

Cloud Pak for Data JumpStart Training

Welcome to the training.

Please try and login to the platform, <https://cp4d.apps.ocp.dmp.qa.bhp.com/> using your BHP user id and Windows password. *Let me know immediately if you are unable to login.*

Download today's training material from,

https://github.com/randyphoaibm/cpd-jumpstart/raw/main/CPD_JumpStart_Training.pdf

https://raw.githubusercontent.com/randyphoaibm/cpd-jumpstart/main/german_credit.csv

Cloud Pak for Data JumpStart Training

Randy Phoa
Senior Data Scientist
IBM Data Science Elite Team



Agenda

1. Introduction to Cloud Pak for Data
2. Projects
3. Prepare data
4. Build models
5. Deploy models
6. Resources

1. Introduction to Cloud Pak for Data

Cloud Pak for Data

Unified, modular, deployable anywhere

App Developers | Business Analysts | Data Engineers | Data Stewards | Data Scientists | Business Users

Integrated User Experience

Extensible: APIs, partner ecosystem, accelerators, and solutions

Collect

- Data virtualization
- SQL and NoSQL databases
- Event ingestion
- Streaming Analytics
- Apache Spark

Organize

- Data transformation
- Data quality and classification
- Policies and rules
- Data cataloging
- Self-service discovery and search

Analyze

- Data science and visualization
- AutoAI
- Model trust and explainability
- Model optimization

Infuse

- Business reporting and visualization
- Financial planning and analysis
- Cloud native AI services
- RegTech and Financial Crimes Insight

Cloud Pak core services
Security, Administration, Operations

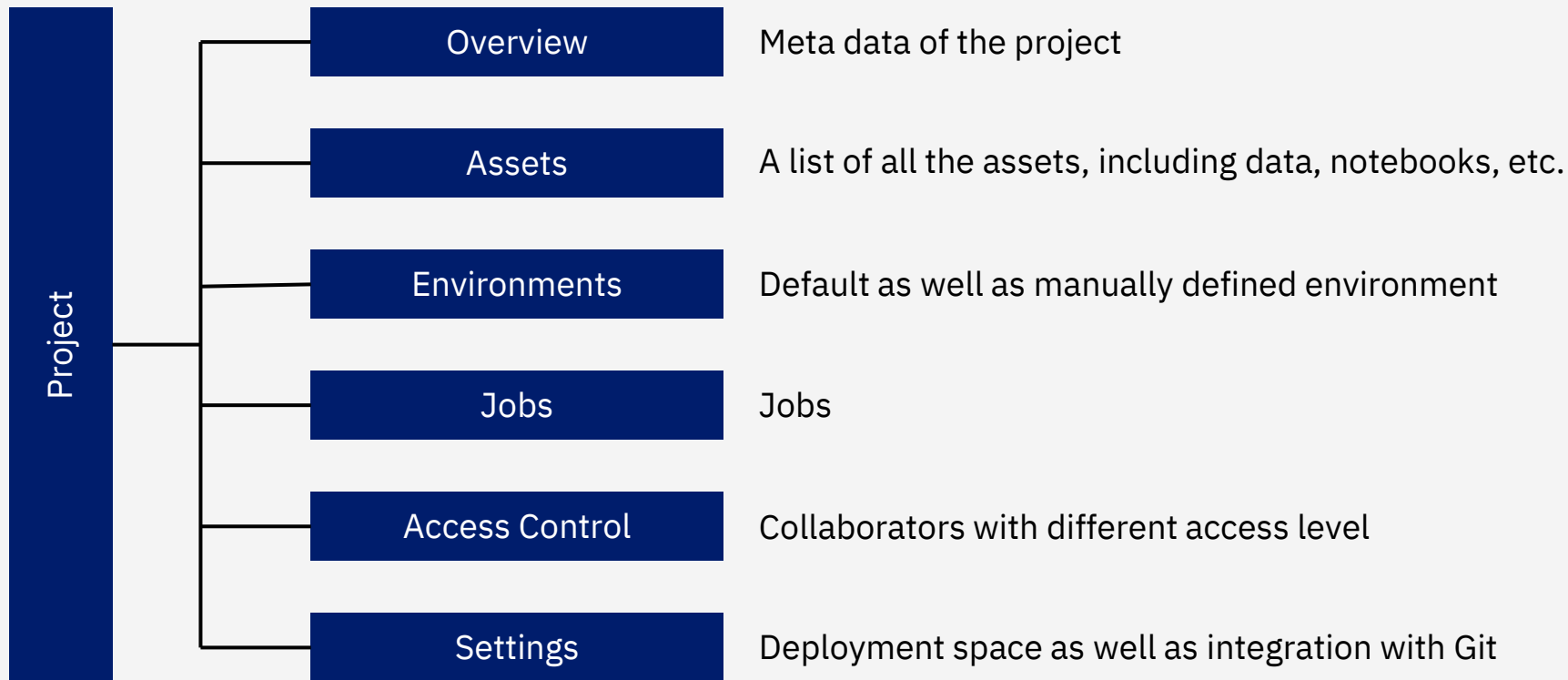
Red Hat OpenShift

IBM Cloud | Amazon Web Services | Microsoft Azure | Google Cloud | Hyperconverged system

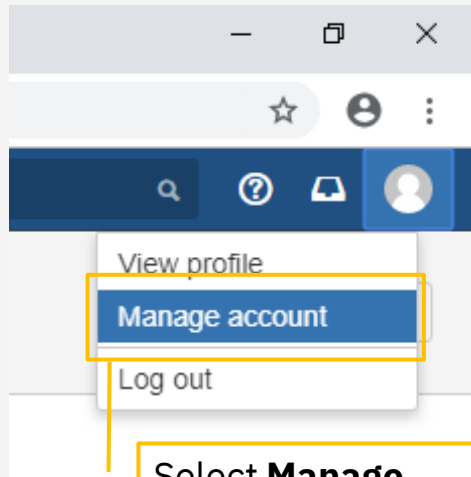
2. Projects

- Project structure
- Git Setup
- Create a project
- Add collaborators
- Custom environments

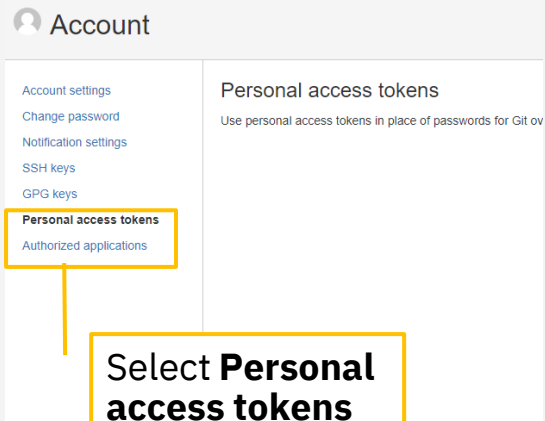
Project structure



Git Setup – Create access token from BitBucket



Select **Manage account** from menu



Select **Personal access tokens**

Create a personal access token

Use personal access tokens in place of passwords for Git over HTTPS, or to authenticate when using the Bitbucket Server REST API. [Learn more.](#)

Token details

Token name CPD Token

Permissions

Tokens are like another password, so their permissions will default to the level of access you have. Because of this, it is recommended that you restrict the token's permission to the level it will need.

Projects Admin

Repositories Admin (inherited)

Summary

This personal access token will allow the supplied third-party application to:

- ✓ Perform pull request actions
- ✓ Update repository settings and permissions
- ✓ Update project settings and permissions
- ✓ Push, pull, clone, and fork repositories
- ✓ Create repositories

Create

Cancel

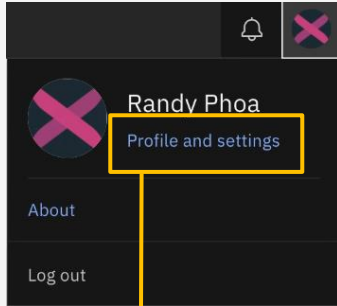
New personal access token created

You will not be able to view this token again.

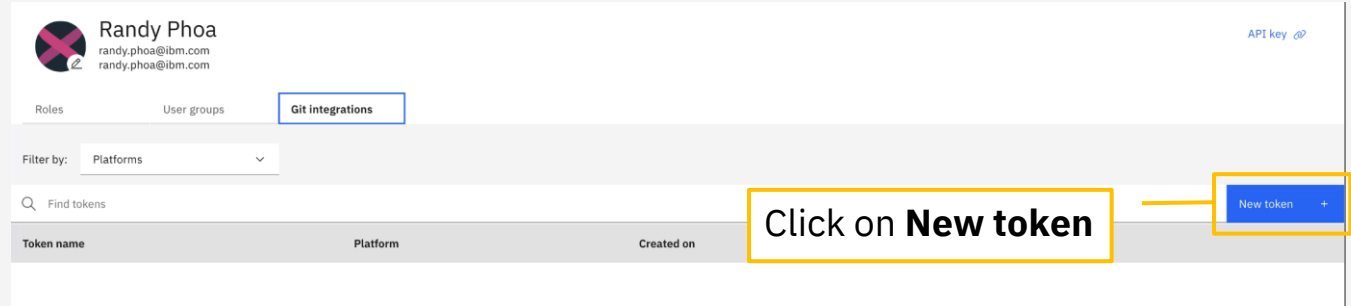
Copy

Continue

Git Setup – Add access token to Cloud Pak for Data

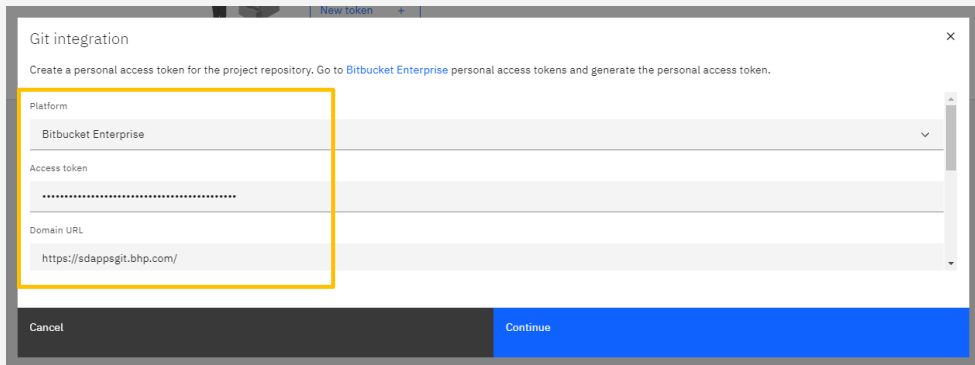


Select **Profile and settings** from menu



Click on **New token**

Git Setup – Add access token to Cloud Pak for Data



Git integration

Create a personal access token for the project repository. Go to [Bitbucket Enterprise](#) personal access tokens and generate the personal access token.

