



AI-Stack Express

Machine Learning Management Platform

User Manual

Version : 4.4

User Guide

1. Introduction	3
2. Login	4
3. Interface	5
4. Service Activation	6
5. Machine Learning Project	8
1. Project Reviewing List	8
2. Project List	9
3. Project Detail	17
4. Container Management	19
4.1 Create Container	19
General Creation	19
Batch Creation	26
4.2 Container List	28
Using container services	30
Provisioning	34
4.3 Delete Container	35
4.4 Create Custom Image	36
5. Job Management	37
5.1 Job List	37
5.2 Job Scheduling	42
6. GPU Utilization Rate	49
7. Storage Management	50
8. Advanced Settings	52
8.1 Custom Image	52

User Guide

Create Custom Image	52
Use Custom Image	55
8.2 Job Template	56
Create Job Template	56
Use Job Template	58
8.3 Key Pair	61
Create Key Pair	61
Import Keys	62
Delete Keys	62
9. Cost Center	63
6. Account Management	64
1. General Information	64
2. Order Record	65
3. Settings	66

1. Introduction

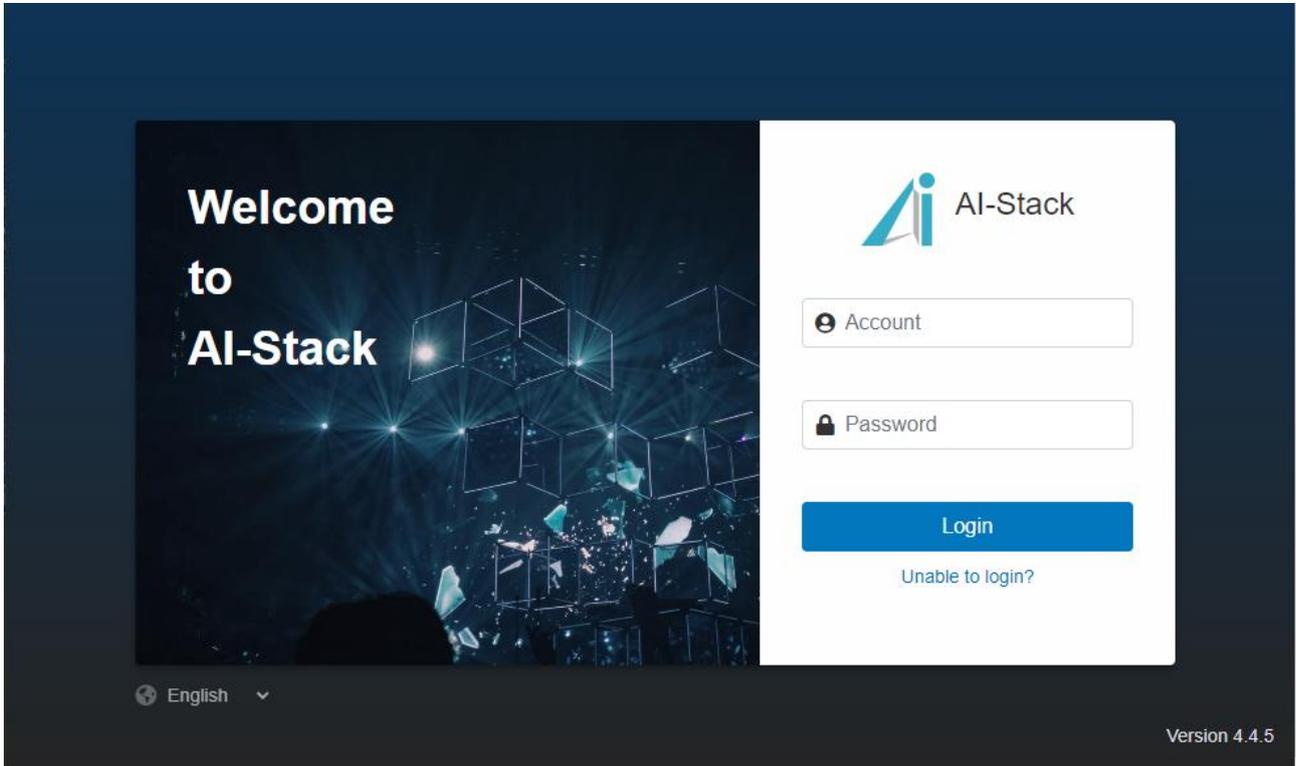
This manual introduces the features and the standard operation of AI-Stack – a Machine Learning Management Platform for platform users. It is recommended to use Chrome versions 60 or above / Firefox versions 60 or above. If the browser version fails to meet the requirements, the system may not function properly.

This system provides users with a self-service environment where they can configure container resources including CPU, GPU, memory, network disk, and AI machine learning frameworks (such as TensorFlow) according to their needs. They can also access other system resource information related to their work.

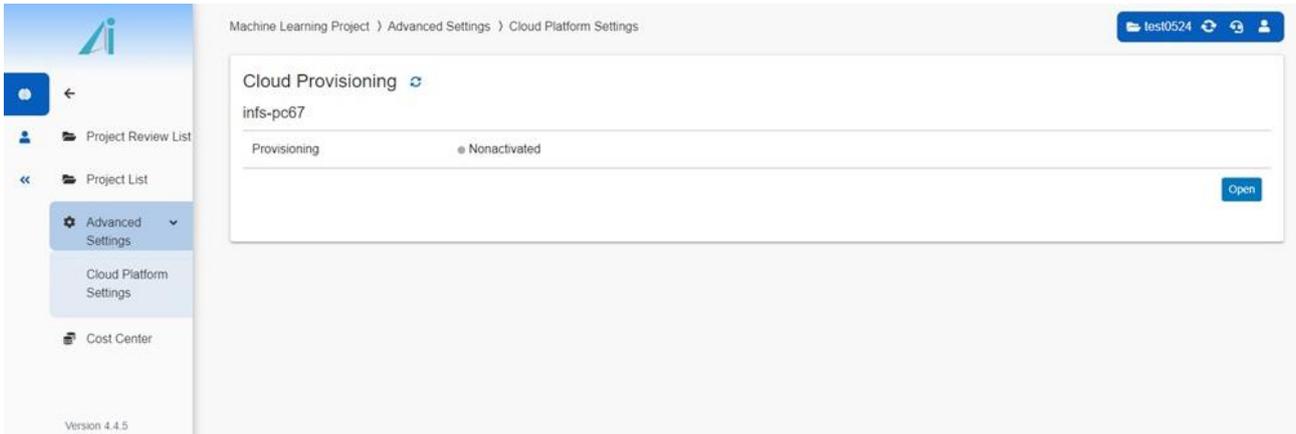
The system operates and organised based on projects which allows users to create project according to specific project goals, departments, small-scale experiments, internal competitions, educational training, event showcases, and various other usage scenarios. This differentiation of projects supports the facilitation of resource allocation based on project requirements. Individual users can also participate in multiple projects simultaneously, ensuring efficient management of resource utilization and budget control.

2. Login

Users enter their account and password at the AI-Stack log in page to log in to the platform.

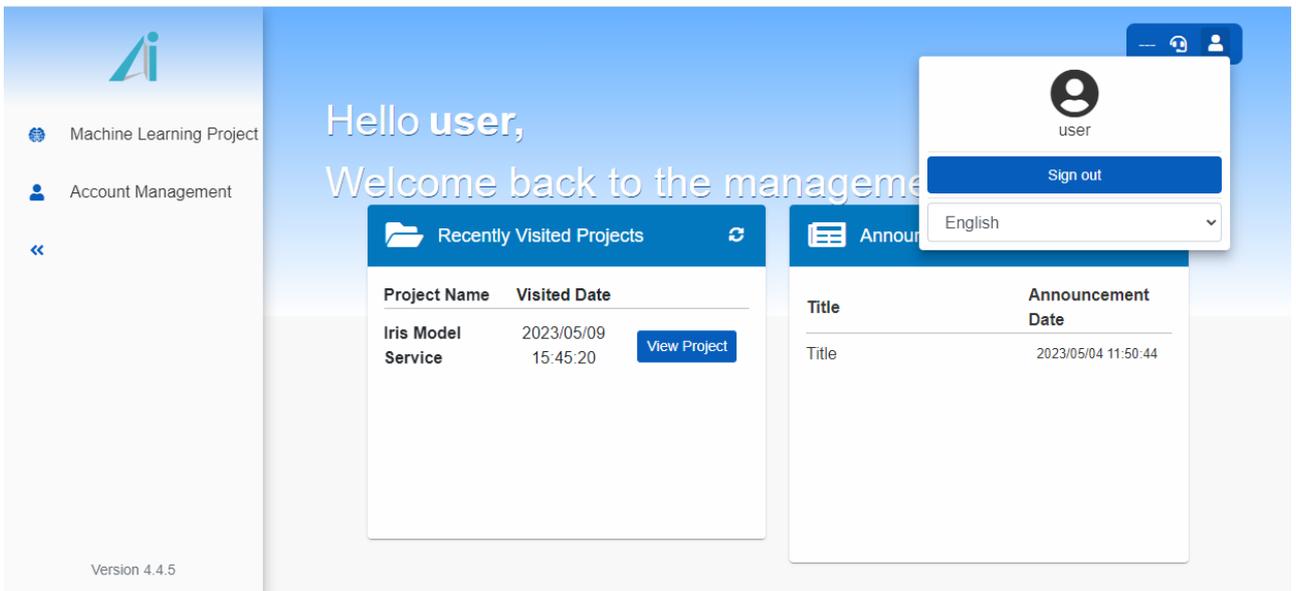


As the platform resources were not activated the first time you logged in, you will land at “Cloud Provisioning” page. To activate the resources, please refer to Section 4 Service Activation.



