

IBM Red Bull Basement Johannesburg 2018

How to start a data science project with IBM Watson Studio

Pre-requisite:

- A IBM Cloud Account link
- Dataset

Cloud <http://ibm.biz/redbullbasement>

Dataset [link to dataset](#)

Create a Watson Studio instance

1. From the dashboard, select **Catalog**

The screenshot shows the IBM Cloud Catalog interface. The top navigation bar includes 'Catalog', 'Docs', 'Support', and 'Manage'. The main content area is titled 'Dashboard' and features several filters: 'RESOURCE GROUP' (All Resources), 'CLOUD FOUNDRY ORG' (All Organizations), 'CLOUD FOUNDRY SPACE' (All Spaces), 'LOCATION' (All Locations), and 'CATEGORY' (All Categories). A search bar labeled 'Filter by resource name...' and a 'Create resource' button are also present.

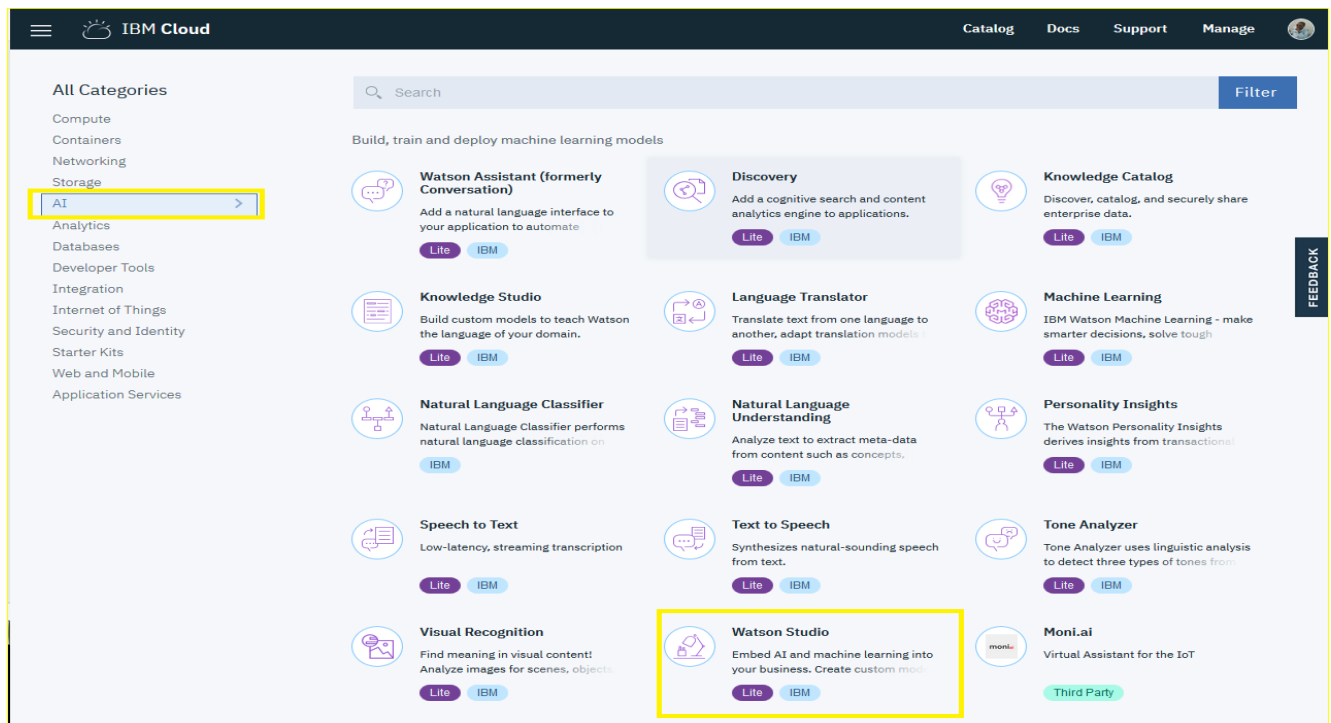
Cloud Foundry Services

Name	Region	CF Org	CF Space	Plan	Service Offering
Msamido_Spark_service	United Kingdom	mapurunyane.samuel.	Dev	Lite	Apache Spark
Msamido_Watson_Studio_Demo	United Kingdom	mapurunyane.samuel.	Dev	Lite	Watson Studio

