

# End-to-End Data Science using IBM's Watson Studio



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

**Bernie Beekman**

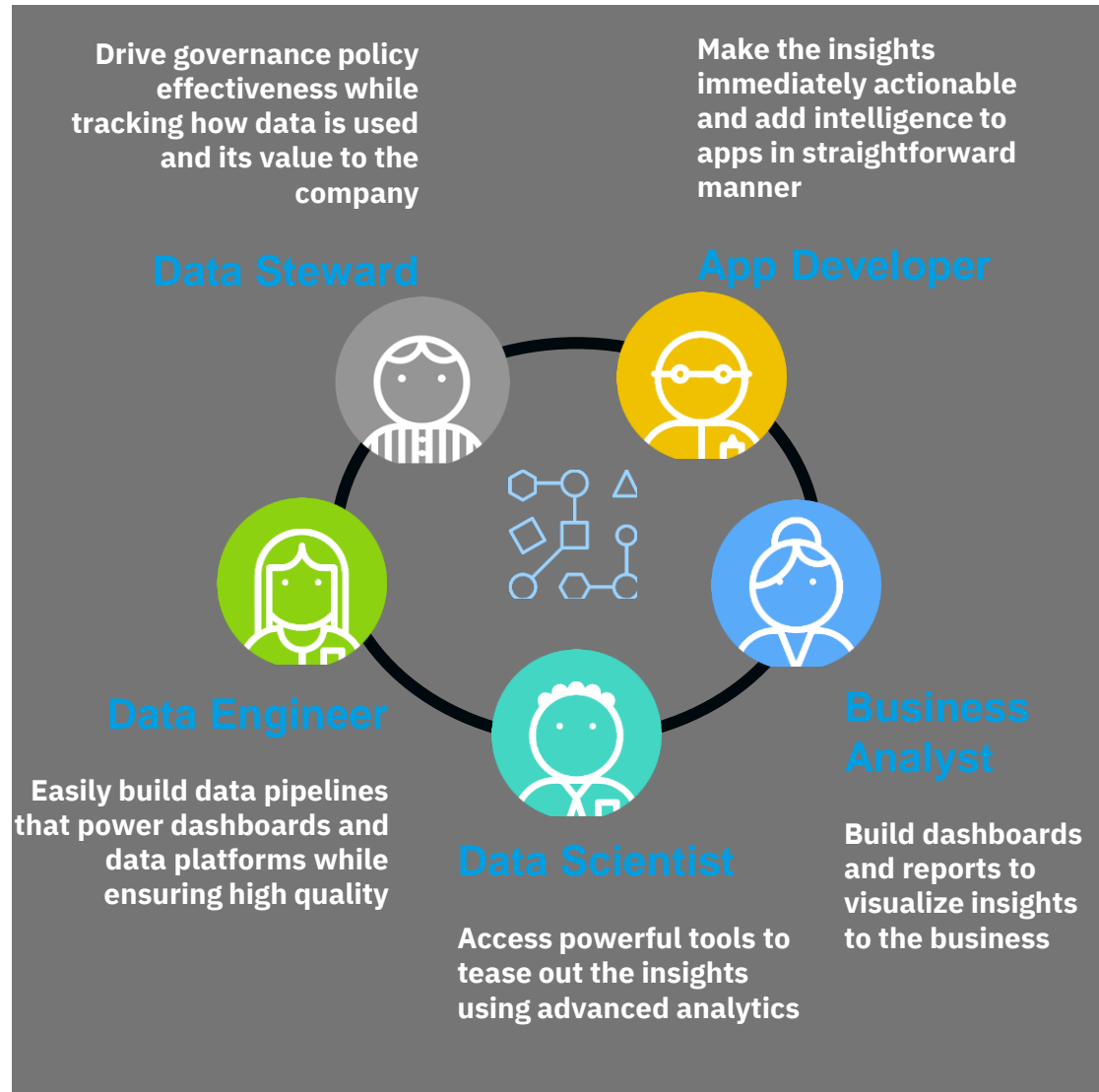
**Michael Cronk**

**James Parry**

**Prithv Rao**

# 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

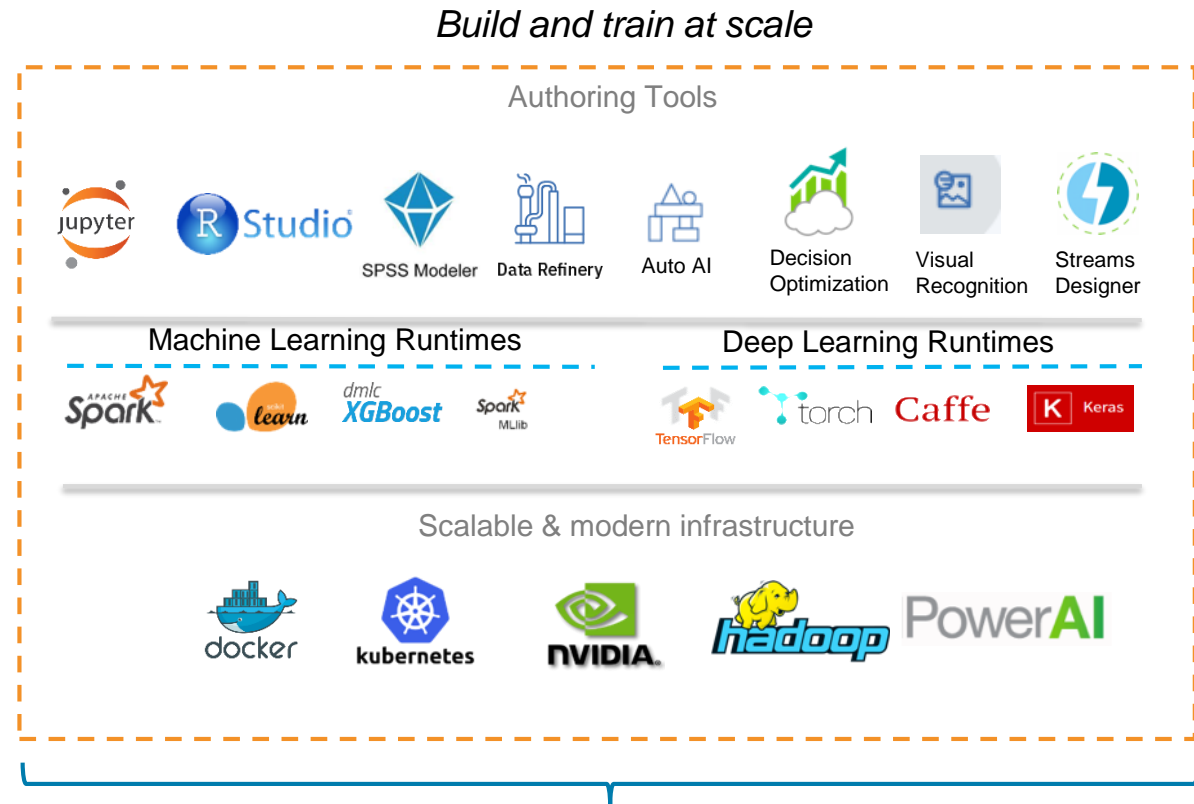
- Watson Studio on IBM Cloud
  - Managed offering provided by IBM
  
