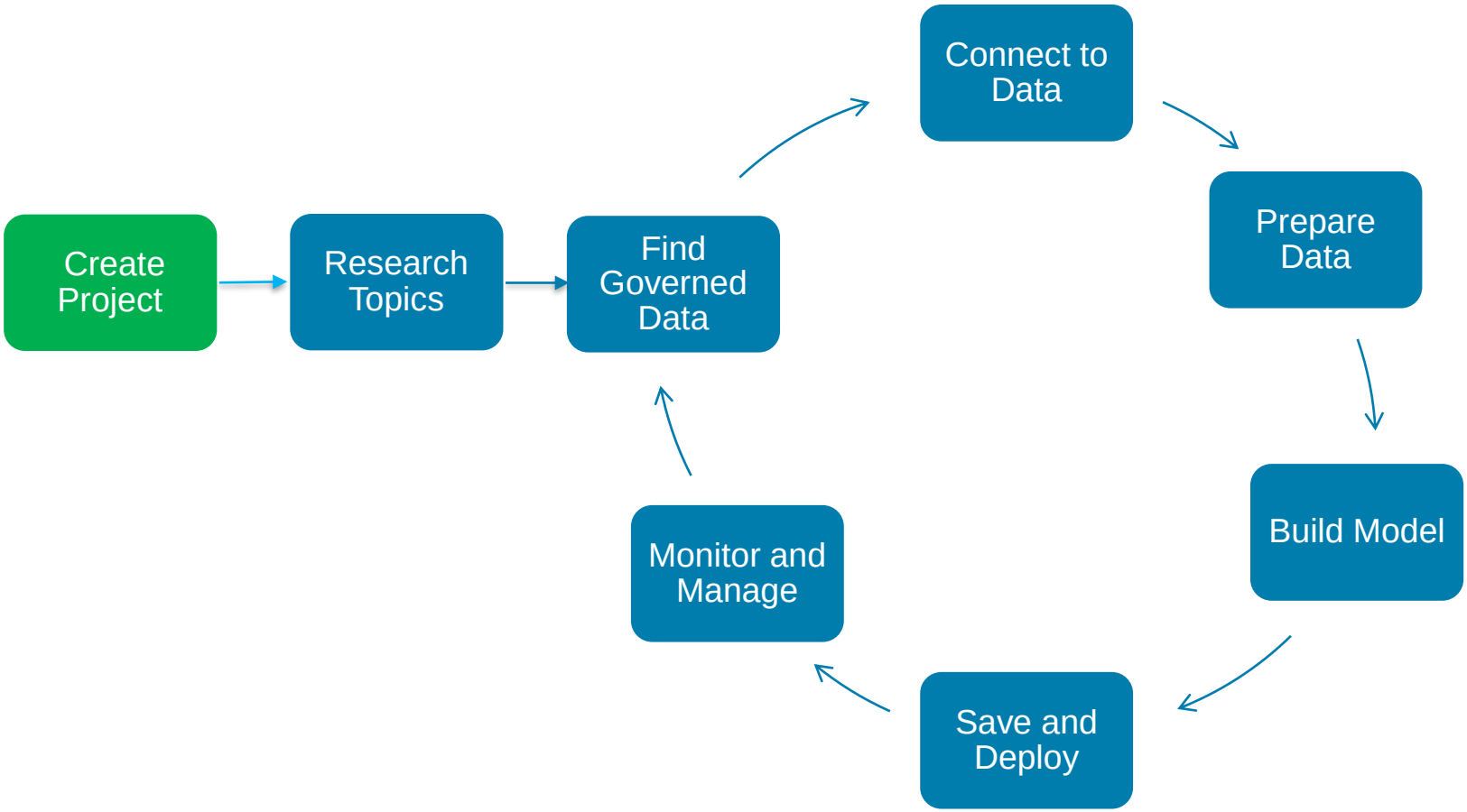
- Container-based resource management
- Elastic cpu/gpu power
- Run on x86, Power, zLinux
- Integrate with Hadoop/Spark Infrastructure

- Train and deploy where your data lives



Watson Studio supports the Data Science Lifecycle

Build, train, deploy, and monitor at scale ML/DL workflows to infuse AI into the enterprise to drive innovation.



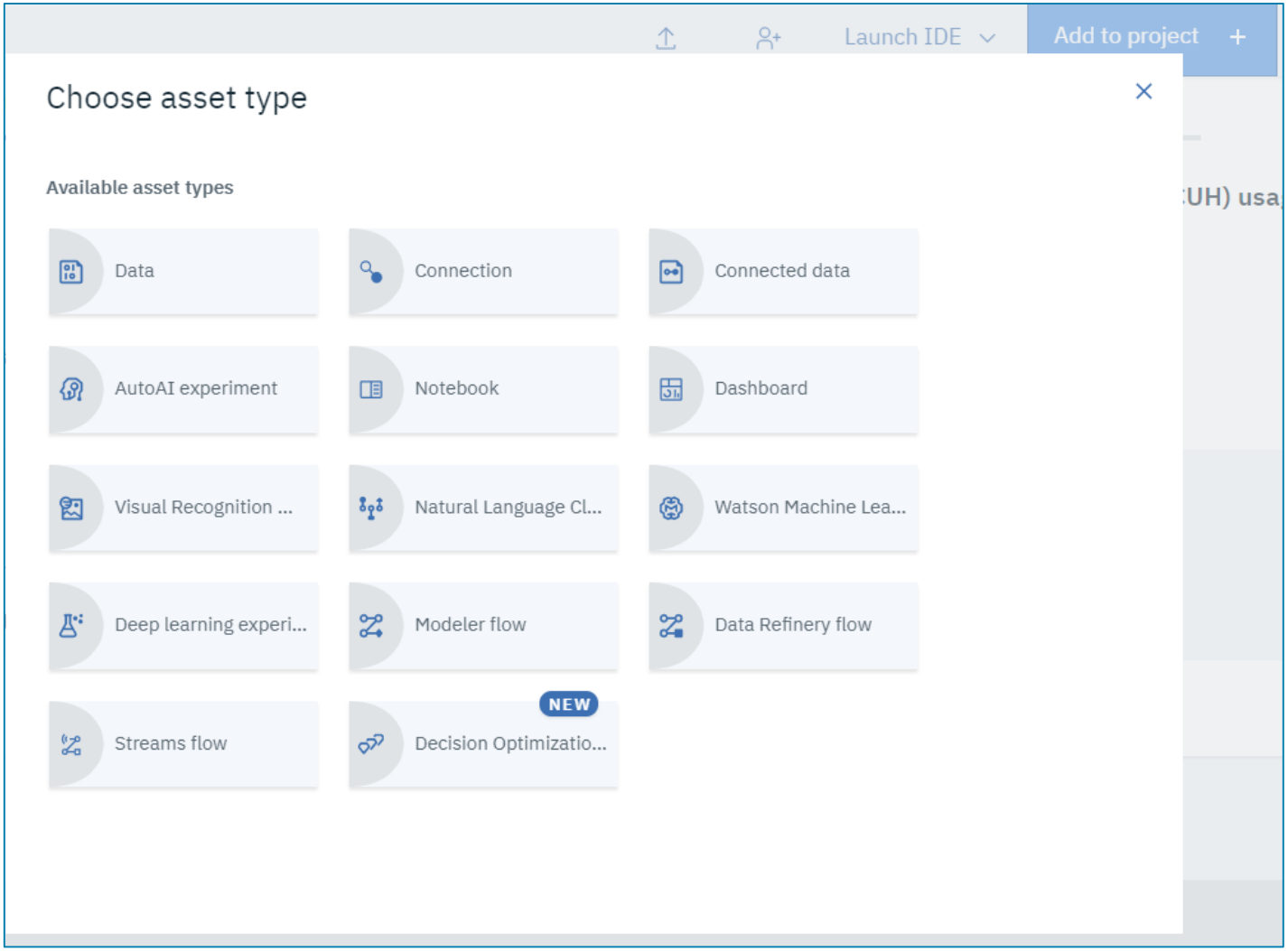
Watson Studio Project Features

Making Data Science a Team Sport

Create Project

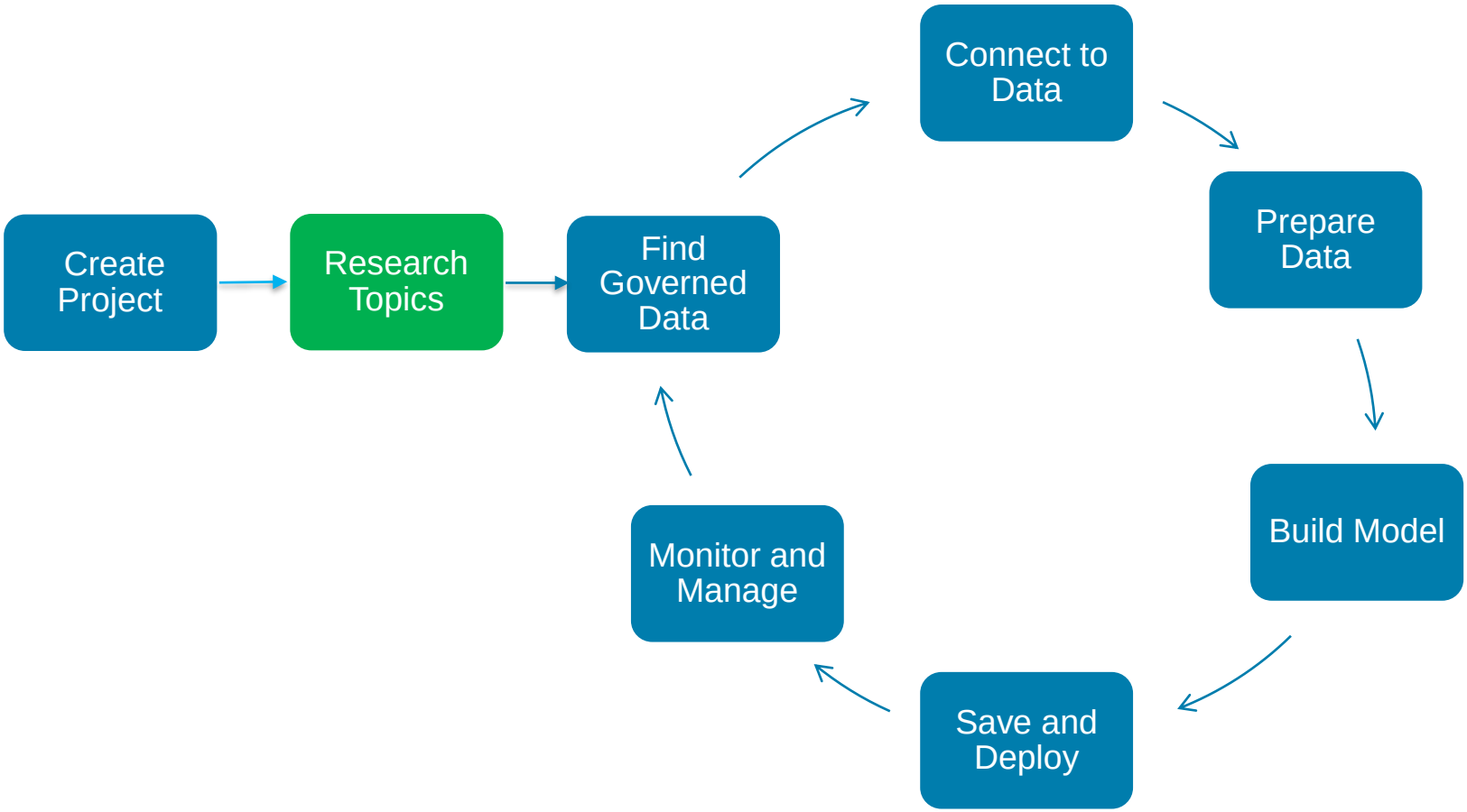
- Organizes resources to achieve a particular data analysis goal
- Support role-based collaboration (Admin, Editor, Viewer)
- Assets from all IDEs can be included in one Watson Studio project: notebooks, data sources, flows, models, etc.
- Export/Import Projects

Add to Project



Watson Studio supports the Data Science Lifecycle

Build, train, deploy, and monitor at scale ML/DL workflows to infuse AI into the enterprise to drive innovation.