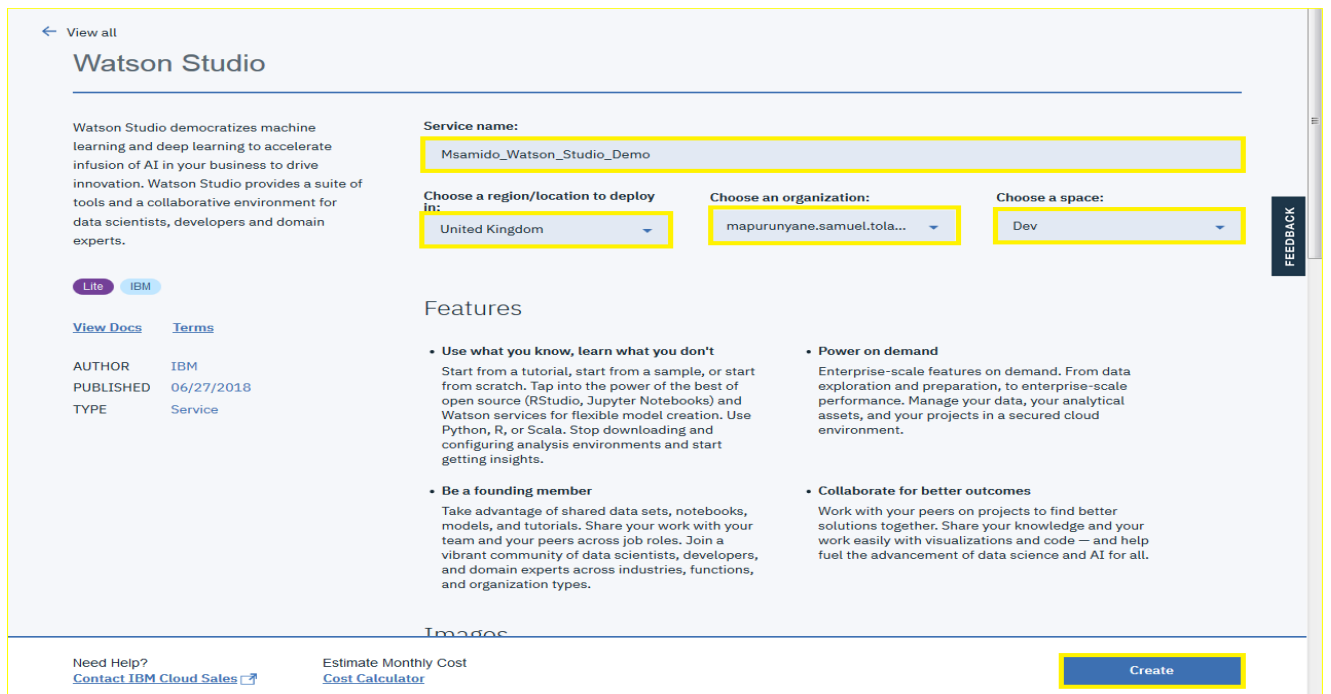
Services

Name	Location	Resource Group	Plan	Details	Service Offering
Msamido_ML	United Kingdom	default	Lite	Provisioned	Machine Learning
cloud-object-storage-pk	global	default	Lite	Provisioned	Cloud Object Storage

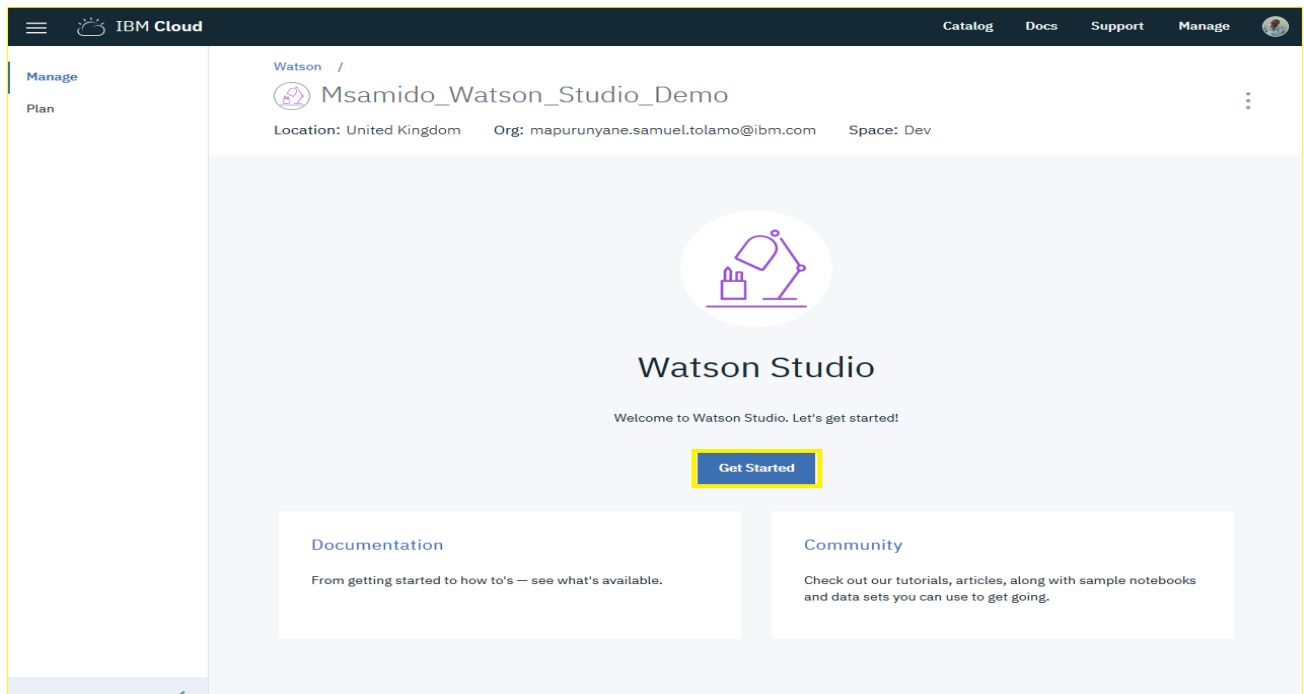
1. Under Watson Services, select **Watson Studio**