Platform

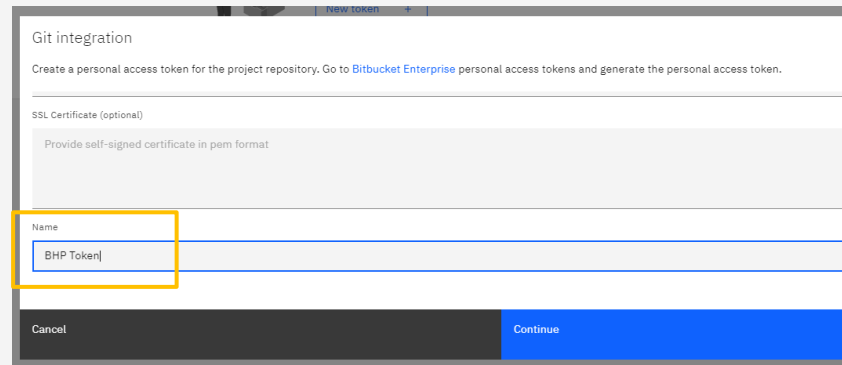
Bitbucket Enterprise

Access token

Domain URL

https://sdappsgit.bhp.com/

Cancel Continue



Git integration

Create a personal access token for the project repository. Go to [Bitbucket Enterprise](#) personal access tokens and generate the personal access token.

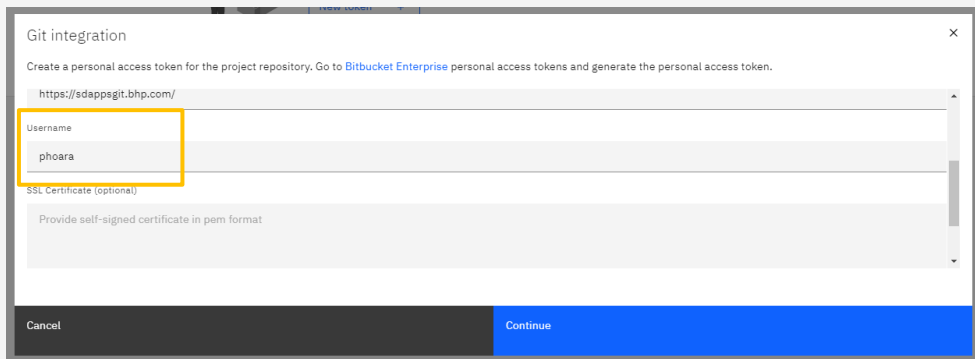
SSL Certificate (optional)

Provide self-signed certificate in pem format

Name

BHP Token

Cancel Continue



Git integration

Create a personal access token for the project repository. Go to [Bitbucket Enterprise](#) personal access tokens and generate the personal access token.

https://sdappsgit.bhp.com/

Username

phoara

SSL Certificate (optional)

Provide self-signed certificate in pem format

Cancel Continue

Fill in required details.
Click **Create** when done.
Remember to save your access token securely.

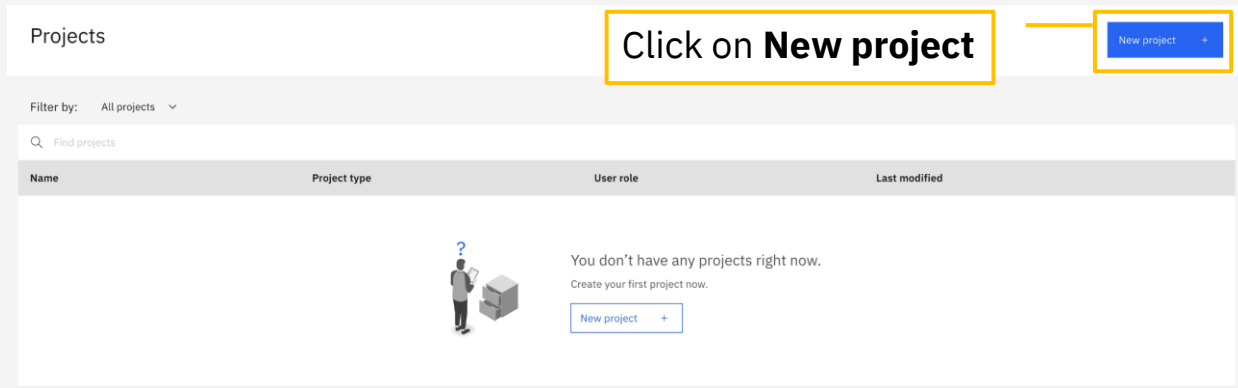
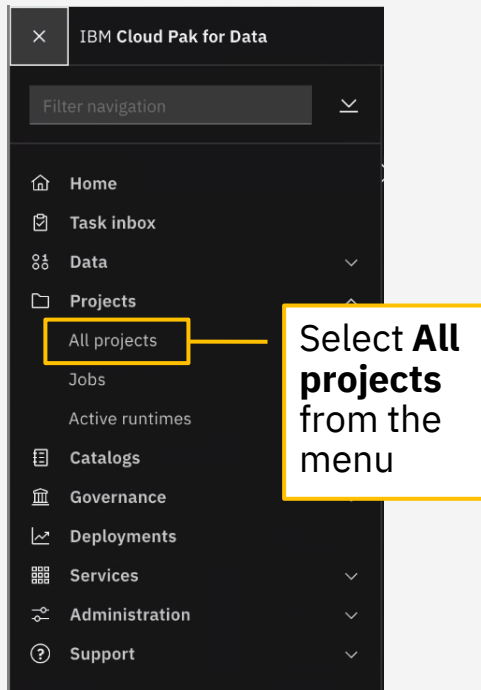
Exercise

- Create your BitBucket access token
- Add your access token to Cloud Pak for Data

Lesson

- Ensure you know how to create a BitBucket access token and how to add it to Cloud Pak for Data

Create a project



Create a project

Create a new project

Select a project type

☒ Analytics project

☐ Data transformation project

☐ Data quality project

Cancel

Next

Select **Analytics project**

Click on **Next**

Create a project

Choose whether to create an empty project or to preload your project with data and analytical assets. Add collaborators and data, and then choose the right tools to accomplish your goals. Add services as necessary.



Create an empty project

Add the data you want to prepare, analyze, or model. Choose tools based on how you want to work: write code, create a flow on a graphical canvas, or automatically build models.

USE TO

Prepare and visualize data
Analyze data in notebooks
Train models



Create a project from a file

Get started fast by loading existing assets. Choose a project file from your system or a Git repository.

USE TO

Learn by example
Build on existing work
Run tutorials
Integrate with Git

Select **Create an empty project**

Create a project – without Git integration

New project

Define project details

Name

CPD JumpStart

Description

Project description

Choose project options

☐ Integrate this project with Git ⓘ

☐ Mark as sensitive ⓘ

☐ Log all project activities ⓘ

Fill in the required details

Create a project – with Git integration

New project

Define project details

Name

CPD JumpStart

Description

Project description

Choose project options

☒ Integrate this project with Git ⓘ

☐ Mark as sensitive ⓘ

☐ Log all project activities ⓘ

Specify the Git repository for on-demand synchronization. All the existing files are deleted from the repository during project creation. If necessary, download or move files before continuing. Afterward, click the sync icon at any time.

Token

BHP Token

[New Token](#) +

Repository URL ✓

https://phoara@sdappsgit.bhp.com/scm/~phoara/cpd-jumpstart.git

Branch ✓

master

☒ Edit notebooks and Python scripts with the JupyterLab IDE ⓘ

Fill in the required details

Add collaborators

Select **Access Control** from the menu

Select **Add collaborators**

Select **Role**

Click **Invite** when done

Search by username or email

The image shows a two-step process for adding collaborators in BHP JumpStart. The first screenshot shows the main interface with the 'Access Control' menu item highlighted in the top navigation bar. A callout points to the 'Add collaborators' button in the top right. The second screenshot shows the 'Add collaborators' dialog box. A callout points to the search input field where 'Denise.Hernandez@ibm.com' is entered. Another callout points to the 'Role' dropdown menu, which is set to 'Editor'. A third callout points to the 'Invite' button at the bottom right of the dialog. A table of collaborators is also visible in the dialog.