Research Topics

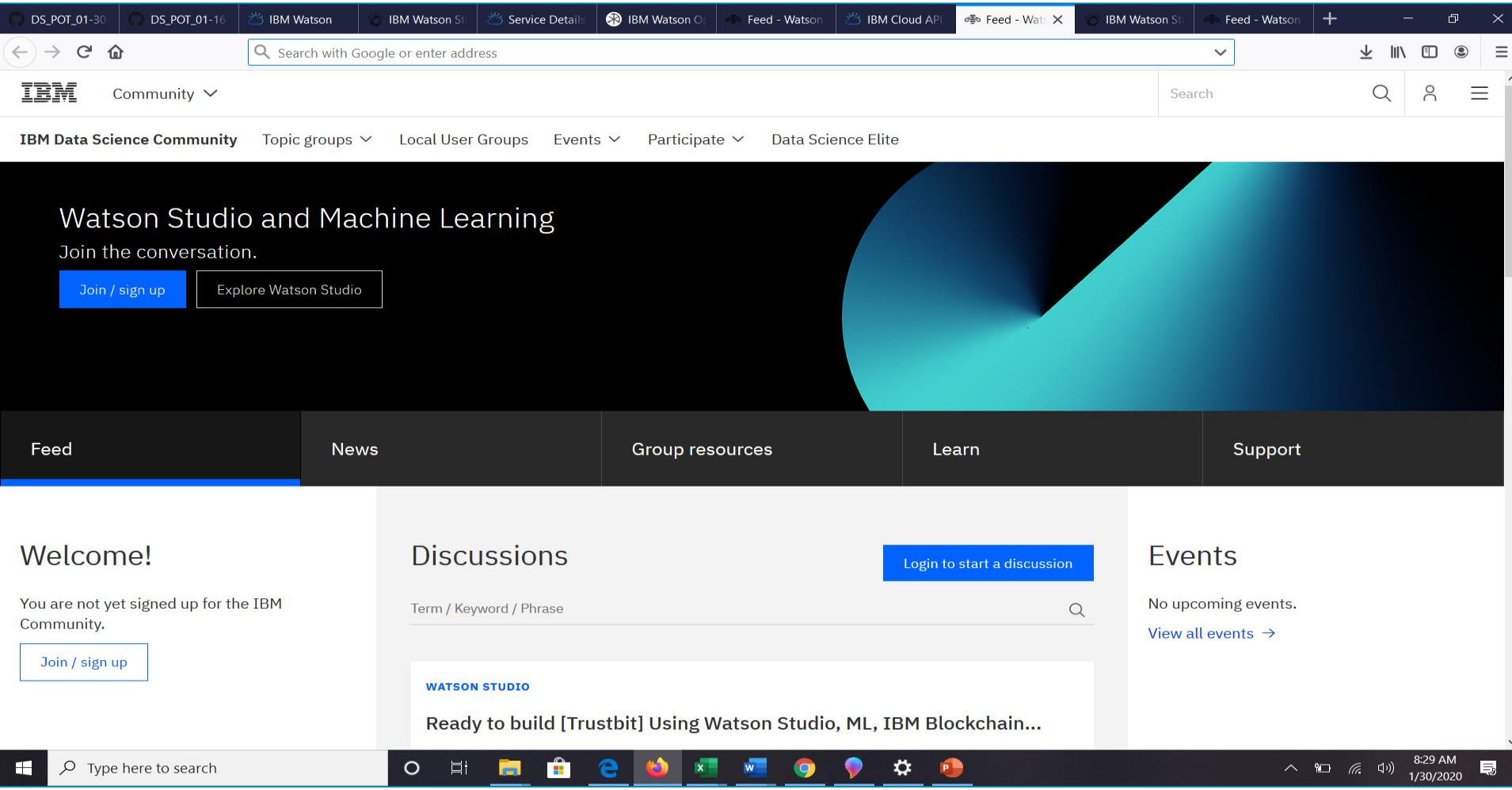
Watson Studio Gallery

Built-in learning to get started

- The Gallery includes notebooks, and data sets
- Copy notebooks or Data Sets into projects
- Continuously updated in IBM's managed service

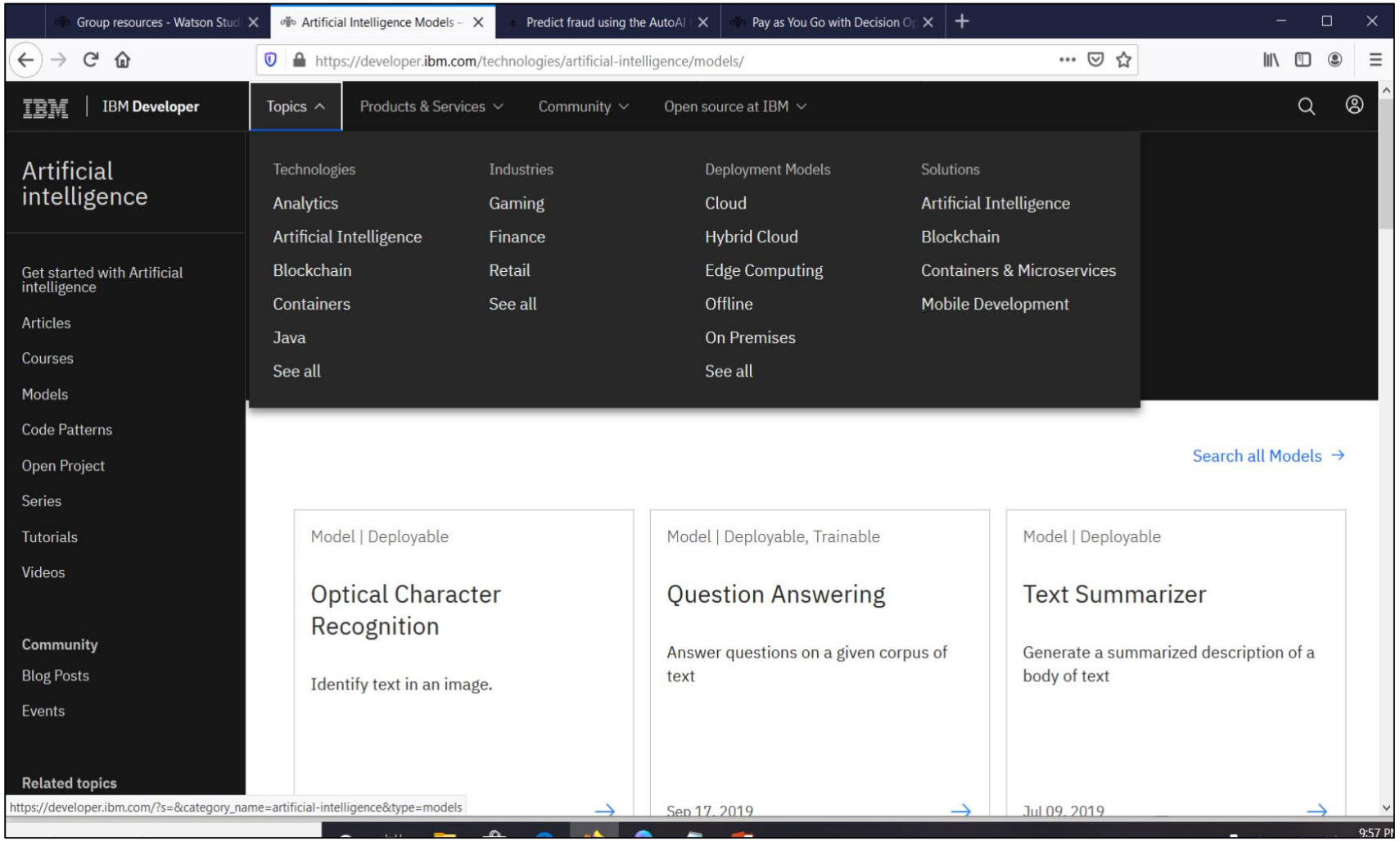
Watson Studio Community

Research Topics



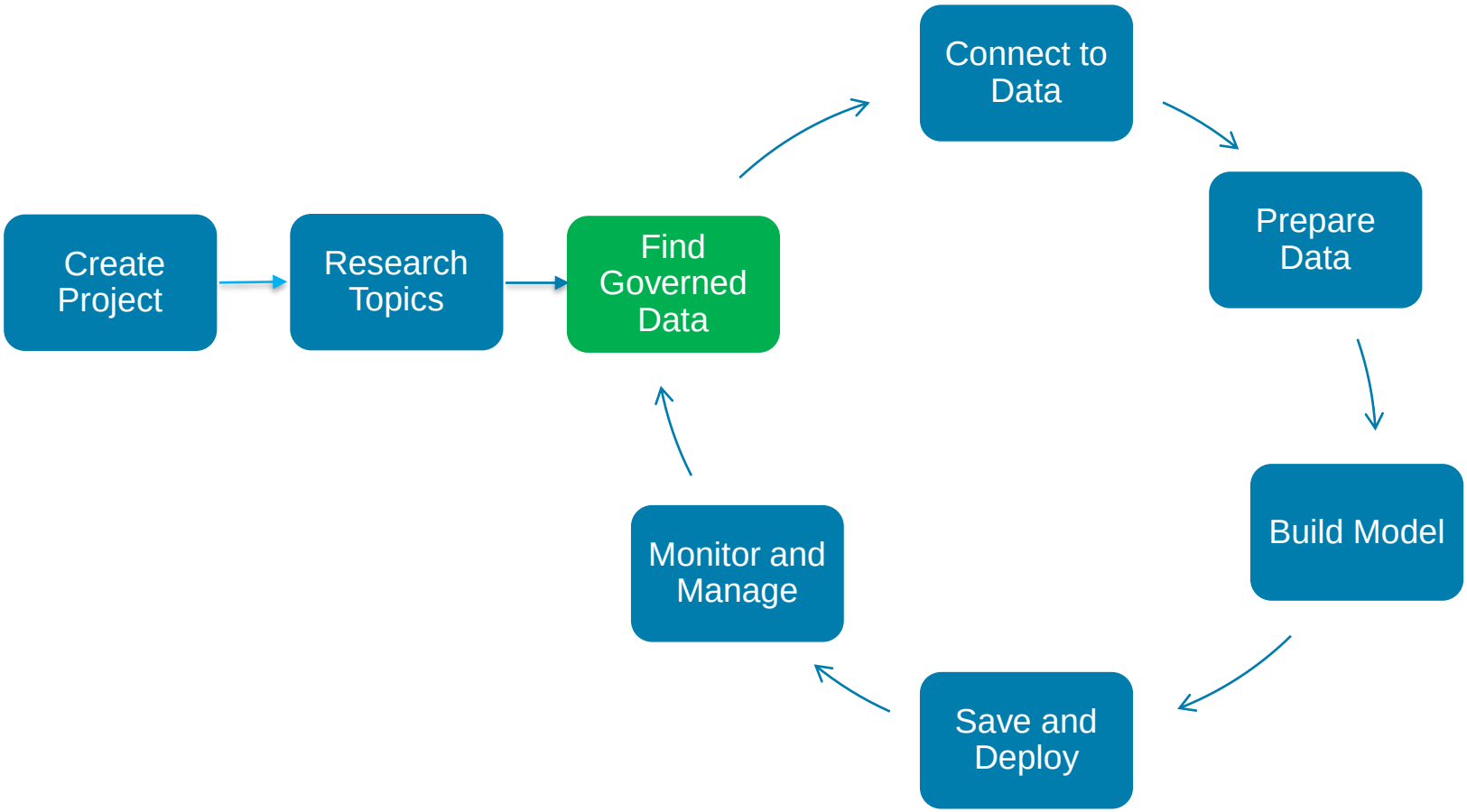
developer.ibm.com

Research Topics



Watson Studio supports the Data Science Lifecycle

Build, train, deploy, and monitor at scale ML/DL workflows to infuse AI into the enterprise to drive innovation.