- Watson Studio Desktop
  
  
- IBM Cloud Private for Data
  - Watson Studio Local

# 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



**IBM**  
IBM Cloud  
Fully Managed

  
On-prem

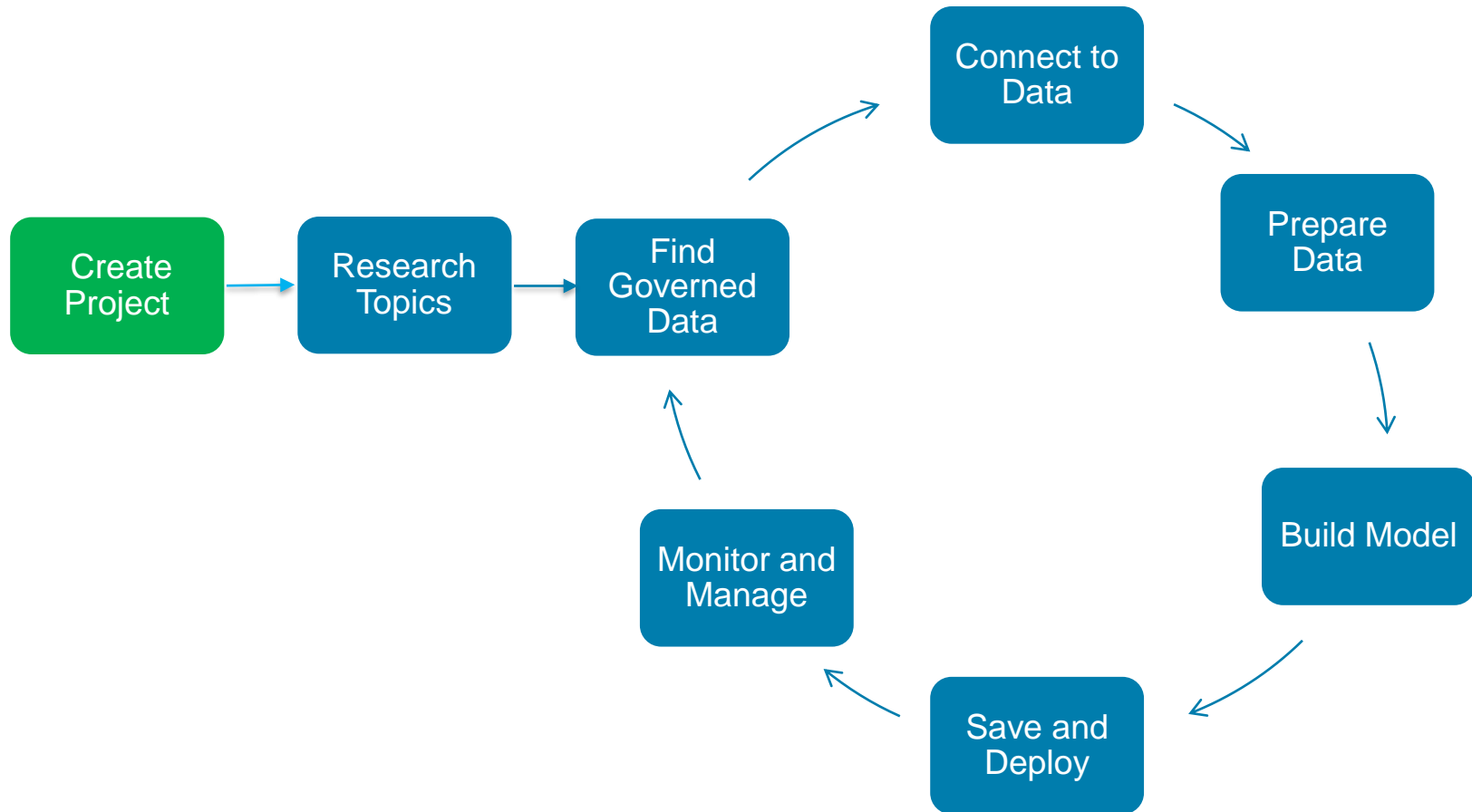
**IBM**  
IBM Cloud

 **amazon**  
web services

 **Azure**

# 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.*



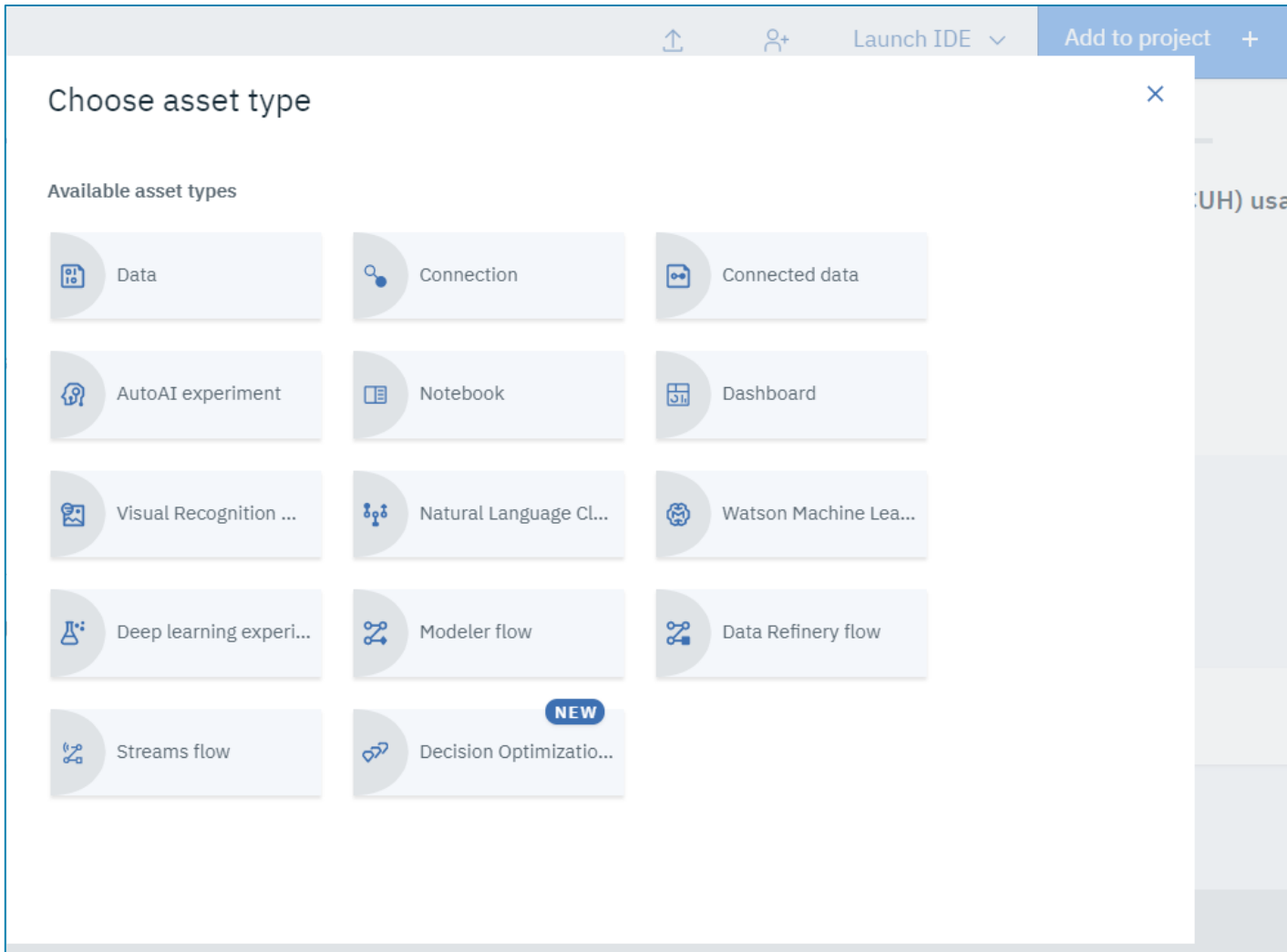
# Watson Studio Project Features

*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



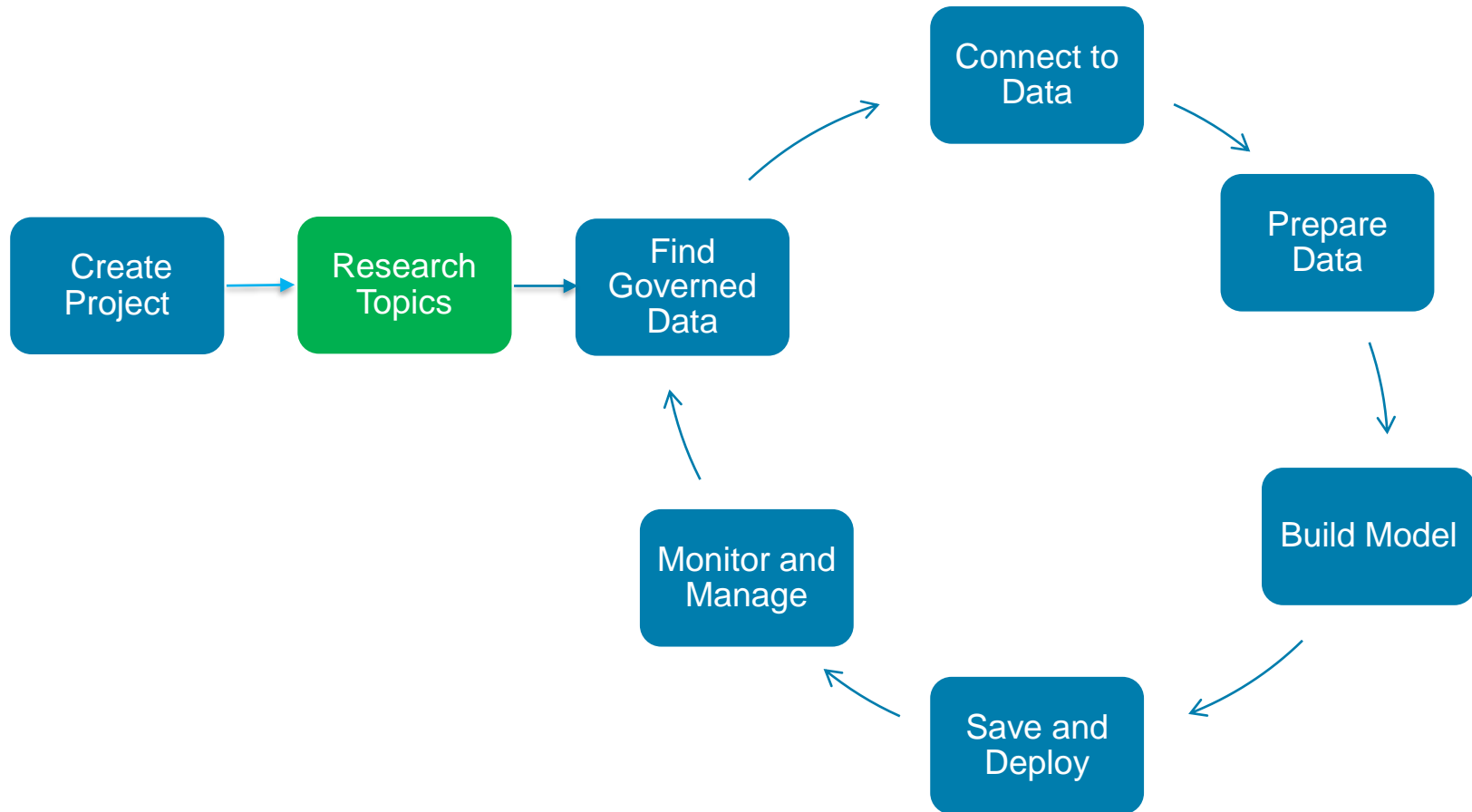
Choose asset type

Available asset types

 Data	 Connection	 Connected data
 AutoAI experiment	 Notebook	 Dashboard
 Visual Recognition ...	 Natural Language CL...	 Watson Machine Lea...
 Deep learning experi...	 Modeler flow	 Data Refinery flow
 Streams flow	 Decision Optimizatio... <b>NEW</b>	

# 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 Gallery

*Built-in learning to get started*


Research  
Topics

- 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](#)

DS\_POT\_01-30DS\_POT\_01-16IBM WatsonIBM Watson StService DetailsIBM Watson OFeed - WatsonIBM Cloud APIFeed - WatXIBM Watson StFeed - Watson+





Search with Google or enter address



IBMCommunity

Search



IBM Data Science CommunityTopic groupsLocal User GroupsEventsParticipateData Science Elite

Watson Studio and Machine Learning

Join the conversation.

Join / sign upExplore Watson Studio

Feed

News

Group resources

Learn

Support

Welcome!

You are not yet signed up for the IBM Community.

Join / sign up

Discussions

Login to start a discussion

Term / Keyword / Phrase

WATSON STUDIO


Ready to build [Trustbit] Using Watson Studio, ML, IBM Blockchain...


Events

No upcoming events.