Name	Email	Permission	Status
Randy Phoa	randy.phoa@ibm.com	Admin	Active

Name	Email	Role
Derek Wk Chan	derek.chan@au1.ibm.com	Admin

Roles

Role	Permissions	Services that contribute permissions	Service that creates the role
Administrator	<ul style="list-style-type: none"> – Administer platform – Create service instances 	Cloud Pak for Data control plane	Cloud Pak for Data control plane
	<ul style="list-style-type: none"> – Integrate and transform data 	DataStage® Edition	
	<ul style="list-style-type: none"> – Analyze data quality – Discover assets – Import metadata – Manage catalogs – Access governance artifacts – Manage governance categories – Manage governance workflows – Manage information assets – Manage data protection rules 	Watson™ Knowledge Catalog	
Business Analyst	<ul style="list-style-type: none"> – Access information assets view – View data quality 	Watson Knowledge Catalog	Watson Knowledge Catalog
Data Engineer	<ul style="list-style-type: none"> – Create service instances 	Cloud Pak for Data control plane	DataStage Edition or Watson Knowledge Catalog
	<ul style="list-style-type: none"> – Integrate and transform data 	DataStage Edition	
	<ul style="list-style-type: none"> – Access catalogs – Discover assets – Import metadata – Access governance artifacts – Manage information assets – View data quality 	Watson Knowledge Catalog	
Data Quality Analyst	<ul style="list-style-type: none"> – Access catalogs – Analyze data quality – Discover assets – Import metadata – Access governance artifacts – Manage information assets 	Watson Knowledge Catalog	Watson Knowledge Catalog
Data Scientist	<ul style="list-style-type: none"> – Access catalogs 	Watson Knowledge Catalog	Watson Knowledge Catalog

Role	Permissions	Services that contribute permissions	Service that creates the role
Data Steward	<ul style="list-style-type: none"> – Access catalogs – Discover assets – Import metadata – Access governance artifacts – Manage information assets – View data quality – Manage data protection rules 	Watson Knowledge Catalog	Watson Knowledge Catalog
Developer	<ul style="list-style-type: none"> – Create service instances 	Cloud Pak for Data control plane	Watson Knowledge Catalog
	<ul style="list-style-type: none"> – Access catalogs 	Watson Knowledge Catalog	
User	<ul style="list-style-type: none"> – Access assigned services 	Cloud Pak for Data control plane	Cloud Pak for Data control plane

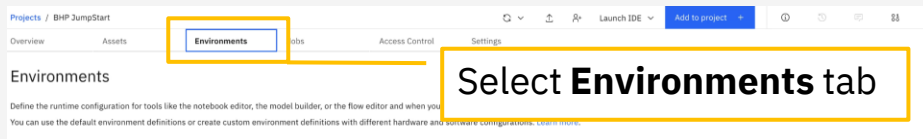
Exercise

- Create a project without Git integration
- Add a few of your colleagues from your team/department
- Ensure they can see your project (and vice versa if they added you)

Lesson

- Ensure you know how to create a project and add collaborators

Custom environments



Which environment are you looking for?

Active environment runtimes

Name	Hardware configuration	Tool	Started at	Owner
No environments are currently active.				

Click on **New environment definition**

New environment definition +

New environment

Define environment details

Name

Randy Env

Description (optional)

Environment description

Define configuration

Type

☒ Default ☐ GPU

Hardware configuration

Specify: 2 vCPU and 8 GB RAM

Software version

Default Jupyterlab 3.7

Customization

```
# Modify the following content to add a software customization to an environment.  
# To remove an existing customization, delete the entire content and click Apply.  
# The customizations must follow the format of a conda environment yaml file.
```

```
# Add conda channels below defaults, indented by two spaces and a hyphen.  
channels:  
  - defaults  
  
# To add packages through conda or pip, remove the # on the following line.  
dependencies:  
  
# Add conda packages here, indented by two spaces and a hyphen.  
# Remove the # on the following line and replace sample package name with your  
package name:  
  - statsmodels  
  
# Add pip packages here, indented by four spaces and a hyphen.  
# Remove the # on the following lines and replace sample package name with your  
package name.  
  - pip:  
    - lightgbm
```

Cancel Apply

Add required packages
and click **Apply**

3. Prepare data

- Manage assets
- Upload files
- Create data connection

Open the **CPD JumpStart Training** project

Projects

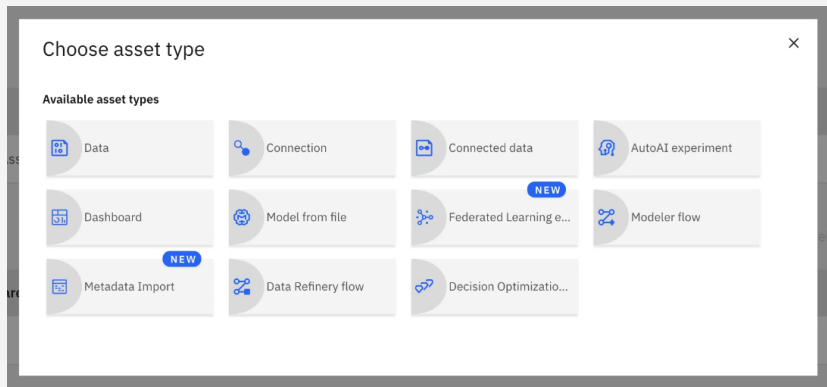
[New project](#) +

Filter by: All projects ▾

🔍 Find projects



Name	Project type	User role	Last modified	
CPD JumpStart Training	Analytics	Admin	Dec 2, 2021 2:44 PM	⋮

Manage assets



▼ Data assets

0 assets selected.

<input type="checkbox"/>	Name	Type	Created by	Last modified	↓
<input type="checkbox"/>	 payload.txt	Data Asset	Randy Phoa	Nov 29, 2021, 11:11 AM	

▼ Notebooks

Name	Shared	Scheduled	Status	Language	Last editor	Last modified
 Test				Python 3.7	Randy Phoa	Nov 29, 2021, 11:11 AM

- Publish to Catalog
- Refine
- Download
- Promote
- Remove

Various options for each asset

Download sample data

Access the below link and download the file to your local computer by clicking on File -> Save As.

https://raw.githubusercontent.com/randyphoaibm/cpd-jumpstart/main/german_credit.csv

Upload data

Local Computer -> Watson Studio

Projects / BHP JumpStart

Overview **Assets** Environments Jobs Access Control Settings

What assets are you looking for?

✓ Data assets

0 assets selected.

<input type="checkbox"/>	Name	Type	Created by	Last modified
<input type="checkbox"/>	CSV german_credit.csv	Data Asset	Randy Phoa	Nov 29, 2021, 11:46 AM

Data

Load **Files** Catalog

Drop files here or [browse](#) for files to upload.

Drag and drop downloaded sample data to the upload area

Exercise

- Download the sample data and add a suffix using your first name, last name and 4 random numbers, i.e, (german_credit_RandyP-8650.csv)
- Upload the file by dragging and dropping to the upload area
- Ensure that you see the file appear in Data Assets

Lesson

- Understand how files are uploaded and shared with collaborators in the project

Create data connection

Add connection


























Create a new connection or select an existing connection from the list of platform connections.

[Supported connection types](#) 





















New

From platform

IBM

- | | | | |
|---|--|--|---|
|  Analytics Engine HDFS |  Data Virtualization Manager for z/OS |  Db2 Hosted |  Informix |
|  Cloud Object Storage |  Databases for PostgreSQL |  Db2 on Cloud |  Netezza (PureData System for Analytics) |
|  Cloud Object Storage (infrastructure) |  Db2 |  Db2 Warehouse |  Planning Analytics |
|  Cloudant |  Db2 Big SQL |  HDFS via Execution Engine for Hadoop |  Storage volume |
|  Cognos Analytics |  Db2 Event Store |  Hive via Execution Engine for Hadoop | |
|  Compose for MySQL |  Db2 for i |  IBM SPSS Analytic Server | |
|  Data Virtualization |  Db2 for z/OS |  Impala via Execution Engine for Hadoop | |