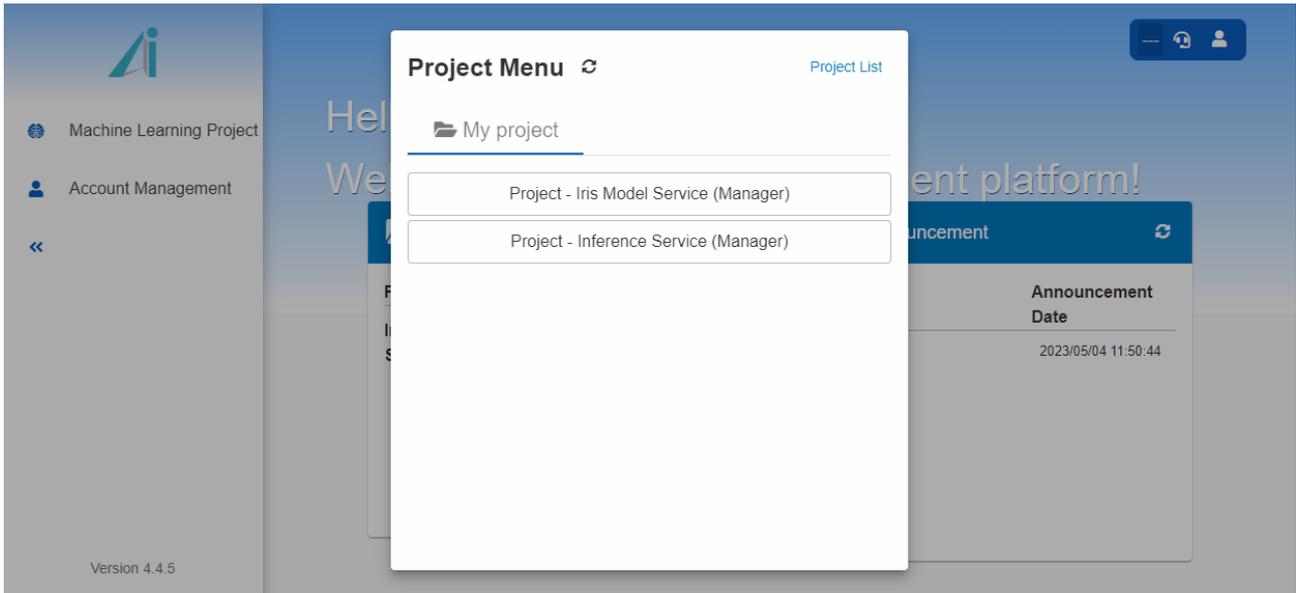
3. Interface

Click on the  logo on the top right corner to view user name, log out the system, or change language settings.



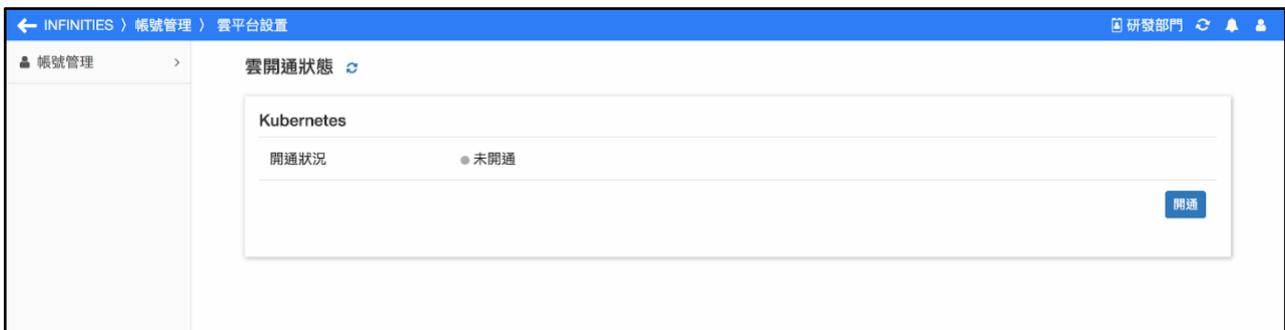
Click  on the top right corner, a user can view the project he is in and see if he is the project manager of the project. Before starting to operate this system, please make sure that you have selected the intended project to avoid mistakenly using resources from different pr

objects with different purposes.

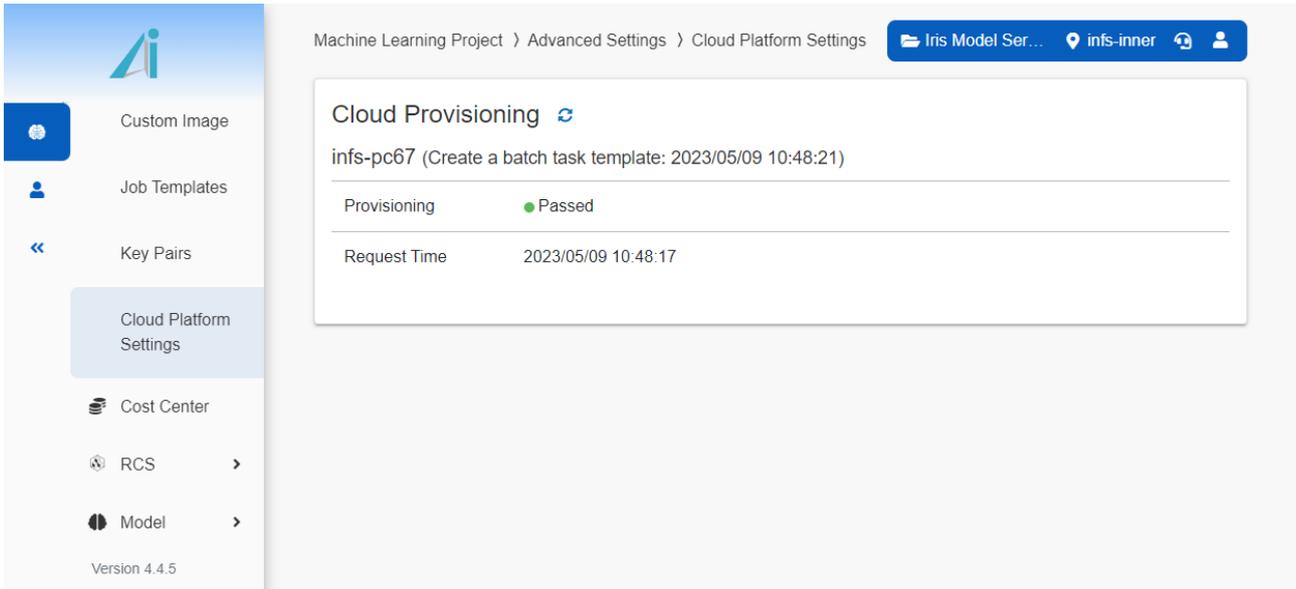


4. Service Activation

The "Cloud Platform Setup" page displays the activation status of the user's resources on this system's platform. As shown in the figure below, when the activation status is displayed as "Not Activated," the user can click the "Activate" button to initiate the resource platform activation process. After clicking the activation button, please wait for the activation process to complete.



Click  to refresh the page until the status becomes "Passed" , and you can start using our services.