[View all events →](#)

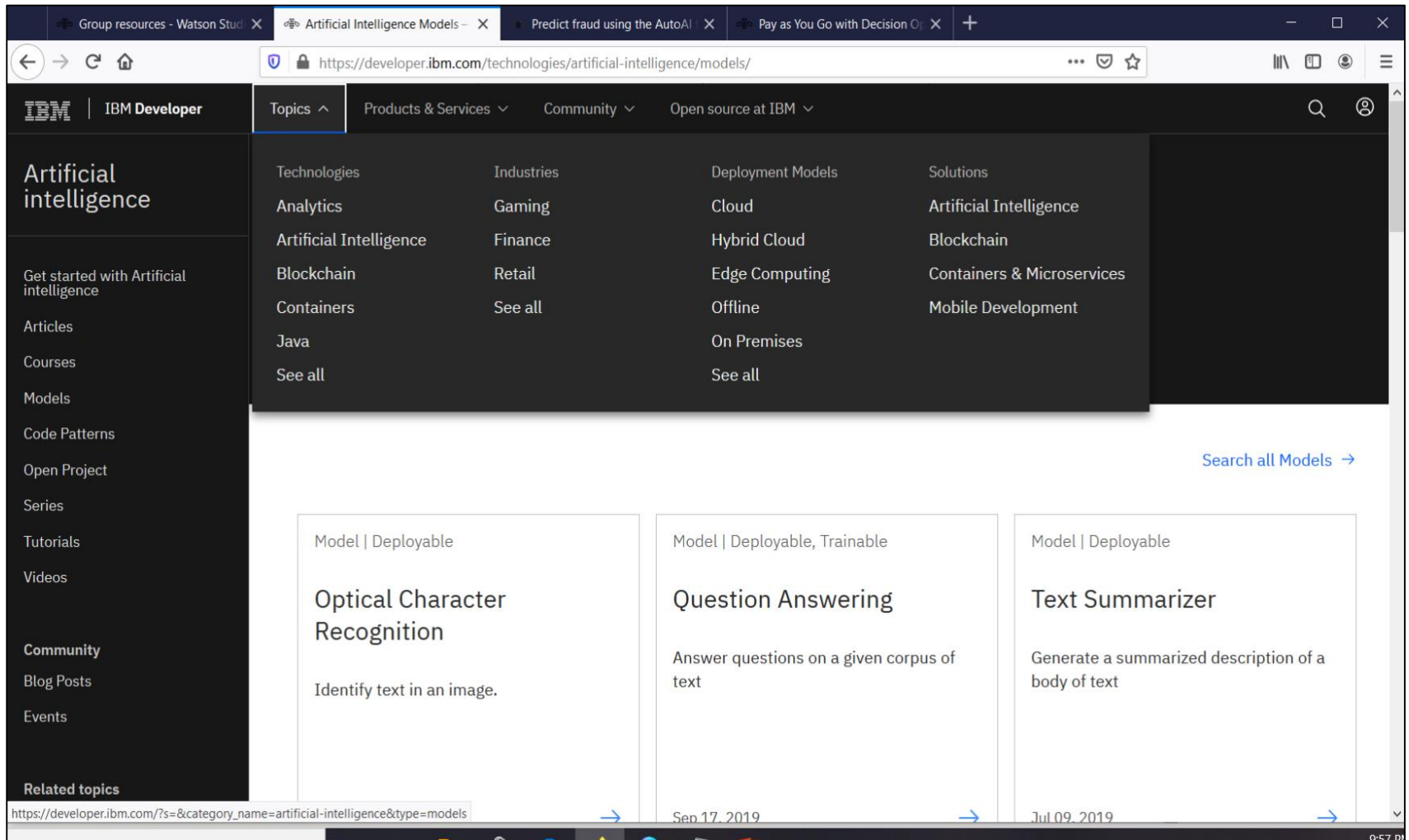
Type here to search





# developer.ibm.com

Research  
Topics



The screenshot displays the IBM Developer website's 'Artificial Intelligence Models' page. The browser's address bar shows the URL <https://developer.ibm.com/technologies/artificial-intelligence/models/>. The page features a dark navigation bar with the IBM logo and 'IBM Developer' text. A sidebar on the left lists various resources under 'Artificial intelligence', including 'Get started with Artificial intelligence', 'Articles', 'Courses', 'Models', 'Code Patterns', 'Open Project', 'Series', 'Tutorials', 'Videos', 'Community', 'Blog Posts', 'Events', and 'Related topics'. The main content area has a dark header with 'Topics' selected, showing a grid of categories: Technologies, Analytics, Artificial Intelligence, Blockchain, Containers, Java, and See all. Below this, a 'Search all Models' link is visible. The main content area displays three model cards: 'Optical Character Recognition' (Model | Deployable), 'Question Answering' (Model | Deployable, Trainable), and 'Text Summarizer' (Model | Deployable). Each card includes a brief description and a date. The footer shows the URL [https://developer.ibm.com/?s=&category\\_name=artificial-intelligence&type=models](https://developer.ibm.com/?s=&category_name=artificial-intelligence&type=models) and the time 9:57 PM.

Group resources - Watson Stud X Artificial Intelligence Models - X Predict fraud using the AutoAI X Pay as You Go with Decision O X +

← → ↺ ↻ ↵

https://developer.ibm.com/technologies/artificial-intelligence/models/

IBM | IBM Developer

Topics ^ Products & Services ^ Community ^ Open source at IBM ^

Artificial intelligence

Get started with Artificial intelligence

Articles

Courses

Models

Code Patterns

Open Project

Series

Tutorials

Videos

Community

Blog Posts

Events

Related topics

Technologies

Analytics

Artificial Intelligence

Blockchain

Containers

Java

See all

Industries

Gaming

Finance

Retail

See all

Deployment Models

Cloud

Hybrid Cloud

Edge Computing

Offline

On Premises

See all

Solutions

Artificial Intelligence

Blockchain

Containers & Microservices

Mobile Development

Search all Models →

Model | Deployable

Optical Character Recognition

Identify text in an image.

Model | Deployable, Trainable

Question Answering

Answer questions on a given corpus of text

Model | Deployable

Text Summarizer

Generate a summarized description of a body of text

https://developer.ibm.com/?s=&category\_name=artificial-intelligence&type=models

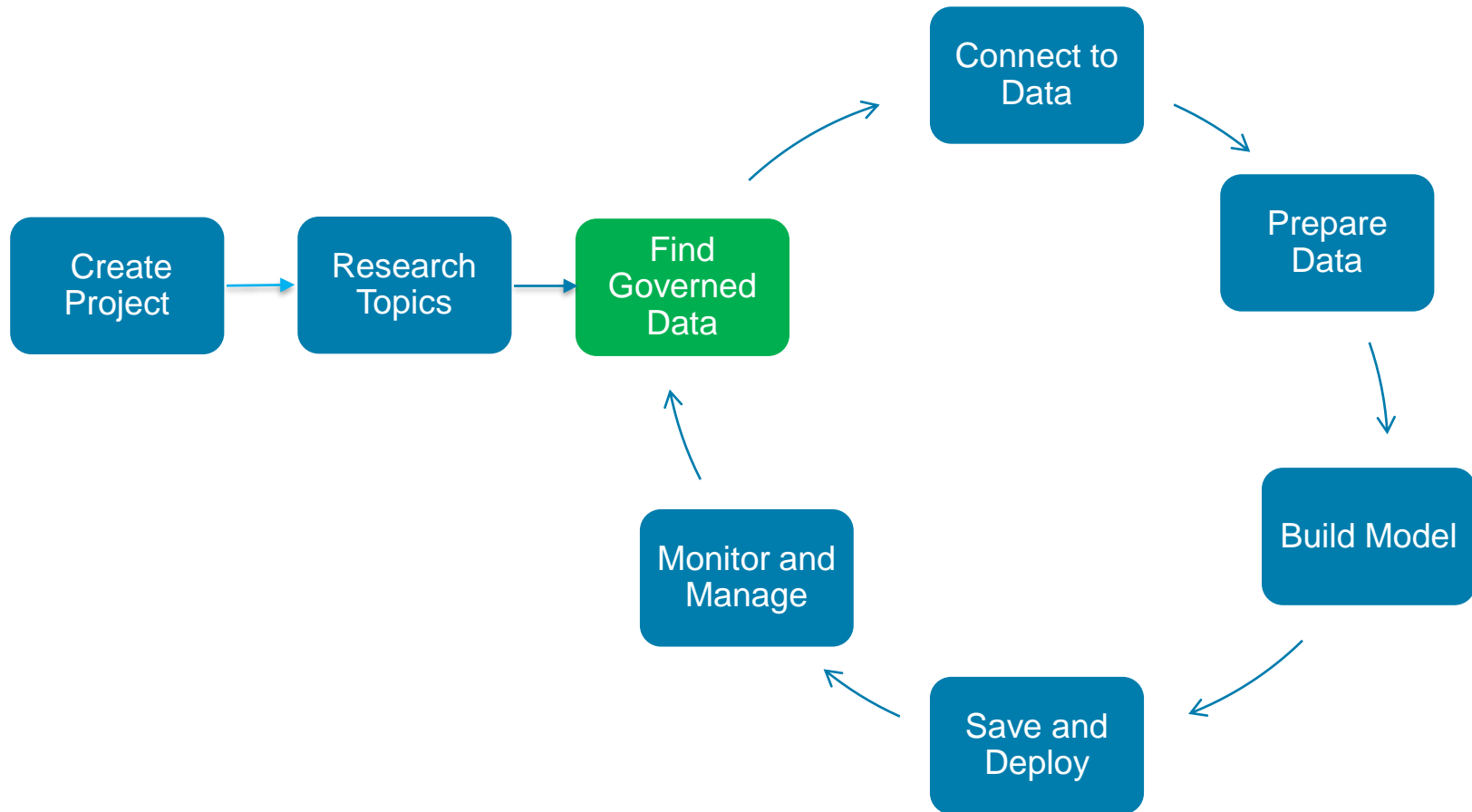
Sep 17, 2019

Jul 09, 2019

9:57 PM

# 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

*Unlock tribal knowledge and unleash knowledge workers*

Find  
Governed  
Data

- **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

10 Data Asset

## female\_human\_trafficking

### Description

There is no description available for this asset.

Added: Jan 31, 2019 10:02 AM  
Format: application/octet-stream  
Size: 347 KB

### Tags

trafficking | female human trafficking

### Reviews

☆☆☆☆☆ 0 reviews

### Connection

Source: [Watson Studio Labs\\_DataCatalog](#)  
Source type: Cloud Object Storage

### Classification

Personally Identifiable Information

Personally identifiable information (PII) is defined as any data that could potentially identify a specific individual. Any information that can be used to distinguish one person from another can be considered PII.