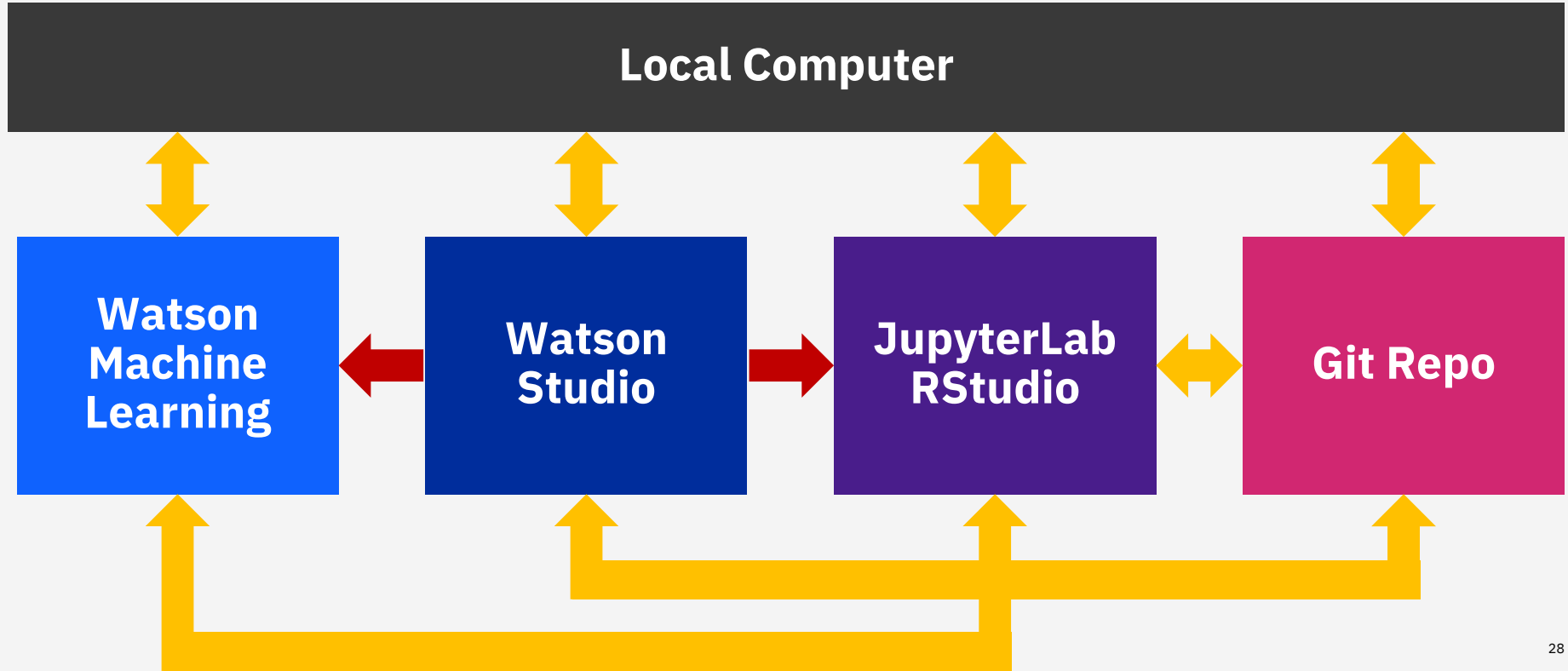
Third-party

- | | | | |
|---|--|---|--|
|  Amazon RDS for MySQL |  Dropbox |  Microsoft Azure Data Lake Store |  Salesforce.com |
|  Amazon RDS for PostgreSQL |  Elasticsearch |  Microsoft Azure SQL Database |  SAP ASE |
|  Amazon Redshift |  FTP |  Microsoft SQL Server |  SAP HANA |
|  Amazon S3 |  Google BigQuery |  MinIO |  SAP IQ |
|  Apache Cassandra |  Google Cloud Storage |  MongoDB |  SAP OData |

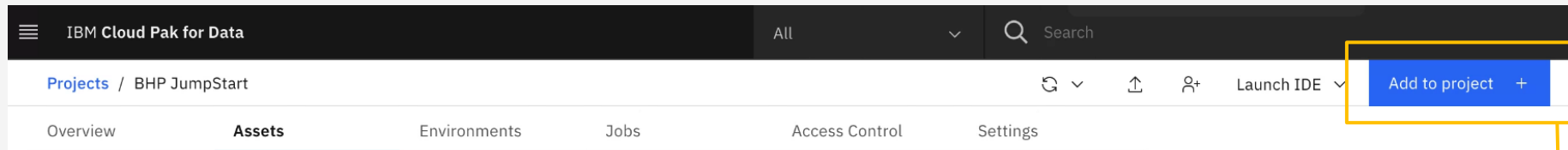
4. Develop models

- AutoAI
- RStudio
- JupyterLab

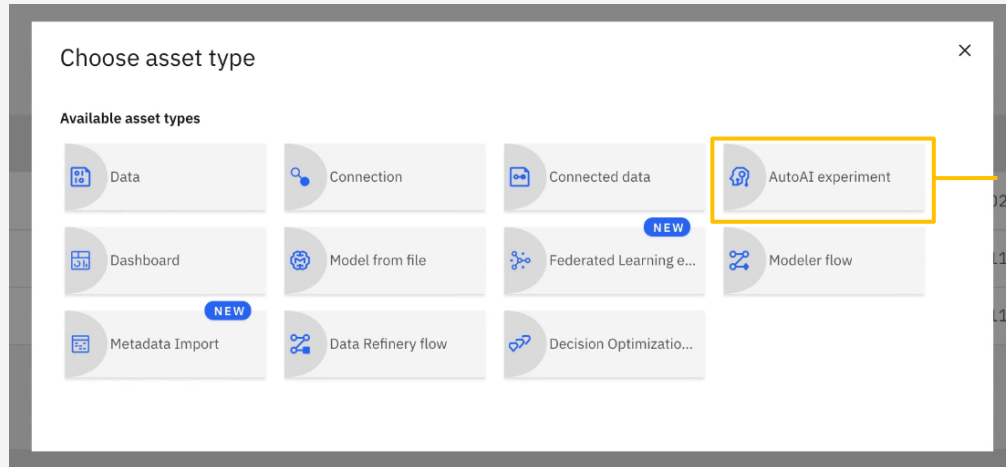
Data flow among the various environments



AutoAI – Create AutoAI experiment



Click on **Add to project**



Select **AutoAI experiment**

AutoAI – Configure

Define details
Name *


Description

Associate services
Compute configuration * ⓘ

4 vCPU and 16 GB RAM

Fill in the required details

Cancel Create



Add data source
Drop or browse for a csv file.

Browse or **Select from project**

Add data source from **Select from project**

Select a .csv file from the list of available data assets for this project.

Search for a CSV asset

File name	Size
<input type="checkbox"/> data.csv	0.00 MB
<input checked="" type="checkbox"/> german_credit.csv	0.69 MB

Items per page: 5 1~2 of 2 items 1 of 1 pages

Cancel Select asset

AutoAI – Settings

Configure details

What do you want to predict?

Prediction column ⓘ

Risk

ExistingCreditsCount

STR Job

INT Dependents

STR Telephone

STR ForeignWorker

STR Risk

Prediction column: Risk

Select **Risk** from Prediction Column

Prediction settings

☐ Algorithm name

☐ Decision Tree Classifier

☐ Extra Trees Classifier

☒ Gradient Boosting Classifier

☐ LGBM Classifier

☒ Logistic Regression

☒ Random Forest Classifier

☐ XGB Classifier

Algorithms to use

AutoAI will test the specified algorithms and use the top performers to create model pipelines. Choose how many top algorithms to apply. Each algorithm generates 4 pipelines and more algorithms increase the runtime.

1 2 3

Select **Algorithms**

Experiment settings

Prediction column
Risk (STR)

Data source

Prediction

Runtime

Prediction settings

Prediction type

Change the prediction type based on data in the prediction column. Changing the type changes other prediction settings.

Binary classification ☒

Classify data into categories. Choose this if your prediction column contains two distinct categories.

Multiclass classification ☐

Classify data into categories. Choose this if your prediction column contains multiple distinct categories.

Regression ☐

Predict values from a continuous set of values. Choose this if your prediction column contains a large number of values.

Positive class


Specify the value in your prediction column to measure performance in to a confusion matrix.

Risk

From **experiment settings**, select the **Prediction** tab and select **Risk** as the **Positive Class**

AutoAI – Run experiment

Configure details

 What do you want to predict?
Prediction column ⓘ

Risk

×

▼

Prediction column: Risk

PREDICTION TYPE ⓘ	POSITIVE CLASS	OPTIMIZED METRIC ⓘ
Binary Classification ⓘ	Risk	Accuracy ⓘ








Experiment settings ⚙

Run experiment

Click on **Run experiment**

AutoAI – Results

Pipeline leaderboard

Rank	↑	Name	Algorithm	Accuracy (Optimized)	Enhancements	Build time
★ 1		Pipeline 2	 Gradient Boosting Classifier	0.805	HPO-1	00:00:36
2		Pipeline 3	 Gradient Boosting Classifier	0.805	HPO-1 FE	00:04:52
3		Pipeline 4	 Gradient Boosting Classifier	0.805	HPO-1 FE HPO-2	00:01:16
4		Pipeline 1	 Gradient Boosting Classifier	0.799	None	00:00:07
5		Pipeline 8	 XGB Classifier	0.789	HPO-1 FE HPO-2	00:01:22
6		Pipeline 7	 XGB Classifier	0.788	HPO-1 FE	00:01:12
7		Pipeline 5	 XGB Classifier	0.783	None	00:00:02

AutoAI – Save as

Save as

Select asset type

Model



Create a Watson Machine Learning model asset that you can test with new data, deploy to generate predictions, and trace lineage activity.

Notebook Tech preview

Create a notebook if you want to view the code that created this model pipeline or interact with with the model programmatically.

Define details

Name

German Credit Risk - P2 Gradient Boosting Classifier

Description (optional)

Enter description here

Tags

Add tags to make assets easier to find.

Add a tag



Exercise

- Create an AutoAI experiment using the sample data provided

Lesson

- Learn how to create and configure an AutoAI experiment

RStudio

The screenshot shows the IBM Cloud Pak for Data web interface. The top navigation bar includes a hamburger menu, the text 'IBM Cloud Pak for Data', a dropdown menu currently showing 'All', and a search bar. Below this, the breadcrumb 'Projects / BHP JumpStart' is visible. A row of tabs includes 'Overview', 'Assets' (which is selected and underlined), 'Environments', 'Jobs', 'Access Control', and 'Settings'. To the right of the tabs is a 'Launch IDE' dropdown menu with an upward arrow, and a blue 'Add to project' button with a plus icon. The 'Launch IDE' menu is open, showing two options: 'JupyterLab' and 'RStudio'. The 'RStudio' option is highlighted with a yellow box. A yellow line extends from this box to a larger yellow box containing text. Below the tabs is a large search input field with a magnifying glass icon and the placeholder text 'What assets are you looking for?'.