Watson Knowledge Catalog Features

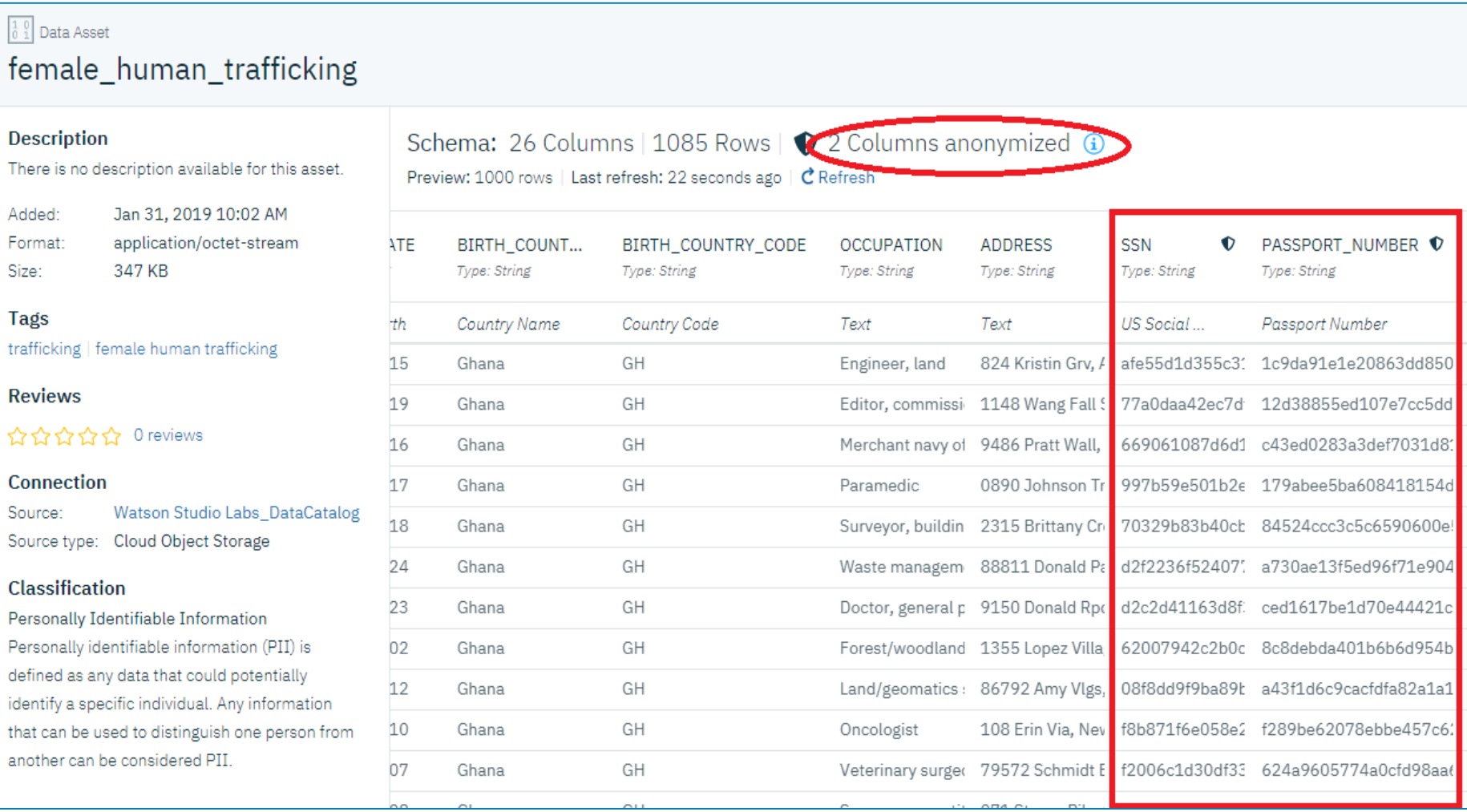
Find Governed Data

Unlock tribal knowledge and unleash knowledge workers

- **Find** data (structured, unstructured) and AI assets (e.g., ML/DL models, notebooks, Watson Data Kits) in the **Knowledge Catalog** with intelligent search and giving the right access to the right users.
- Discover assets, profiling, classification
- Policy, rule authoring
- Policy, rule enforcement
- Asset Usage Statistics

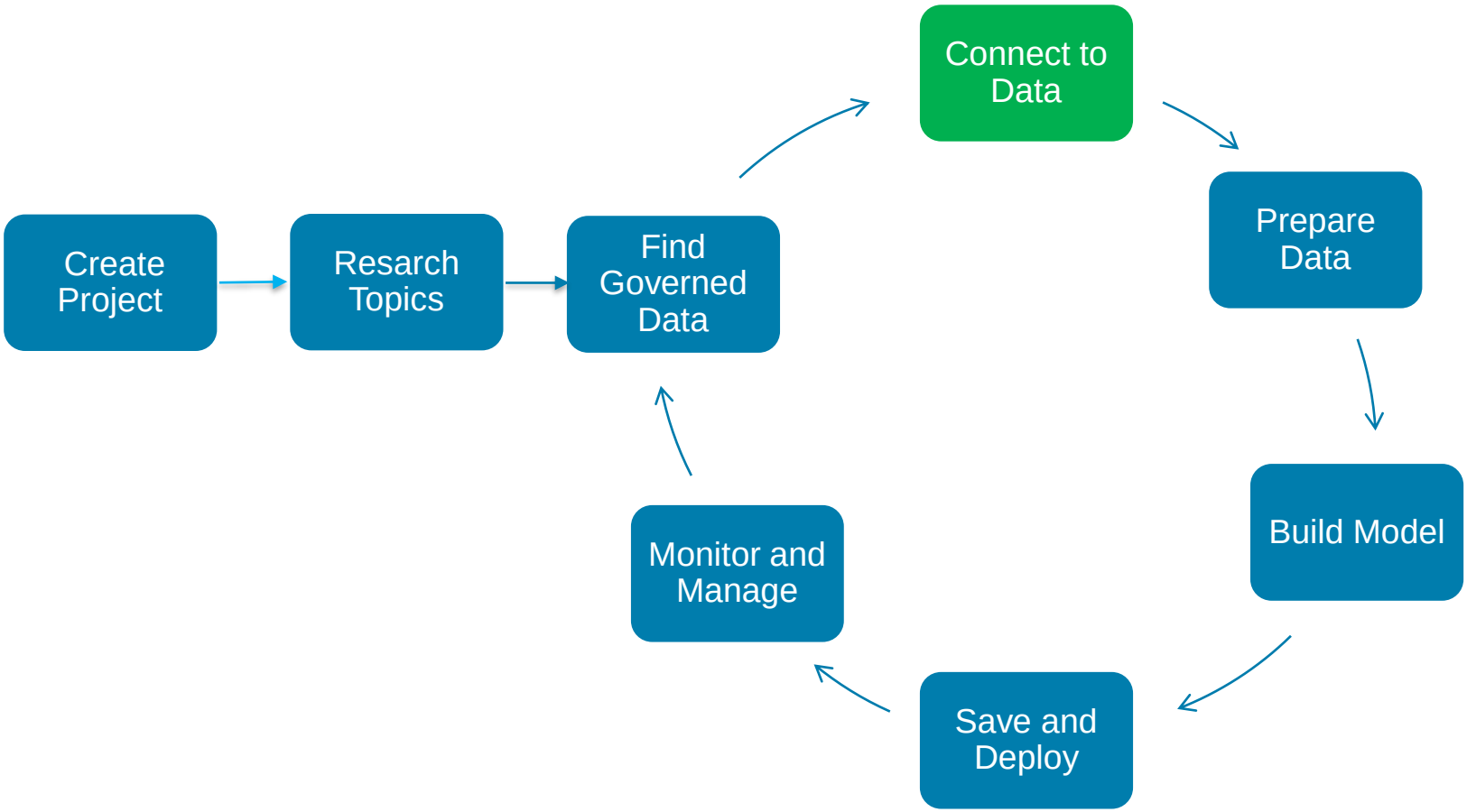
Watson Knowledge Catalog Features

Find Governed Data



Watson Studio supports the Data Science Lifecycle

Build, train, deploy, and monitor at scale ML/DL workflows to infuse AI into the enterprise to drive innovation.



Watson Studio Connection Features

Connect to Data

- Upload files
- Connectors to Structured and Unstructured, On-prem and Cloud data sources.
- Wizard based connection definition and code generation

Connection Options

Connect to Data

New connection

IBM services

- BigInsights HDFS
- Cloud Object Storage
- Cloud Object Storage (infrastructure)
- Cloudant
- Cognos Analytics
- Compose for MySQL
- Compose for PostgreSQL
- Db2
- Db2 Big SQL
- Db2 for i
- Db2 for z/OS
- Db2 Hosted
- Db2 on Cloud
- Db2 Warehouse
- Informix
- Object Storage OpenStack Swift (infrastructure)
- PureData for Analytics
- Watson Analytics