The screenshot displays the 'Cloud Platform Settings' page within the INFINITIES interface. The breadcrumb trail at the top reads: 'Machine Learning Project > Advanced Settings > Cloud Platform Settings'. The page title is 'Cloud Provisioning' with a refresh icon. Below the title, the identifier 'infs-pc67' is shown with a note: '(Create a batch task template: 2023/05/09 10:48:21)'. A table below provides the provisioning status:

Provisioning	● Passed
Request Time	2023/05/09 10:48:17

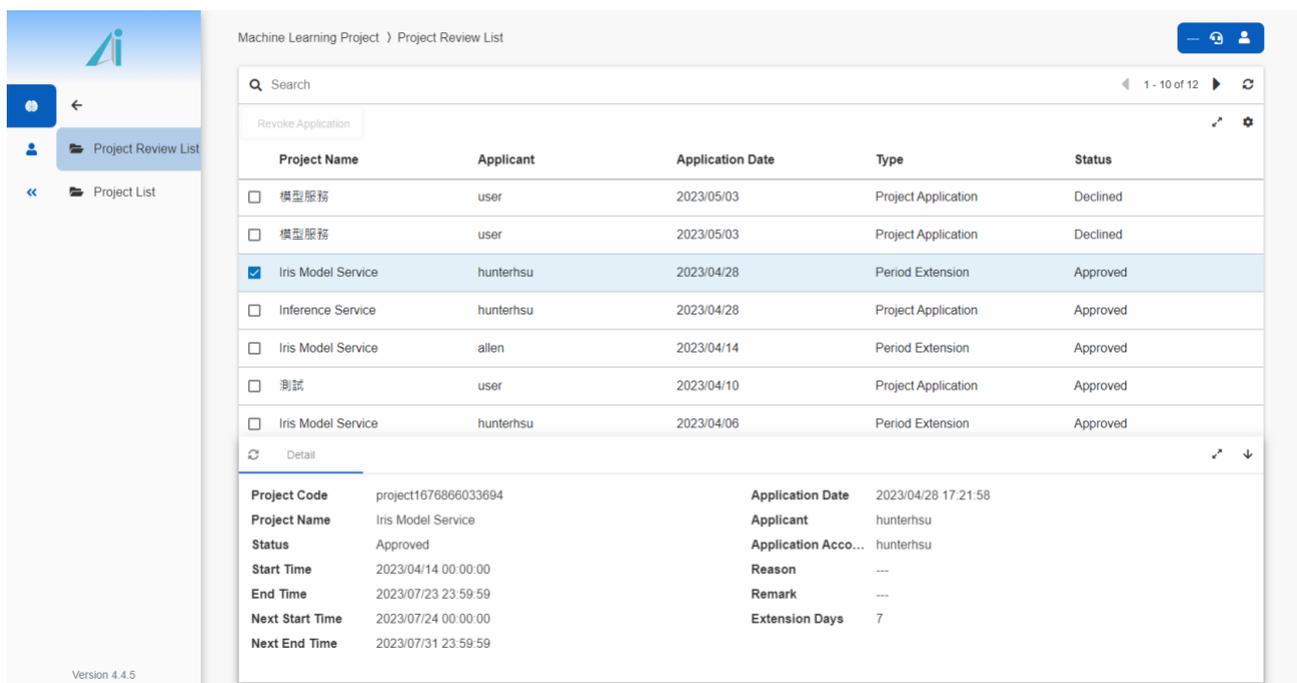
The left sidebar contains navigation options: Custom Image, Job Templates, Key Pairs, Cloud Platform Settings (highlighted), Cost Center, RCS, and Model. The version number 'Version 4.4.5' is visible at the bottom of the sidebar.

5. Machine Learning Project

Users can use the project overview page to view project general information and resource usage status for each project. Projects have start and end dates, and prior to the project end date, the project manager will receive email notifications to extend the project. If the project is not extended after the end date, the platform will release all the resources allocated to this project, which includes but not limited to containers, job, and storage cluster.

1. Project Reviewing List

Users can view their applications in the project reviewing list, including project application, extension application and GPU quota application.

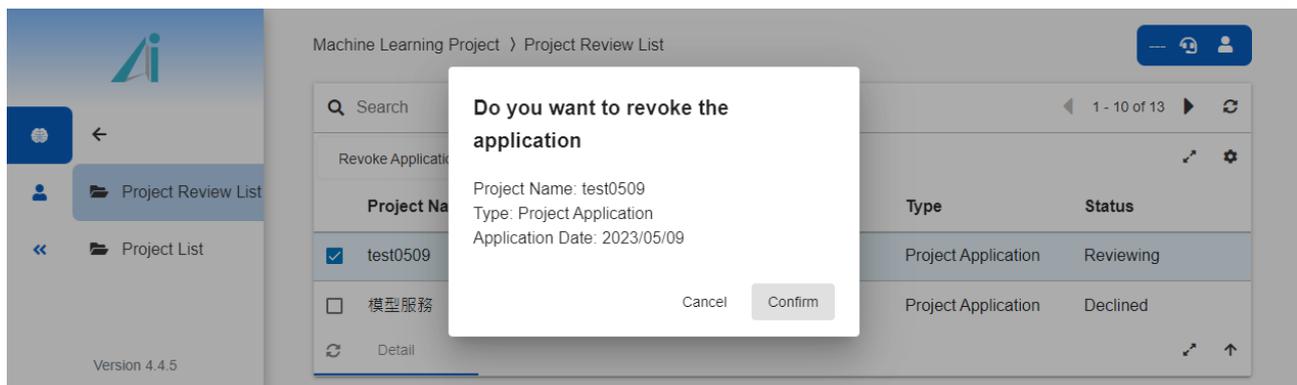


The screenshot shows the 'Project Review List' interface. At the top, there is a search bar and a 'Revoke Application' button. Below is a table with columns: Project Name, Applicant, Application Date, Type, and Status. The table contains several rows, with 'Iris Model Service' by 'hunterhsu' selected. Below the table is a 'Detail' section for the selected application.

Project Name	Applicant	Application Date	Type	Status
<input type="checkbox"/> 模型服務	user	2023/05/03	Project Application	Declined
<input type="checkbox"/> 模型服務	user	2023/05/03	Project Application	Declined
<input checked="" type="checkbox"/> Iris Model Service	hunterhsu	2023/04/28	Period Extension	Approved
<input type="checkbox"/> Inference Service	hunterhsu	2023/04/28	Project Application	Approved
<input type="checkbox"/> Iris Model Service	allen	2023/04/14	Period Extension	Approved
<input type="checkbox"/> 測試	user	2023/04/10	Project Application	Approved
<input type="checkbox"/> Iris Model Service	hunterhsu	2023/04/06	Period Extension	Approved

Field	Value	Field	Value
Project Code	project1676866033694	Application Date	2023/04/28 17:21:58
Project Name	Iris Model Service	Applicant	hunterhsu
Status	Approved	Application Acco...	hunterhsu
Start Time	2023/04/14 00:00:00	Reason	---
End Time	2023/07/23 23:59:59	Remark	---
Next Start Time	2023/07/24 00:00:00	Extension Days	7
Next End Time	2023/07/31 23:59:59		

User can choose to revoke the application if the application status is still at "reviewing