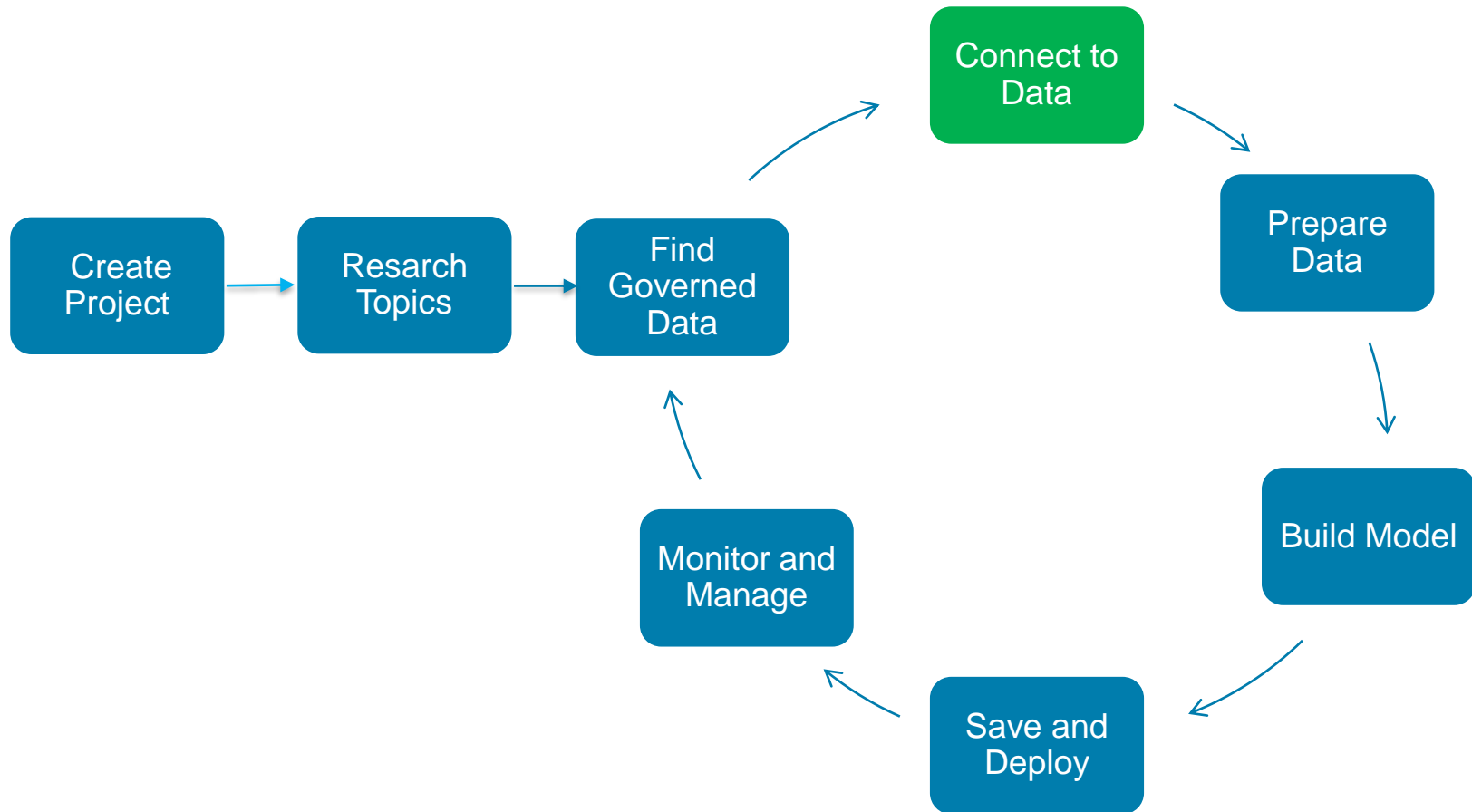
Schema: 26 Columns | 1085 Rows | 2 Columns anonymized ⓘ

Preview: 1000 rows | Last refresh: 22 seconds ago | [Refresh](#)

DATE	BIRTH_COUNT...	BIRTH_COUNTRY_CODE	OCCUPATION	ADDRESS	SSN	PASSPORT_NUMBER
	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String
th	Country Name	Country Code	Text	Text	US Social ...	Passport Number
15	Ghana	GH	Engineer, land	824 Kristin Grv, A	afe55d1d355c3:	1c9da91e1e20863dd850
19	Ghana	GH	Editor, commissi	1148 Wang Fall S	77a0daa42ec7d	12d38855ed107e7cc5dd
16	Ghana	GH	Merchant navy of	9486 Pratt Wall,	669061087d6d1	c43ed0283a3def7031d8:
17	Ghana	GH	Paramedic	0890 Johnson Tr	997b59e501b2e	179abee5ba608418154d
18	Ghana	GH	Surveyor, buildin	2315 Brittany Cr	70329b83b40cb	84524ccc3c5c6590600e:
24	Ghana	GH	Waste managem	88811 Donald Pa	d2f2236f52407:	a730ae13f5ed96f71e904
23	Ghana	GH	Doctor, general p	9150 Donald Rpe	d2c2d41163d8f:	ced1617be1d70e44421c
02	Ghana	GH	Forest/woodland	1355 Lopez Villa	62007942c2b0c	8c8debd401b6b6d954b
12	Ghana	GH	Land/geomatics :	86792 Amy Vlgs,	08f8dd9f9ba89t	a43f1d6c9cacdfa82a1a1
10	Ghana	GH	Oncologist	108 Erin Via, Nev	f8b871f6e058e2	f289be62078ebbe457c6:
07	Ghana	GH	Veterinary surger	79572 Schmidt E	f2006c1d30df33	624a9605774a0cfd98aa

# 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](#)

The screenshot shows the IBM Analytics Notebook interface. The top menu bar includes File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. The right side of the top bar shows 'Trusted', 'Python3.6', and tabs for 'Local', 'Remote', and 'Other'. Below the menu is a toolbar with icons for file operations, running, and saving. The main area contains two code cells. The first cell displays a schema for a table with various integer and string fields. The second cell, titled 'Read in CLAIMS Table', contains Python code using PySpark to connect to a remote data source and load data into a DataFrame.

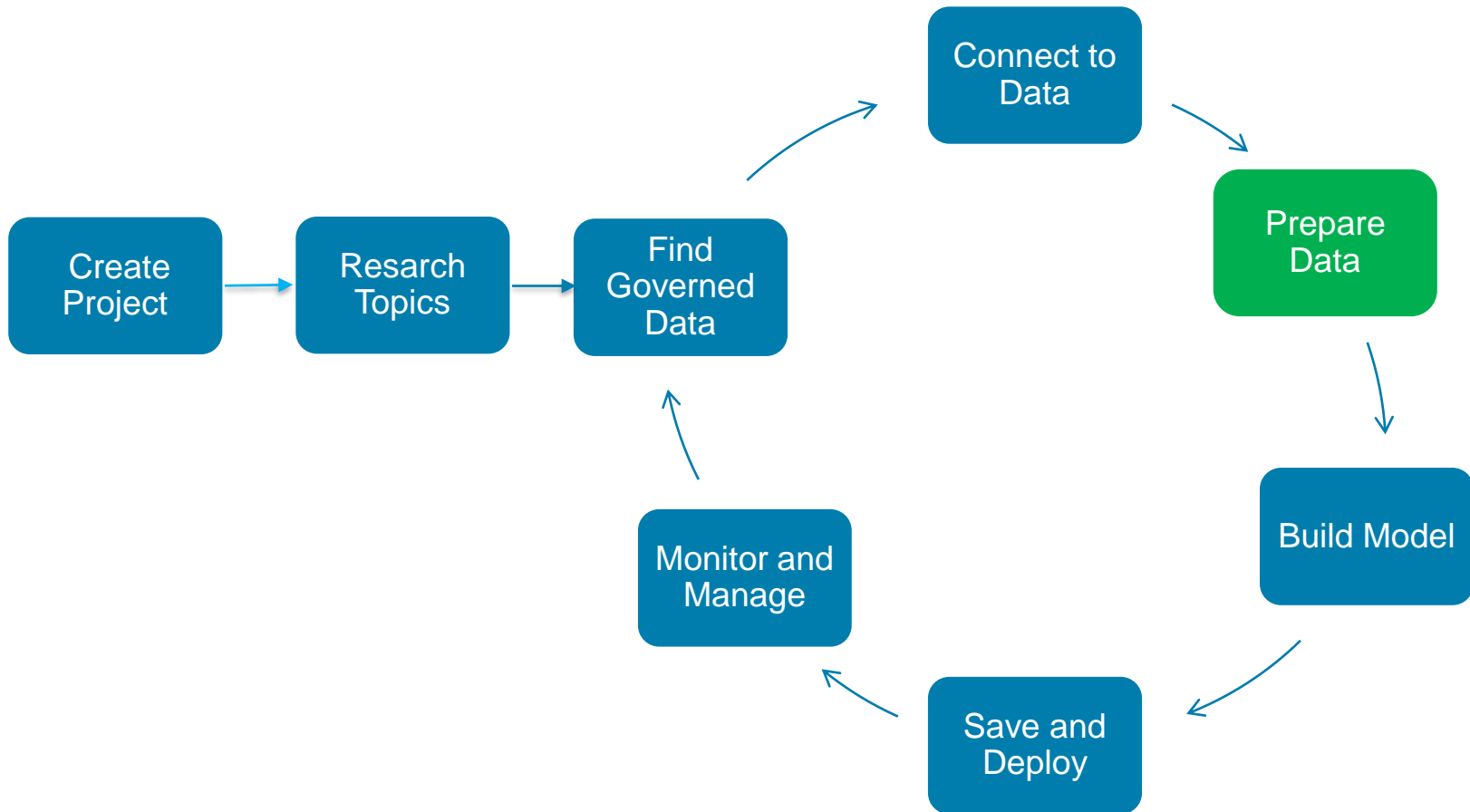
```
-- PROCEDURE_PERCT_RANK: integer (nullable = true)
-- PROCEDURE_RISK_GROUP: string (nullable = true)
-- QUANTITY_INDEX: integer (nullable = true)
-- SERVICE_TYPE: string (nullable = true)
-- SUBMIT_CHG: integer (nullable = true)
-- SUBMITTED_CHG_INDEX: integer (nullable = true)
-- TOTAL_CHARGES_INDEX: integer (nullable = true)
-- TOTAL_CHARGES_PER_PROCEDURE: integer (nullable = true)
-- USER_DEFINED_FLAG_0: string (nullable = true)
-- SUBMITTED_CHARGE_AMOUNT: integer (nullable = true)
-- CLAIM_NUMBER: string (nullable = true)
-- IS_FRAUD: string (nullable = true)
```

