

ASEANZK Cloud Pak for Data – Practicum Scenario:

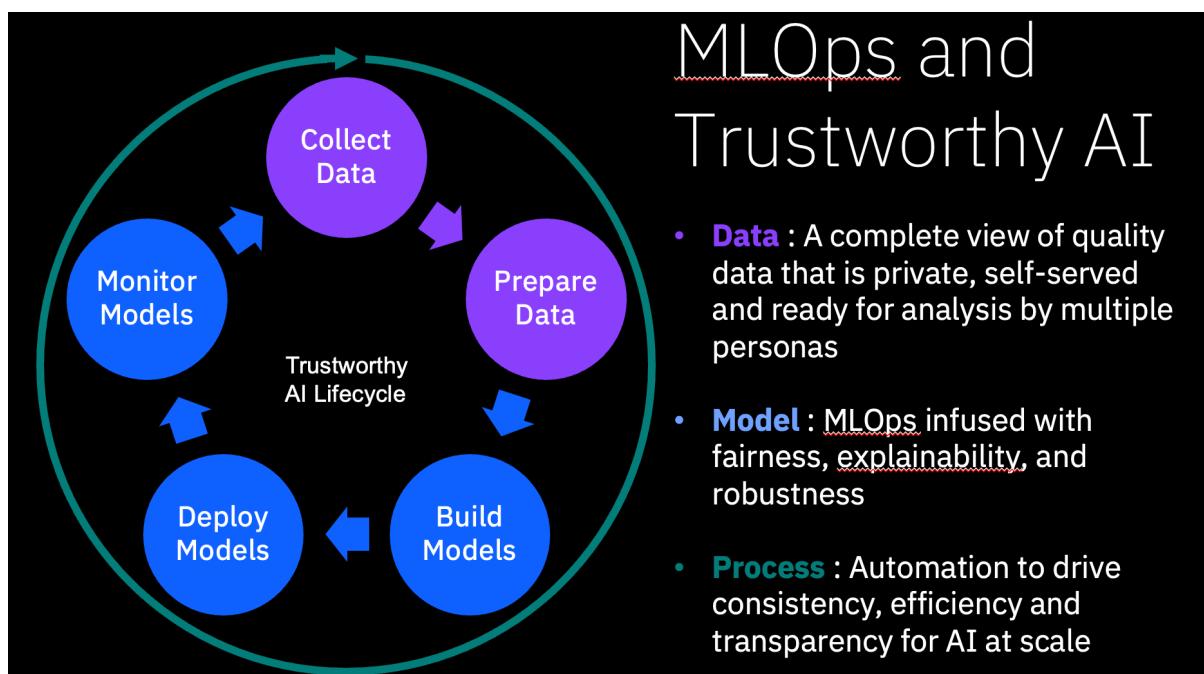
Author:

Teerarat Siwapathomchai (Toon), Customer Success Manager

Satit Pongbundit (Oat), Customer Success Manager

Use Case

MLOps and Trustworthy AI (focusing on Model Development / Model Deployment / Monitoring)



Scenario Description

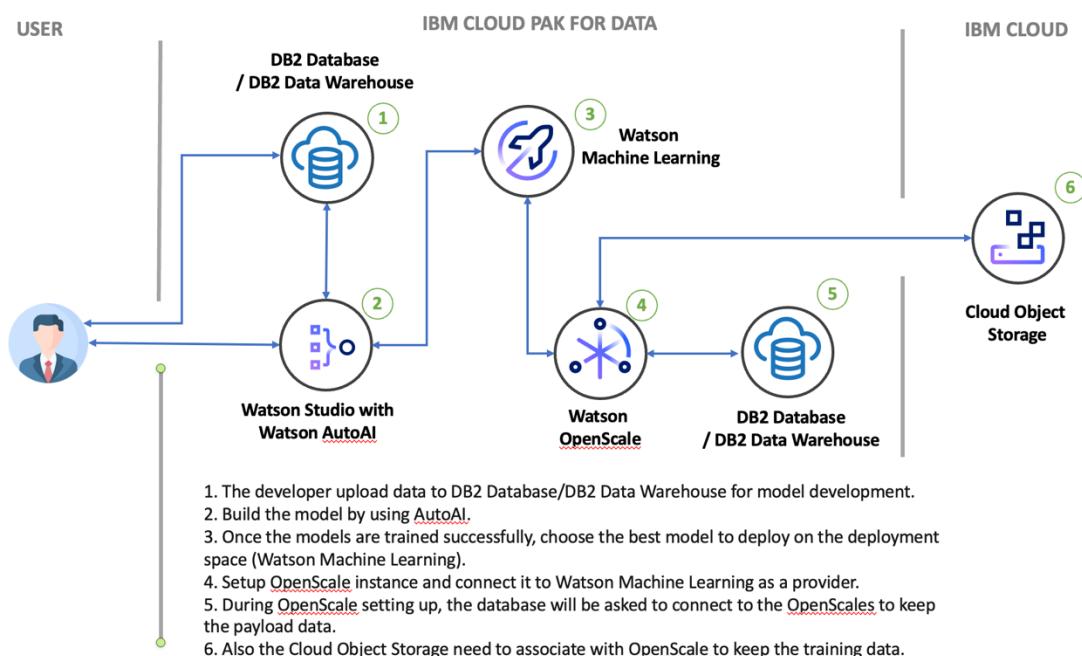
Credit risk is significant for the banking business. If the borrower fails to repay, it results in a loss. To minimize loan from the bank's perspective, ABC Bank needs a decision rule regarding to give approval of the loan and who not to. They decided to implement Application Score (A-Score) model to access loan application based on application's demographic and economic profiles. Loan analyst who has no programming skill has been asked to develop the model. The saved model need to be deployed machine learning engineer and generate the API for the front-end application. Due to economic uncertainty can cause the customer behaviour change which affects to the model. To avoid the opportunity cost in business from dropping in model accuracy. The team also has been asked to set up the instance to monitor the model after deploying.

The company has recently acquired Cloud Pak for Data v4 and are planning to utilize its key features to achieve the following goals :

- i. Accelerate model development without programming skill.
- ii. Save and deploy model seamlessly to reduce friction between model developer and machine learning engineer.
- iii. The model need to be an "on-line" deployment with API for the front end application.
- iv. After deployment successfully, you also have been asked to set up instance to monitor the model to see there are any changes affect to the model (Drift and Fairness for example).
- v. Use the Python Client to test the model and check the payload is logging in the OpenScale.

High Level Architecture

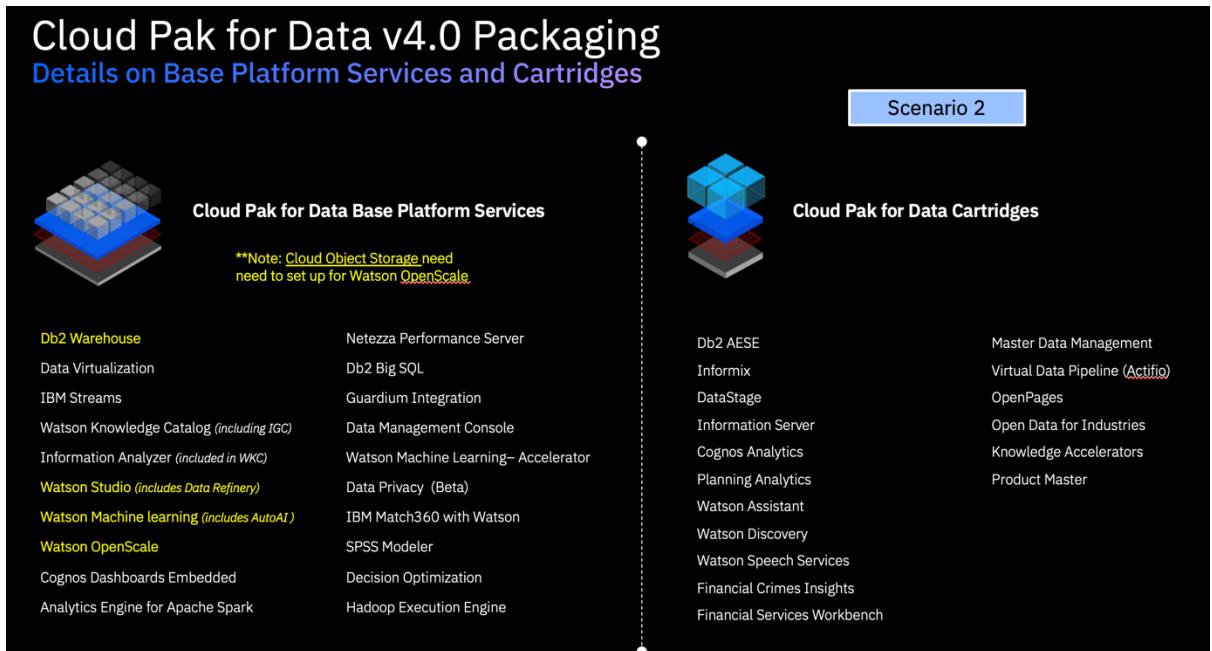
Below is the high level architecture that we plan to implement as part of this scenario.



Product Used

Your objective is to use IBM Cloud Pak for Data features to develop, deploy and monitor model under the MLOps and trustworthy AI concept. The key products you will be used for this project are here;

- 1x Watson Studio (WSL) with AutoAI
- 1x Watson Machine Learning (WML)
- 1x Watson OpenScale instance.
- 2x DB2 Database or DataWarehouse (1x for Training data storing, 1x for Watson OpenScale set up)
- 1x Cloud Object Storage (for Watson OpenScale to keep training data)
- 1x Data Management Console



Task Allocation by Roles

SI No	Tasks for Scenario 2	Role
1	Provision all required Services	CP4D Administrator
2	Upload data to the database	Data Engineer/Database Admin
3	Develop the Model using AutoAI	Data Scientist
4	Deploy the model	Machine Learning Engineer
5	Configure and monitor the model	Model validator / CP4D Administrator

Implementation Step

Task 0: Preparing Steps

- 1) For this lab, we need the number of services as shown below;
 - 1x Watson Studio (WSL) with AutoAI
 - 1x Watson Machine Learning (WML)
 - 1x Watson OpenScale instance.
 - 2x DB2 Database or Data Warehouse (1x for Training data storing, 1x for Watson OpenScale set up)
 - Your Preferred Browser (If you are using Chrome, please use the incognito mode)

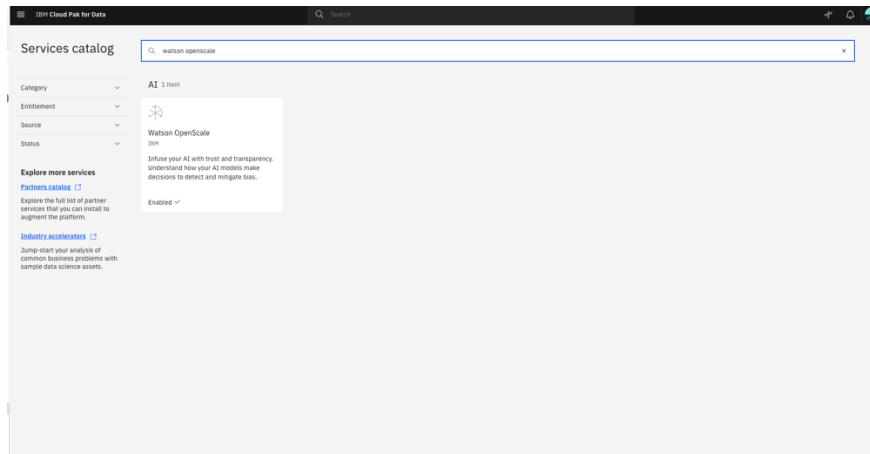
- 2) Please check all the services above are enabled or available, from the navigation bar go to services catalog. Check all the services are indicated as enabled like the example below. If there are any services above have no label “Enabled” or “Available”, please contact your facilitators.