Third-party services

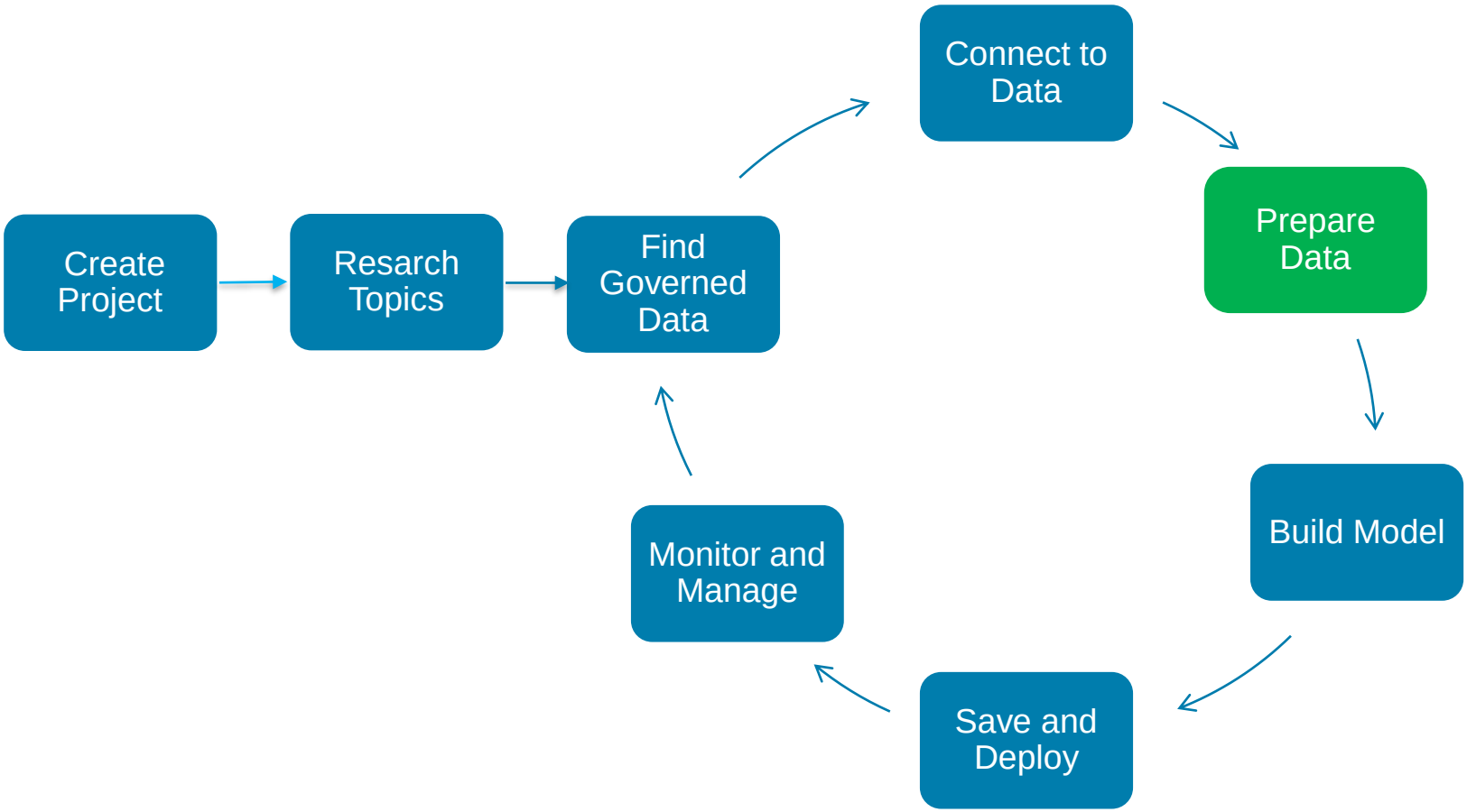
- Amazon Redshift
- Amazon S3
- Apache Hive
- Cloudera Impala
- Dropbox
- FTP
- Google BigQuery
- Google Cloud Storage
- Hortonworks HDFS
- Looker
- Microsoft Azure Data Lake Store
- Microsoft Azure SQL Database
- Microsoft SQL Server
- MySQL
- Oracle
- Pivotal Greenplum
- PostgreSQL
- Salesforce.com
- Sybase
- Sybase IQ
- Tableau
- Teradata

Notebook Screenshot

Connect to Data

Watson Studio supports the Data Science Lifecycle

Build, train, deploy, and monitor at scale ML/DL workflows to infuse AI into the enterprise to drive innovation.



Watson Studio Data Refinery Features

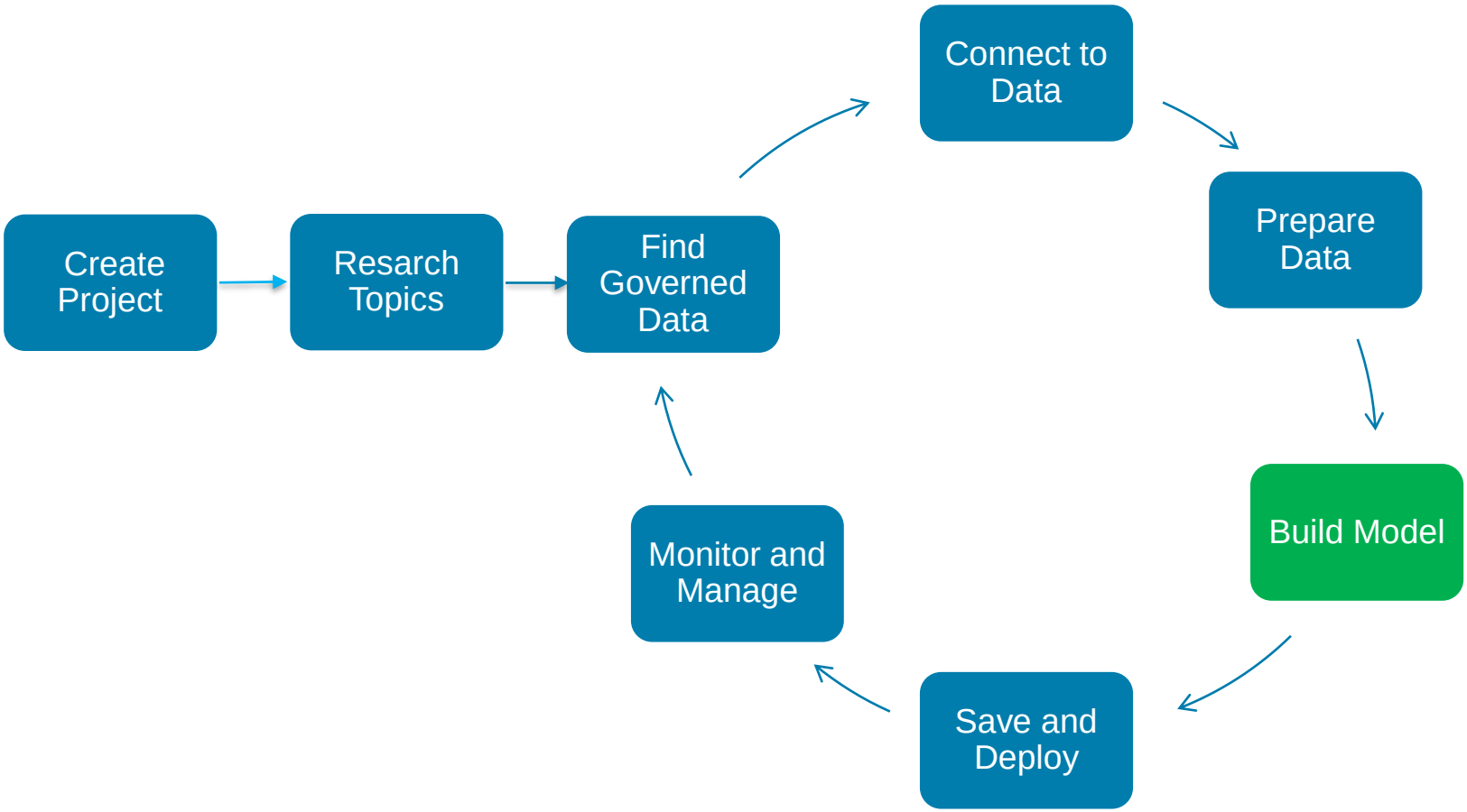
Making Data fit for use

Prepare Data

- Data Refinery tool to profile, visualize, and shape data.
- Create data preparation pipelines via point and click capability on subset of data
  - ✓ Cleanse the data: fixing or removing data that is incorrect, incomplete, improperly formatted, or duplicated
  - ✓ Shape the data: customize data by filtering, sorting, combining, or removing columns, and performing operations
- Run the pipeline on all the data
  - Manually (on demand)
  - Automated (scheduled)

Watson Studio supports the Data Science Lifecycle

Build, train, deploy, and monitor at scale ML/DL workflows to infuse AI into the enterprise to drive innovation.



Watson Studio Model Building Features

Build Model

The best of open source and IBM Watson tools to create start-of-the-art data products

Open Source Tools

- Jupyter Notebooks**
- RStudio and Shiny
- Libraries- scikit-learn**, XGBoost**, Spark, TensorFlow**, Caffe, Keras, PyTorch

IBM Tools

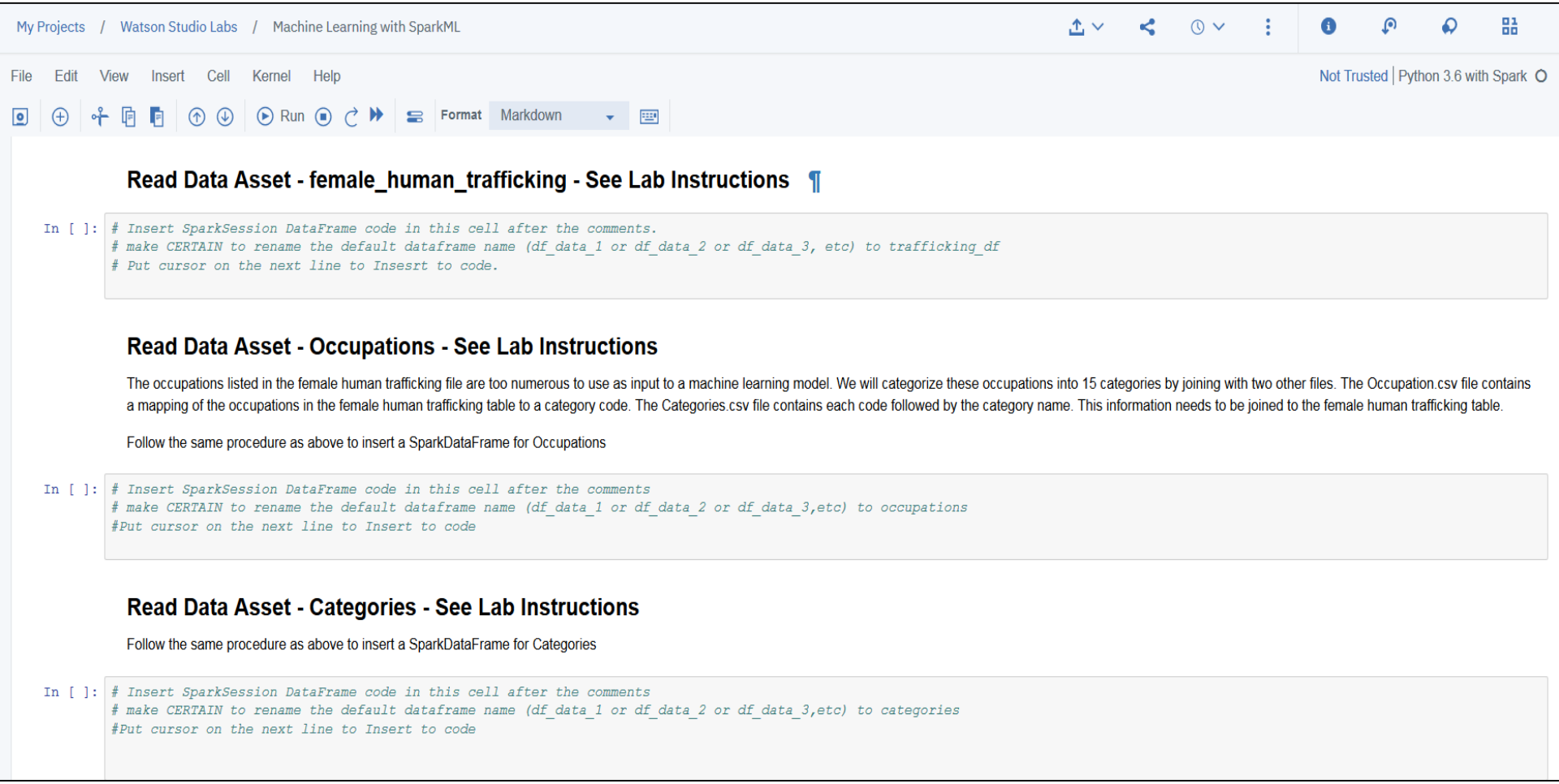
- AutoAI **
- SPSS Modeler**
- Neural Network Modeler**
- Experiment Builder**
- Natural Language Classifier Model
- Visual Recognition Model