### Read in CLAIMS Table

```
In [5]:
import dsx_core_utils, requests, os, io
from pyspark.sql import SparkSession
# Add asset from remote connection
df7 = None
dataSet = dsx_core_utils.get_remote_data_set_info('CLAIMS')
dataSource = dsx_core_utils.get_data_source_info(dataSet['datasource'])
sparkSession = SparkSession(sc).builder.getOrCreate()
# Load JDBC data to Spark dataframe
dbTableOrQuery = ''' + (dataSet['schema'] + '`.`' if len(dataSet['schema'].strip()) != 0 else '') + dataSet['table'] + '''
if (dataSet['query']):
```

# 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

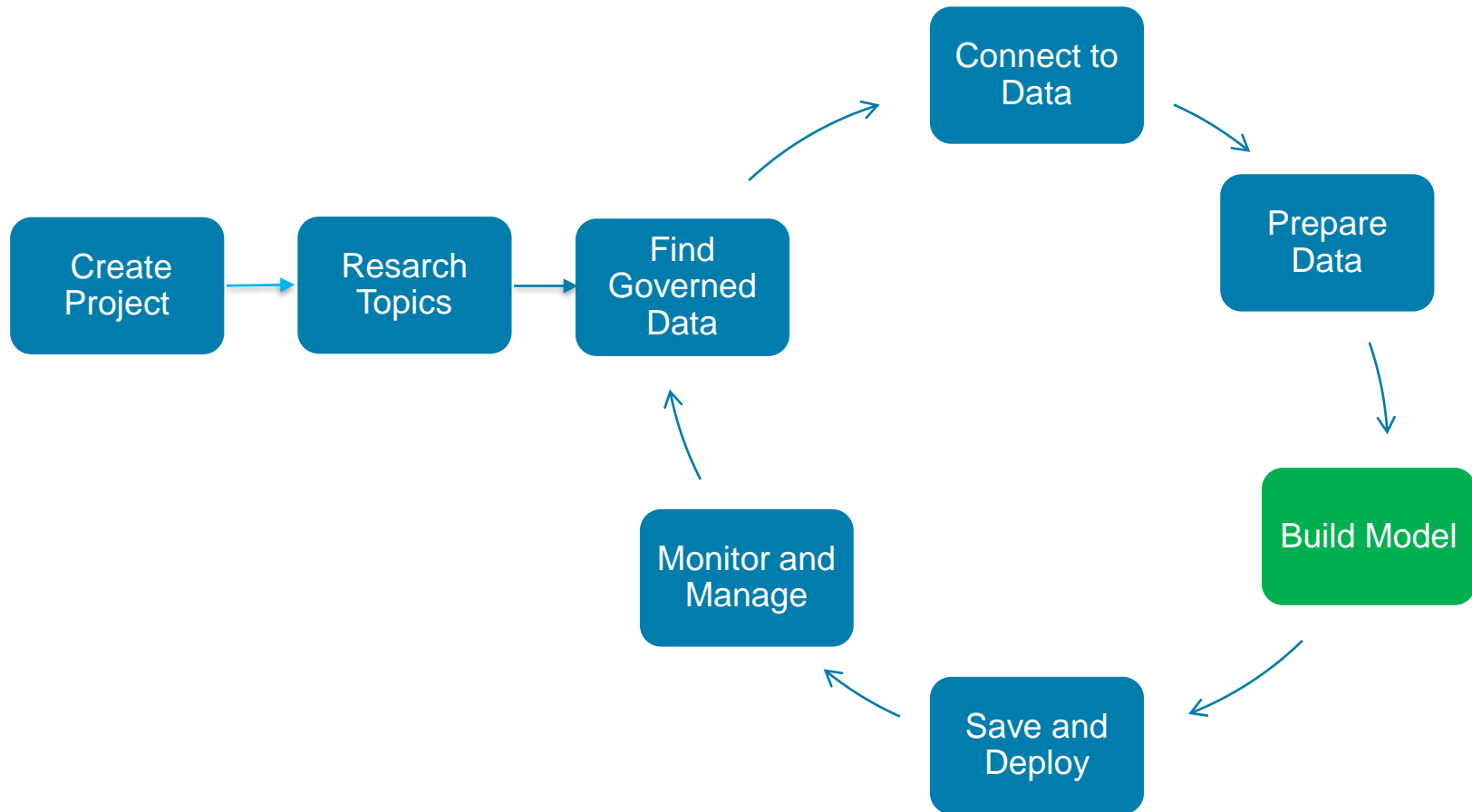
*Making Data fit for use*

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

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

Build Model

## 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
- Decision Optimization

\*\* in hands-on labs


# Jupyter Notebook

[Build Model](#)

My Projects / Watson Studio Labs / Machine Learning with SparkML

File Edit View Insert Cell Kernel Help

Not Trusted | Python 3.6 with Spark           Format Markdown 

## Read Data Asset - female\_human\_trafficking - See Lab Instructions

```
In [ ]: # Insert SparkSession DataFrame code in this cell after the comments.  
# make CERTAIN to rename the default dataframe name (df_data_1 or df_data_2 or df_data_3, etc) to trafficking_df  
# Put cursor on the next line to Insert to code.
```

## Read Data Asset - Occupations - See Lab Instructions

The occupations listed in the female human trafficking file are too numerous to use as input to a machine learning model. We will categorize these occupations into 15 categories by joining with two other files. The Occupation.csv file contains a mapping of the occupations in the female human trafficking table to a category code. The Categories.csv file contains each code followed by the category name. This information needs to be joined to the female human trafficking table.

Follow the same procedure as above to insert a SparkDataFrame for Occupations

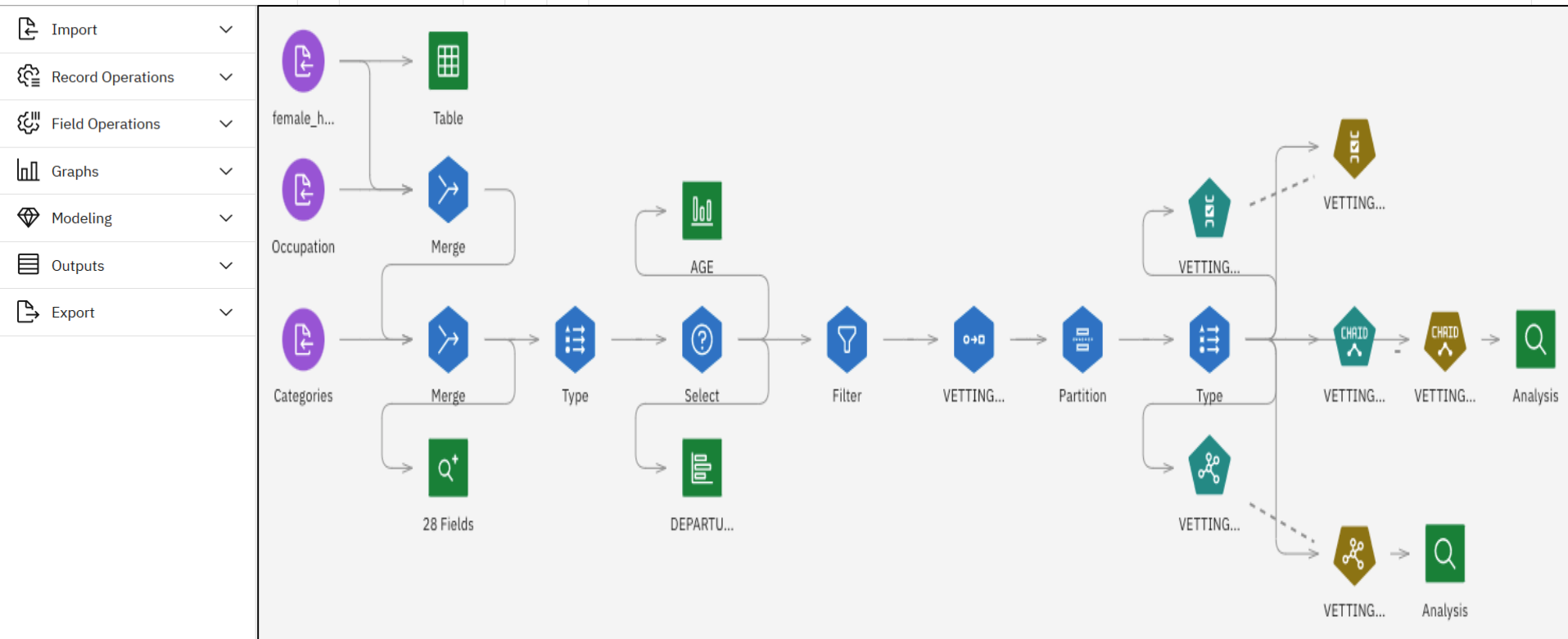
```
In [ ]: # Insert SparkSession DataFrame code in this cell after the comments  
# make CERTAIN to rename the default dataframe name (df_data_1 or df_data_2 or df_data_3,etc) to occupations  
#Put cursor on the next line to Insert to code
```

## Read Data Asset - Categories - See Lab Instructions

Follow the same procedure as above to insert a SparkDataFrame for Categories

```
In [ ]: # Insert SparkSession DataFrame code in this cell after the comments  
# make CERTAIN to rename the default dataframe name (df_data_1 or df_data_2 or df_data_3,etc) to categories  
#Put cursor on the next line to Insert to code
```