IBM Cloud Pak for Data

All

Search

Projects / BHP JumpStart

Overview Assets Environments Jobs Access Control Settings

Launch IDE ^

Add to project +

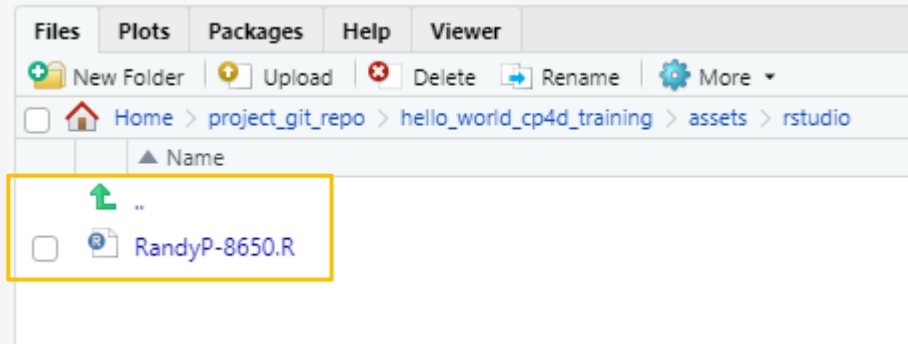
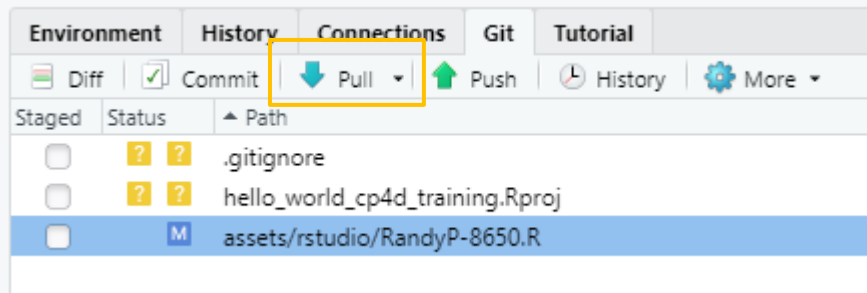
JupyterLab

RStudio

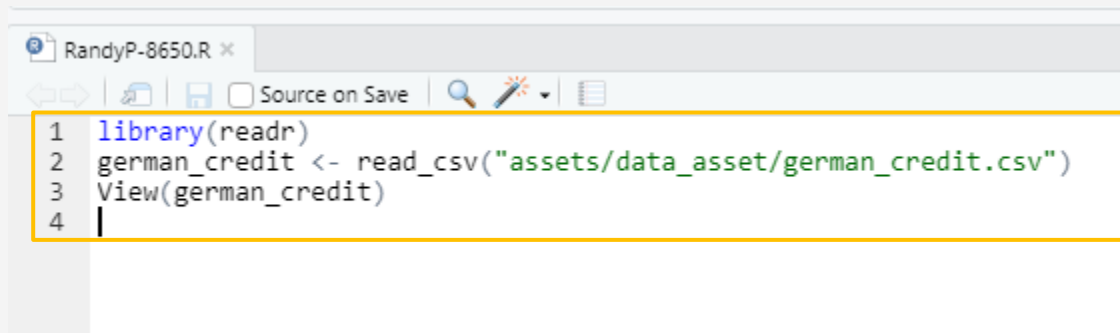
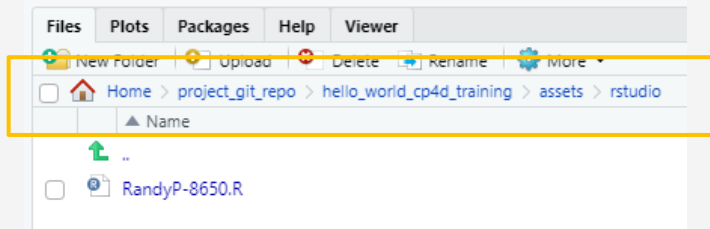
What assets are you looking for?

Click on **Launch IDE**
and select **RStudio**

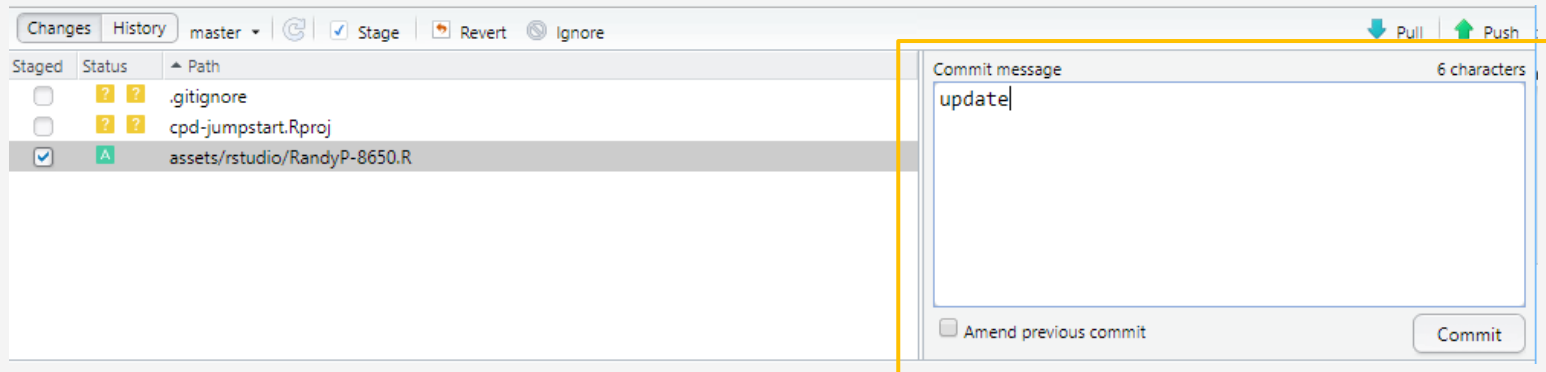
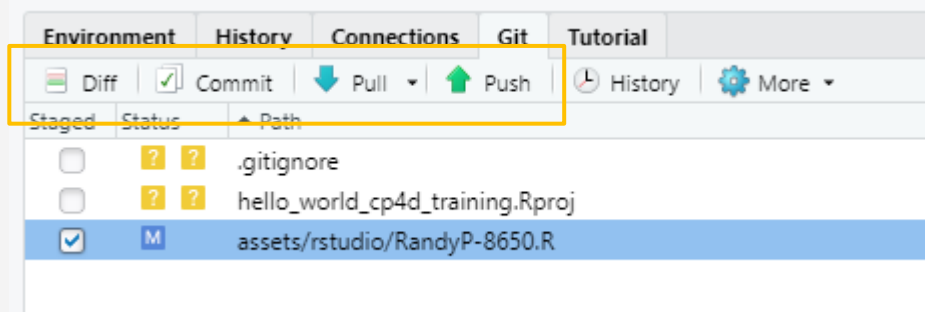
RStudio – Pull from repo



RStudio – Add script



RStudio – Push to repo



Exercise

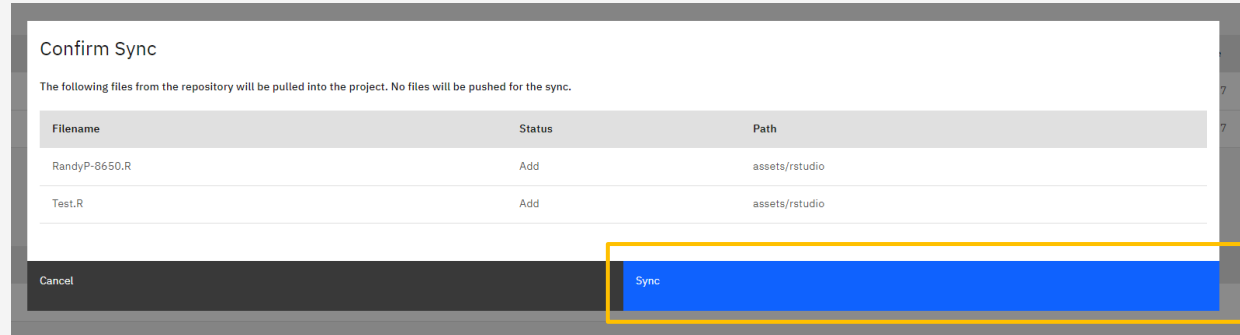
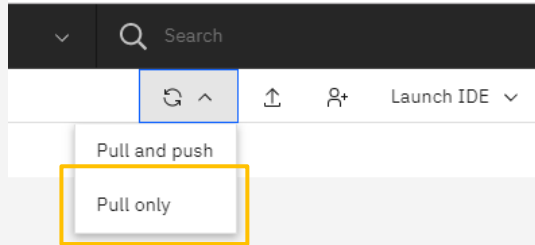
- Launch the RStudio IDE
- Pull from repo
- Add a R script
- Push to repo

Lesson

- Understand how the different file environments (Watson Studio, RStudio and Git repo work together)

Watson Studio – Check R Scripts

R Studio -> Git repo -> Watson Studio



Scripts

Name	Language	Software specification	Created by	Last modified ↓
Test.R	R	R 3.6	Randy PHOA	Dec 01, 2021, 03:23 PM
RandyP-8650.R	R	R 3.6	Randy PHOA	Dec 01, 2021, 03:20 PM

Run jobs in Watson Studio – R Scripts

✓ Scripts

Name	Language	Software specification	Created by	Last modified	↓
Test.R	R	R 3.6	Randy PHOA	Dec 01, 2021, 03:23 PM	
RandyP-8650.R	R	R 3.6	Randy PHOA	Dec 01, 2021, 03:20 PM	

Promote

Create job

RandyP-8650.R

Create a job

● Define details

○ Configure

○ Schedule

○ Review and create

Define details

Associated asset

 RandyP-8650.R

Name

How will you identify this job?

Description (optional)

What's the purpose of this job?