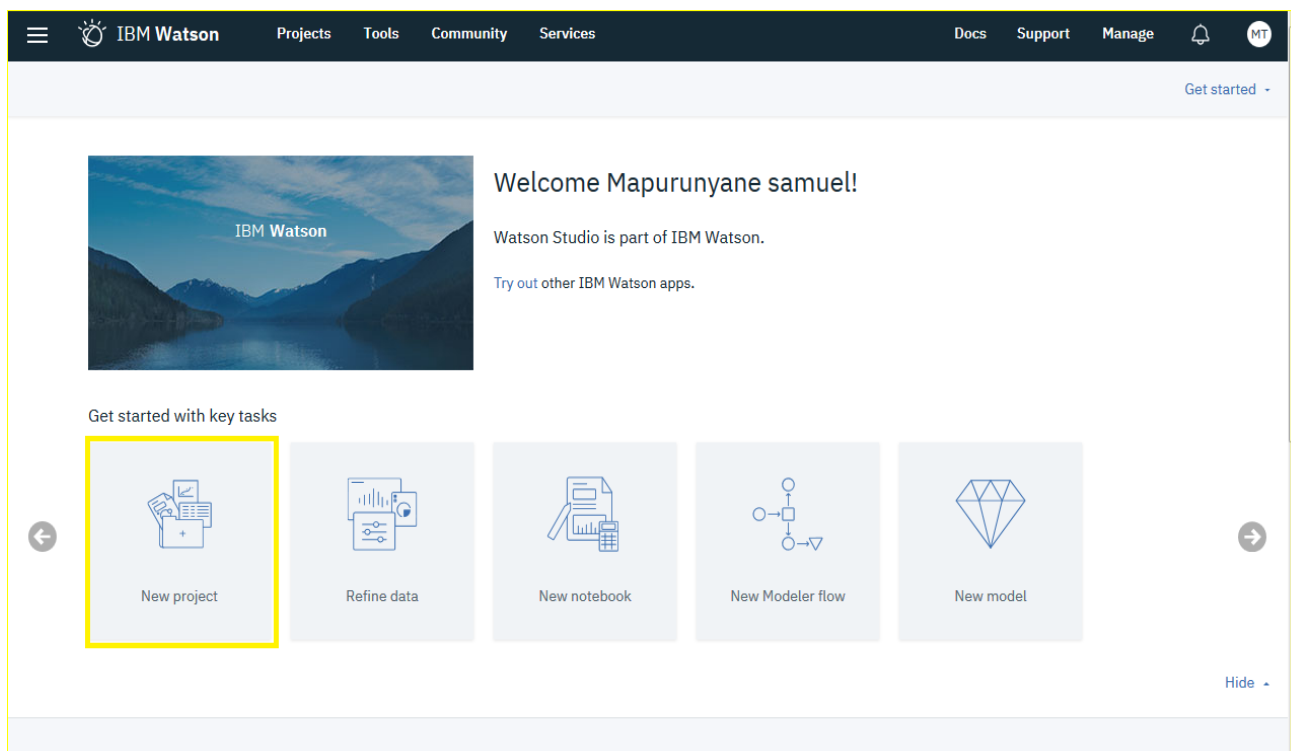
Give your service a name, and click **Create**



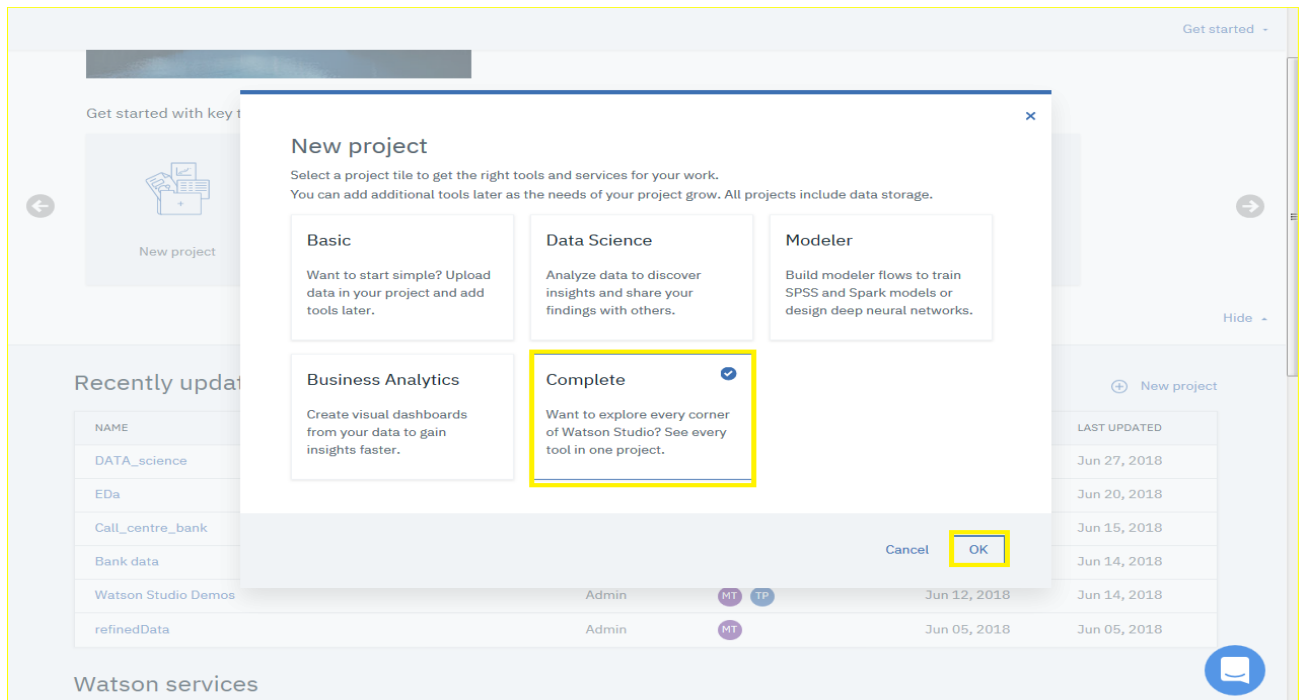
You will be redirected to the Watson Service start page, click the **Get Started** button