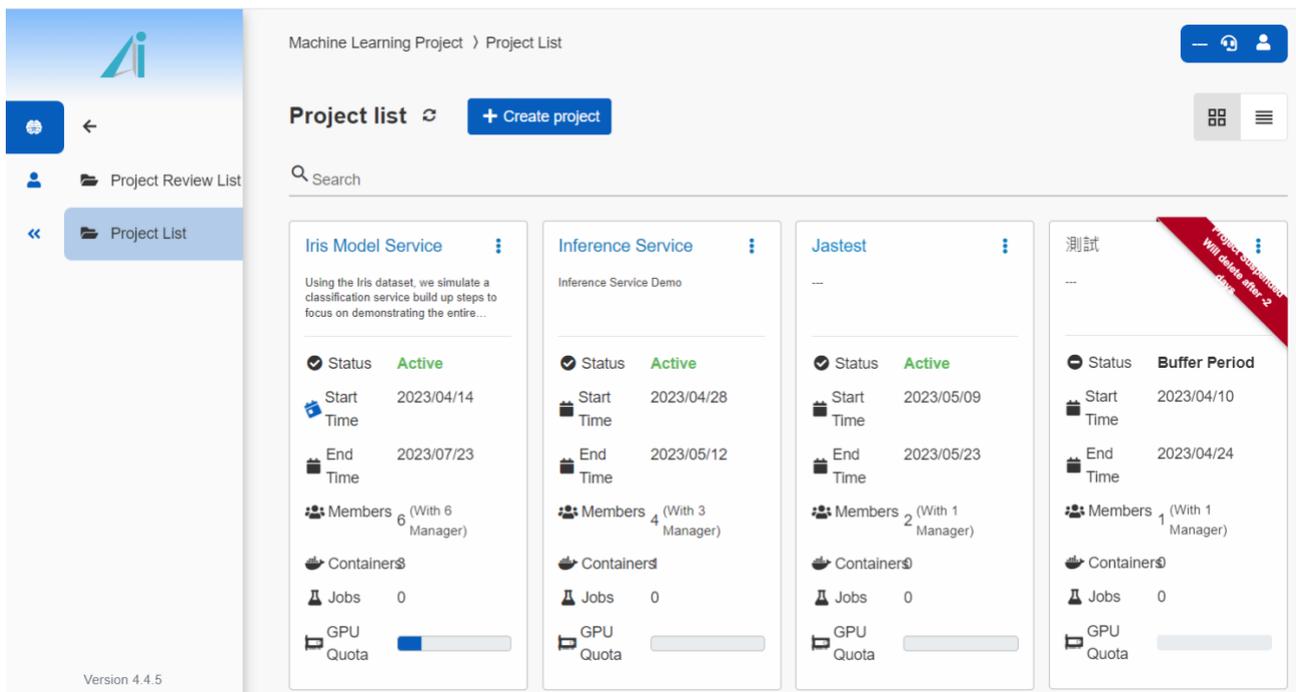


The screenshot shows the 'Project Review List' interface with a confirmation dialog box overlaid. The dialog box asks 'Do you want to revoke the application' and provides details for the application to be revoked: Project Name: test0509, Type: Project Application, Application Date: 2023/05/09. The dialog box has 'Cancel' and 'Confirm' buttons.

2. Project List

A. Project Overview

Clicked on project list to inspect the overall status of the projects. Default to card view, each project card records the general information and the resource utilization status of each project. When creating projects, the applications have to be approved before you can start using the resources. And before the project application is approved, the project status will remain as "reviewing".



Switch to table view by clicking on the  icon on the top right.

User Guide

Machine Learning Project > Project List

Project list ↻ + Create project

Q Search ◀ 1 - 6 of 6 ▶ ↻

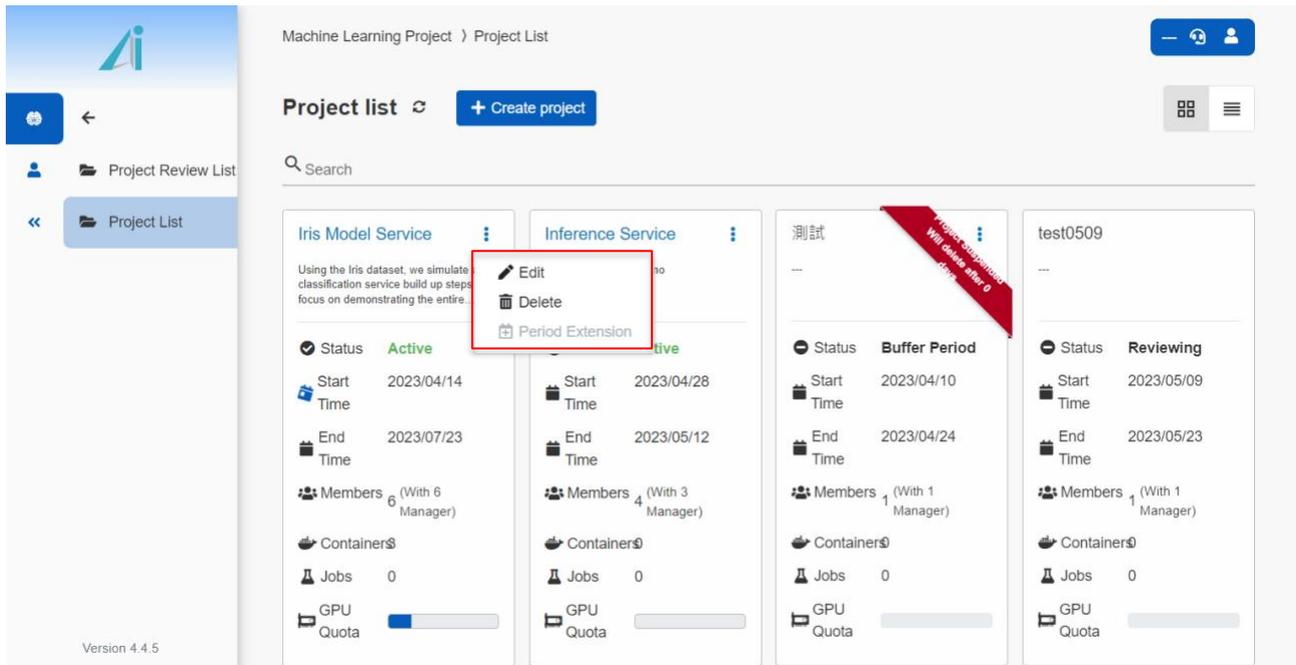
🗑️ ✎ 📅 Period Extension ⚙️

<input type="checkbox"/>	Project Na...	StartTime	Endtime	Members	Container	Job	GPU Quota	Project St...
<input type="checkbox"/>	Iris Model S...	2023/04/...	2023/07/23	6 (With 6 M...	3	0	0.625 / 3	Active
<input type="checkbox"/>	Inference S...	2023/04/28	2023/05/12	4 (With 3 M...	0	0	0 / 3	Active
<input type="checkbox"/>	測試	2023/04/10	2023/04/24	1 (With 1 M...	0	0	---	Buffer P... !
<input type="checkbox"/>	test0509	2023/05/09	2023/05/23	1 (With 1 M...	0	0	---	Reviewing
<input type="checkbox"/>	模型服務	2023/05/03	2023/05/17	1 (With 1 M...	0	0	---	Declined
<input type="checkbox"/>	模型服務	2023/05/03	2023/05/17	1 (With 1 M...	0	0	---	Declined