Exercise

- Sync project with repo
- Ensure you can see your R Script
- Create a Job

Lesson

- Understand how the different file environments (Watson Studio, RStudio and Git repo work together)
- Learn how to create and configure a Job

JupyterLab

The screenshot displays the IBM Cloud Pak for Data web interface. At the top, a dark navigation bar contains the IBM logo, the text "IBM Cloud Pak for Data", a dropdown menu set to "All", and a search bar. Below this, a breadcrumb trail shows "Projects / BHP JumpStart". A horizontal menu bar includes tabs for "Overview", "Assets" (which is selected), "Environments", "Jobs", "Access Control", and "Settings". To the right of these tabs are icons for refresh, upload, and user management, followed by a "Launch IDE" button with a dropdown arrow and an "Add to project" button. The "Launch IDE" dropdown menu is open, showing "JupyterLab" and "RStudio" options. A yellow box highlights the "Launch IDE" button and the "JupyterLab" option, with a text overlay that reads "Click on **Launch IDE** and select **JupyterLab**". Below the menu bar is a search bar with the placeholder text "What assets are you looking for?". In the lower-left corner, a "JupyterLab configuration" dialog box is open. It has a close button in the top right. Inside, there are two sections: "Select environment" with a dropdown menu showing "Randy Env", and "Select Git token" with a dropdown menu showing "BHP JumpStart". Both dropdown menus are highlighted with yellow boxes. A yellow text box with the text "Select **Environment** and **Git token**" points to these two sections. At the bottom right of the dialog is a blue "Start" button.

IBM Cloud Pak for Data

All

Search

Projects / BHP JumpStart

Overview Assets Environments Jobs Access Control Settings

Launch IDE ^

Add to project +

JupyterLab

RStudio

What assets are you looking for?

JupyterLab configuration

Select environment

Randy Env

Select Git token

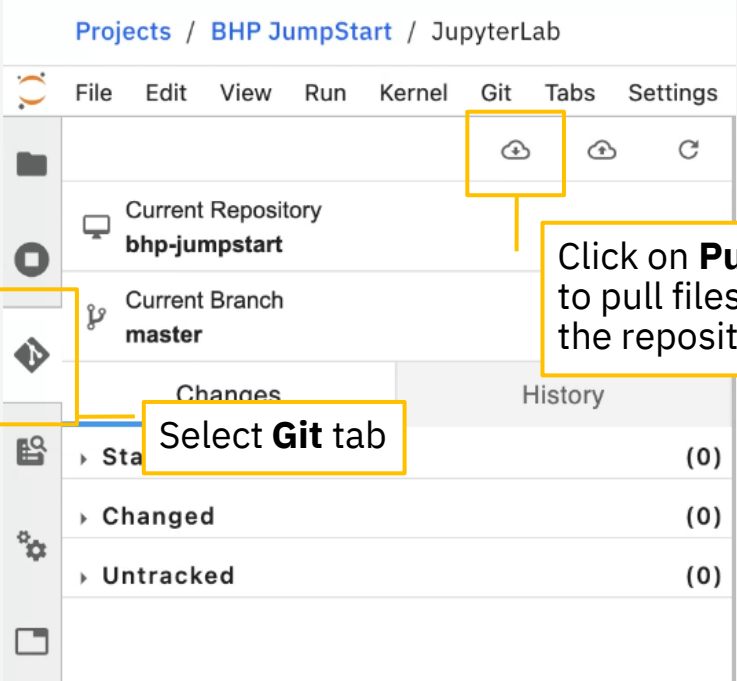
BHP JumpStart

Start

Click on **Launch IDE** and select **JupyterLab**

Select **Environment** and **Git token**

JupyterLab – Pull from repo



Projects / BHP JumpStart / JupyterLab

File Edit View Run Kernel Git Tabs Settings

Current Repository
bhp-jumpstart

Current Branch
master

Changes History

Sta (0)

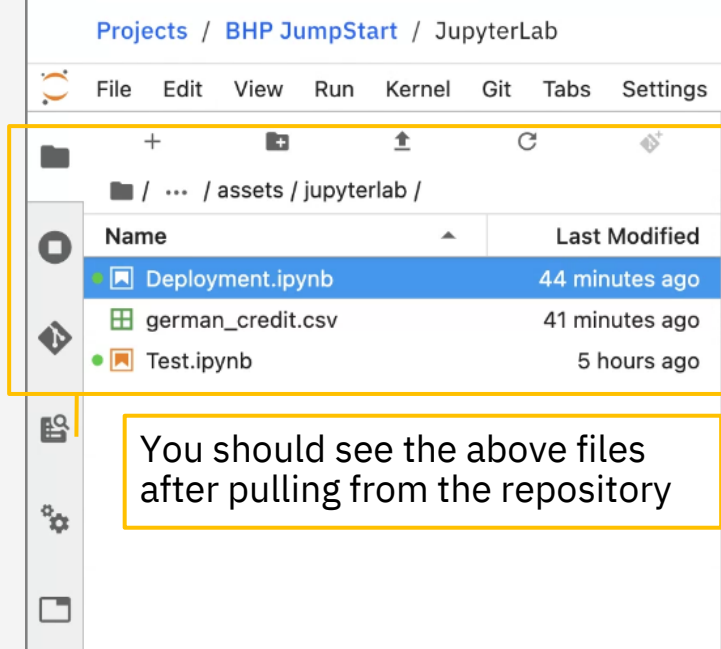
Changed (0)

Untracked (0)

Click on **Pull** icon to pull files from the repository

Select **Git** tab

An orange box highlights the 'Git' tab in the top menu. Another orange box highlights the 'Pull' icon (a cloud with a downward arrow) in the Git toolbar. A third orange box highlights the 'Git' tab in the left sidebar. A yellow arrow points from the 'Pull' icon to the right-hand screenshot.



Projects / BHP JumpStart / JupyterLab

File Edit View Run Kernel Git Tabs Settings

/ ... / assets / jupyterlab /

Name	Last Modified
Deployment.ipynb	44 minutes ago
german_credit.csv	41 minutes ago
Test.ipynb	5 hours ago

You should see the above files after pulling from the repository

The right-hand screenshot shows the file browser view after pulling. The 'Git' tab is selected in the top menu. The file list shows three files: 'Deployment.ipynb' (44 minutes ago), 'german_credit.csv' (41 minutes ago), and 'Test.ipynb' (5 hours ago). An orange box highlights the file list area. A yellow box highlights the text 'You should see the above files after pulling from the repository'.

JupyterLab – Add notebook

Projects / CPD JumpStart Training / JupyterLab

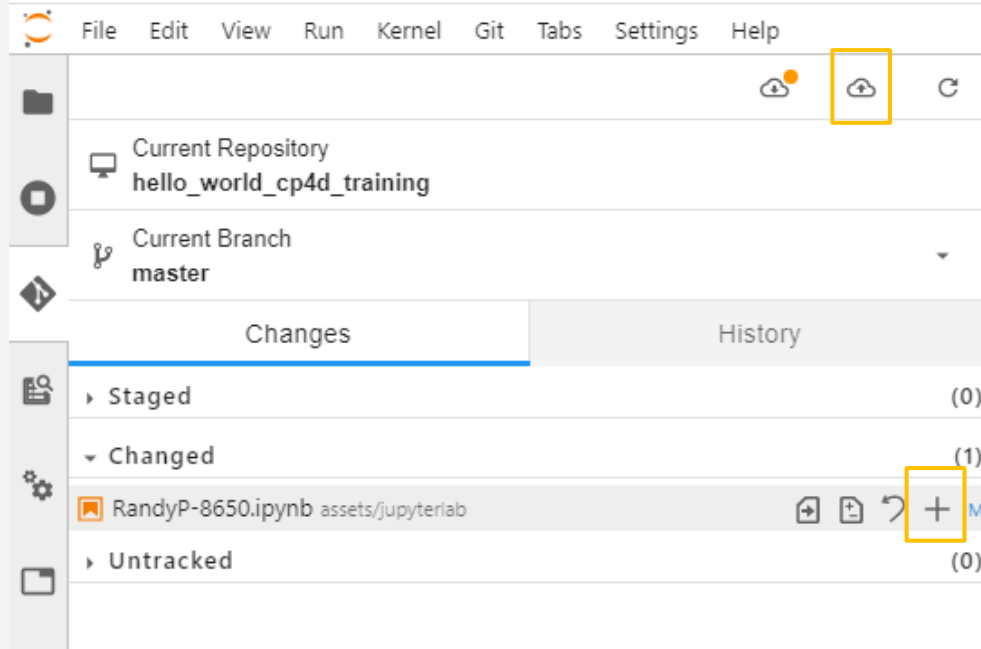
File Edit View Run Kernel Git Tabs Settings Help

+

/ ... / assets / jupyterlab /

Name	Last Modified
• Build-Deploy.ipynb	an hour ago
• RandyP-8650.ipynb	an hour ago

JupyterLab – Push to repo



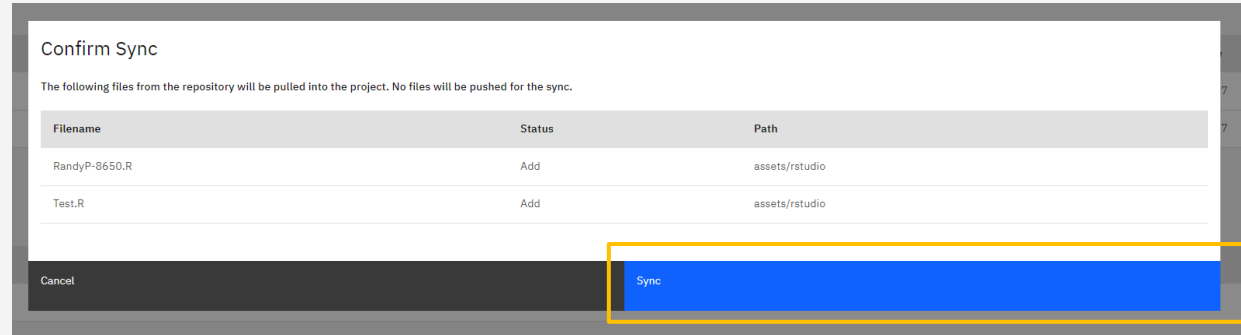
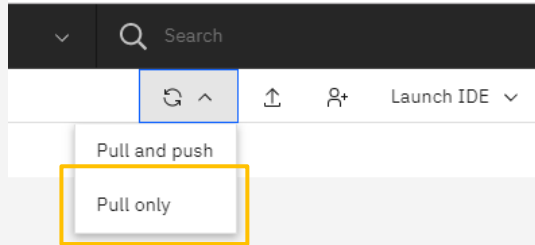
Exercise