Select **New Project**



From the Pop-up Window select **Complete**, then click **Ok**.



From the project Detail window, You are required to create object storage service if you don't have one.

New project

Define project details

Name
Msamido_Machine_Learning

Description
my 1st Watson studio machine learning project

Choose project options

☐ Restrict who can be a collaborator ⓘ

Project will include integration with [Cloud Object Storage](#) for storing project assets.

Define storage

① Select storage service
[Add](#)

By clicking add from the above screen, it will take you to object storage page.

Creating an object storage

Cloud Object Storage

ExistingNew

Cloud Object Storage

IBM Cloud Object Storage is a highly scalable cloud storage service, designed for high durability, resiliency and security. Store, manage and access your data via our self-service portal and RESTful APIs. Connect applications directly to Cloud Object Storage use other IBM Cloud Services with your data.

Features

Storage for the IBM Cloud

IBM Cloud Object Storage provides unstructured data storage for cloud applications. Libraries and SDKs support a common set of S3 API functions for connecting new applications to scalable cloud storage and integrating your data into other services on the IBM Watson and Cloud Platform.

Data storage classes for Active, Less Active, Archive and Dynamic workloads

Choose storage classes for frequently accessed data, occasionally accessed data and long-term data retention with Standard, Vault, and Cold Vault. Or, choose Flex class for dynamic data access needs that fluctuate month to month.

IAM Policies - Bucket level access management

IBM Identity and Access Management (IAM) integration allows for granular access control at the bucket level using role-based policies.

Encryption management

All data is encrypted at-rest and in-flight by default. Keys are automatically managed by default, but can optionally be self-managed or managed using IBM Key Protect*. (*Key Protect is only available for buckets created in the US South (Dallas) and EU GB (London) regions.)

Regional and Cross Region resiliency options

Select the best resiliency option for your data. Choose "Cross Region" to store unstructured data across three regions, or choose "Regional" resiliency to store your data within a single region.

Pricing Plan: Monthly Prices shown above reflect the: United States

PLAN	FEATURES	PRICING
<input checked="" type="radio"/> Lite	<div>1 COS Service Instance</div> <div>Storage up to 25 GB/mo.</div> <div>Up to 20,000 GET requests/mo.</div> <div>Up to 2,000 PUT requests/mo.</div> <div>Up to Data Retrieval 10 GB/mo.</div> <div>Up to 5GB Public Outbound</div> <div>Applies to aggregate total across all storage bucket classes</div>	Free
<input type="radio"/> Standard	There is no minimum fee, so you pay only for what you use.	-

Cancel

Create

Cloud Object Storage

ExistingNew

Cloud Object Storage

IBM Cloud Object Storage is a highly scalable service, designed for high durability, resiliency and security. Store, manage and access your data via our self-service portal and RESTful APIs. Connect applications directly to Cloud Object Storage use other IBM Cloud Services with your data.

Confirm Creation

Plan

Lite

Resource group

default

Service name

cloud-object-storage-pk

Cancel

Confirm

after adding object storage, click refresh

New project

my 1st Watson studio machine learning project

2955

Choose project options

☐ Restrict who can be a collaborator ⓘ

Project will include integration with [Cloud Object Storage](#) for storing project assets.

Define storage

① Select storage service

Add

Add an object storage instance and then return to this page and click Refresh.

② Refresh

After adding object storage, **click** create

Choose project options

☐ Restrict who can be a collaborator ⓘ

Project will include integration with [Cloud Object Storage](#) for storing project assets.

Storage

cloud-object-storage-pk

Cancel

Create

After creating project, you will be presented with overview page. **click Assets**

The screenshot shows the 'Overview' page of a project named 'Msamido_Machine_Learning'. The 'Assets' tab is highlighted in the top navigation bar. The page displays statistics: 0 Assets, 0 Bookmarks, and 1 Collaborator. Below these, there are sections for 'Date created' (Jun 28 2018), 'Description' (my 1st watson studio machine learning project), 'Storage' (0% of 25 GB used), and 'Collaborators' (View all (1)). A 'Recent activity' section is also present, showing a placeholder for alerts.

My Projects / Msamido_Machine_Learning

+ Add to project

Overview **Assets** Environments Bookmarks Deployments Access Control Settings

Msamido_Machine_Learning

Last Updated: Jun 28 2018

[Readme](#)

0 Assets

0 Bookmarks

1 Collaborators

Date created
Jun 28 2018

Description
my 1st watson studio machine learning project

Storage
0% of 25 GB used

Collaborators [View all \(1\)](#)

Recent activity

Alerts related to this project will show here when the project is active.