Version 4.4.5



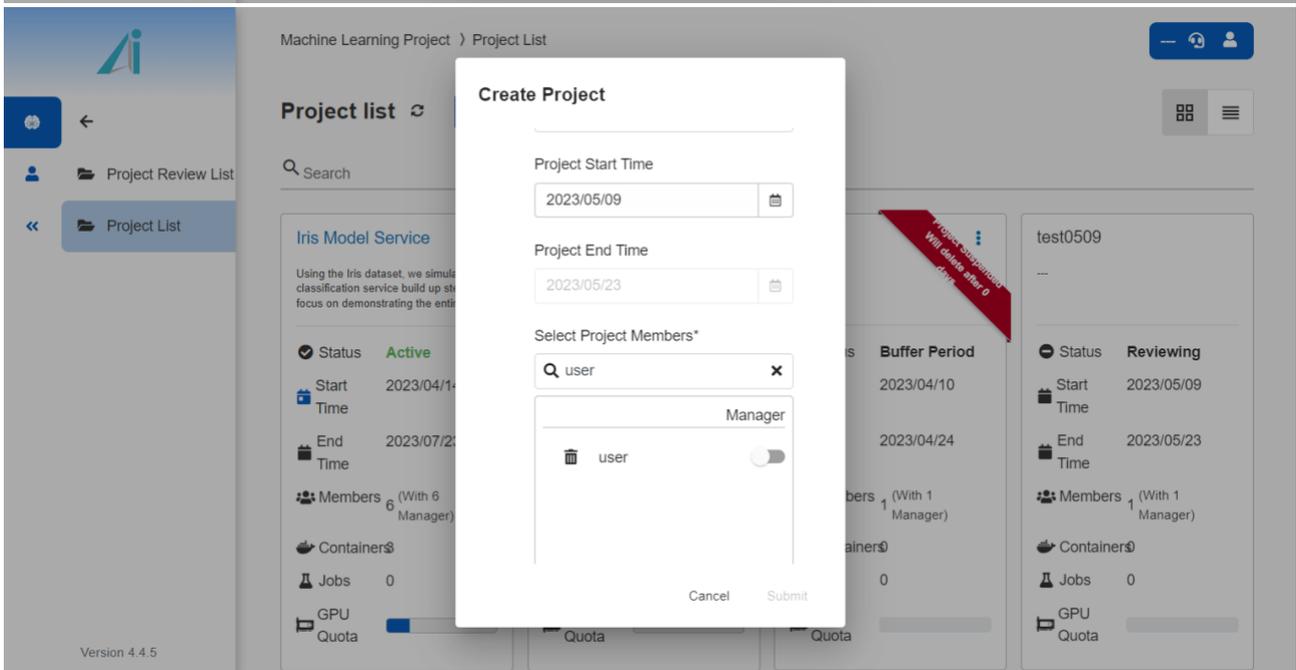
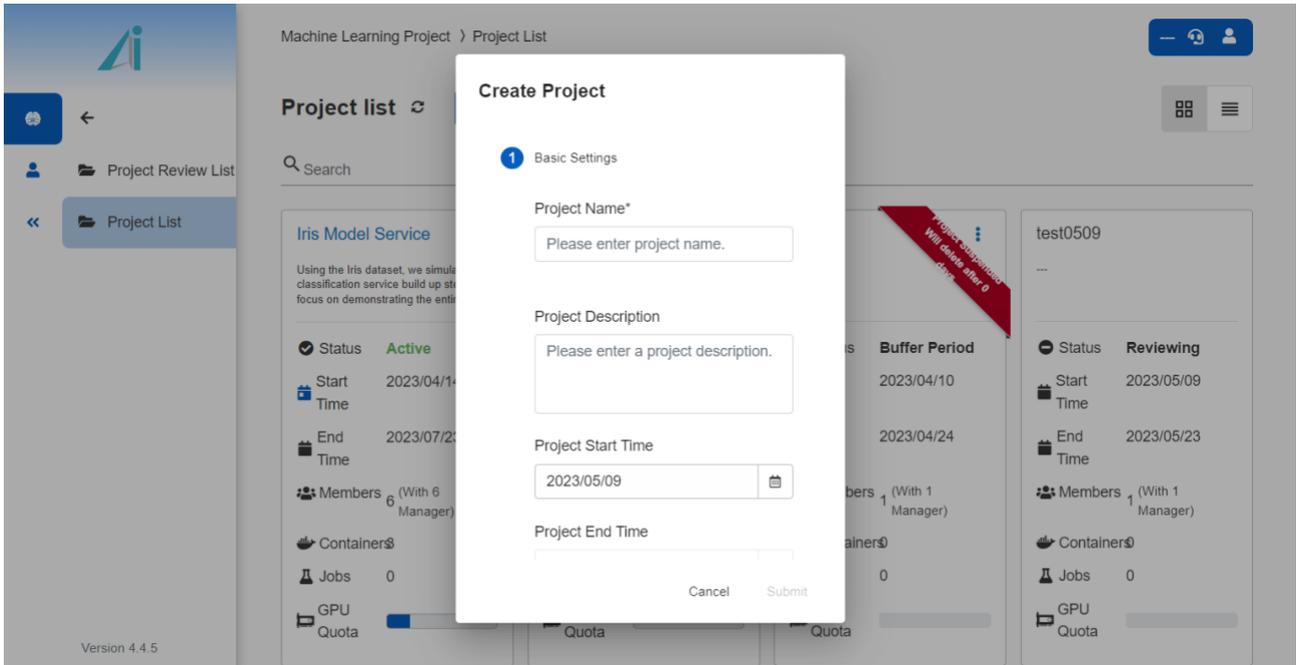
The project manager can edit a project, delete a project, or extend the project duration by clicking the  icon on the top right of each project card.



B. Create Project

Submit application to create new project. Once the project is approved by the platform administrators, you can start utilizing the resources such as creating a container.

- Project Name: Enter a project name.
- Start and End Date: Enter the dates required.
- Select Project members: you have to assign at least one project manager to the project.
- Project purpose: Choose according to the project requirements (This will affect the resource availability)



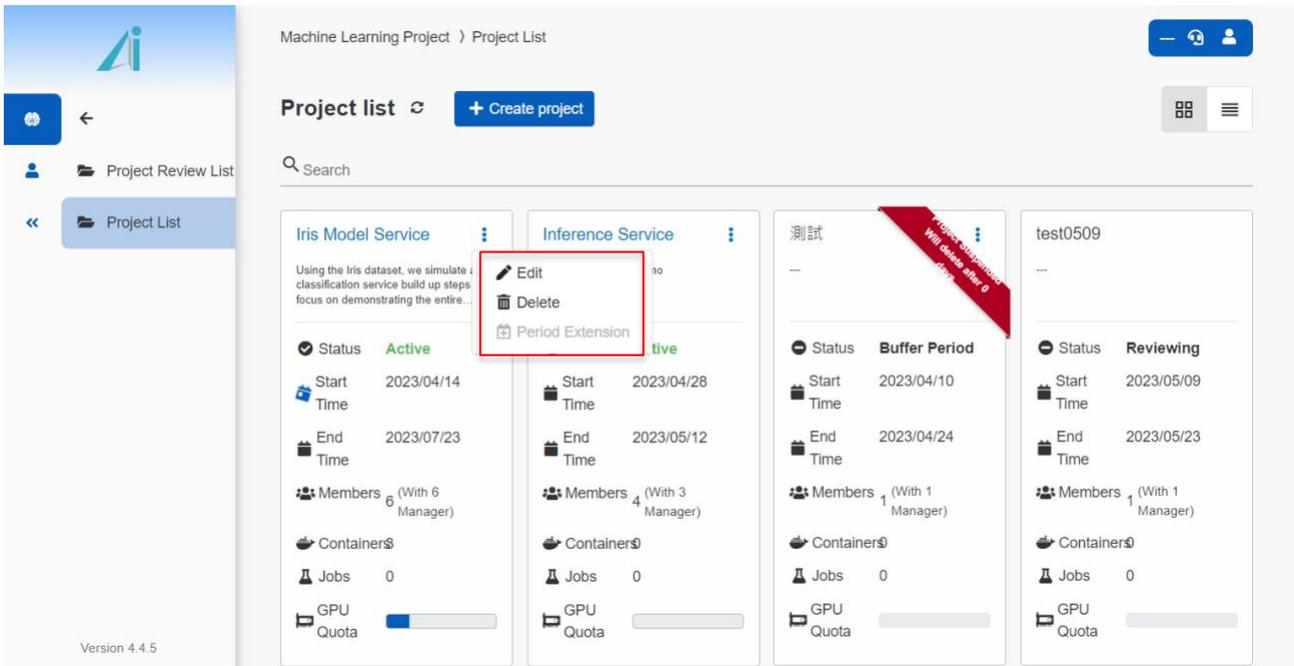
※ If the project duration was not extended after the buffering period, the platform will release all the resources allocated to this project, which includes but not limited to containers, job, and storage cluster.

The applicant will receive an email notification such as the below figure once the project application is approved.

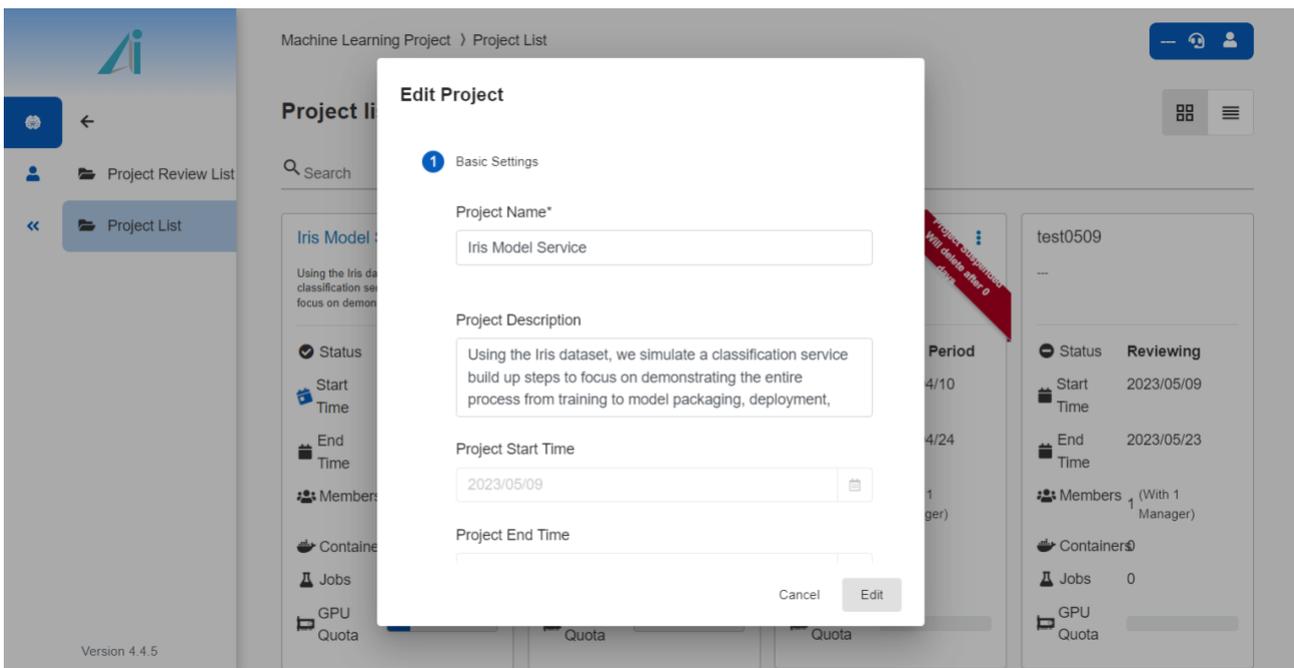
User Guide

C. Edit Project

Project managers can edit a project by clicking on the  icon on the top right of the project card.