 Watson Studio IBM Unleash the power of your data. Build custom models and infuse your business with AI and machine learning. Enabled ✓	 Db2 Data Management Console IBM Administer, monitor, manage, and optimize your Db2 databases from a single web-based console. Available ↗
--	---

- 3) If all services are already enabled, Watson Studio (WSL) and Watson Machine Learning (WML) are already to use automatically.
- 4) For Watson OpenScale instance, Data Management Console and DB2 Database or Data Warehouse, you need to provision the instances by yourself. Please note that Data Management Console can have only single instance in the platform. You can go check it at navigation bar -> “Instance” menu. If you see the Data Management Console already created, you no need to create the second one.

Name	Type	Created by	vCPU requests	Memory requests (GiB)	Users	Status	Created on
data-virtualization	dv	admin	11.50	38.50 Gi	3	Green	Dec 22, 2022
Db2-1	db2oltp	admin	2.20	5.75 Gi	3	Green	Dec 22, 2022
openscale-defaultinstance	aios	admin	-	-	2	Green	Dec 21, 2022
data-management-console Data Management Console	dmc	admin	4.20	10.62 Gi	3	Green	Dec 21, 2022

- 5) To provision Watson OpenScale.
1. Go to service catalog and search for “Watson OpenScale”. Click on it.



- On the left hand side, click drop down and create the instance.

On the left hand side, click drop down and create "New Instance"

- Set up the name and choose the name space "cpd-instance", then click next.

New service instance

Name: Openscale

Namespace: cpd-instance

- See the summary and click create.

New service instance
Watson OpenScale

Instance de... Summary

Summary
Review the following summary. If all the values are correct, click Create.

Instance details

Instance name	OpenScale
Service type	aios
Service version	4.5.3

- Once the OpenScale instance is created successfully. The status will show as below.

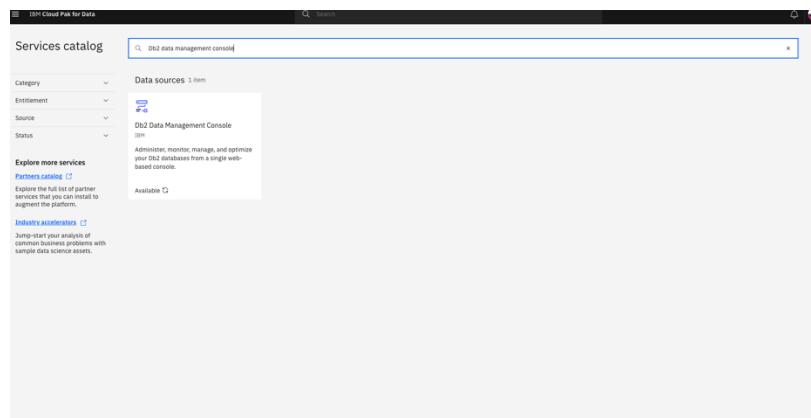
Filter by: Type Status

Find instances

Name	Type	Created by	vCPU requests	Memory requests (GiB)	Users	Status	Created on
openscale	aios	admin	-	-	1	Success	Dec 21, 2022

6) To provision DB2 Data Management Console.

- Go to service catalog and search for “DB2 Data Management Console”. Click on it.



- In the DB2 Data Management Console page, at the top right-hand side click “Provision Instance”.

The screenshot shows the Db2 Data Management Console interface. On the left, there's a sidebar with basic information about the service: Type Application, Version 4.5.3, Provider IBM, Category Data sources, and Related links Docs. The main area is titled 'Summary' with the sub-instruction: 'Administrator, monitor, manage, and optimize your Db2 databases from a single web-based console.' Below this is the 'Databases' dashboard, which displays various metrics for different database connections. The 'Provision instance' button is highlighted with a red box in the top right corner of the dashboard area.

3. Choose Medium Size of Instance for compute.

The screenshot shows the 'New service instance' wizard, step 2: Compute. The 'Compute' section is active, with the 'Medium' radio button selected. The 'Documentation' link is visible. At the bottom right, the 'Next' button is highlighted in blue.

4. Choose storage class to be “ocs-storagecluster-crphfs” and Storage size to be 10.

The screenshot shows the 'New service instance' wizard, step 3: Storage. The 'Storage' section is active, showing the 'ocs-storagecluster-crphfs' storage class selected in the dropdown and '10' entered in the 'Storage size (Gi)' input field. The 'Next' button is highlighted in blue.

5. Review the summary and click “Create”.

New service instance
Db2 Data Management Console

Compute Storage Summary

Summary
Review the following summary. If all the values are correct, click Create.

Compute summary

Data Management Console Version 4.5.3

Plan Medium

Storage summary

Persistent storage class ocs-storagecluster-cephfs

Persistent storage size 10 Gi

Create

6. Review the summary and click “Create” and wait until the service is ready.
The status will show the green sign like the picture below.

Name	Type	Created by	vCPU requests	Memory requests (GiB)	Users	Status	Created on
data-management-console IBM Db2 Data Management Console	dmc	admin	5.20	21.40 Gi	1	Green	Dec 23, 2022
DB2-1 Service instance for db2oltp-167176877844254	db2oltp	admin	2.20	5.75 Gi	1	Green	Dec 23, 2022

7) To provision DB2.

1. Go to service catalog and search for “DB2”. Click on it.

The screenshot shows the IBM Cloud Pak for Data Services catalog. A search bar at the top contains the text 'Db2'. Below it, under the 'Analytics' section, is a card for 'Db2 Big SQL' from IBM. Under the 'Data sources' section, there are four cards: 'Db2' (selected and highlighted with a red box), 'Db2 Data Gate', 'Db2 Data Management Console', and 'Db2 Warehouse'. Each card has a brief description and an 'Enabled' status indicator.

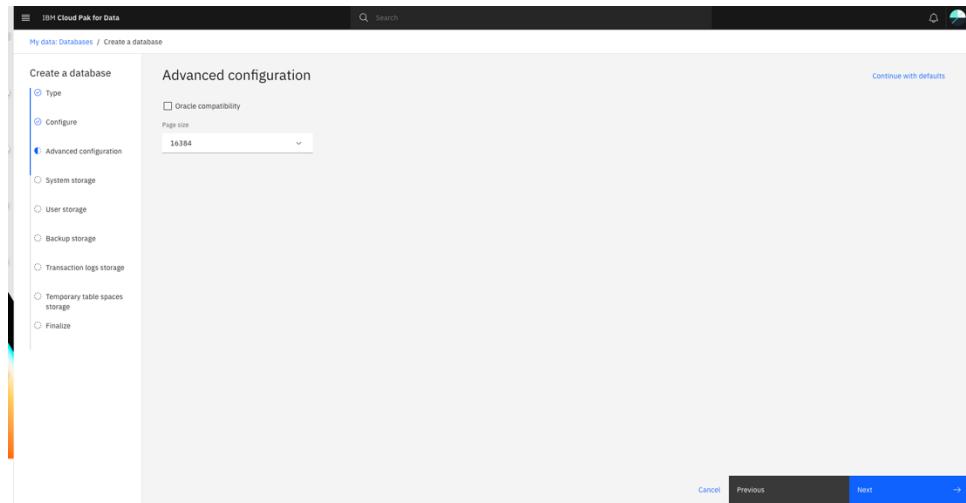
2. On the left hand side, click drop down and create the instance.

The screenshot shows the 'Db2' instance creation page. On the left, there's a sidebar with 'Db2' selected. The main area shows a summary of the Db2 service, including its type (Database), version (11.5.7.0-cn7-x86_64), provider (IBM), and category (Data sources). A red box highlights the 'New instance' button in the 'Instances' dropdown menu. Another red box highlights the 'About' link in the sidebar.

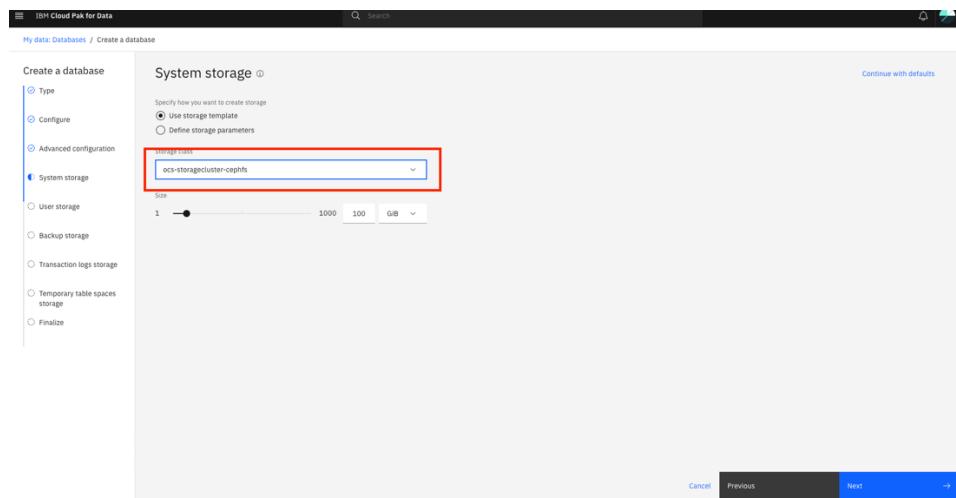
3. Configure, keep everything as default.

The screenshot shows the 'Configure' step of the database creation wizard. It includes fields for 'Database name' (set to 'BLUDB'), 'Number of nodes' (set to '1'), 'CPU per node for Db2' (set to '2.1'), 'Memory per node for Db2' (set to '5.5'), and a checkbox for 'Deploy database on dedicated nodes' which is unchecked. There are also sections for 'Namespace' (using 'cpd-instance') and 'Storage structure' (selecting 'Separate locations for all data'). At the bottom, there are 'Cancel', 'Previous', 'Next', and a blue 'Continue with defaults' button.