- Launch the JupyterLab IDE
- Pull from repo
- Add a Jupyter Notebook
- Push to repo

Lesson

- Understand how the different file environments (Watson Studio, RStudio and Git repo work together)

Watson Studio – Check Jupyter Notebook



▼ Notebooks


New Notebook +

Name	Shared	Scheduled	Status	Language	Last editor	Last modified
 Build-Deploy				Python 3.7	Randy PHOA	Dec 02, 2021
 RandyP-8650				Python 3.7	Randy PHOA	Dec 02, 2021

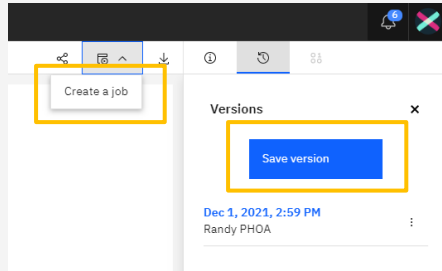
Run jobs in Watson Studio – Jupyter Notebook

▼ Notebooks

Name

 Build-Deploy

 RandyP-8650



Test

Create a job

☒ Define details

☐ Configure

☐ Schedule

☐ Review and create

Define details

Associated asset

 Test

Name

How will you identify this job?

Description (optional)

What's the purpose of this job?

Exercise

- Sync project with repo
- Ensure you can see your Jupyter Notebook
- Create a Job

Lesson

- Understand how the different file environments (Watson Studio, RStudio and Git repo work together)
- Learn how to create and configure a Job

Watson Studio – View job at project level

Projects / cpd-jumpstart

Overview

Assets

Environments

Jobs

Access Control

Settings

Launch IDE

Add to project

Jobs

2

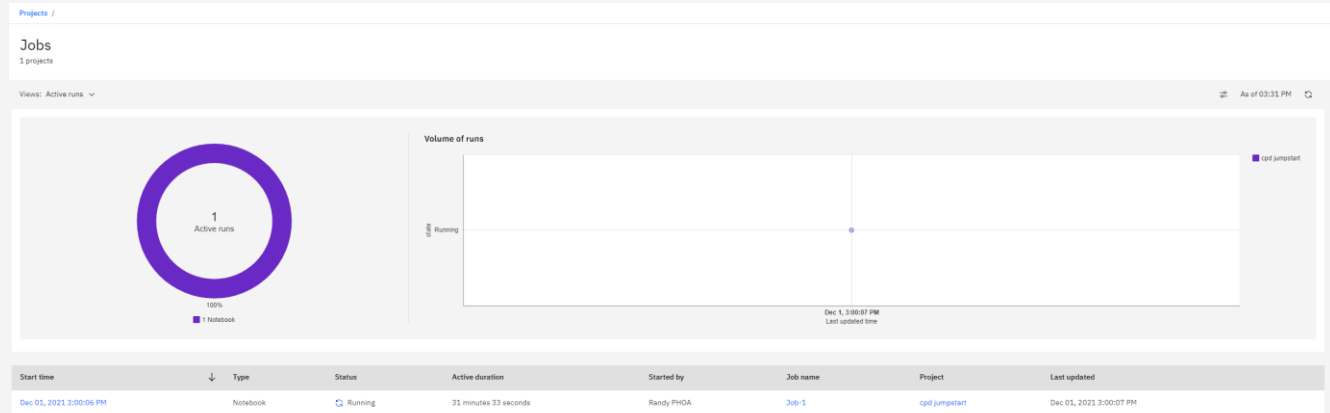
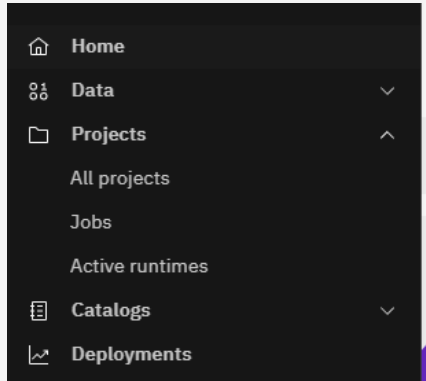
Total Jobs

A job is a way of running an asset like a notebook or flow. You can run a job immediately or on a schedule.

Which job are you looking for?

Job name	Associated asset	Date created	Created by
Job-2	Script	Dec 01, 2021, 03:24 PM	Randy PHOA
Job-1	Notebook	Dec 01, 2021, 03:00 PM	Randy PHOA

Watson Studio – View jobs at global level



Exercise

- View the jobs that you have created earlier

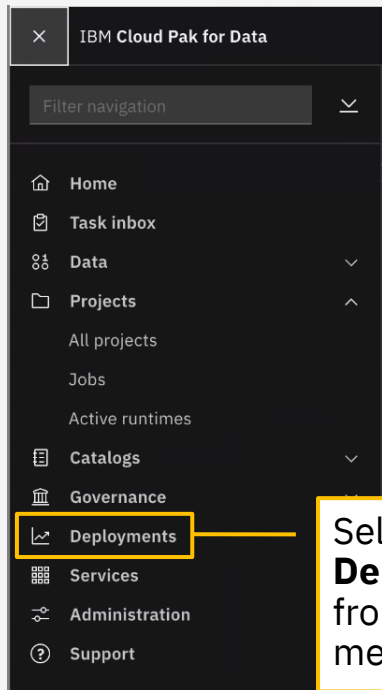
Lesson

- Learn the different ways of viewing job status

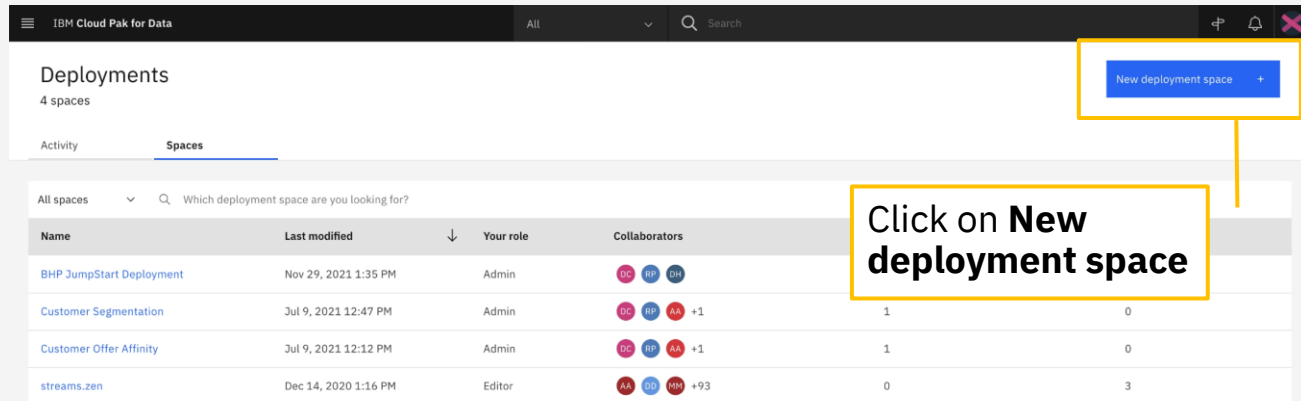
5. Deploy models

- Deployment spaces
- Promote assets
- Online scoring
- Batch scoring
- Schedule jobs

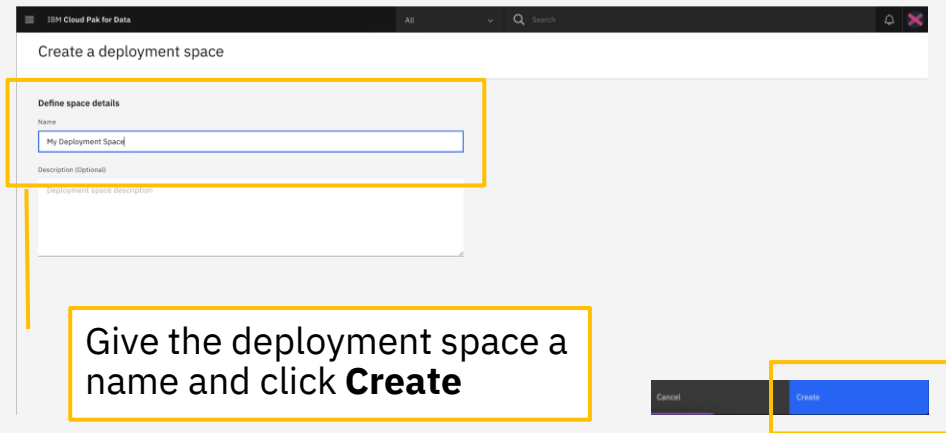
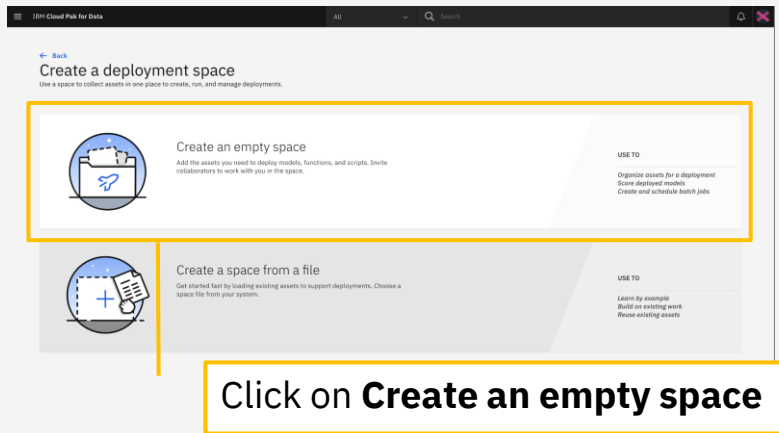
Create a deployment space



Select **Deployments** from the menu



Create a deployment space



Exercise

- Create a deployment space

Lesson

- Understand what is a deployment space

Promote assets

▼ Data assets New data asset +

0 assets selected.