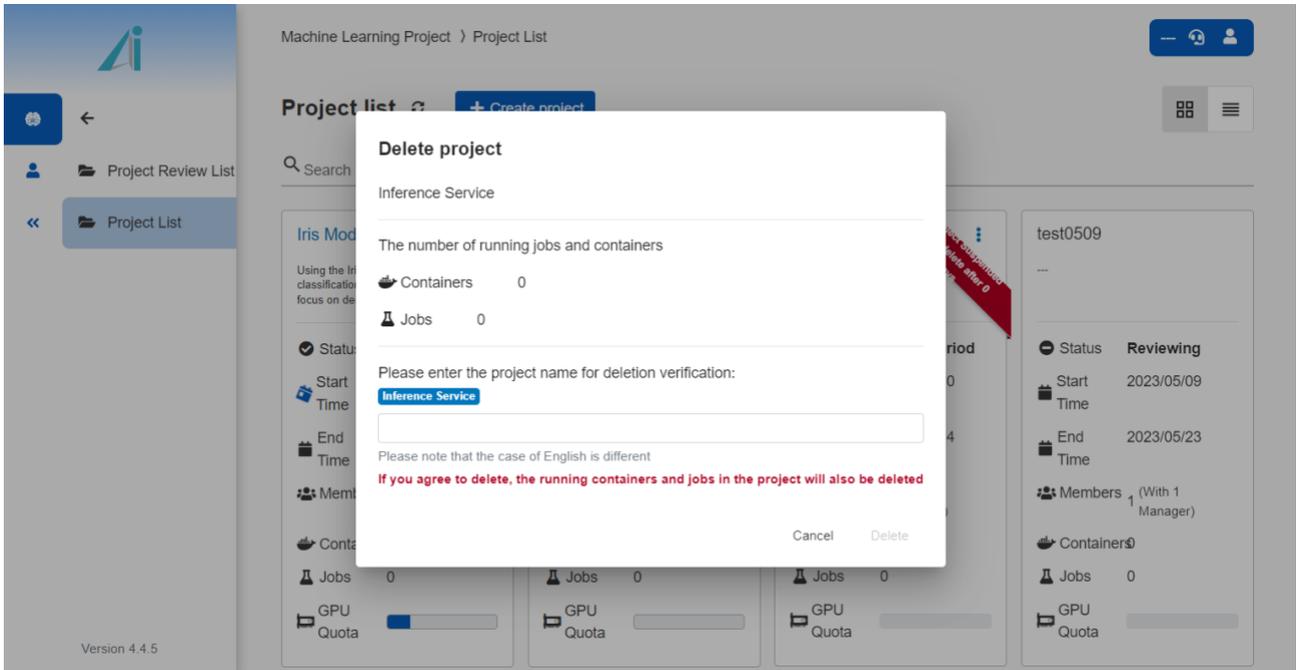


Project managers can edit project name, edit project description, add project members and assign other project managers.



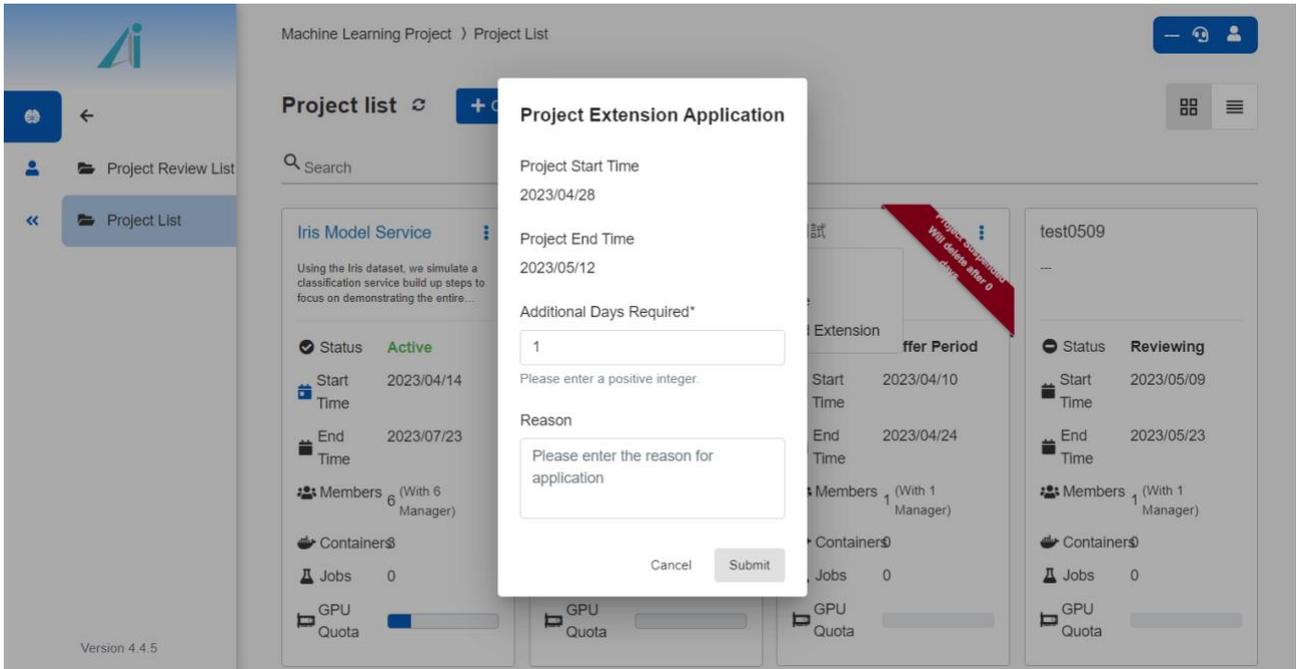
D. Delete Project

Project managers can delete a project along with the resources allocated to this project. You will have to enter a verification code to confirm the deletion.

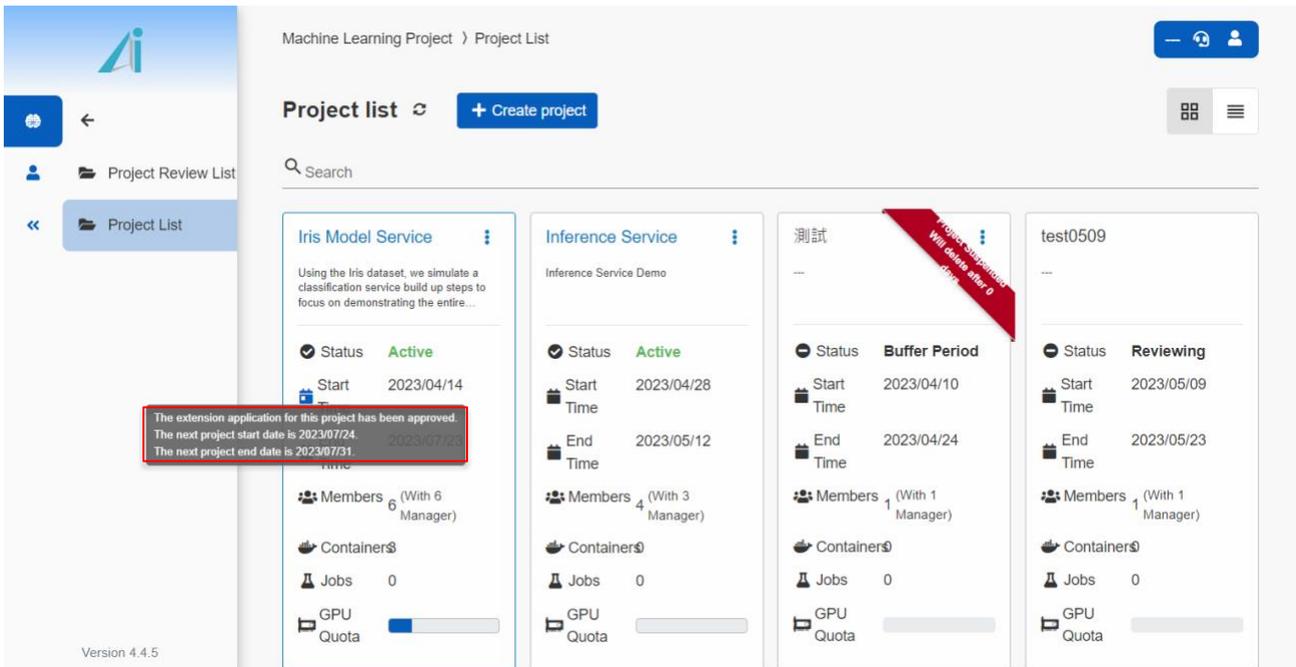


E. Project Extension Application

Project managers can extend the project duration by clicking on the  icon and press extension. Submit the application through the popup window after you enter the days required.