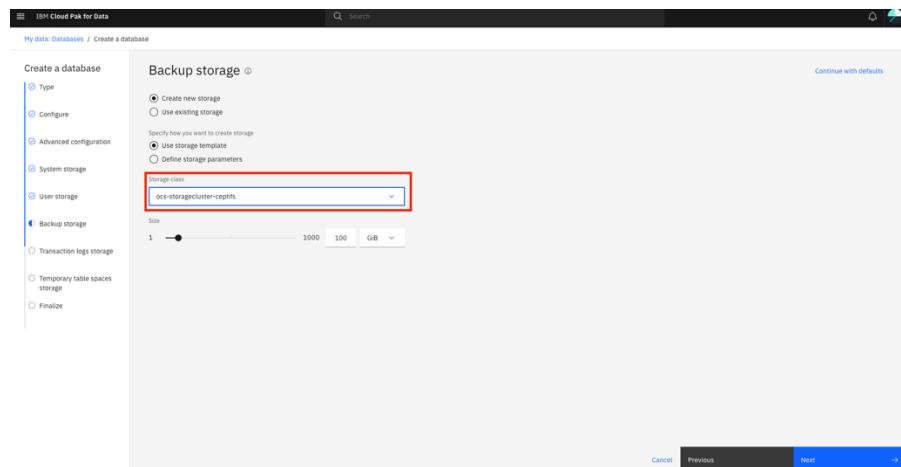
4. Advance configuration, keep as default



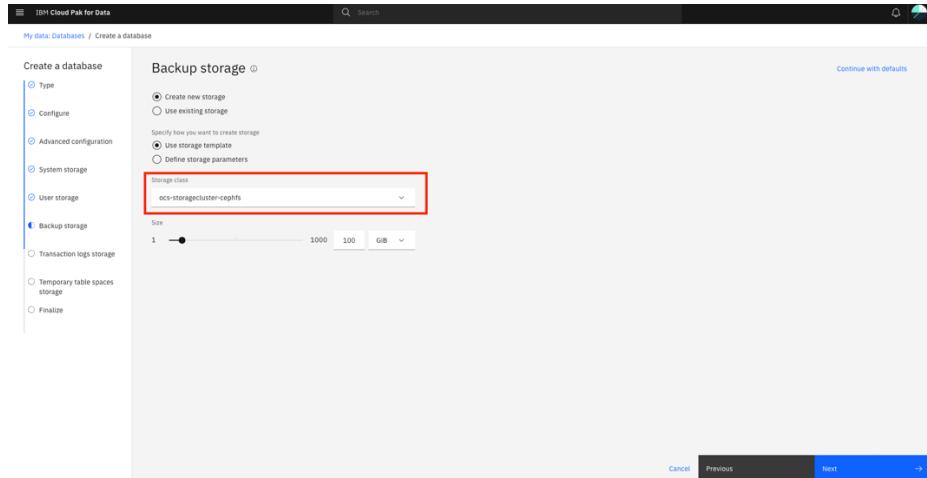
5. System storage , change the storage class to “ocs-storagecluster-cephfs” and keep others as default.



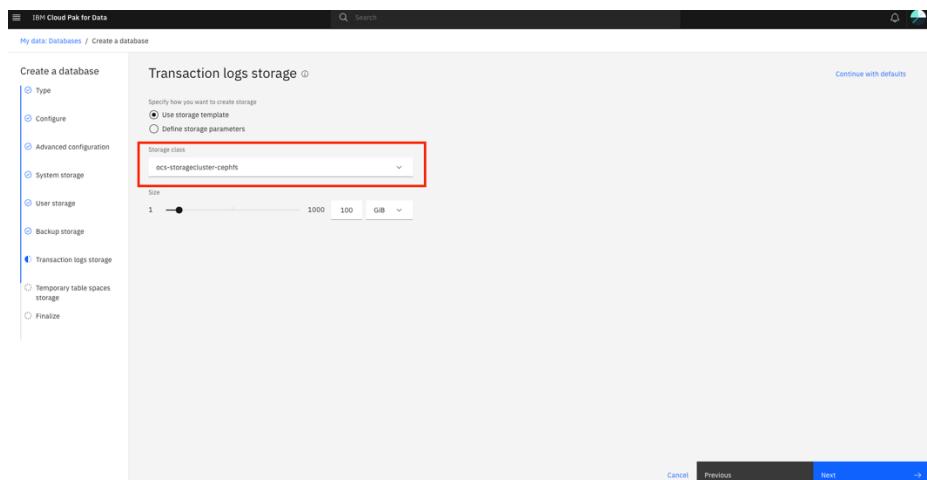
6. User storage , change the storage class to “ocs-storagecluster-cephfs” and keep others as default.



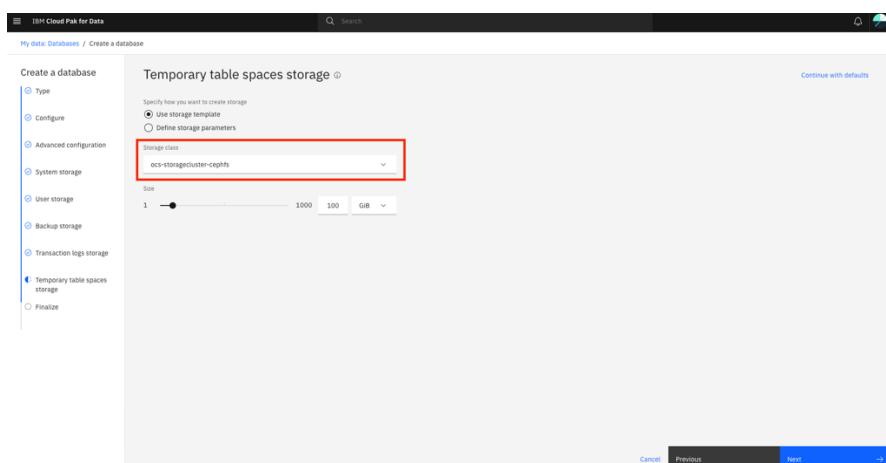
7. Backup storage , change the storage class to “ocs-storagecluster-cephfs” and keep others as default.



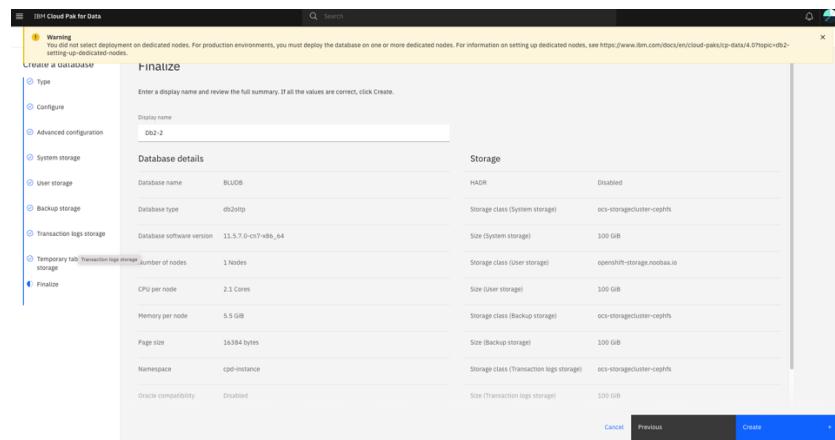
8. Transaction logs Storage , change the storage class to “ocs-storagecluster-cephfs” and keep others as default.



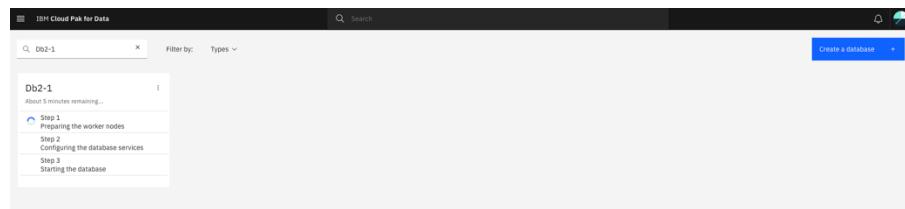
9. Temporary table spaces storage , change the storage class to “ocs-storagecluster-cephfs” and keep others as default.



10. Finalize, please review database and storage details information. Then, create.



11. The database is going to be created. Wait until it completes.



12. Create another one database by replete the step 1-11.

- 2) In reality, you need to create a Cloud Object Storage for Watson OpenScale to keep training data by following the step below. However, in this practicum you can skip this task. The cloud object storage is created and training data is already placed there. So, you can use the provided credential to configure OpenScale in the later step.

1. Going to <https://cloud.ibm.com/objectstorage>
2. Login to IBM console, if you don't have account, please create one.
3. After login, go to the "catalog" at the top bar. Click "catalog", in the search menu type "Object".
4. Select "Object Storage by IBM"
5. Choose the lite plan and create it.

Loading the training data to the data warehouse for further model development.

- 1) Go to https://github.com/IBM-CSM-SG/ASEANZK-CP4D-Practicum/blob/main/scenario2/data/training_data.csv and loading training_data to your local repository.
- 2) Go back to IBM Cloud Pak for Data.
- 3) From navigation menu, go to Data > Database.

The screenshot shows the 'Instances' page in the IBM Cloud Pak for Data interface. On the left, there's a sidebar with various navigation options like Home, Task Index, Data, Projects, Catalog, Services, and Administration. The main area displays a table of data instances. The columns include Type, Created by, vCPU requests, Memory requests (GiB), Users, Status, and Created on. There are five instances listed:

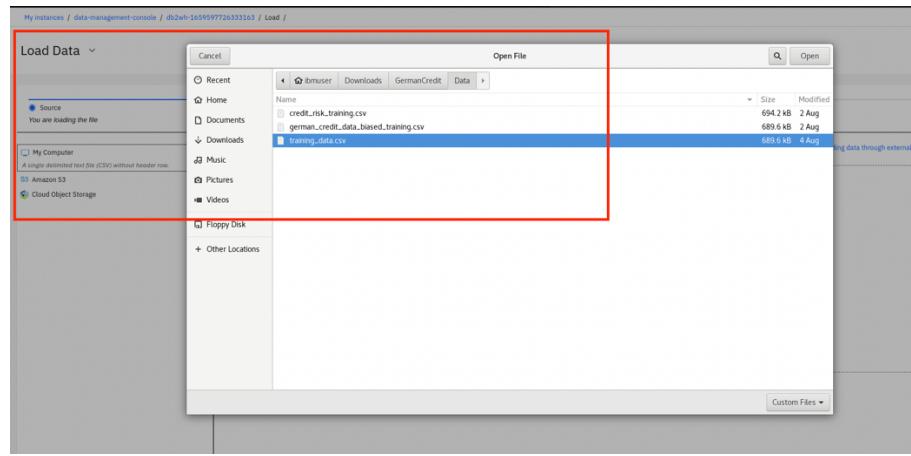
Type	Created by	vCPU requests	Memory requests (GiB)	Users	Status	Created on
db2wh	admin	6.20	18.25 Gi	1	Green	Aug 4, 2022
db2wh	admin	6.20	18.25 Gi	1	Green	Aug 4, 2022
dv	admin	11.60	38.70 Gi	1	Green	Apr 5, 2022
dmc	admin	3.70	9.50 Gi	1	Green	Apr 5, 2022
aos	admin	-	-	1	Green	Apr 5, 2022

- 4) Open the first data warehouse you can see the “Summary” in the left hand-side. Click drop down and go to “Load”. Under “Load”, click load data.