From assets, you scroll down, to create a new model

The screenshot shows the 'Assets' page of the project. The 'Assets' tab is highlighted in the top navigation bar. The page displays a search bar and four sections: 'Data assets', 'Visual recognition models', 'Notebooks', and 'Dashboards'. Each section has a table with columns for NAME, TYPE, SERVICE, CREATED BY, LAST MODIFIED, and ACTIONS. The 'Visual recognition models' section also includes a 'New visual recognition model' button. The 'Notebooks' section includes a 'New notebook' button. The 'Dashboards' section includes a 'New dashboard' button. A 'Load' button is visible in the top right corner.

My Projects / Msamido_Machine_Learning

+ Add to project

Overview **Assets** Environments Bookmarks Deployments Access Control Settings

What assets are you looking for?

Data assets

NAME	TYPE	SERVICE	CREATED BY	LAST MODIFIED	ACTIONS
You currently have no data assets					

Visual recognition models [New visual recognition model](#)

NAME	MODEL ID	SERVICE INSTANCE	LAST MODIFIED	ACTIONS
You currently have no visual recognition models				

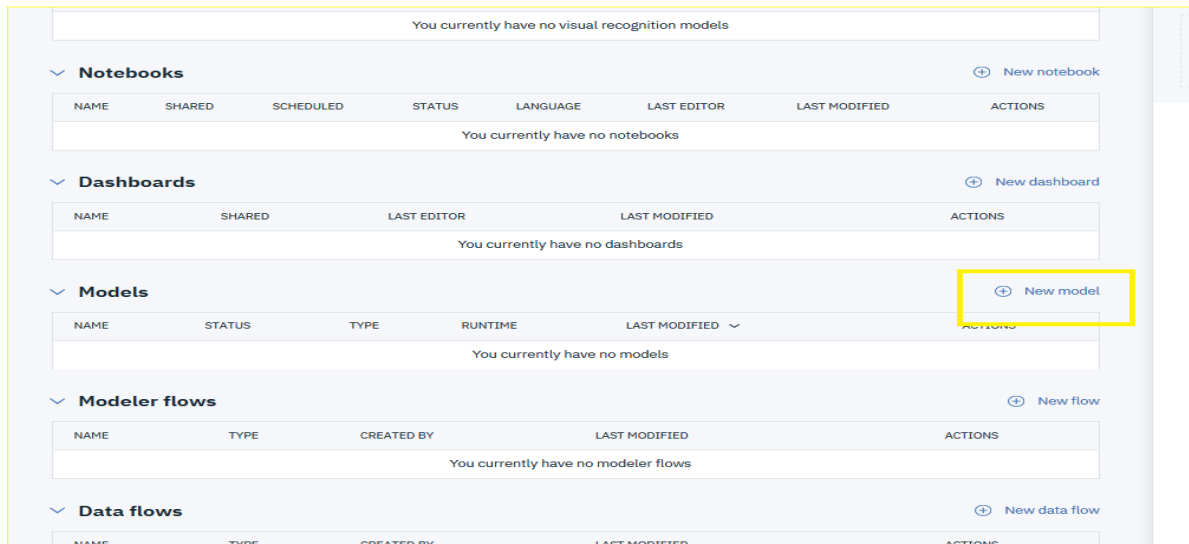
Notebooks [New notebook](#)

NAME	SHARED	SCHEDULED	STATUS	LANGUAGE	LAST EDITOR	LAST MODIFIED	ACTIONS
You currently have no notebooks							

Dashboards [New dashboard](#)

NAME	SHARED	LAST EDITOR	LAST MODIFIED	ACTIONS
You currently have no dashboards				

Drop files here or browse for files to upload.



After creating a new model, you will be presented with the below page, to update model requirements.

From the model, U will realize that you need two services which are:

- Machine learning instance
- Spark analytics instance

The 'New model' page is divided into two main columns. The left column, 'Define model details', contains a 'Name' field (highlighted with a yellow box) and a 'Description' text area (circled in blue). The right column, 'Select model type', has radio buttons for 'Model builder' (selected), 'From file', and 'From sample'. Below this is the 'Spark Service' section, which includes a 'Reload' button. At the bottom left, the 'Machine Learning Service' section is visible, with a red arrow pointing to it. The bottom right features 'Automatic' and 'Manual' tabs, and a footer with 'Cancel' and 'Create' buttons.

Creating a ML service

setting Up machine learning service

Machine Learning

Existing **New**

Machine Learning

IBM Watson Machine Learning is a full-service IBM Cloud offering that makes it easy for developers and data scientists to work together to integrate predictive capabilities with their applications. The Machine Learning service is a set of REST APIs that you can call from any programming language to develop applications that make smarter decisions, solve tough problems, and improve user outcomes.

Features

- Machine Learning features**
Take advantage of machine learning model management (continuous learning system) and deployment (online, batch, streaming). Select any of widely supported machine learning frameworks: TensorFlow, Keras, Caffe, PyTorch, Spark MLlib, scikit-learn, xgboost and SPSS.
- Wide choice of interfaces**
Use the convenient Java interface and Python client to manage your artifacts. Extend your application with artificial intelligence through the Watson Machine Learning REST API.
- Integration with Watson Studio**
Create and train machine learning models with the best tools and the latest expertise in a social environment built by and for data scientists.