Train at scale on **GPUs** and **distributed** compute

** in hands-on labs

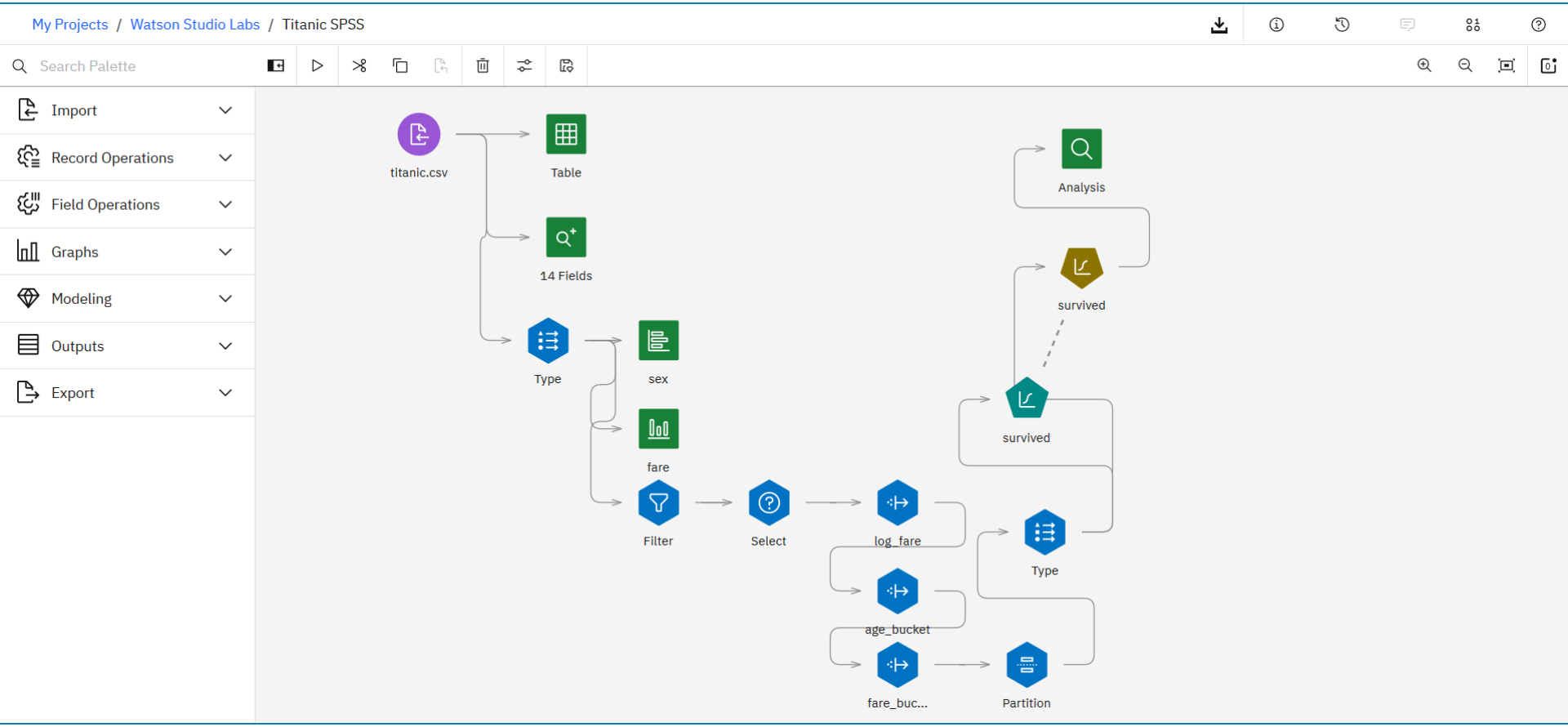
Jupyter Notebook

Build Model



SPSS Modeler

Build Model



Neural Network Modeler

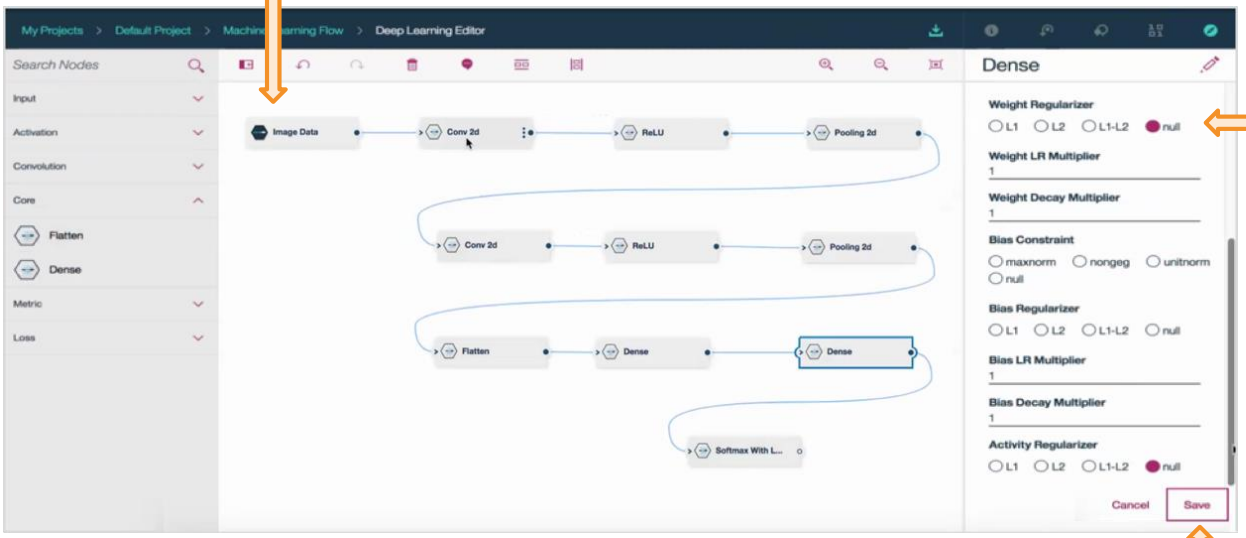
Build Model

An intuitive drag-and-drop, no-code interface for designing neural network structures using the most popular deep learning frameworks. Quickly capture your network design then single click export for experimental optimization.

Supported Frameworks

TensorFlow™ PYTORCH Caffe Keras

Real-time validation of network flow



Drag-and-drop network layers

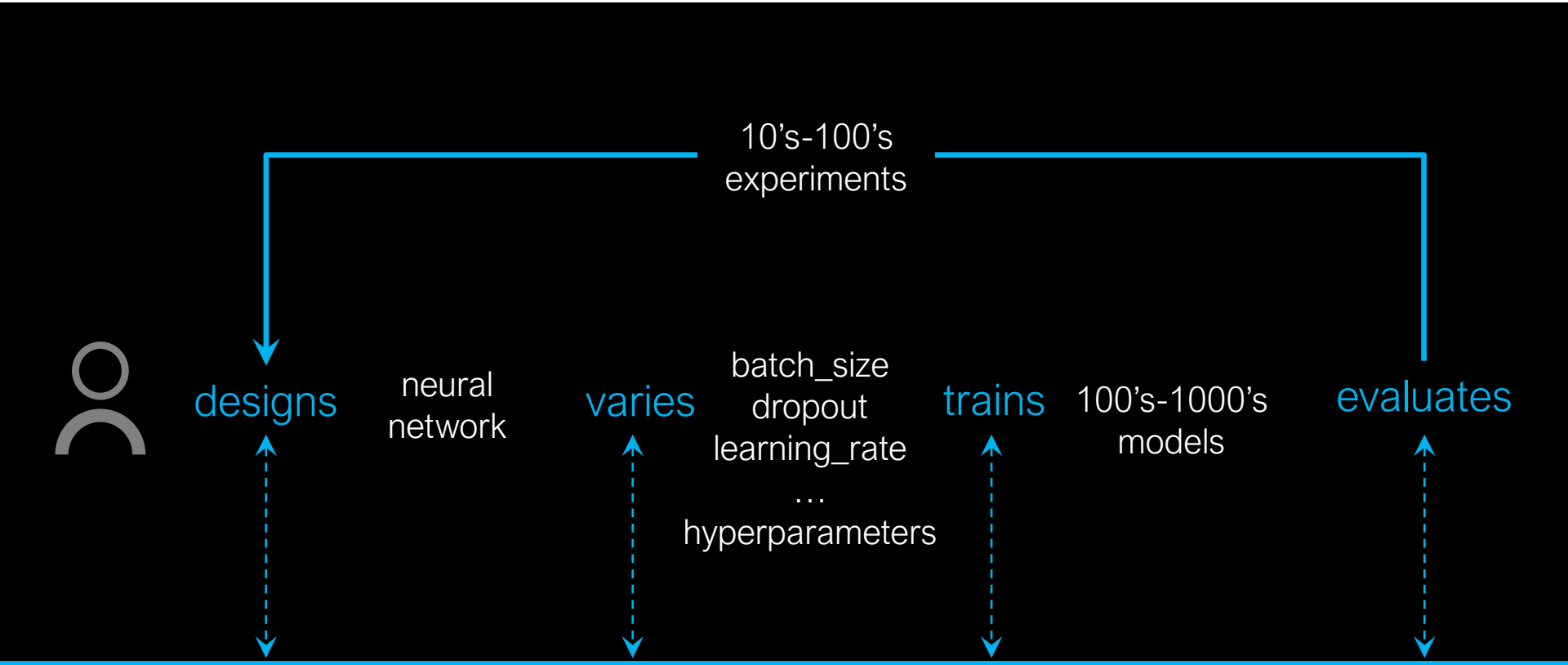
- Define layer configuration
- Choose optimizer params

- Generate CPU or GPU compatible code

- Save as popular framework code
- Export as a python notebook
- Execute as batch experiment