The screenshot shows the 'Monitor' section for a specific instance, with the 'Load' tab selected in the left sidebar. The main area is divided into several performance metrics:

- Responsiveness:** Shows 0 statements and 0ms response time over the last hour. A histogram shows statement distribution by response time.
- Contention:** Shows 0.00 min lock waits and 6.00 concurrent connections over the last hour. A line chart shows the number of lock waits over time.
- Resource Usage:** Shows 9.56% memory, 2% log space, and 4% storage (70G / 19...).

- 5) From the Load Data GUI, browse file and upload training_data to DB2 data warehouse.

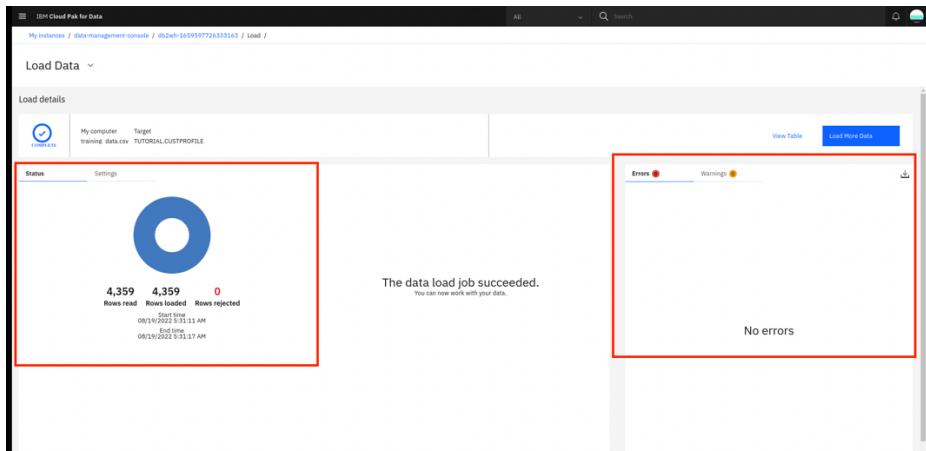


6) Create new schema name “TUTORIAL” and Create new table name “CUSTPROFILE”

7) Click next and review columns name. Click next and begin load.

	LOANAMOUNT SMALLINT	SEX VARCHAR(6)	AGE SMALLINT	HOUSE VARCHAR(4)	CREDITCARD SMALLINT	JOB VARCHAR(10)	DEPENDENT SMALLINT	TELEPHONE VARCHAR(4)	FOREIGNER VARCHAR(3)	RISK VARCHAR(7)
1	1889	female	32	own	1	skilled	1	none	yes	No Risk
2	462	female	37	own	2	skilled	1	none	yes	No Risk
3	250	male	28	own	2	skilled	1	yes	no	No Risk
4	3693	male	32	own	1	skilled	1	none	yes	No Risk
5	6235	male	57	own	2	skilled	1	none	yes	Risk
6	9604	male	57	free	2	skilled	2	yes	yes	Risk
7	3109	female	36	own	2	skilled	1	none	yes	No Risk
8	7138	male	49	free	2	skilled	2	yes	yes	Risk
9	4302	male	34	free	1	skilled	1	none	yes	No Risk
10	3310	male	40	free	1	skilled	1	yes	yes	No Risk

8) In the final page, check the data is download completely.



Create an analytics project

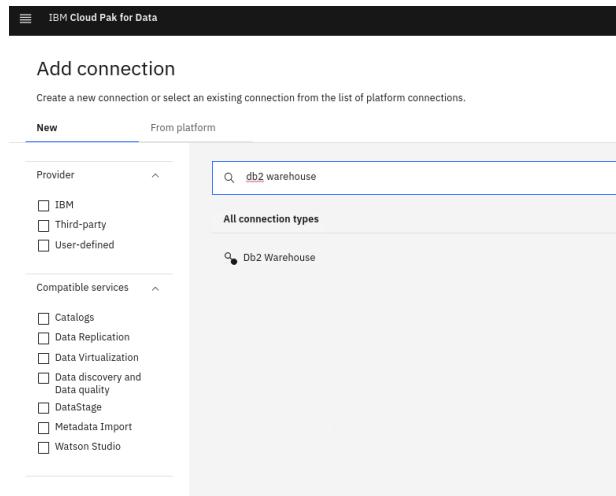
- 1) If you have an existing project, open it. If you don't have an existing project, click **Create a project** on the home page or click New project on your **Projects** page.
- 2) Select **Analytics Project** and Next.
- 3) **Create an empty** project.
- 4) On the Create a project screen, add a name and optional description for the project.
- 5) Click Create.

Task 1: Build Model

Task 1.1: Connect to the Data Warehouse.

Before you build the model using Auto AI, you need to connect the data to the project first. So, the data we will use is already located in the DB2 Data Warehouse which we have already loaded data in the previous step. To connect the Data Warehouse, follow the steps below;

- 1) Go to the project,
- 2) Set up the project connection first for CP4D communicate with DB2 Data Warehouse to get data for model development.
 - o Add to project.
 - o Go to connection and add connection.
 - o In the connection page find “DB2 Warehouse”, click and select.



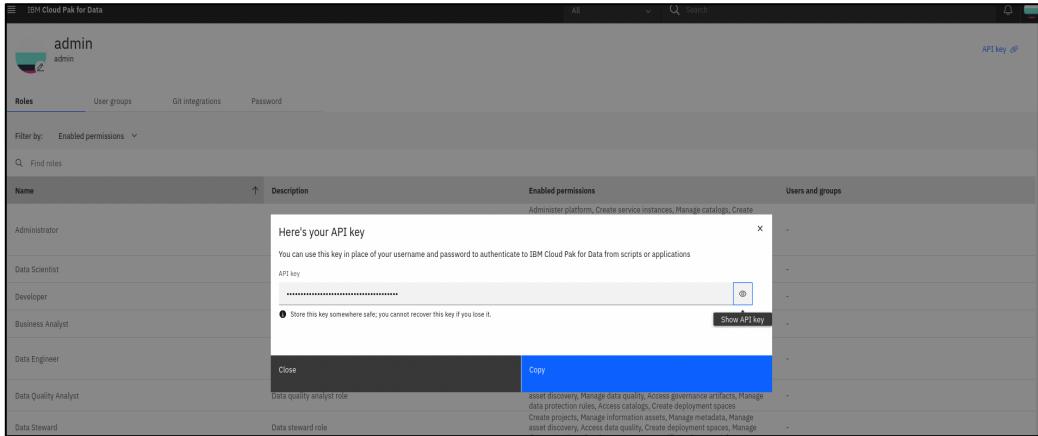
- Open another window and navigate to DB2 data warehouse instance. Click on the instance name that used for storing training_data and click to see the details.

About this database		Storage	
Database name	BLUDB	Storage class (System storage)	managed-nfs
Database type	db2wh	Size (System storage)	100 GiB
Database software version	11.5.7.0-cn3-x86_64	Storage class (User storage)	managed-nfs
Processor	x86-64	Size (User storage)	100 GiB
Deployment id	db2wh-165957726333163	Storage class (Backup storage)	managed-nfs
Created on	Aug 4, 2022 3:22 AM	Size (Backup storage)	100 GiB
Status	Available	Storage class (Transaction logs storage)	managed-nfs
Nodes		Size (Transaction logs storage)	100 GiB
HOSTNAME	CPU	Storage class (Temporary table spaces storage)	managed-nfs
worker2	6.1 cores 18.0 GiB	Size (Temporary table spaces storage)	100 GiB

- Use the information fill in to the “Create connection: DB2 Warehouse”. After filling all information, then click create.
- If the data warehouse is connected to data based properly, you will see the screen like this.

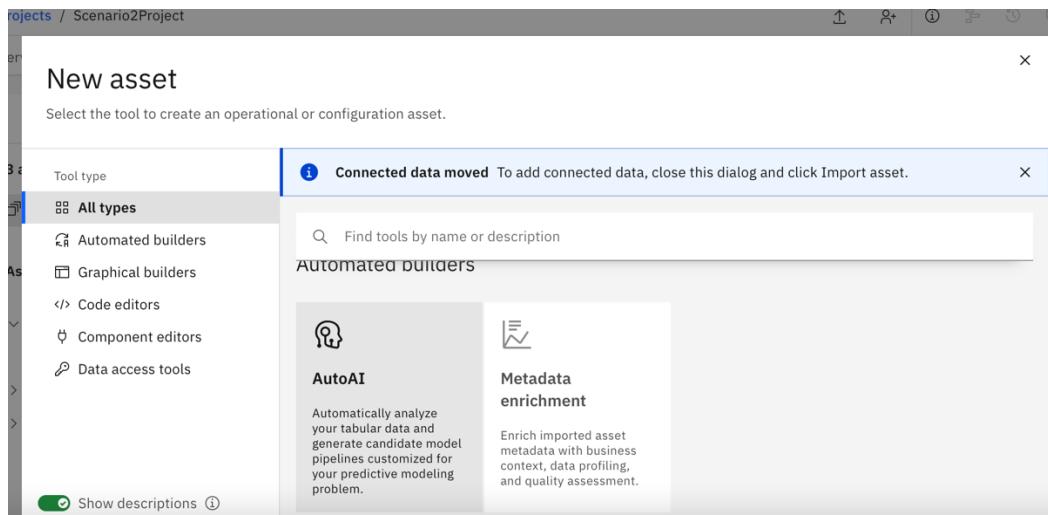
Name	Type	Created by	Last modified
Data Warehouse	Connection	admin	Aug 21, 2022, 10:42 PM

Note: If you need to connect via using API key, please generate the API key by going to your “profile and setting” which located in the top right corner. Then generate API key.