Pricing Plan: Monthly Process shown above reflect the: **United States**

PLAN	FEATURES	PRICING
<input checked="" type="radio"/> Lite	Service Instance (5 models per instance) 5,000 predictions 50 capacity unit-hours: Compute Tier: k80 = 2 capacity units for 1 training hour Compute Tier: k80d2 = 4 capacity units for 1 training hour Compute Tier: k80d4 = 8 capacity units for 1 training hour Otherwise 1 capacity unit for 1 computation hour Max 8 k80 GPUs (Deep Learning Training)	Free
<input type="radio"/> Professional	Service Instance 2,000,000 predictions included and then billed per 1,000 predictions 1,000 capacity unit-hours included and then billed per capacity unit-hour: Compute Tier: k80 = 2 capacity units for 1 training hour Compute Tier: k80d2 = 4 capacity units for 1 training hour Compute Tier: k80d4 = 8 capacity units for 1 training hour Compute Tier: p200 = 5 capacity units for 1 training hour Compute Tier: p200d2 = 10 capacity units for 1 training hour Compute Tier: v200 = 8 capacity units for 1 training hour Compute Tier: v200d2 = 16 capacity units for 1 training hour Otherwise 1 capacity unit for 1 computation hour Unlimited elastic compute environments	
<input type="radio"/> Standard	Predictions Capacity unit-hours: Compute Tier: k80 = 2 capacity units for 1 training hour Compute Tier: k80d2 = 4 capacity units for 1 training hour Compute Tier: k80d4 = 8 capacity units for 1 training hour Compute Tier: p200 = 5 capacity units for 1 training hour Compute Tier: p200d2 = 10 capacity units for 1 training hour Compute Tier: v200 = 8 capacity units for 1 training hour Compute Tier: v200d2 = 16 capacity units for 1 training hour Otherwise 1 capacity unit for 1 computation hour Unlimited elastic compute environments	

The lite plan instance of the IBM Watson Machine Learning service provides you with a maximum of 5 deployed models, 5,000 predictions per month, and 50 capacity unit-hours per month during which model can be trained, evaluated, and deployed to be available to accept prediction events, with a minimum of 1 minute per training job.

Cancel **Create**

Compute Tier: k80 = 2 capacity units for 1 training hour
Compute Tier: k80d2 = 4 capacity units for 1 training hour
Compute Tier: k80d4 = 8 capacity units for 1 training hour
Compute Tier: p200 = 5 capacity units for 1 training hour
Compute Tier: p200d2 = 10 capacity units for 1 training hour
Compute Tier: v200 = 8 capacity units for 1 training hour
Compute Tier: v200d2 = 16 capacity units for 1 training hour
Otherwise 1 capacity unit for 1 computation hour
Unlimited elastic compute environments

Confirm Creation

Plan
Lite

Resource group
default

Service name
Msamido_ML

Cancel **Confirm**

Define model details

Name

Msamido_Model

87

Description

My 1st machine learning model and I can't even write a single line of code.

226

Machine Learning Service

No Machine Learning service instances associated with your project.

[Associate a Machine Learning service instance](#) with your project on the project settings page, then click the reload button below to refresh the instances available for association with your new model builder instance.

Reload

Select model type

☒ Model builder

☐ From file

☐ From sample

Spark Service

No Spark instances associated with your project.

[Associate an IBM Analytics for Apache Spark instance](#) with your project on the project settings page, then click the reload button below to refresh the instances available for association with your new model builder instance.

Reload

Automatic

Prepare my data and create a model automatically

Manual

Let me prepare my data and select which models to train

Need something more flexible? Create a [notebook](#) or design a [Modeler flow](#)

Cancel

Create

Click reload after creating machine learning service

Define model details

Name

Msamido_Model

87

Description

My 1st machine learning model, and i cant even write a sigle line of code.

226

Machine Learning Service

Msamido_ML

Select model type

☒ Model builder

☐ From file

☐ From sample

Spark Service

No Spark instances associated with your project.

[Associate an IBM Analytics for Apache Spark instance](#) with your project on the project settings page, then click the reload button below to refresh the instances available for association with your new model builder instance.

Reload

Automatic

Prepare my data and create a model automatically

Manual

Let me prepare my data and select which models to train

Need something more flexible? Create a [notebook](#) or design a [Modeler flow](#)

Creating an Apache service

- setting Apache spark service

Existing

New

Apache Spark

Apache Spark is an open source cluster computing framework optimized for extremely fast and large scale data processing, which you can access via the newly integrated notebook interface IBM Analytics for Apache Spark. You can connect to your existing data sources or take advantage of the on-demand big data optimization of Object Storage. Spark plans are based on the maximum number of executors available to process your analytic jobs. Executors exist only as long as they're needed for processing, so you're charged only for processing done.

Features

Incredibly Fast
Apache Spark delivers 100x the performance of Apache Hadoop for certain workloads because of its advanced in-memory computing engine.

Easy to Use and Powerful
Apache Spark's Streaming and SQL programming models backed by MLlib and GraphX make it incredibly easy for developers and data scientists to build apps that exploit machine learning and graph analytics. Because the service is 100% compatible with Apache Spark, developers can build their apps and run them against the IBM managed service to benefit from operational, maintenance, and hardware excellence.