Experiment Builder

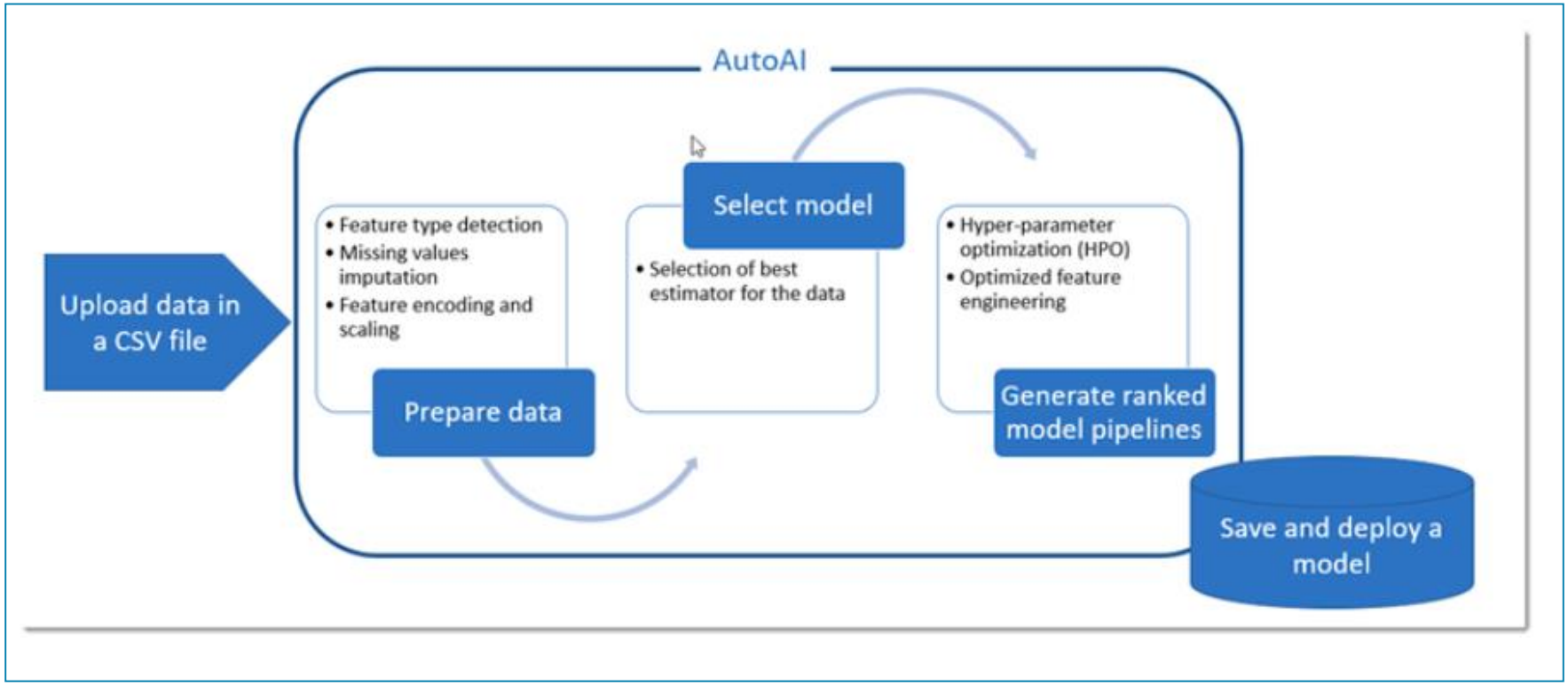
Build Model



Experiment Builder
supports the end-to-end workflow

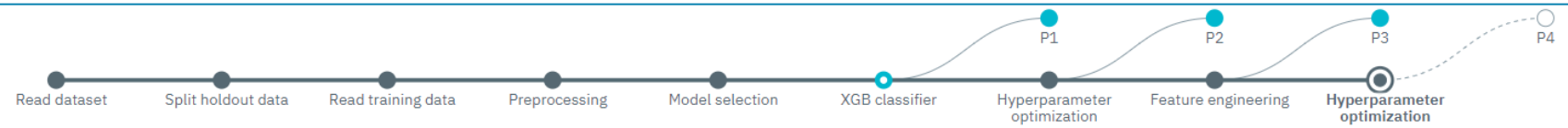
AutoAI

Build Model



AutoAI

Build Model



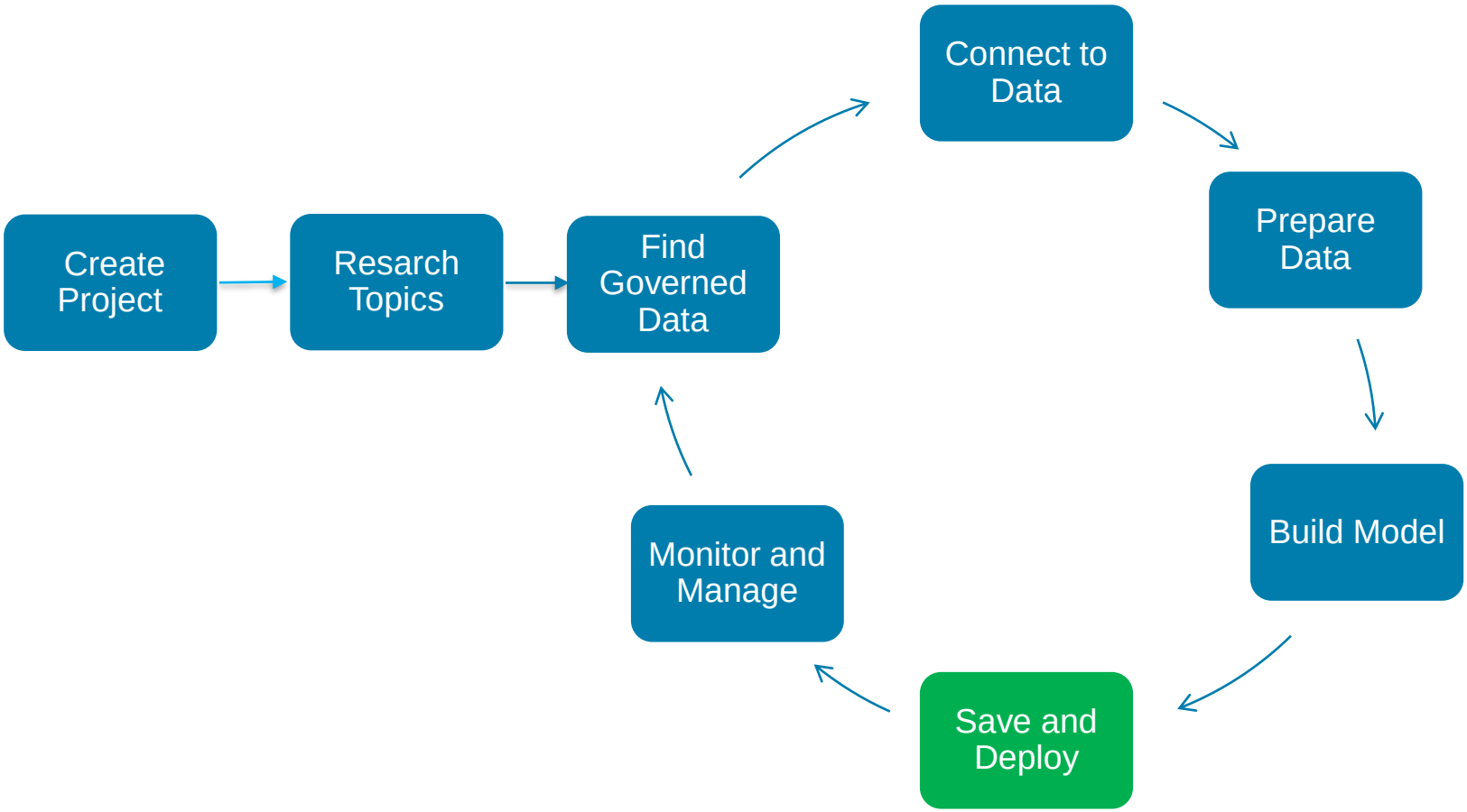
Pipeline leaderboard

Compare pipelines | Ranking based on: Accuracy

	RANK	ACCURACY	PIPELINE INFORMATION		
>	1	0.897	P3 - XGB classifier estimator Transformers (8): Preprocessing > Standard scaler > Univariate feature selection > Sine > Univariate feature selection > Tangent > ...	View details	Save as model
>	2	0.884	P1 - XGB classifier estimator Transformers (2): Preprocessing > XGB classifier estimator	View details	Save as model
>	3	0.884	P2 - XGB classifier estimator Transformers (2): Preprocessing > XGB classifier estimator	View details	Save as model

Watson Studio supports the Data Science Lifecycle

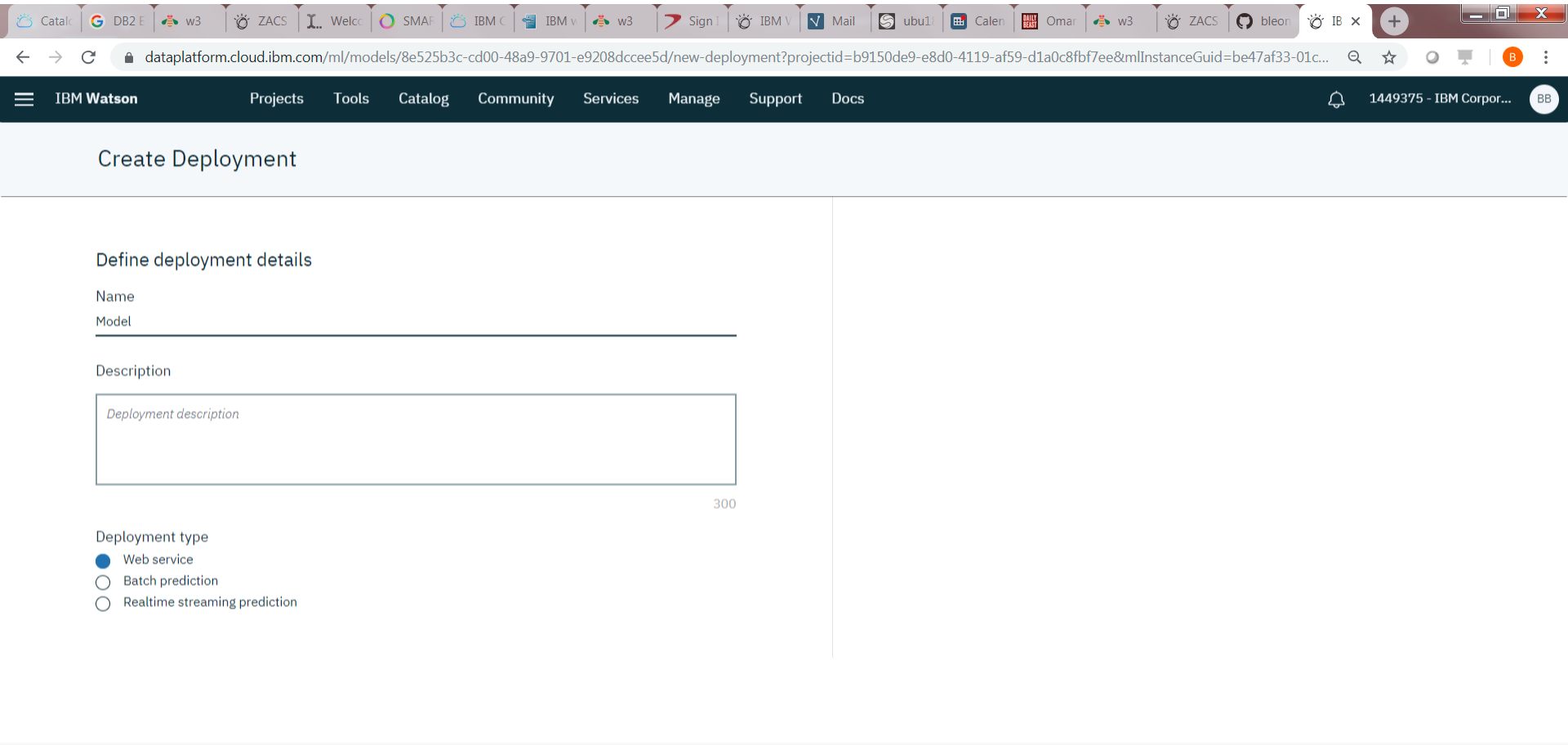
Build, train, deploy, and monitor at scale ML/DL workflows to infuse AI into the enterprise to drive innovation.