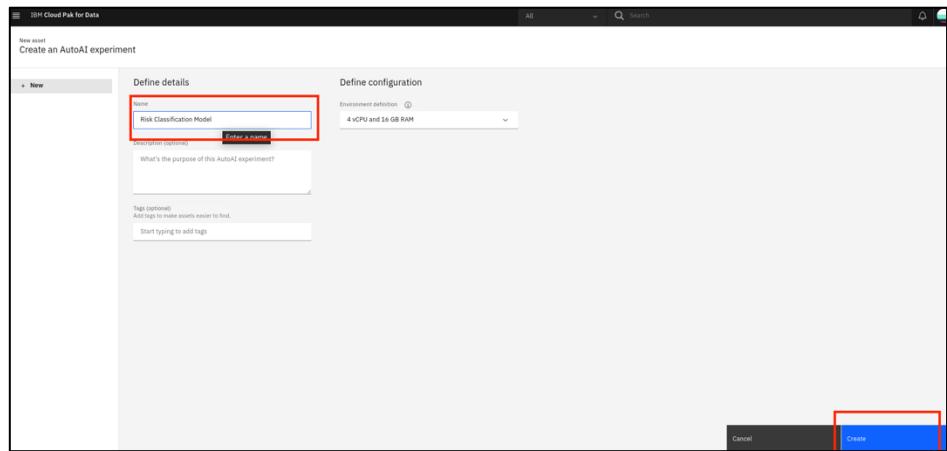


Task 1.2: Build the model using Auto AI

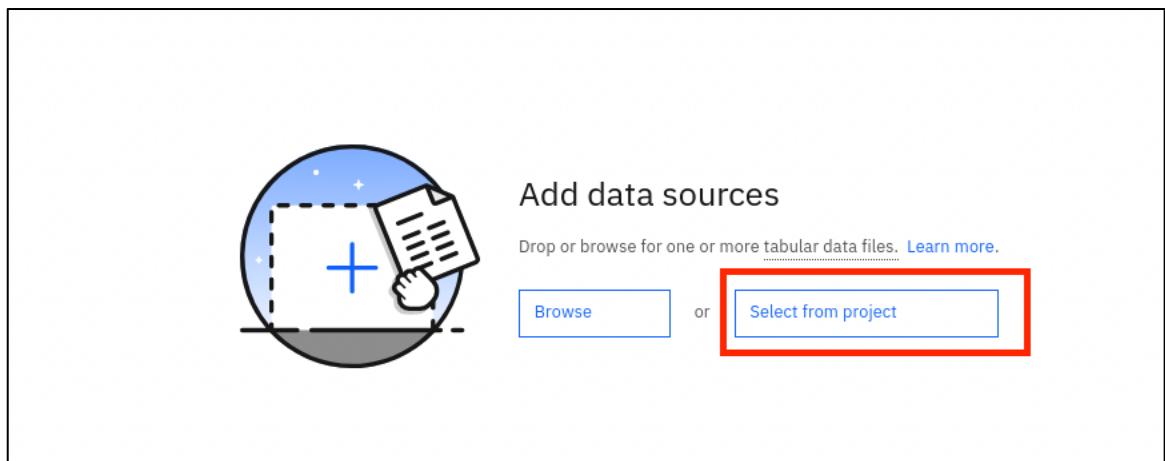
- 1) In the project page under asset tab, go to the right hand side click “ADD to Project”
- 2) Add AutoAI Experiment to the project.



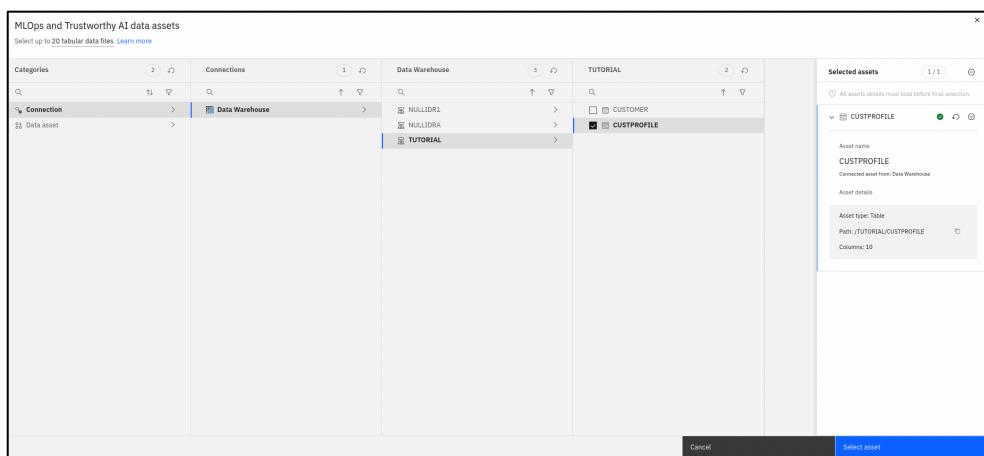
- 3) Give the name of the AutoAI Experiment and then create.



- 4) Add the data source, in our case we will use the connection from our project. So, choose select from project.

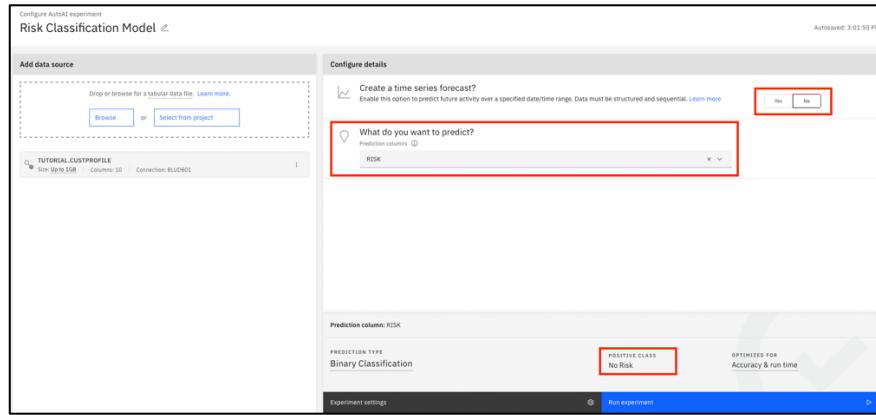


- 5) Select “Connection” > “Data Warehouse” > “TUTORIAL” > “CUSTOMERPROFILE” and select asset.



- 6) In the configuration page
- o Create a time series forecasting? Select “No”, because we will do the classification model to classify the customer “Risk”

- What do you want to predict? Choose the target variable in this case is “Risk”
- Select the Positive Class to be “No RISK”. This is what we are interested in.
- Then run the experiment.



- 7) During the AutoAI Experiment has been run, you can see the Experiment Summary. Please explore the AutoAI Page and list down any questions you have.
- 8) In the pipeline leader board, you can see the eight model pipeline are generated and they're already ranked by performance (Accuracy).

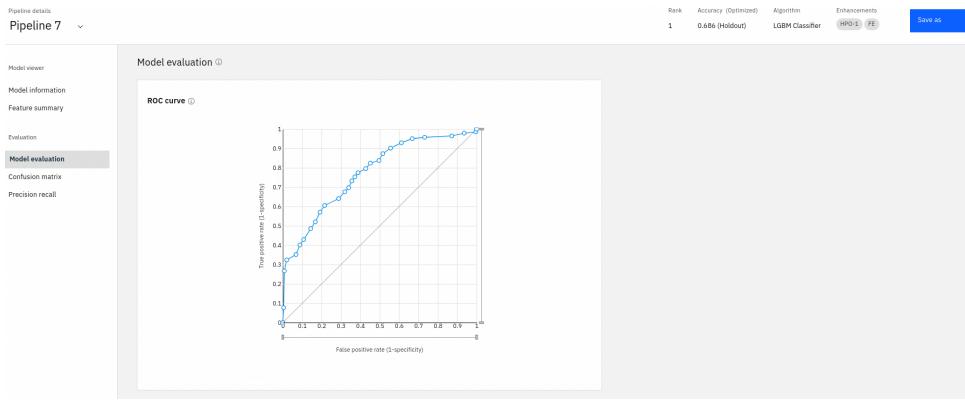
Rank	Name	Algorithm	Accuracy (Optimized) Cross Validation	Enhancements	Build time
1	Pipeline 7	LGBM Classifier	0.733	HPO-1 FE	00:00:17
2	Pipeline 8	LGBM Classifier	0.733	HPO-1 FE HPO-2	00:00:28
3	Pipeline 3	Snap SVM Classifier	0.731	HPO-1 FE	00:00:14
4	Pipeline 4	Snap SVM Classifier	0.731	HPO-1 FE HPO-2	00:00:01
5	Pipeline 6	LGBM Classifier	0.727	HPO-1	00:00:16
6	Pipeline 1	Snap SVM Classifier	0.720	None	00:00:01
7	Pipeline 2	Snap SVM Classifier	0.720	HPO-1	00:00:01
8	Pipeline 5	LGBM Classifier	0.705	None	00:00:01

- 9) Under name click the top pipeline in your leader board to see the information of the model.

Pipeline leaderboard ▾

Rank	Name	Algorithm	Accuracy (Optimized) Cross Validation	Enhancements	Build time
1	Pipeline 7	LGBM Classifier	0.733	HPO-1 FE	00:00:17
2	Pipeline 8	LGBM Classifier	0.733	HPO-1 FE HPO-2	00:00:28
3	Pipeline 3	Snap SVM Classifier	0.731	HPO-1 FE	00:00:14

- 10) You can see the details of model performance. The model information, feature summary, model evaluation and metric are generated for data scientist to examine.



Congratulation!!! You have done for the model building part.

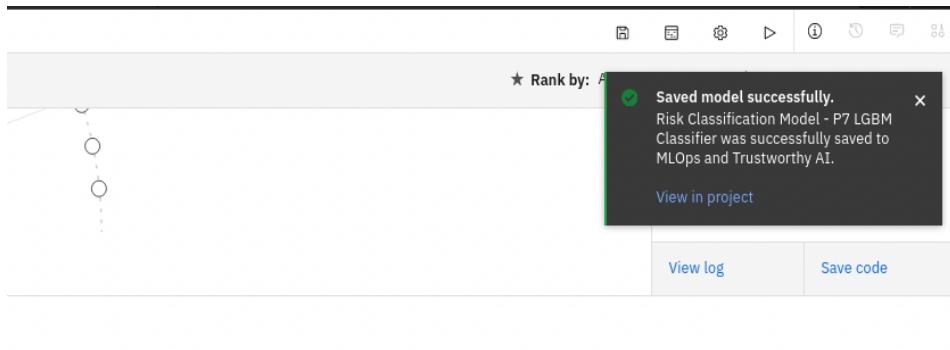
Task 2: Save and Deploy Model

- 1) Go back to the model leader board.
- 2) Click “Save as” on the top rank model.

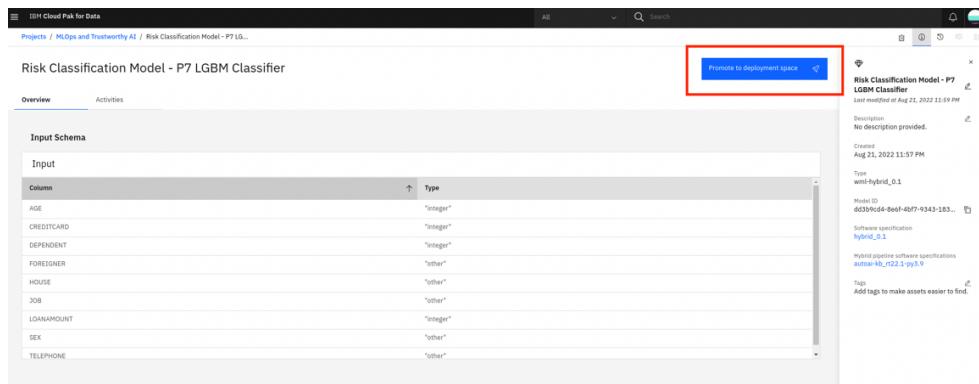


- 3) Go back to the model leader board.
- 4) Click “Save as” on the top rank model.
- 5) We keep everything default. We save the model as Model and Use the default name which generated by system. Click Create.