Convenient Data Storage
Object Storage enables a convenient way to upload your data from a file for immediate use by your Spark instance. You can set up Object Storage directly from the Spark service interface.

Pricing Plan: Monthly Process shown above reflect the: United States

PLAN	FEATURES	PRICING
<input checked="" type="radio"/> Lite	2 Spark Executors	Free
An entry level plan to run programs using up to 2 Spark executors		
<input type="radio"/> Reserved Enterprise	30 Spark Executors	-

Cancel

Create

Computing frame
sing, which yo
analytics for A
ke advantage
plans are base
our analytic jo
ng, so you're c

Confirm Creation

Organization: mapurunyane.samuel.tolamo@ibm.com

Plan
Lite

Space
Dev

Service name
Msamido_Spark_service

Cancel

Confirm

: Data Storage
e enables a convenient v
ata from a file for immed
: instance. You can set up
tly from the Spark service

FEATURE
2 Spark
ns using up to 2 Spark executors

Reload after adding apache spark service

Define model details

Name

Msamido_Model

Description

My 1st machine learning model, and i cant even write a sigle line of code.

Machine Learning Service

Msamido_ML

Select model type

Model builder

From file

From sample

Spark Service

No Spark instances associated with your project.

Associate an IBM Analytics for Apache Spark instance with your project on the project settings page, then click the reload button below to refresh the instances available for association with your new model builder instance.

Reload

Automatic

Prepare my data and create a model automatically

Manual

Let me prepare my data and select which models to train

Need something more flexible? Create a [notebook](#) or design a [Modeler flow](#)

Cancel

Create

Button is now enabled, **click** Create, to create Your machine learning model.

New model

Define model details

Name

Msamido_Model

Description

My 1st machine learning model, and i cant even write a sigle line of code.

Machine Learning Service

Msamido_ML

Select model type

Model builder

From file

From sample

Spark Service

Msamido_Spark_service

Automatic

Prepare my data and create a model automatically

Manual

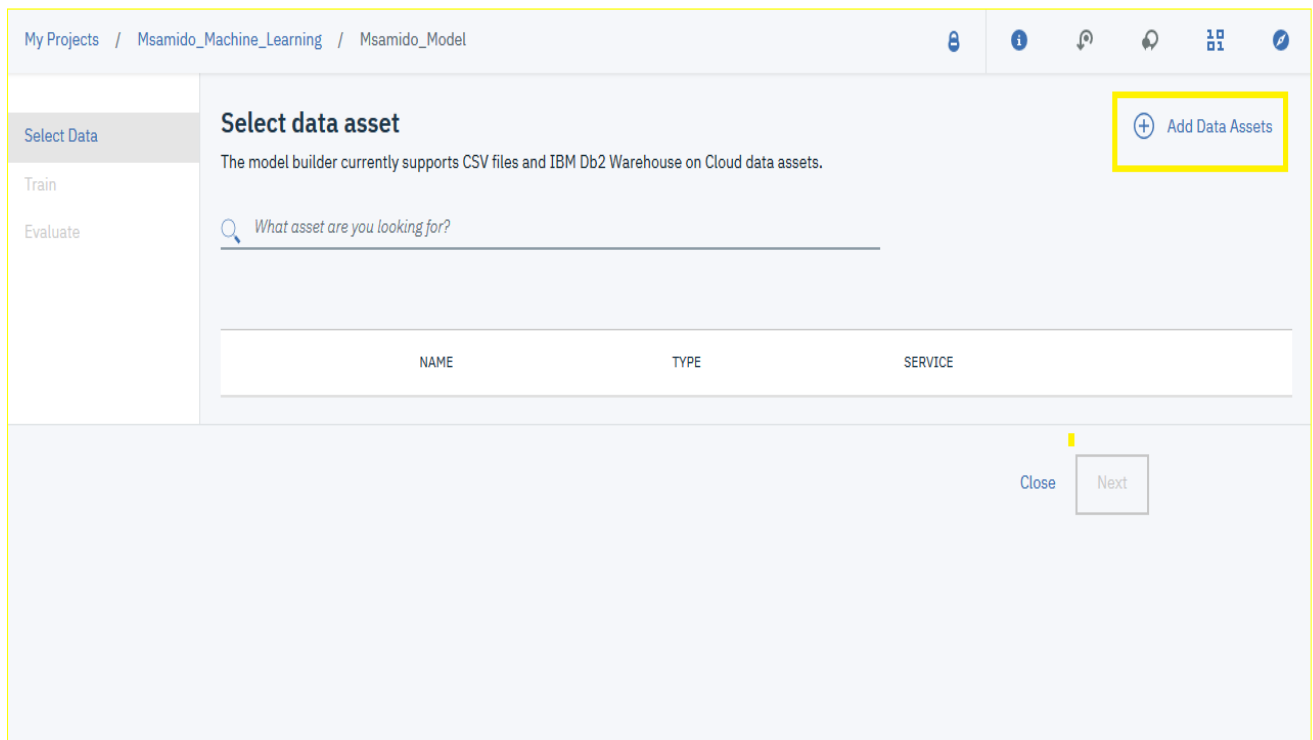
Let me prepare my data and select which models to train