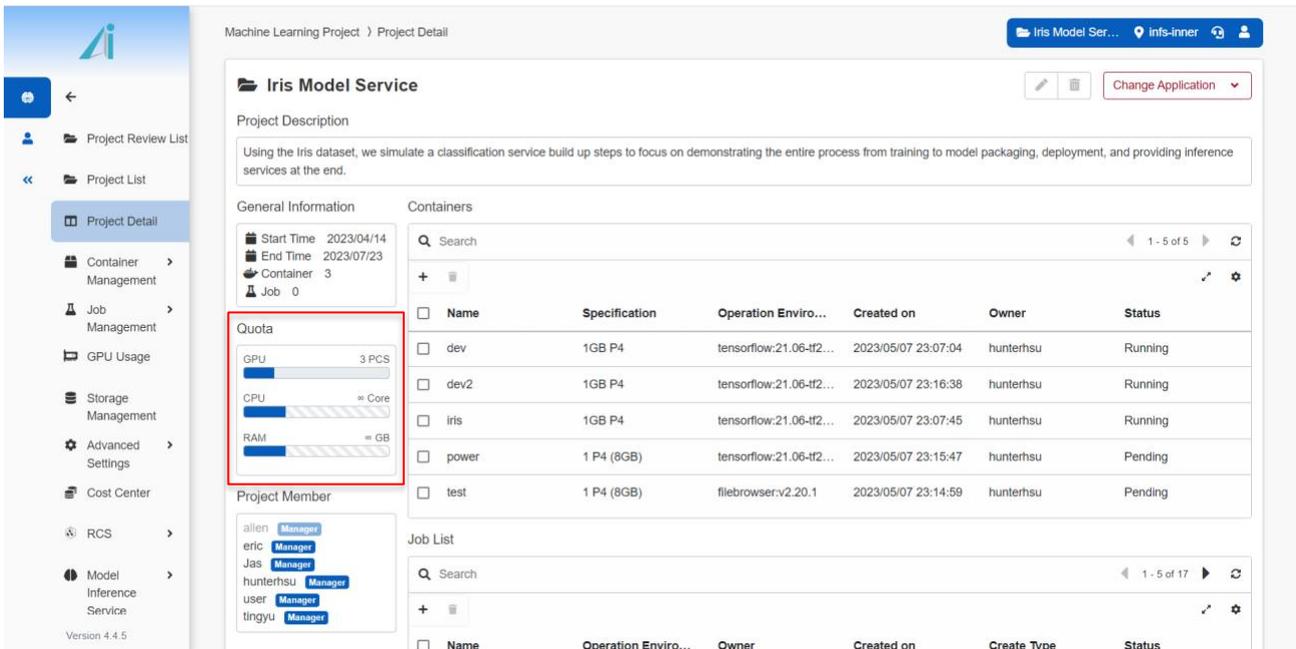


Once the project extension is approved. The icon for the start date will turn blue. Hover over the icon and you can see the next start and end date for the project.



3. Project Detail

Click on the project to enter “Project Details” Page. Users can view containers created in the project*, resource quota, project members and current jobs. General information such as the start date, end date, container and job counts are also shown at the top left side of the page.

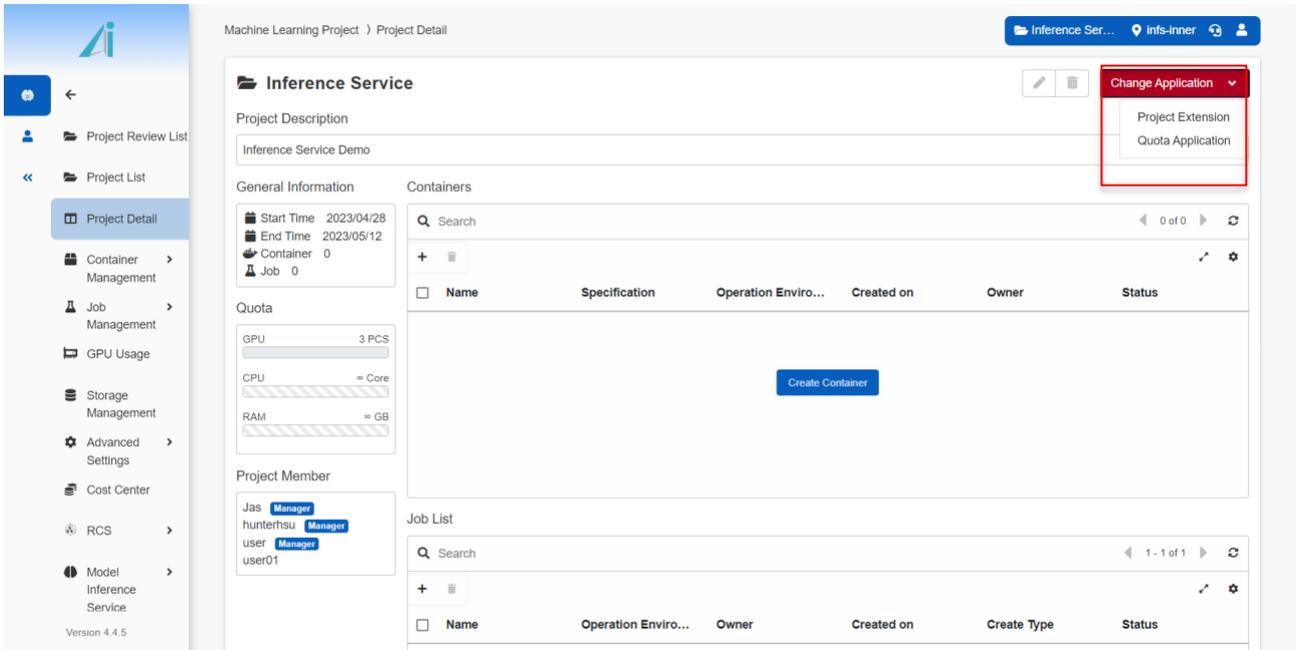


The area in the red box shown in the above figure records resource quota and availability for GPU, CPU and RAM quota of the project.

*Remar: As a default, project managers can see all the containers in a project, and members can only see the containers they created.