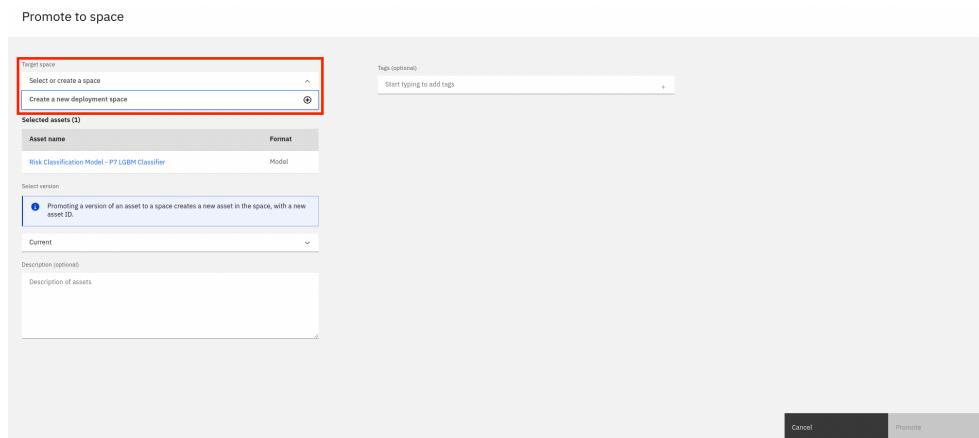
- 6) Once you save, it will back to AutoAI page, click view in project.



- 7) Review input schema and model information. On the top right corner, click "Promote to deployment space" to promote our model to deployment space (Watson Machine Learning).



- 8) Under target space, please click to create a new deployment space.



- 9) Give the deployment space name and then create.

Create a deployment space

Use a space to collect assets in one place to create, run, and manage deployments

Define space details

Name	Credit Risk
Description (optional)	Deployment space description

Deployment space tags (optional) (1)

Add a tag

Create

- 10) Back to promote to space page, select the target space that we just already created and keep everything default. Then click promote to promote our model.

Promote to space

Target space

Credit Risk

Go to the model in the space after promoting it

Selected assets (1)

Asset name	Format
Risk Classification Model - P7 LGBM Classifier	Model

Select version

Promoting a version of an asset to a space creates a new asset in the space, with a new asset ID.

Current

Description (optional)

Promote

- 11) Once you promote it, it will lead you to the project page. You can see your model is saved and on the top right corner click "go to the deployment space"

IBM Cloud Pak for Data

Projects / ML Ops and Trustworthy AI

Overview Assets Environments Jobs Access control Settings

All Search

What assets are you looking for?

Data assets

0 assets selected.

Name	Type	Created by	Last modified
Data Warehouse	Connection	admin	Aug 21, 2022, 10:42 PM

AutoAI experiments

Name	Status	Model type	Last modified
Risk Classification Model	Completed	Binary Classification	Aug 21, 2022, 11:22 PM

Models

Watson Machine Learning models

Name	Type	Software specification	Last modified
Risk Classification Model - P7 LGBM Classifier	wml-hybrid_0.1	hybrid_0.1	Aug 21, 2022

Success message

Successfully promoted Risk Classification Model - P7 LGBM Classifier to the deployment space. Go to the deployment space to explore the assets for deployment.

- 12) On the deployment space under asset tab, you can see your model is registered to the repository. Click at your model name.

The screenshot shows the IBM Cloud Pak for Data interface. The top navigation bar includes 'IBM Cloud Pak for Data', 'All', a search bar, and various icons. Below the navigation is a breadcrumb trail: 'Deployments / Credit Risk'. The main content area has tabs for 'Overview', 'Assets' (which is selected), 'Deployments', 'Jobs', and 'Manage'. A search bar says 'What assets are you looking for?'. Under the 'Models' section, there is one entry: 'Risk Classification Model - P7 LGBM Classifier' (Type: wml-hybrid_0.1, Software specification: hybrid_0.1, Last modified: Aug 22, 2022 1:49 AM). On the right side, there is an 'Import model' button and a file upload area with instructions: 'Drop files here or browse for files to upload.' A note below says 'Stay on the page until upload completes. Incomplete uploads are cancelled.'

13) Click new deployment to deploy your model.

The screenshot shows the deployment details for the 'Risk Classification Model - P7 LGBM Classifier'. The top navigation bar shows 'Deployments / Credit Risk'. The main content area displays the model's details: 'Risk Classification Model - P7 LGBM Classifier' (Created: Aug 22, 2022 1:49 AM, Type: wml-hybrid_0.1, Model ID: 63070794-c195-4ab3-9ba4-d923..., Software specification: hybrid_0.1). It also shows 'Hybrid pipeline software specifications' and 'Description' (No description provided). Below the details, there is a 'Source asset details' section. At the top right, there is a 'New deployment' button, which is highlighted with a red box.

14) Under create a deployment page.

- Choose deployment type to be “Online” because we need the model to return result on-line (real-time) upon inference.
- Give the name of this deployment
- Create.

The screenshot shows the 'Create a deployment' dialog. At the top, it says 'Create a deployment'. Below that, it shows the 'Associated asset' as 'Risk Classification Model - P7 LGBM Classifier'. The 'Deployment type' section has two options: 'Online' (selected) and 'Batch'. The 'Online' option is described as 'Run the model on data in real-time, as data is received by a web service.' The 'Batch' option is described as 'Run the model against data as a batch process.' Below the deployment type, there is a 'Name' field containing 'Credit Risk Model LGBM Deployment', which is highlighted with a red box. Further down, there are fields for 'Deployment serving name', 'Description', 'Deployment description', and 'Tags'. At the bottom, there is a 'Software specification' dropdown set to 'hybrid_0.1'. On the far right, there are 'Cancel' and 'Create' buttons, with the 'Create' button highlighted with a red box.

15) Wait until the model is deployed successfully and then please click at the name of the model to see the details of deployment.

The screenshot shows the 'Deployments' section of the IBM Cloud Pak for Data interface. On the left, there's a navigation bar with 'Deployments / Credit Risk /'. Below it, a table lists 'DEPLOYMENT TYPES' with one 'Online' entry and zero 'Batch' entries. The 'Online' row has a single item labeled 'Credit Risk Model LGBM Deployment' with a status of 'Deployed' and a timestamp of 'Aug 22, 2022 2:02 AM'. To the right of the table, detailed information about the deployment is provided, including its name ('Risk Classification Model - P7 LGBM Classifier'), type ('wmn-hybrid_0.1'), model ID ('63076794-c195-4ab3-9ba4-d823...'), software specification ('hybrid_0.1'), and hybrid pipeline software specifications ('autoai-k8_r22.1-py3.9'). There are also sections for description ('No description provided.') and tags ('Add tags to make assets easier to find.'), along with a 'Source asset details' dropdown.

- 16) Under deployment page, you can see the detail of the deployment.

Explore API Reference and Test

- API Reference – The endpoint and code snippets are created for further use in the other front-end system.
- Test – The testing area to test the API and model inference.
- Next step in the test tab, we will test the endpoint and inference. Please enter input data as follow;
 - LOANAMOUNT : 2000
 - SEX : male
 - AGE : 32
 - HOUSE : own
 - CREDITCARD : 1
 - JOB : skilled
 - DEPENDENT : 1
 - TELEPHONE : yes
 - FOREIGNER :no
- Add to list > Predict.
- Check the model can give you the result properly or not.

Congratulation!!! You have done for the model deployment and testing part.

Task 3: Set up Model Monitoring with Watson OpenScales

- 1) Go back to the IBM Cloud Pak for Data Navigation bar. Service > Instance
- 2) Check the OpenScale instance is already created. If not, please contact your facilitator.

Instances							
Filter by: Type ▾ Status ▾							
Name	Type	Created by	vCPU requests	Memory requests (GiB)	Users	Status	Created on
Db2 Warehouse-1 Db2 Warehouse 11.5.7.0-cn3-x86_64	db2wh	admin	6.20	18.25 Gi	1	Green	Aug 4, 2022
Db2 Warehouse-2 Db2 Warehouse 11.5.7.0-cn3-x86_64	db2wh	admin	6.20	18.25 Gi	1	Green	Aug 4, 2022
data-virtualization	dv	admin	11.60	38.70 Gi	1	Green	Apr 5, 2022
data-management-console Data Management Console	dmc	admin	3.70	9.50 Gi	1	Green	Apr 5, 2022
openscale-defaultinstance IBM Watson OpenScale	aios	admin	-	-	1	Green	Apr 5, 2022

- 3) On the right hand-side of OpenScale instance row, you will see triple dot. Click “Open” to the OpenScale Application.
 - 4) On the OpenScale landing page, click “Manual setup”
 - 5) In the system seup, first set up the database.
- o Choose the database type to be “DB2”. We will use the second DB2 Data warehouse which is created for OpenScale (not the same as we use to keep training data)
 - o Go back to service > instance to get the data warehouse credential.

Details: Db2 Warehouse-2																	
My data: Databases / Db2 Warehouse-2 / Details																	
Deployment id	db2wh-1659596976865976																
Created on	Aug 4, 2022 3:09 AM			Storage class (Backup storage)	managed-nfs												
Status	Available			Size (Backup storage)	100 GiB												
Nodes	<table border="1"> <thead> <tr> <th>HOSTNAME</th> <th>CPU</th> <th>MEMORY</th> <th>Storage class (Transaction logs storage)</th> <th>managed-nfs</th> </tr> </thead> <tbody> <tr> <td>master1</td> <td>6.1 cores</td> <td>18.0 GiB</td> <td>Size (Transaction logs storage)</td> <td>100 GiB</td> </tr> </tbody> </table>							HOSTNAME	CPU	MEMORY	Storage class (Transaction logs storage)	managed-nfs	master1	6.1 cores	18.0 GiB	Size (Transaction logs storage)	100 GiB
HOSTNAME	CPU	MEMORY	Storage class (Transaction logs storage)	managed-nfs													
master1	6.1 cores	18.0 GiB	Size (Transaction logs storage)	100 GiB													
	<table border="1"> <thead> <tr> <th>Storage class (Temporary table spaces storage)</th> <th>managed-nfs</th> </tr> </thead> <tbody> <tr> <td>Size (Temporary table spaces storage)</td> <td>100 GiB</td> </tr> </tbody> </table>							Storage class (Temporary table spaces storage)	managed-nfs	Size (Temporary table spaces storage)	100 GiB						
Storage class (Temporary table spaces storage)	managed-nfs																
Size (Temporary table spaces storage)	100 GiB																
	<table border="1"> <thead> <tr> <th>Access information</th> </tr> </thead> <tbody> <tr> <td>JDBC Connection URL</td> <td>jdbc:db2://<CLUSTER_ACCESSIBLE_IP>:32236/BLUDB;user=-;password=-<password>;securityMechanism=0;encryptionAlgorithm=2</td> </tr> <tr> <td>JDBC Connection URL (SSL)</td> <td>jdbc:db2://<CLUSTER_ACCESSIBLE_IP>:30057/BLUDB;user=-;password=-<password>;securityMechanism=0;sslConnection=true;encryptionAlgorithm=2</td> </tr> </tbody> </table>							Access information	JDBC Connection URL	jdbc:db2://<CLUSTER_ACCESSIBLE_IP>:32236/BLUDB;user=-;password=-<password>;securityMechanism=0;encryptionAlgorithm=2	JDBC Connection URL (SSL)	jdbc:db2://<CLUSTER_ACCESSIBLE_IP>:30057/BLUDB;user=-;password=-<password>;securityMechanism=0;sslConnection=true;encryptionAlgorithm=2					
Access information																	
JDBC Connection URL	jdbc:db2://<CLUSTER_ACCESSIBLE_IP>:32236/BLUDB;user=-;password=-<password>;securityMechanism=0;encryptionAlgorithm=2																
JDBC Connection URL (SSL)	jdbc:db2://<CLUSTER_ACCESSIBLE_IP>:30057/BLUDB;user=-;password=-<password>;securityMechanism=0;sslConnection=true;encryptionAlgorithm=2																

- o Go back to the openpage data base configuration page and fill in the credential information. Then get connect.