## Build Model





# Neural Network Modeler

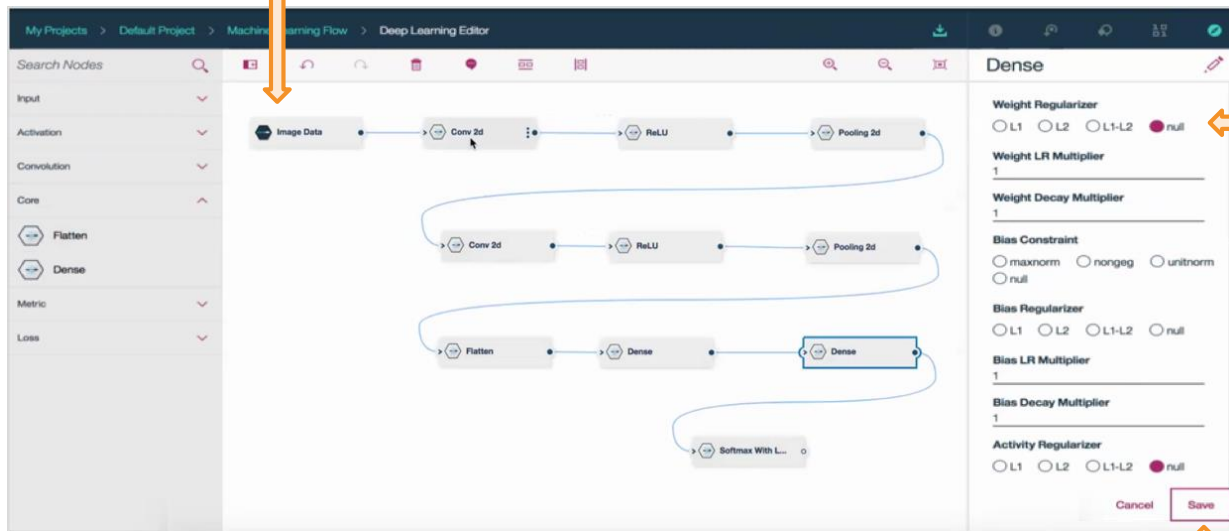
[Build Model](#)

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

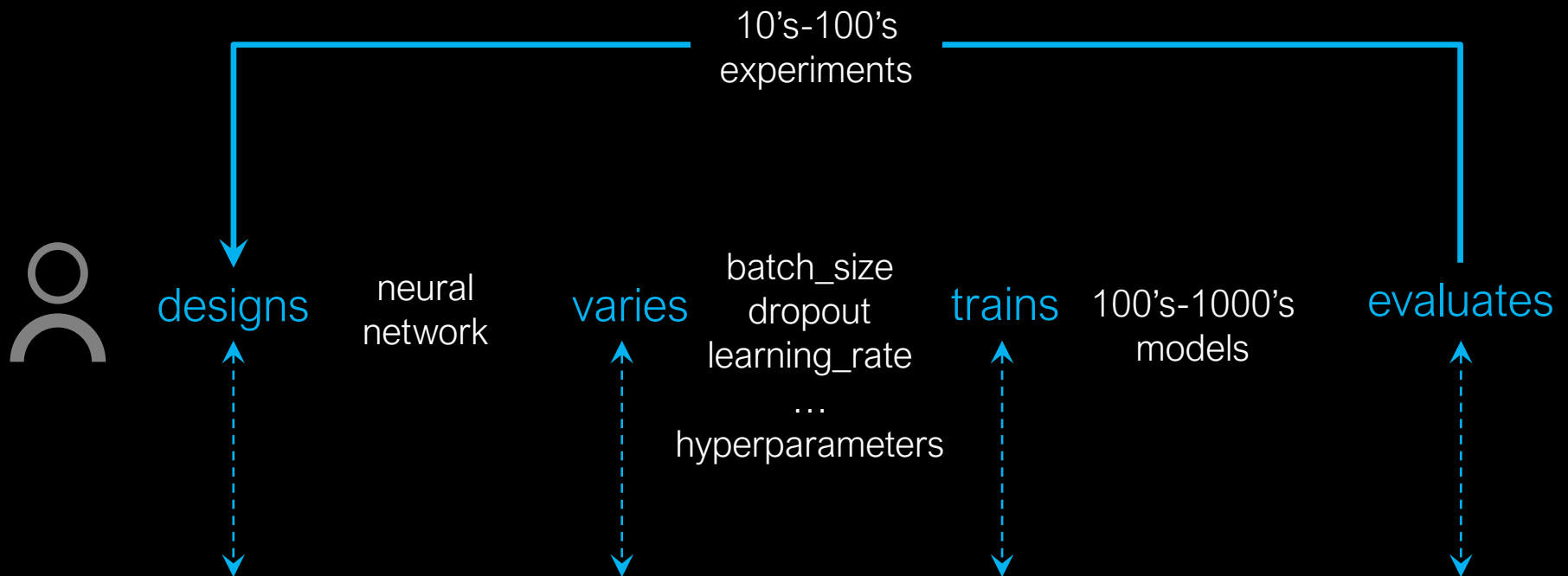
 KerasDrag-and-drop  
network layersReal-time validation of network  
flow

- Define layer configuration
- Choose optimizer params

- Generate CPU or GPU compatible code

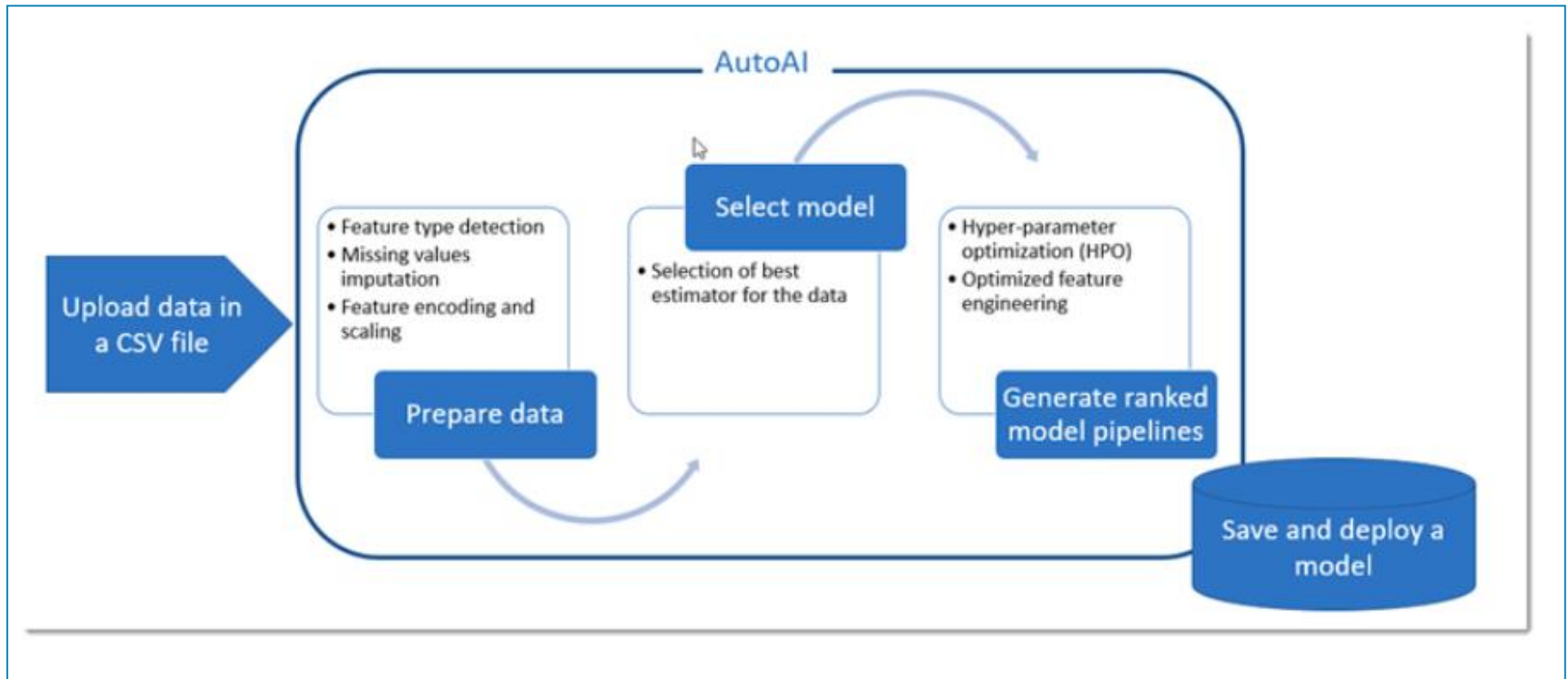
- Save as popular framework code
- Export as a python notebook
- Execute as batch experiment

# Experiment Builder

[Build Model](#)

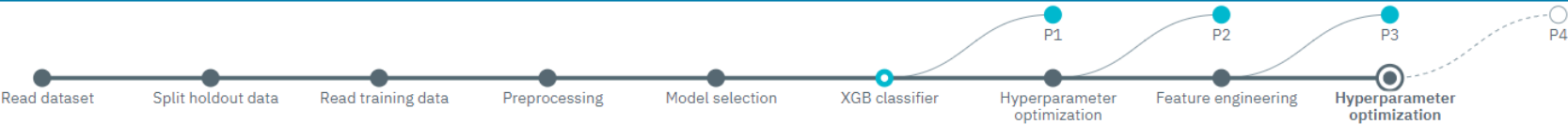
Experiment Builder  
supports the end-to-end workflow

# 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 > ...	<a href="#">View details</a>	<a href="#">Save as model</a>
>	2	0.884	P1 - XGB classifier estimator Transformers (2): Preprocessing > XGB classifier estimator	<a href="#">View details</a>	<a href="#">Save as model</a>
>	3	0.884	P2 - XGB classifier estimator Transformers (2): Preprocessing > XGB classifier estimator	<a href="#">View details</a>	<a href="#">Save as model</a>

# Decision Optimization

[Build Model](#)

**Decision Optimization (DO)** enables data science teams to capitalize on the power of *prescriptive analytics* and build solutions using a combination of techniques like optimization and machine learning.

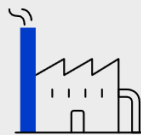
Integrated with Watson Studio, Decision Optimization can combine optimization techniques with coding and non-coding tools, model management and deployment – as well as other data science capabilities.

Decision Optimization evaluates millions of possibilities – balancing trade-offs and business constraints to find the best possible solution.