Need something more flexible? Create a [notebook](#) or design a [Modeler flow](#)

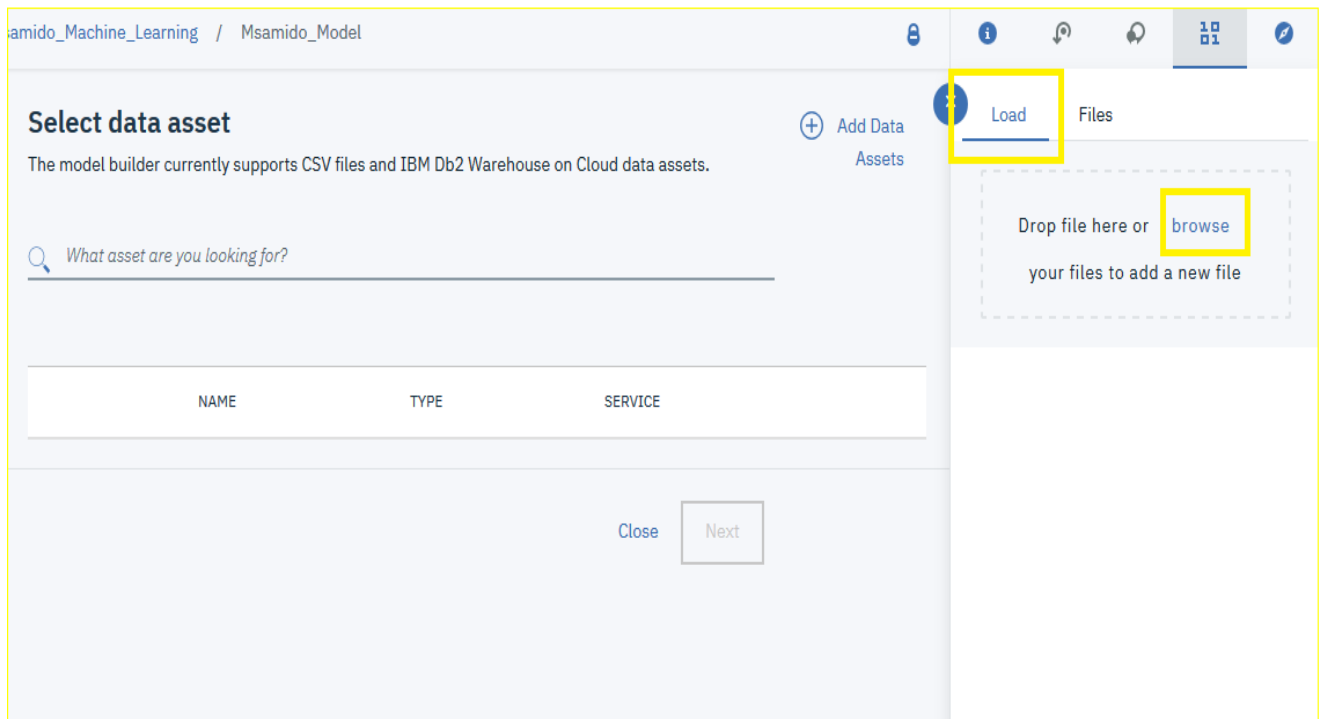
Cancel

Create

Adding dataset to a model



Browse dataset from your local machine.



create a machine learning model from the dataset, **note** in this project we will create a linear regression model since our data is continuous.

Select a technique

Column value to predict (Label Col)
MV (Decimal)

Feature columns
All (default)

Binary Classification
Classify new data into defined categories based on existing data. Choose if your label column contains two distinct categories.

Multiclass Classification
Classify new data into defined categories based on existing data. Choose if your label column contains a discrete number of categories.

Regression
Predict values from a continuous set of values. Choose if your label column contains a large number of values.

Validation Split
Train: 60 Test: 20 Holdout: 20

Close Previous **Next**

Wait for a model to train successfully, then **click** save

Select model

ESTIMATOR TYPE	STATUS	ROOT MEAN SQUARED ERROR	MEAN SQUARED ERROR	R2	EXPLAINED VARIANCE	MEAN ABSOLUTE ERROR	LAST EVALUATION
<input checked="" type="radio"/> LinearRegression	Trained & Evaluated	4.78513	22.89744	0.68217	45.71322	3.12736	28 Jun 2018, 3:05 PM

Close Previous **Save**

Msamido_Model

[Overview](#)[Evaluation](#)[Deployments](#)

Summary

Machine learning service	Msamido_ML
Model Type	wml-1.1
Runtime environment	spark-2.1
Training date	28 Jun 2018, 3:07 PM
Label column	MV
Latest version	a0462c07-d519-44b5-be52-af53b99b9212
Model builder details	View

Input Schema

COLUMN	TYPE
CRIM	decimal(31,6)
ZN	decimal(31,6)

Model deployment

We are going to deploy our model as web service.

Create Deployment

Define deployment details

Name
Msamido

Description
Deploy

Deployment type
☒ Web service
☐ Batch prediction
☐ Realtime streaming prediction

Cancel Save

Msamido_Model

Overview Evaluation Deployments

+ Add Deployment

NAME	STATUS	DEPLOYMENT TYPE	ACTIONS
Msamido	DEPLOY_SUCCESS	Web Service	