System setup

Connect to a database, machine learning providers, and integrated services. Optionally enable batch support.

Database

Machine learning providers

Batch support (optional)

Integrations (optional)

Users & roles

Required

Database

Select a database to store your model transactions and model evaluation results. For Db2 options that are part of your cluster, see [Services](#), [Data Sources](#) where you find options, such as Db2 Warehouse and Db2 Advanced Enterprise Server Edition. For an external database, you can use IBM DB2 Database.

Database type

Db2

Database location

Other

Hostname or IP address

master1

Port

32236

Use SSL

Database

BLUDB

Username

admin

Password

Connect

- Check the database is connected to OpenScale successfully.
- Then scroll down a bit, under schema choose Auto-create a new schema, then save.

Database

BLUDB

Username

admin

Password

Connect

Schema

No schema found

Auto-create a new schema

Cancel

Save

- Finish OpenScale database setup.

6) Setup machine learning provider.

- On the machine learning providers page, click “add machine learning provider”

System setup

Connect to a database, machine learning providers, and integrated services. Optionally enable batch support.

Database

Machine learning providers

Batch support (optional)

Integrations (optional)

Users & roles

Required

Machine learning providers

Description

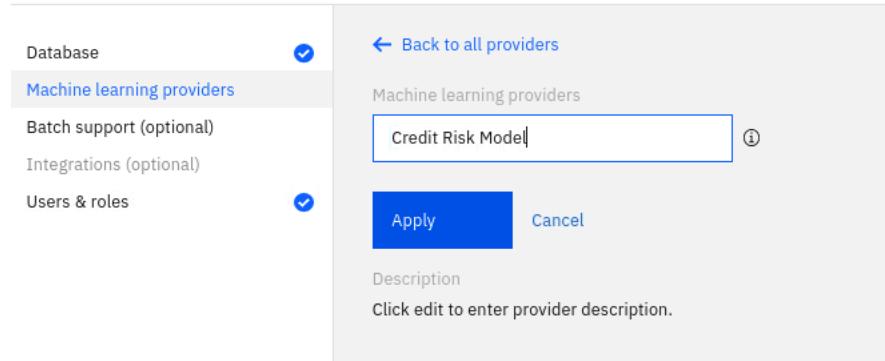
Watson OpenScale connects to deployed models stored in a machine learning environment.

Add machine learning provider

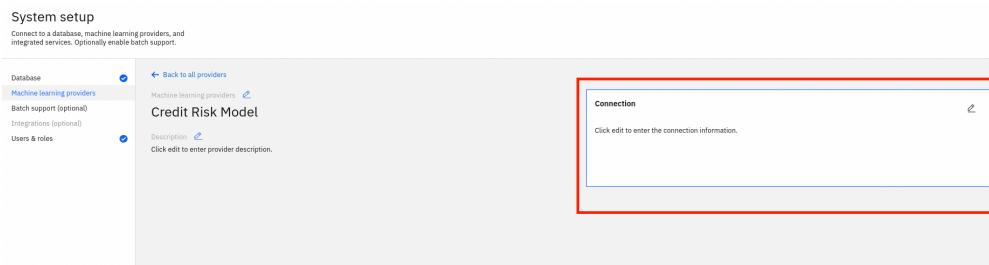
- Give the machine learning provider name.

System setup

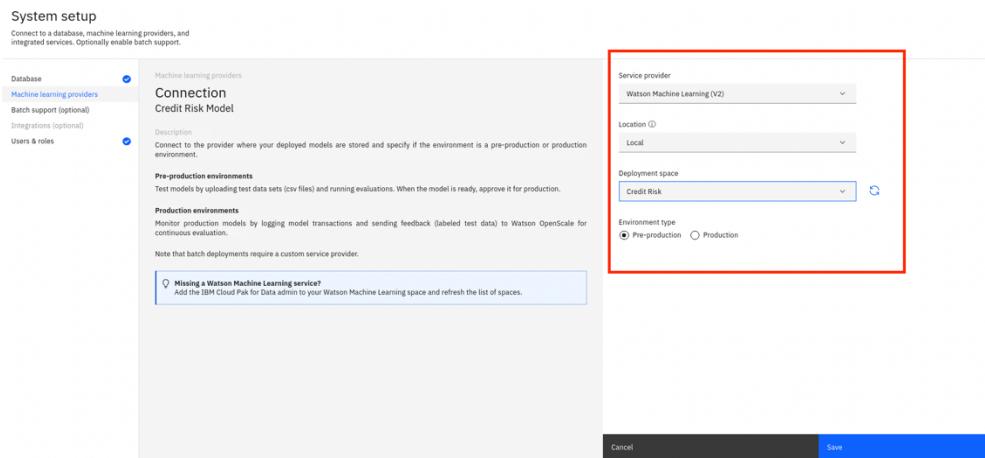
Connect to a database, machine learning providers, and integrated services. Optionally enable batch support.



- Click connection to set up the connection with machine learning provider.

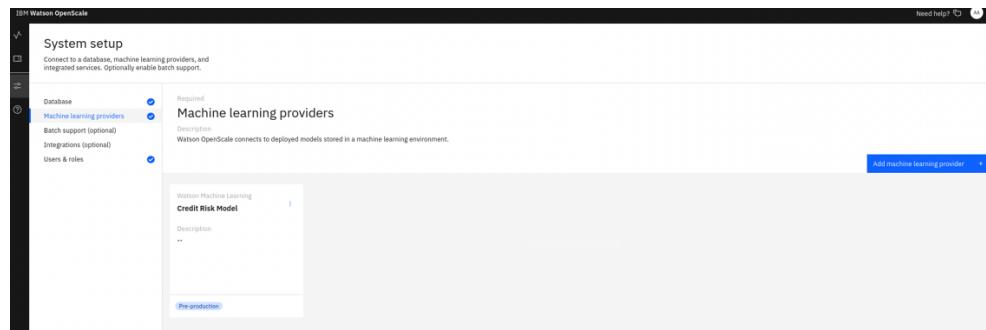


- Under connection set up page, please configure it as follows;
 - Service provider : Watson Machine Learning (V2)
 - Location : Local
 - Deployment space : <your deployment space name>
 - Save.



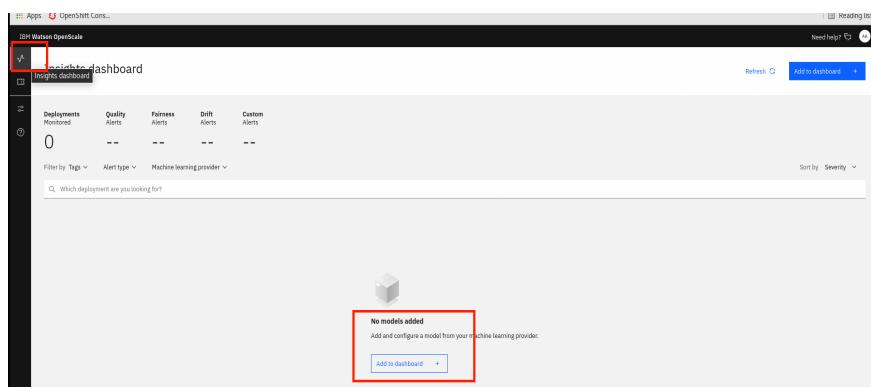
- Machine Learning Provider is setup successfully.

-

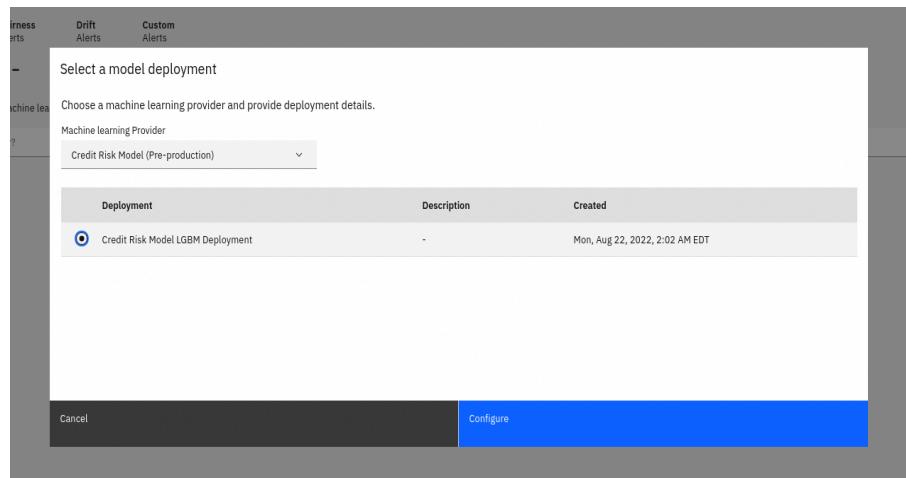


7) Setup OpenScale dashboard.

- Click insight dashboard on the left hand-side of the menu bar.
- From the insight dashboard page, click add to dash board.



- Add our machine learning provider to the dashboard.



- Our machine learning provider is added to the dashboard successfully.

The screenshot shows the 'Insights dashboard' for IBM Watson OpenScales. At the top, there are tabs for 'Deployments Monitored', 'Quality Alerts', 'Fairness Alerts', 'Drift Alerts', and 'Custom Alerts'. Below these, a summary table shows the count '1' for each category. A search bar at the bottom allows filtering by 'Tags', 'Alert type', and 'Machine learning provider'. On the left, a sidebar displays deployment details for 'Credit Risk Model Credit Risk Model LGBM Dep...'. It lists 'Issues' such as 'Quality', 'Fairness', 'Drift', and 'Custom', all marked as '---'. Buttons for 'Pre-production' and 'Evaluation pending' are also present.