Watson Studio Save and Deploy Models

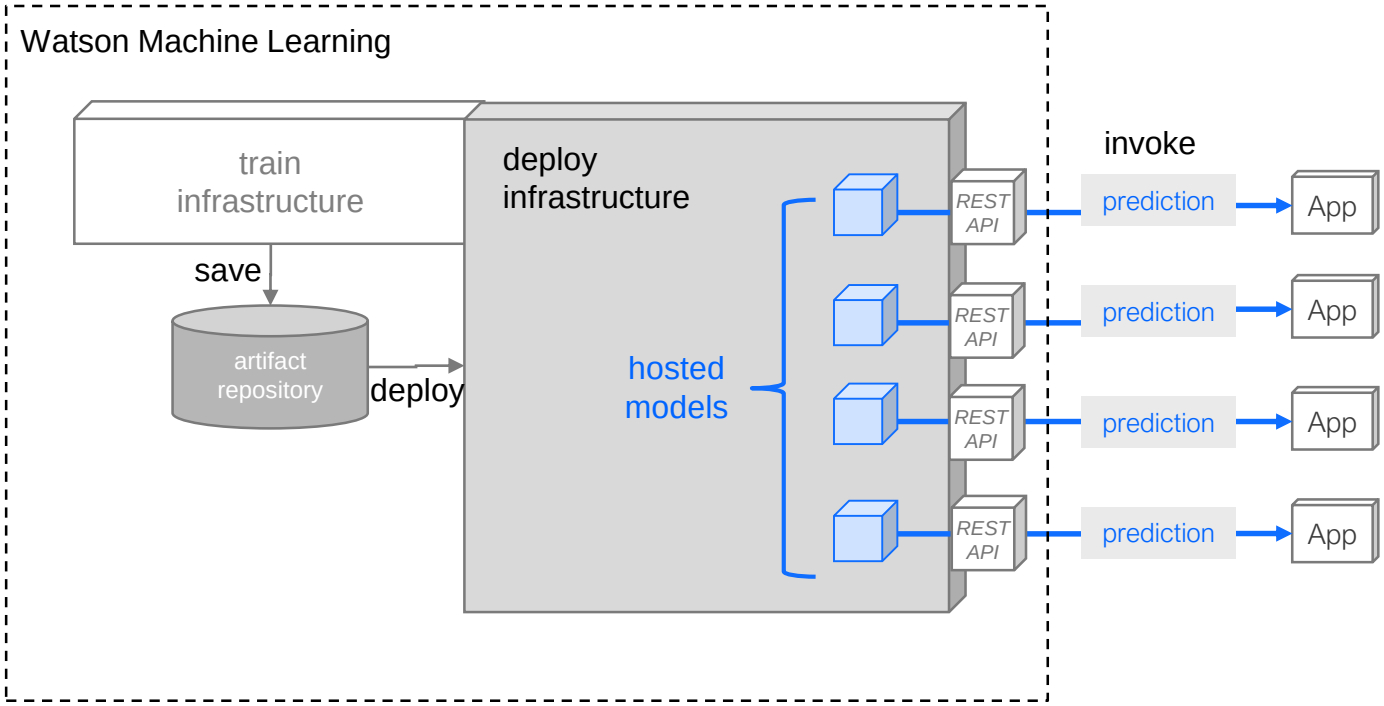
Save and Deploy Models with Watson Machine Learning

Save and Deploy



Watson Studio Save and Deploy Trained Models

Save and Deploy Models with Watson Machine Learning



Watson Studio Save and Deploy Features

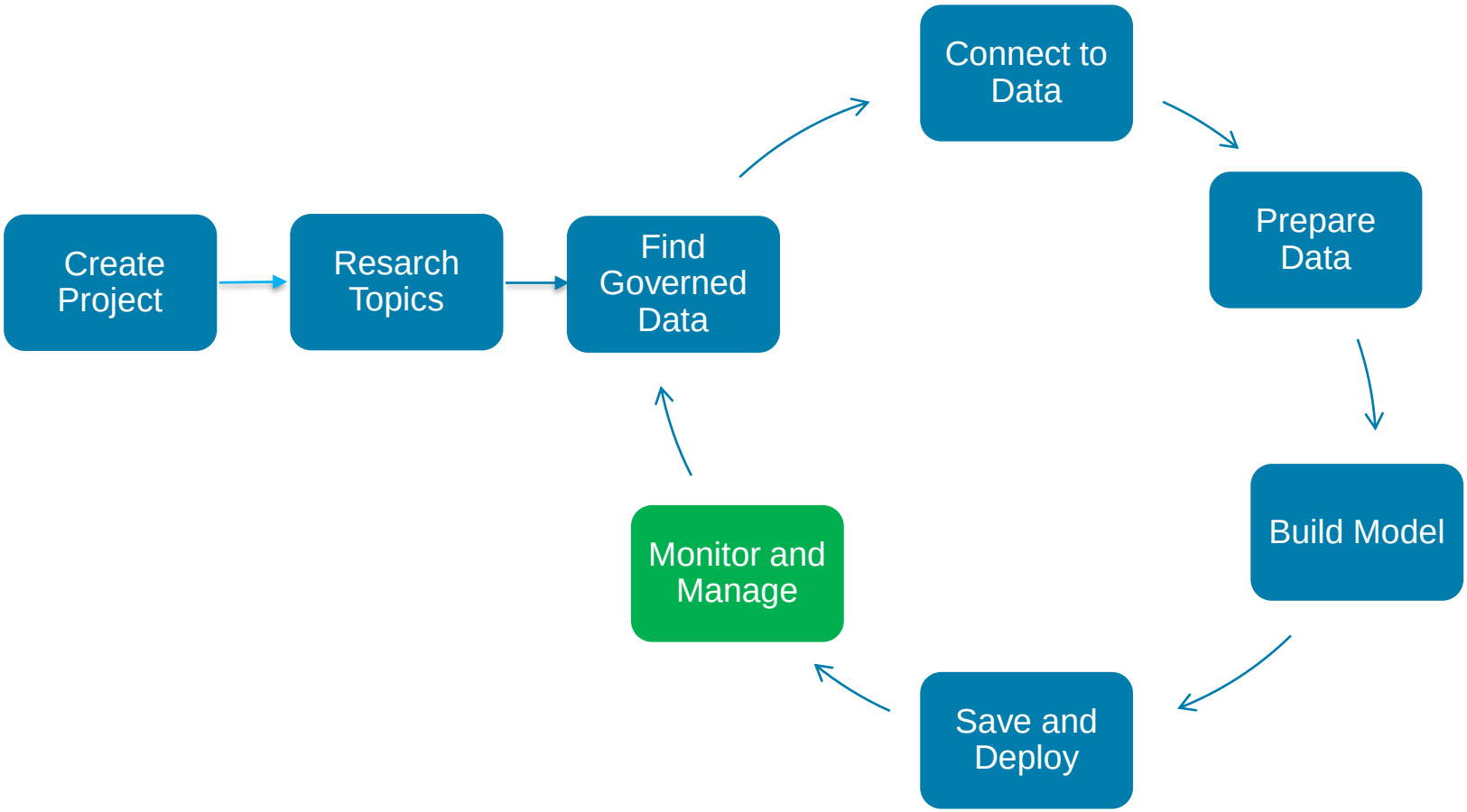
Save and Deploy

Save and Deploy Models with Watson Machine Learning

- Watson Machine Learning API to save/load models to/from repository
- Watson Machine Learning API to deploy saved models easily and have them scale automatically.
- Watson Machine Learning API to invoke deployed models

Watson Studio supports the Data Science Lifecycle

Build, train, deploy, and monitor at scale ML/DL workflows to infuse AI into the enterprise to drive innovation.



Our vision for Trusted AI

Pillars of trust, woven into the lifecycle of an AI application

Monitor and Manage

Is it accurate?

Is it fair?

Is it easy to understand ?

Did anyone tamper with it?

Watson OpenScale

Monitor and Manage

Trust and Transparency

- Intelligently delivers bias mitigation help
- Provides traceability & auditability of AI predictions made in production applications
- Tracks AI accuracy in applications
- Explains an outcome in business terms
- Drift Detection

Automation

- Automatically detects and mitigates bias in model output, without affecting currently deployed model or outcomes
- *NeuNetS (beta) – automatically generate Neural Networks

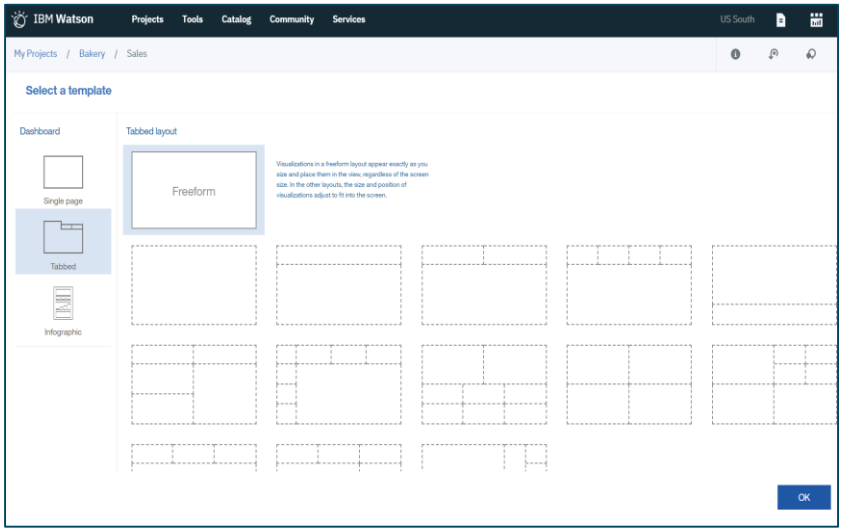
Open By Design

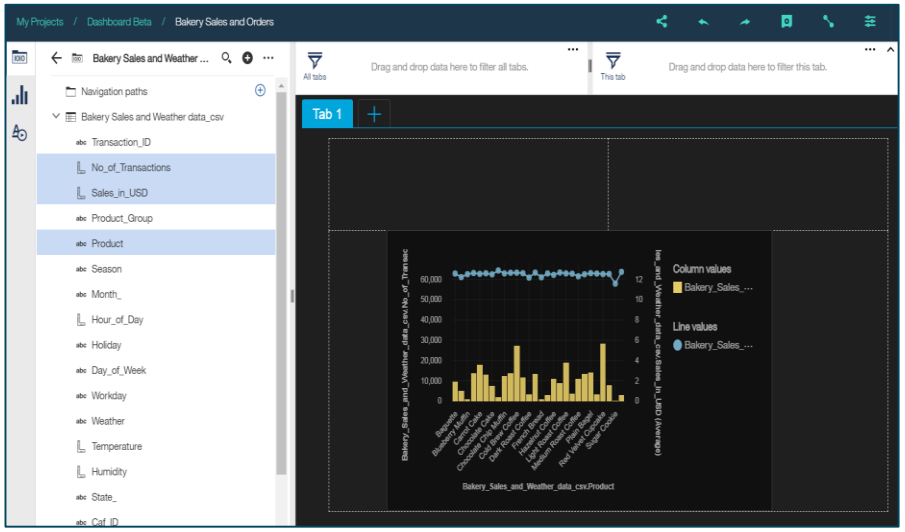
- Monitor models deployed on third party model server engines
- Deploy behind enterprise firewall or on IaaS provider