<input type="checkbox"/>	Name	Type	Created by	Last modified	↓
<input type="checkbox"/>	CSV german_credit.csv	Data Asset	Randy Phoa	Nov 29, 2021, 11:46 AM	<div><div>Publish to Catalog</div><div>Refine</div><div>Download</div><div>Promote</div><div>Remove</div></div>

> AutoAI experiments

> Notebooks

> Scripts

Select asset and
click **Promote**

Promote assets

Promote to space

Target space
BHP JumpStart Deployment

Selected assets (1)

Asset name	Format	Actions
german_credit.csv	text/csv	

Description (optional)
Description of assets

New space +

Tags (optional)
Start typing to add tags

Select Target space and click **Promote**

Cancel

Promote

Exercise

- Promote your sample data file and your R or Python Script

Lesson

- Understand how the deployment space (Watson Machine Learning) environment relates to Watson Studio, IDEs (RStudio and JupyterLab) and Git repo.

Online Scoring

BHP JumpStart Deployment

Assets Deployments Jobs Environments Access control Settings

🔍 What assets are you looking for?

▼ Models (1) [Import model +](#)

Name	Type	Software specification	Last modified	↓
💎 Iris SVM Model	scikit-learn_0.23	default_py3.7 ⚠️	Nov 29, 2021 3:26 PM	

Select model and click **Create deployment**


[Deployments](#) / [BHP JumpStart Deployment](#) / [Iris SVM Model](#)


Iris SVM Model

Create deployment 

Online Scoring


Create a deployment ×

 Associated asset
Iris SVM Model

 **Associated asset uses a deprecated framework**
Software specification default_py3.7 is deprecated and will be removed in a future release. Use default_py3.7_opense instead. For details, see [Software specifications](#).

Deployment type

Deployment type

Online 

Run the model on data in real-time, as data is received by a web service.

Batch

Run the model against data as a batch process.

Core ML for iOS

Download a model in Core ML format for use in iOS apps.

Name

Description

Cancel

Create

Ensure Deployment type is **Online**, give your deployment a **name** and click **Create**

Online Scoring

My Online Deployment ✔ Deployed Online

API reference

Test

Enter input data

Body

```
{ "input_data": [  
  {  
    "values": [  
      [1.1, 1.2, 1.3, 1.4],  
      [2.1, 2.2, 2.3, 2.4]  
    ]  
  }  
]}
```

Click on the **Test** tab, input sample data and click **Predict**

Predict

Result

```
0 {  
1   "predictions": [  
2     {  
3       "fields": [  
4         "prediction"  
5       ],  
6       "values": [  
7         [  
8           0  
9         ],  
10        [  
11          0  
12        ]  
13      ]  
14    }  
15  ]  
16 }
```


Exercise

- Create an online scoring deployment

Lesson

- Learn how to configure an online scoring deployment and test using sample input data

Batch Scoring

BHP JumpStart Deployment

[Assets](#) [Deployments](#) [Jobs](#) [Environments](#) [Access control](#) [Settings](#)

🔍 What assets are you looking for?

▼ Models (1) [Import model +](#)

Name	Type	Software specification	Last modified	↓
💎 Iris SVM Model	scikit-learn_0.23	default_py3.7 ⚠️	Nov 29, 2021 3:26 PM	

Select model and click **Create deployment**

[Deployments](#) / [BHP JumpStart Deployment](#) / [Iris SVM Model](#)

Iris SVM Model

Create deployment 

Batch Scoring

Create a deployment ✕

Deployment type

Deployment type

Online

Run the model on data in real-time, as data is received by a web service.

Batch

Run the model against data as a batch process.

Core ML for iOS

Download a model in Core ML format for use in iOS apps.

Name

My Batch Deployment

Description

Deployment description

Software specification

[default_py3.7](#)

The software specification is predefined for the asset type. You can update or customize the software specification programmatically. [Learn more](#)

Hardware definition

1 CPU and 4 GB RAM

Cancel

Create

Ensure Deployment type is **Batch**, give your deployment a **name**, select the required **hardware definition** and click **Create**

Batch Scoring

[Deployments](#) / [BHP JumpStart Deployment](#) / [Iris SVM Model](#) / Iris SVM Deployment-batch



Iris SVM Deployment-batch ✓ Deployed Batch

Create job +

Click on **Create job**

Create new job

- ☒ Define details
- ☐ Configure
- ☐ Schedule
- ☐ Choose data
- ☐ Review and create

Define details

Name

Job1

Description

Deployment job description

Create new job

- ☒ Define details
- ☒ Configure
- ☐ Schedule
- ☐ Choose data
- ☐ Review and create

Configure

Hardware specification

2 CPU and 8 GB RAM

☐ Environment variables (optional)

Batch Scoring

Create new job

- Define details
- Configure
- Schedule**
- Choose data
- Review and create

☒ Schedule to run

Time Zone: **GMT+0800 (Singapore Standard Time)**

☒ Start On

11/30/2021 at 22:00 of 24-hr time

☒ Repeat

Every

day at 22:00 of 24-hr time

☒ Exclude Days

Click to exclude

Sun Mon Tue Wed **Thur** Fri Sat

☒ End On

12/30/2021 at 22:00 of 24-hr time

Create new job

- Define details
- Configure
- Schedule
- Choose data**
- Review and create

Choose data

Select the data asset(s) to use during each run. This step is required to run the job on a schedule.

Input Data asset

input.csv

german_credit.csv

input.csv

output.csv

output.csv

Output Create new

Name

job1-output.csv

Description

Output asset description

Batch Scoring

Create new job


- ✓ Define details
- ✓ Configure
- ✓ Schedule
- ✓ Choose data
- Review and create

Review and create

Details 


Name
Job1

Description
[Add description](#)


Configuration 

Hardware specification
2 CPU and 8 GB RAM

Schedule 

 Start on
11/30/2021


Time
10:00 PM

 Repeat
Every day at 10:00 PM

Except on
Mon, Tue, Wed, Thur, Fri


 Until
12/30/2021

Time
10:00 PM

Data assets 

 Input
input.csv



 Output
job1-output.csv

Cancel

Create

Batch Scoring – Check job status

Iris SVM Deployment-batch ✓ Deployed Batch

Create job +

Jobs (2)

Job name	Date created	Created by
Job-1	Nov 29, 2021 4:43 PM	Randy Phoa
WML-Deployment-Job-60c083ca-b796-44...	Nov 29, 2021 3:26 PM	Randy Phoa

Batch Scoring – Check job status

Deployments / BHP JumpStart Deployment / Job-1



Job-1

Associated Asset
 DEPLOYMENT
[Iris SVM Deployment-batch](#)

Scheduled to run

[Edit](#)

No Schedule Created


Environment definition

[Edit](#)

2 CPU and 8 GB RAM

Input	Edit	Output	Edit
input.csv		output.csv	

Runs (1)

Start Time	↓	Status	Duration	Started By	
Nov 29, 2021 4:43 PM		✔ Completed	1 second	Randy Phoa	

Batch Scoring – Check job status

Run log

```
{
  "deployment": {
    "id": "60c083ca-b796-4419-807c-be6aa9e5f64b"
  },
  "hardware_spec": {
    "id": "e7ed1d6c-2e89-42d7-aed5-863b972c1d2b"
  },
  "platform_job": {
    "job_id": "ad96a534-720e-4b10-9ff0-b7c244116d97",
    "run_id": "01a48b79-794c-4feb-a885-2c40a7011610"
  },
  "scoring": {
    "input_data_references": [
      {
        "connection": {},
        "location": {
          "href": "/v2/assets/8c678000-d9f8-467d-baf7-b7309eb355db?space_id=a6f3c567-ff47-4388-a9b8-5d91dd8a2b62"
        },
        "type": "data_asset"
      }
    ],
    "output_data_reference": {
      "connection": {},
      "location": {
        "description": "",
        "href": "/v2/assets/3538bfcf-f4e0-4944-aa7b-3de2eed81694?space_id=a6f3c567-ff47-4388-a9b8-5d91dd8a2b62",
        "name": "output.csv"
      },
      "type": "data_asset"
    },
    "status": {
      "completed_at": "2021-11-29T08:43:28.275853Z",
      "running_at": "2021-11-29T08:43:27.422854Z",
      "state": "completed"
    }
  }
}
```

Show less ^

Exercise

- Create a batch scoring deployment

Lesson

- Learn how to configure a batch scoring deployment and test using sample input data
- Learn how to check a batch job status

6. Resources

- Documentation – <https://www.ibm.com/docs/en/cloud-paks/cp-data/3.5.0>
- Examples – https://github.com/IBM/watson-machine-learning-samples/tree/master/cpd3.5/notebooks/python_sdk