## Insights that drive optimal decisions to complex problems



Determine location  
and capacity  
of warehouses



Determine which plant  
should manufacture  
which product



Build financial  
portfolios by balancing  
risks and rewards

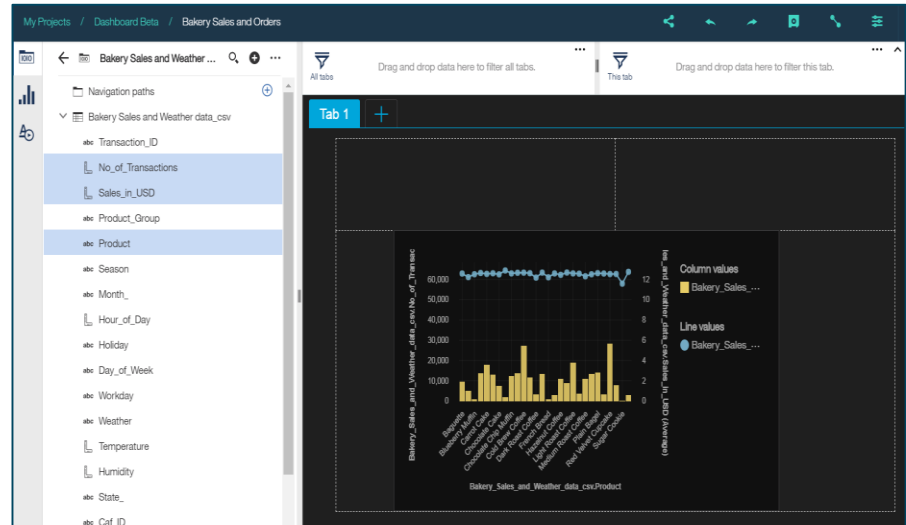
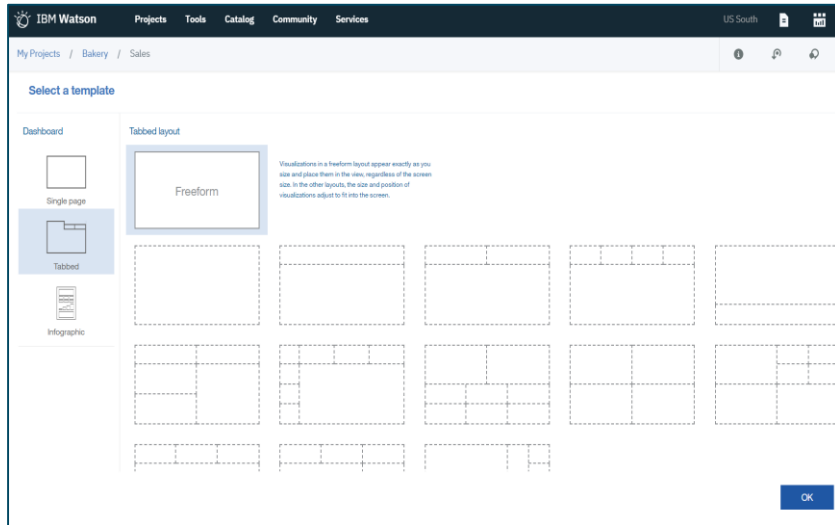


Allocate aircraft  
and crew to flights

# Watson Studio Dynamic Dashboards

*Making insights available to all*

Build Model



My Projects / Bakery Sales

0 assets selected.

NAME	TYPE	SERVICE	CREATED BY	LAST MODIFIED	ACTIONS
UNdata_agri_value_add.csv	Data Asset	Project	Alex Jones	7 Mar 2018, 9:37:13 am	
EuropeanCountryStats.csv	Data Asset	Project	Alex Jones	7 Mar 2018, 9:37:12 am	
Bakery Sales and Weather data.csv	Data Asset	Project	Alex Jones	8 Feb 2018, 3:07:05 pm	

0 notebooks selected.

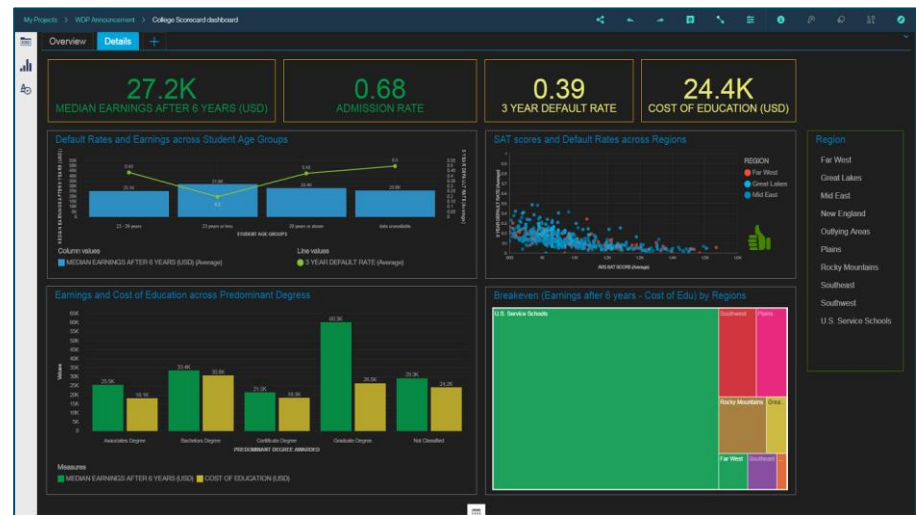
NAME	SHARED	SCHEDULED	STATUS	LANGUAGE	LAST EDITOR	LAST MODIFIED	ACTIONS
Sales Predictions					Alex Jones	7 Mar 2018	

0 streams flows selected.

0 dashboards selected.

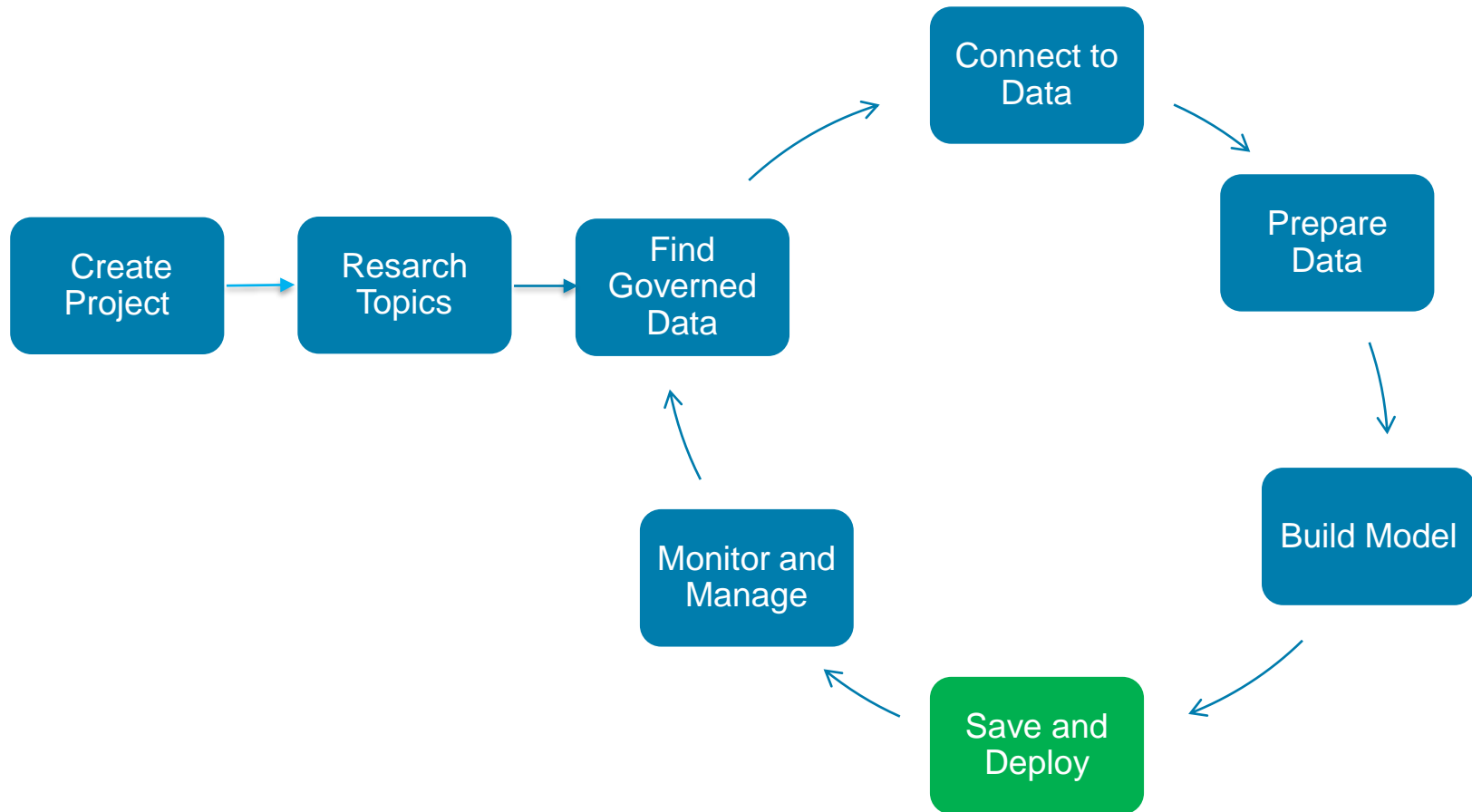
NAME	SHARED	LAST EDITOR	LAST MODIFIED	ACTIONS
Bakery Dashboard		Alex Jones	9 Feb 2018, 4:58:46 pm	

0 models selected.



# 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



dataplatform.cloud.ibm.com/ml/models/8e525b3c-cd00-48a9-9701-e9208dccee5d/new-deployment?projectid=b9150de9-e8d0-4119-af59-d1a0c8fbf7ee&mlInstanceGuid=be47af33-01c...