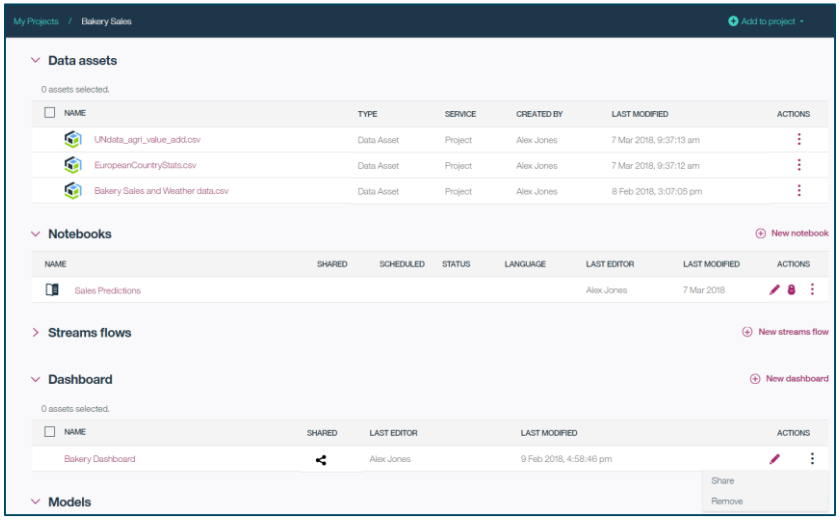
* https://arxiv.org/abs/1901.06261

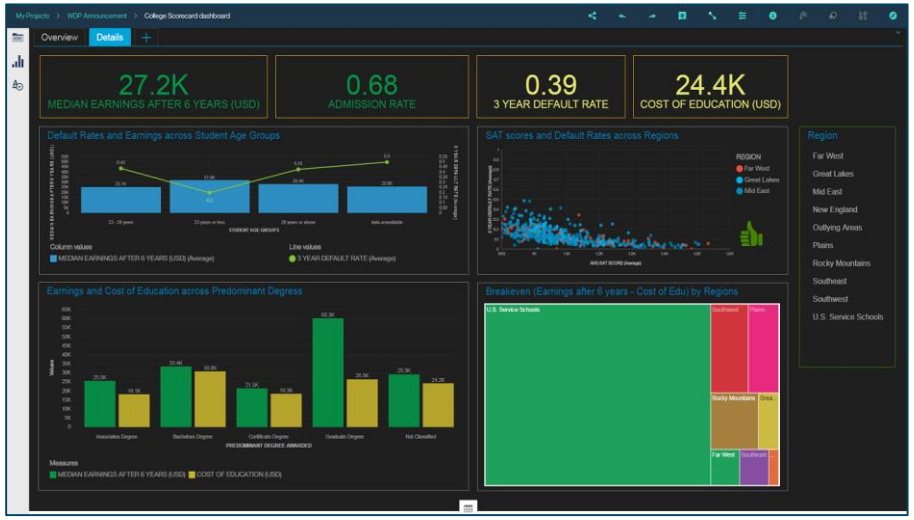
Watson Studio Dynamic Dashboards

Making insights available to all









Watson Studio Takeaways

Integrated Collaboration Environment

- Data Scientists, Subject Matter experts, Business Analysts & Developers all in one environment to accelerate innovation, collaboration and productivity
- Built-in learning to get started or go the distance with advanced tutorials

Choice of Tools for the full AI lifecycle

- Best in-breed open source and IBM tools that support the end-to-end AI lifecycle
- Choice of code or no-code tools to build and train your own ML/DL models or easily train and customize pre-trained Watson APIs

Support for all levels of expertise

- Use Watson smarts and recommendations for the best algorithms to use given your data, OR
- Use the rich capabilities and controls to fine tune your models

Multiple Deployment Options

- Watson Studio on IBM Cloud – Managed offering
- Watson Studio Local – Private Cloud, Public Cloud-(IBM, Azure, AWS)
- Watson Studio Desktop

Model lifecycle & management

- Deploy models into production then monitor them to evaluate performance.
- Capture new data for continuous learning and retrain models so they continually adapt to changing conditions.

Integrated with Knowledge Catalog

- Intelligent discovery of data and AI assets that enables reuse & improves productivity
- Seamlessly integrated for productive use with Machine Learning and Data science
- Powerful governance tools to control and protect access to data

40

Lab Overview

Lab Tips

- Watson Studio url: datascience.ibm.com
- Labs are in www.github.com/bleonardb3/ML_POT_03-26-2020 repository.
- Instructions for each Lab are in the README file in the respective Lab folder.
- Cloud development enables making frequent improvements in the user interface. We reviewed the lab instructions and made screen updates so they should be pretty faithful to the user interface. Small differences may occur but shouldn't get in the way of successfully completing the labs.
- Do not use Internet Explorer or Edge as the browser. For Mac users do not use Safari.
- All of the Labs should be done in the project that you created in Lab-1
- For Lab-1 make sure that you uncheck the "restrict who can be a collaborator" checkbox when creating the project.

Lab Tips

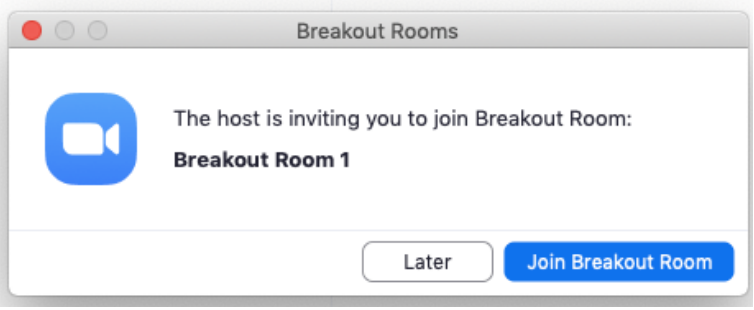
- Lab-7: GPU Cluster for non-production workloads is overloaded. It may take a while for the Neural Network job to start processing. Move on to Lab-8 if the model does not get out of the queueing stage in 2-3 minutes.

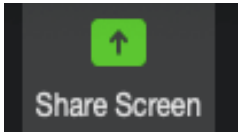
- Lab-7- use Firefox. Make sure you switch Region to us-geo.

- Lab-8: Confidences are not displaying correctly.

IBM Analytics

Breakout Rooms





Lab-1: Set up Environment

Create Project

Introduction:

This lab will set up the Watson Studio environment for subsequent labs and introduce you to the Project features of Watson Studio.

Objectives:

Upon completing this lab, you will know how to:

- Create a project
- Create an object storage instance and associate it with the project
- Associate an existing Watson Machine Learning instance with the project
- Add collaborators to the project

Labs: 2,3,4 Titanic Data

- **Variable Descriptions:**

survival	Survival
	(0 = No; 1 = Yes)
pclass	Passenger Class
	(1 = 1st; 2 = 2nd; 3 = 3rd)
name	Name
sex	Sex
age	Age
sibsp	Number of Siblings/Spouses Aboard
parch	Number of Parents/Children Aboard
ticket	Ticket Number
fare	Passenger Fare
cabin	Cabin
embarked	Port of Embarkation
	(C = Cherbourg; Q = Queenstown; S = Southampton)

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.25		S
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.925		S
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1	C123	S
5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.05		S
6	0	3	Moran, Mr. James	male		0	0	330877	8.4583		Q
7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	51.8625	E46	S
8	0	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909	21.075		S
9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0	2	347742	11.1333		S
10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	1	0	237736	30.0708		C

Lab-2: Introduction to the Data Refinery

Prepare Data

Introduction:

In this lab, you will use the Watson Studio Data Refinery to profile data, visualize data, and prepare data for modeling.

Objectives:

Upon completing the lab, you will know how to:

- Profile the data
- Visualize the data to gain a better understanding
- Prepare the data for modeling
- Run the sequence of data preparation operations on the entire data set.

Lab-3: SPSS Modeler

Build Model

Save and Deploy

Introduction:

In this lab, you will use the Watson Studio SPSS Modeler capability to explore, prepare, and model trafficking data. The SPSS Modeler is a drag and drop capability to build machine learning pipelines.

Objectives:

Upon completing this lab, you will have:

- Become familiar with the Watson Studio SPSS Modeler capability
- Profiled the data set
- Explored the data set with visualizations
- Transformed the data
- Trained/Evaluated a machine learning mode.

Lab-4: AutoAI

Build Model

Save and Deploy

Introduction:

In this lab, you will use IBM's Watson Machine Learning GUI to train, evaluate, and deploy a Watson Machine Learning model based on the Titanic dataset.

Objectives:

Upon completing the lab, you will:

- Become familiar with the AutoAI feature of Watson Studio.
- Train/Evaluate a machine learning model
- Deploy a machine learning model.

Lab-5: Heart Disease Notebook

Build Model

Save and Deploy

Introduction:

In this lab, you will use a Jupyter Notebook to train a model using the XGBoost library to classify whether a person has heart disease or not. In addition to training a model, the notebook also explains how to persist a trained model to the IBM Watson Machine Learning repository, and deploy the model as a REST service.

Objectives:

Upon completing the lab, you will know how to:

- Load a CSV file into Pandas DataFrame.
- Prepare data for training and evaluation.
- Create, train, and evaluate a XGBoost model.
- Visualize the importance of features that were used to train the model.
- Use cross validation to select optimal model hyperparameters based on a parameter grid
- Persist best model in Watson Machine Learning repository using Python client library.
- Deploy the model for online scoring using the Watson Machine Learning's REST APIs

Lab 6: Watson OpenScale

Monitor and Manage

Introduction:

IBM Watson OpenScale is an open platform that helps remove barriers to enterprise-scale AI. In this lab you will configure Watson OpenScale to monitor quality, fairness,and drift and to provide the factors that explain a deployed model's classification.

Objectives:

Upon completing the lab, you will

- Import and Deploy a machine learning model
- Provision a Watson OpenScale service
- Configure Watson OpenScale for Payload Logging, Quality, Fairness, and Drift.
- Submit Feedback and View Quality Metrics
- Score Data and View Fairness Metrics
- Explain a Transaction.

Lab-7: Recognizing Handwritten Digits

Build Model

Introduction:

This lab will use the MNIST computer vision data set to train a convolutional neural network (CNN) model to recognize handwritten digits. The Watson Studio neural network flow editor, Watson Studio experiment builder and the Watson Machine Learning component will be used to build, train, and save the trained model.

Objectives:

Upon completing the lab, you will know how to:

- Create Cloud Object Storage buckets to contain the input and result files
- Create a neural network design from an example using the flow editor
- Use the experiment builder used to set up a training definition to train the neural network model
- Monitor the training progress and results.
- Save the trained model.
- Test the model

Lab 8: ART in Action

Monitor and Manage

- **Create a Notebook in Watson Studio**
- **Upload the Lab-8 Notebook file using provided URL**
- **Run through cells**

- **Overview**
  - Load a Tensorflow trained model
  - Create an ART classifier object using the loaded model
  - Perform an adversarial attack
  - Perform a defense to make sure manipulated images can still be classified correctly