8) Configure monitors.

- From the machine learning provider, click triple dots and configure monitor.

This screenshot shows the deployment configuration interface for the 'Credit Risk Model Credit Risk' deployment. The main panel displays deployment statistics: 'Deployments Monitored' (1), 'Quality Alerts' (---), 'Fairness Alerts' (---), 'Drift Alerts' (---), and 'Custom Alerts' (---). A search bar at the bottom asks 'Which deployment are you looking for?'. Below the main panel, a sidebar for 'Issues' shows a list with one item: 'Configure monitors', which is highlighted with a red rectangle. Other options in the sidebar include 'View details', 'Remove deployment', and buttons for 'Pre-production' and 'Evaluation pending'.

- Model details configuration.

- Model input:
 - Data type : Numeric/categorical
 - Algorithm type : Binary Classification
 - Save and continue.
- Training data:
 - In reality, you need to create the cloud object storage to keep the training data for Watson OpenScales as following actions. However, in this practicum workshop we already done for you and check your service credential in the next step.
 - Going to <https://cloud.ibm.com> and login to IBM console.
 - From the navigation bar left hand-side, select the resource list.
 - Find Storage and select your object storage.
 - Create bucket

- Choose quickly get start
- Give a bucket name for example “trainingdataforwos”. Please note that the cloud object storage name must be unique across IBM cloud. If your bucket name is not unique, please change to be an unique one.
- Next and the bucket will be created.
- Upload the training data to the “trainingdataforwos” bucket
- Back to service credential to see the credential information.
- Object Storage service credential:


```
{
  "apikey": "AHrxCLY3jMGTAfzVZ1nxOaNsOFwvH4a8ucsW-FTmGTVH",
  "endpoints": "https://control.cloud-object-storage.cloud.ibm.com/v2/endpoints",
  "iam_apikey_description": "Auto-generated for key crn:v1:bluemix:public:cloud-object-storage:global:a/432ff2d283604e4f8a4e4c230de28f62:e611df9d-a172-478d-8d25-642063ad96b8:resource-key:c9ee16e2-e488-4ff5-ba0c-1777818100d7",
  "iam_apikey_name": "trainingdatafor-practicum",
  "iam_role_crn": "crn:v1:bluemix:public:iam::::serviceRole:Writer",
  "iam_serviceid_crn": "crn:v1:bluemix:public:iam-identity::a/432ff2d283604e4f8a4e4c230de28f62::serviceid:ServiceId-314f8d63-b8c1-433a-b544-580374d9663b",
  "resource_instance_id": "crn:v1:bluemix:public:cloud-object-storage:global:a/432ff2d283604e4f8a4e4c230de28f62:e611df9d-a172-478d-8d25-642063ad96b8::"
}
```
- Back to model detail > Training data. Then configure and fill in Object Storage credential
 - Storage type : Database or cloud storage
 - Location : Cloud object Storage
 - Resource instance id : <your object storage resource instance id>
 - API Key : <your object storage API key>
 - Click connect
 - Choose the bucket
 - Choose the data set.
 - Next

x

Storage type

Database or cloud storage

Location

Cloud Object Storage

Resource instance ID

crn:v1:bluemix:public:cloud-object-storage:global:a/432ff2d283604e4f8a4e4c

API key

.....

Bucket

trainingdataforwos

Data set

training_data.csv

x v

Cancel Next

- Select label columns
 - Choose “Risk”

- **Next**

Select the label column	
Features (10)	Type
AGE	81
CREDITCARD	81
DEPENDENT	81
FOREIGNER	A
HOUSE	A
JOB	A
LOANAMOUNT	81
RISK	A
SEX	A
TELEPHONE	A

Back Next

- **Select the training features**

- Choose all feature.
- **Next**

Select the training features Selected features 9

<input checked="" type="checkbox"/> Features (9)	Type	Categorical
<input checked="" type="checkbox"/> AGE	continuous	<input type="checkbox"/>
<input checked="" type="checkbox"/> CREDITCARD	continuous	<input type="checkbox"/>
<input checked="" type="checkbox"/> DEPENDENT	continuous	<input type="checkbox"/>
<input checked="" type="checkbox"/> FOREIGNER	categorical	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> HOUSE	categorical	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> JOB	categorical	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> LOANAMOUNT	continuous	<input type="checkbox"/>
<input checked="" type="checkbox"/> SEX	categorical	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> TELEPHONE	categorical	<input checked="" type="checkbox"/>

Items per page: 25 | 1 - 9 of 9 items | 1 of 1 page | < | >

Back Next

- Configure Scoring
 - Scoring Method : JSON payload
 - Download JSON_Payload_Example from Github
https://github.com/IBM-CSM-SG/ASEANZK-CP4D-Practicum/blob/main/scenario2/data/JSON_Payload_Example
 - Open the file and Copy the content.
 - Paste the content to the blank under enter a sample request and responses.
 - Click send now
 - Next.

Scoring method

JSON payload

Enter a sample request and response. Replace the fields and values to match your model input and output. [Learn more.](#)

```
{
  "values": [
    {
      "2000",
      "male",
      32,
      "own",
      1,
      "skilled",
      1,
      "yes",
      "no",
      [0.7541479304432869,
       0.24585206955671313
     ],
      "No RISK"
    }
  ]
}
```

Send now

Back **Next**

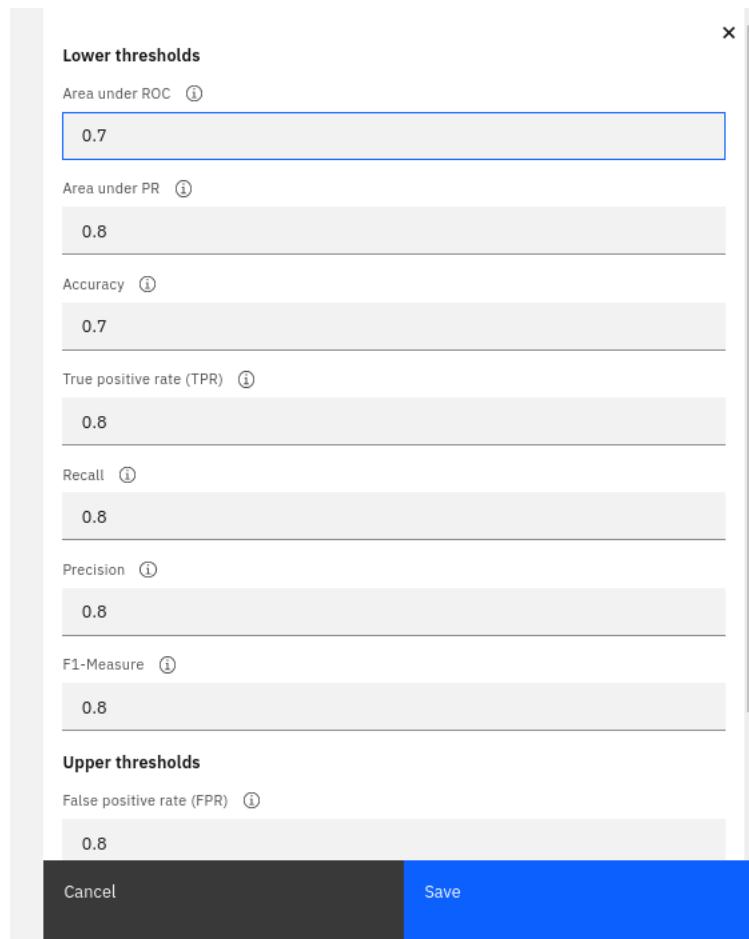
- Specify the model output.
 - Ensure you select prediction and probability.
 - Save

Use the checkboxes to make one or more selections

Features (2)	Type	Prediction	Probability
prediction	A	<input checked="" type="checkbox"/>	
probability	B		<input checked="" type="checkbox"/>

- Explainability
 - Explainability will be configured automatically, once you completed the previous steps correctly.
- Quality
 - Click Quality
 - Quality threshold > edit
 - The lower threshold and upper threshold is the configuration if the monitor metric gets lower up higher than the threshold we set, it will be alerted. Try to change some lower threshold (the
 - Area under ROC : 0.7
 - Accuracy : 0.7

- Save



- Sample size is the size of sample to get evaluate. We will keep 1,000 as default.
- Drift
 - Click Drift
 - Drift Model > edit
 - Training option
 - Choose Train in Watson OpenScale (Suitable when you connected your training data to Watson OpenScale and it is less than 500 MB)
 - Drift thresholds
 - There are two drift metrics, drop in accuracy and drop in data consistency. You can set the thresholds to get alert. We will set to 5percent., so it will give alert if drop in accuracy or data consistency greater than 5 percent.
 - Drop in accuracy : 5%
 - Drop in data consistency 5%

- Sample size is the size of sample to get evaluate. We will keep 1,000 as default.
- Save
- Fairness
 - Click Fairness
 - Favourable Outcome > edit
 - Add 0 and 1 value
 - Set 0 to favourable outcome
 - Set 1 to unfavourable outcome
 - Save

Select the favorable outcomes

1	-	+	Add value
			Favorable values 1 Unfavorable values 1
<input type="text"/>			
Values	Favorable	Unfavorable	
0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
1	<input type="checkbox"/>	<input checked="" type="checkbox"/>	

Cancel	Save
--------	------

- Sample size is the size of sample to get evaluate. We will keep 1,000 as default.
- Select the fields to monitoring
 - Select SEX (We will monitor bias based on SEX attributes)

Select one or more fields

Selected fields 1

Fields	Recommended	Type
<input type="checkbox"/> LOANAMOUNT		
<input checked="" type="checkbox"/> SEX		
<input type="checkbox"/> AGE		
<input type="checkbox"/> HOUSE		
<input type="checkbox"/> CREDITCARD		
<input type="checkbox"/> JOB		
<input type="checkbox"/> DEPENDENT		
<input type="checkbox"/> TELEPHONE		
<input type="checkbox"/> FOREIGNER		

- Specify the monitored groups for [SEX]
 - Add custom value
 - Type “male”, then add value. The check reference for male.
 - Type “female”, then add value. The check monitored for female.
 - Keep fairness alert threshold as default.
 - Save

Select the groups to monitor [SEX]

Add custom value

female Add value

Monitored group 1 Reference group 1

Values	Monitored	Reference	Recommended
female	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
male	<input type="checkbox"/>	<input checked="" type="checkbox"/>	

Set fairness alert threshold [SEX]

80

Congratulation!!! You have done for the model monitoring set up part.

Please note that we will not see any alert in the dashboard yet because there are no any new payload load to the model. Once, the payload reached out the threshold sample that we set (1,000 sample), it will evaluate and show the results.