IBM Watson Projects Tools Catalog Community Services Manage Support Docs 1449375 - IBM Corpor... BB

## Create Deployment

### Define deployment details

Name

Model

Description

Deployment description

300

Deployment type

- ☒ Web service  
☐ Batch prediction  
☐ Realtime streaming prediction

Cancel

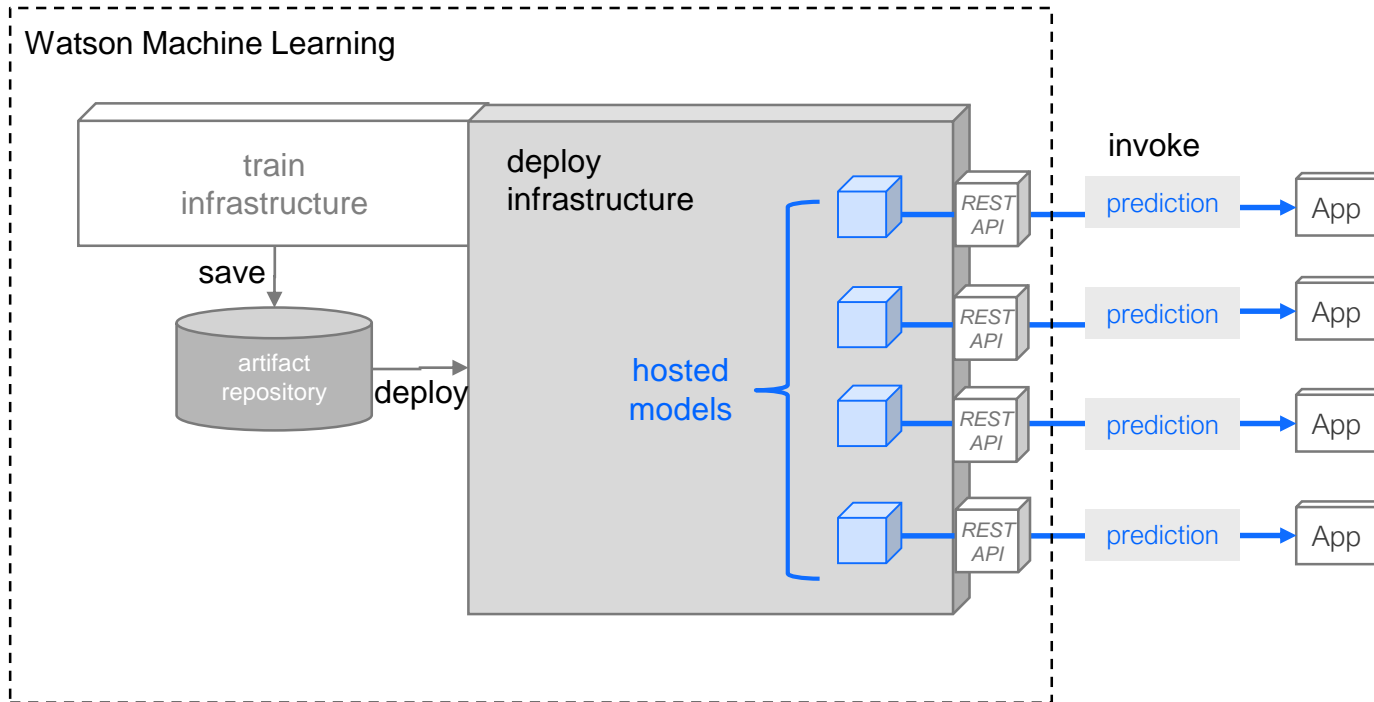
Save





# Watson Studio Save and Deploy Trained Models

*Save and Deploy Models with Watson Machine Learning*



# Watson Studio Save and Deploy Features

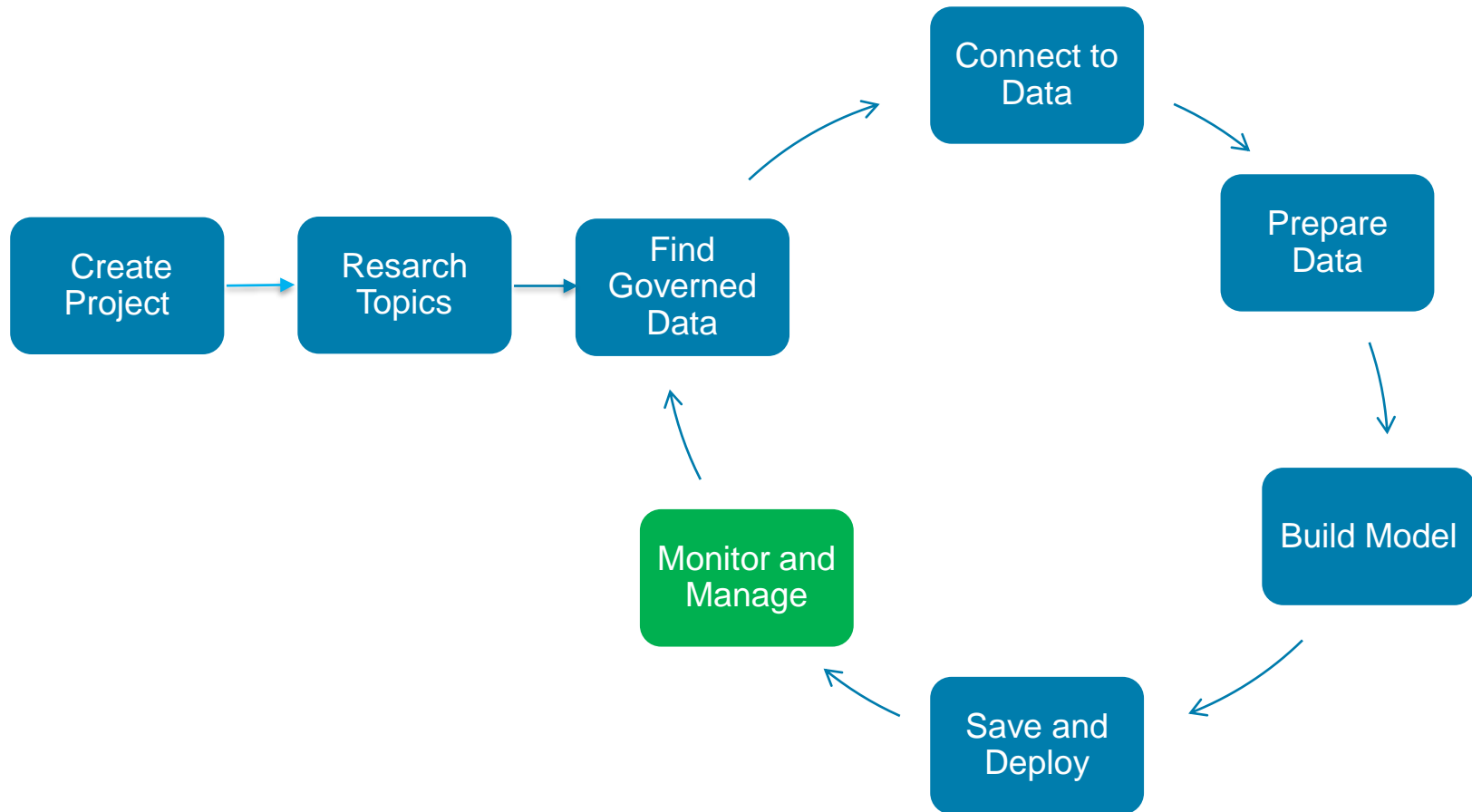
*Save and Deploy Models with Watson Machine Learning*

Save and  
Deploy

- 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*

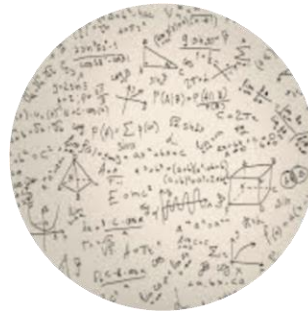
Monitor and  
Manage



**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

## Open By Design

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

# 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

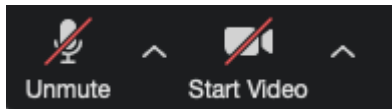
- Deploy models into production then monitor them to evaluate performance.
- Capture new data for continuous learning and retrain models so they continually adapt to changing conditions.

## 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

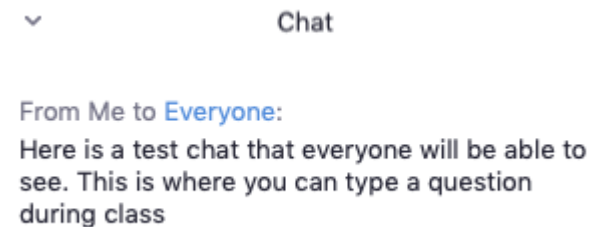
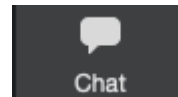
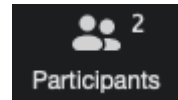
# Lab Overview

## The Zoom Tool Bar



We will stay on mute during the main sessions. You may unmute yourself during breakouts.

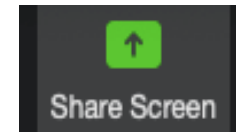
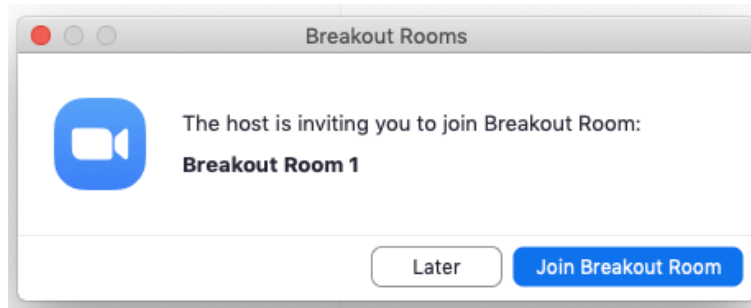
We will not use video to preserve bandwidth during our class today.



The Share Screen function has been deactivated during the lectures



## Breakout Rooms



Note: you will need to un-mute when you join the breakout room



Note: please make sure that you are not displaying any proprietary information on the screen before joining a breakout session.

## Lab Tips

- Watson Studio url: [dataplatform.ibm.com](https://dataplatform.ibm.com)
- Labs are in [www.github.com/bleonardb3/ML\\_POT\\_07-23-2020](https://www.github.com/bleonardb3/ML_POT_07-23-2020) 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. Review the rest of the lab and then move on to Lab-8 if the model does not get out of the queueing stage in 2-3 minutes.
- Lab-7- Make sure you switch Region to us-geo. If you get a 502 Gateway error, clear the browser cache.

# 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
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
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](#)[Save and Deploy](#)

## 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