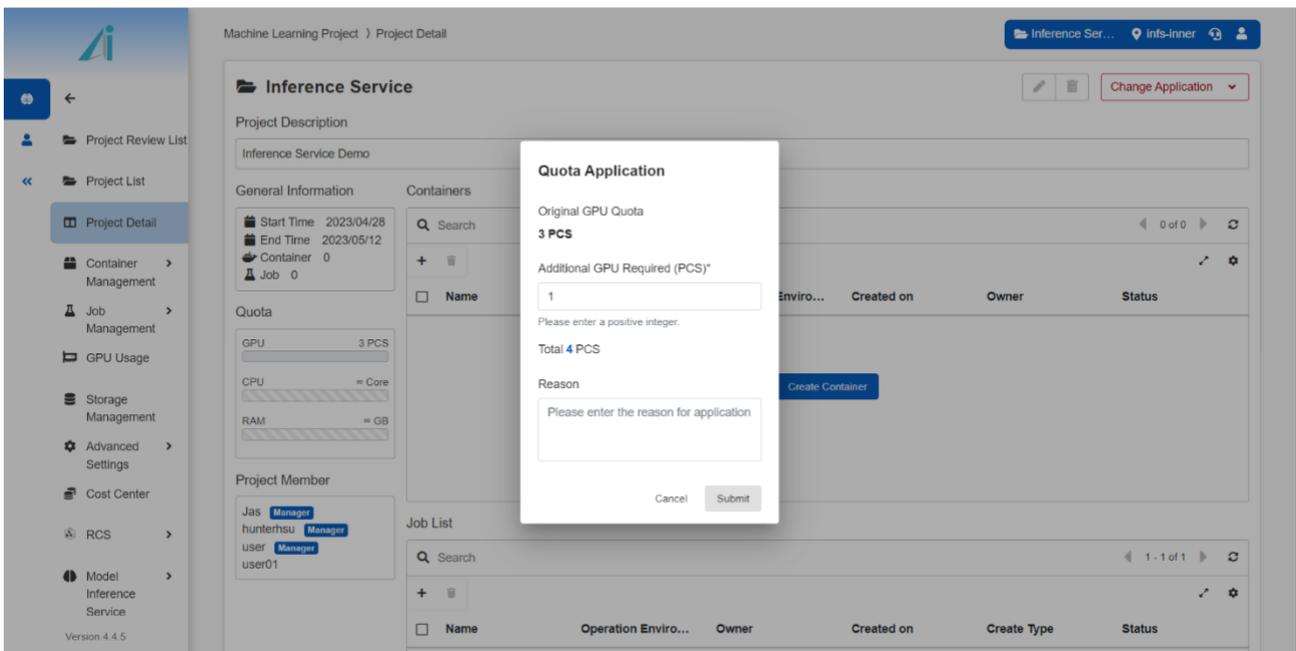
A. Project Change Application

In change applications, you are able to apply for a duration extension and quota increase, both operations are performed by project managers.



B. Quota Application

Quota application form allowed project managers to input the additional pieces of GPU required to complete the form and start the application process. The application requires the platform administrator to approve before taking effect.



C. Project Expiring Notice

When the project expires without extending the duration of the project. The platform will notify the project manager by email. The project will enter a 14-day buffering period, during which you are unable to access any resources.

4. Container Management

AI-Stack provides a website for users to use machine learning services at ease and choose machine learning frameworks such as Tensorflow or Pytorch according to your needs.

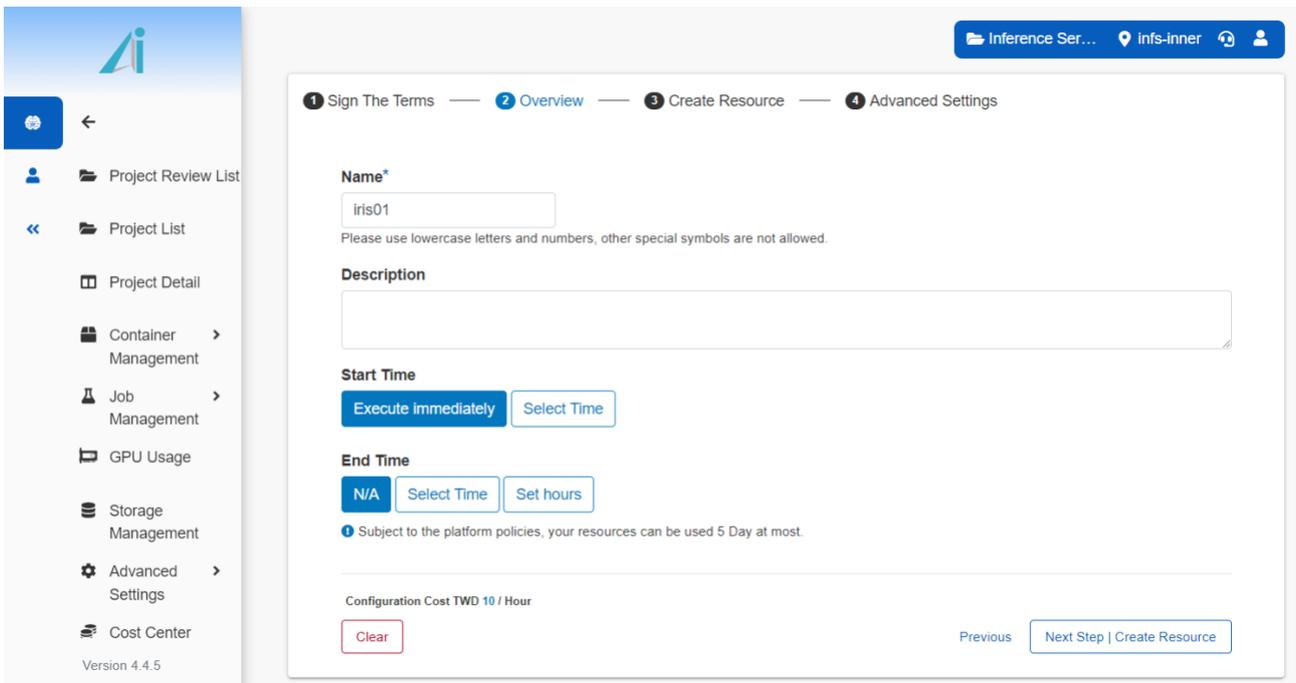
4.1 Create Container

General Creation

- (1) Go to "Container List" and click on "+"
- (2) Enter the Name
- (3) You can choose the start time of the container. It could start immediately or at a specific time. "Start immediately" means that the container will be created immediately after submission. Alternatively, you can choose a specific date and time for the container to be created.

Specifying an end time determines whether a specific time should be assigned to the container. You can also choose not to specify an end time. If you select a specific end time, the container will be deleted at the specified time. If you select a specific duration, the container will be deleted as requested.

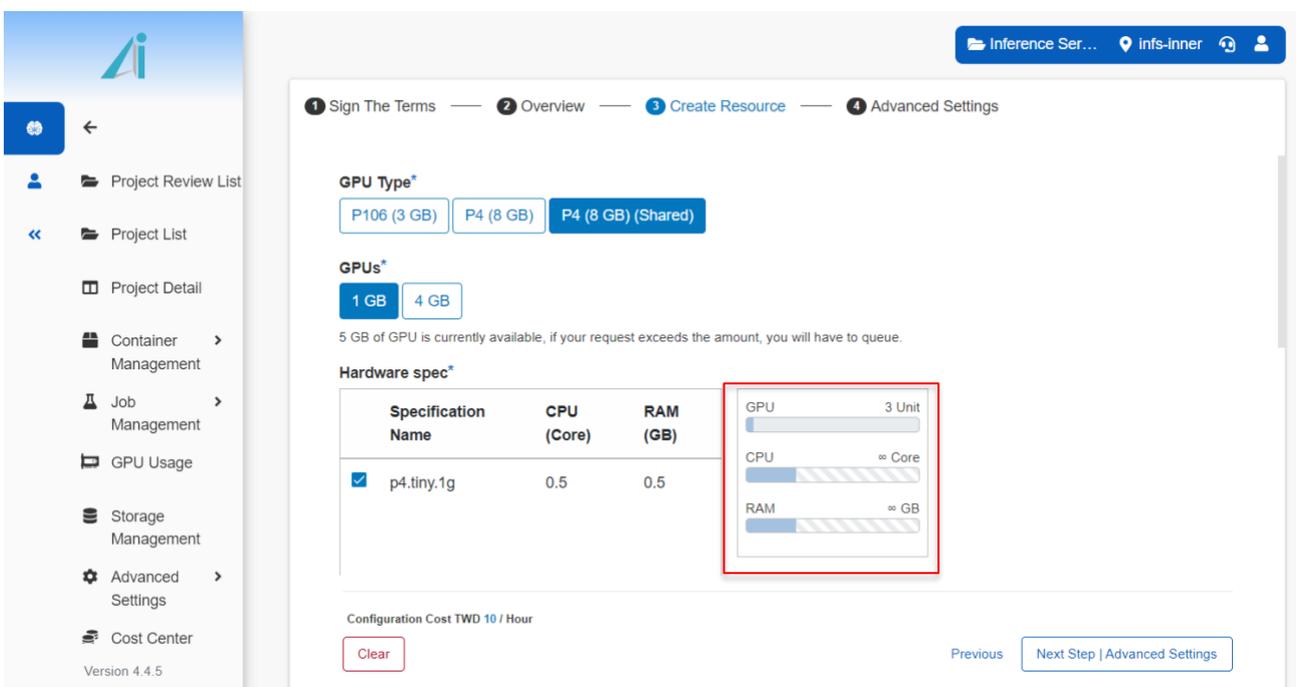
Click "Next Step | Create Resource" to continue



(4) Select the GPU type and quantity required. (Set by the platform administrator)

(5) Select the preferred hardware specification, including the number of CPU cores and memory.

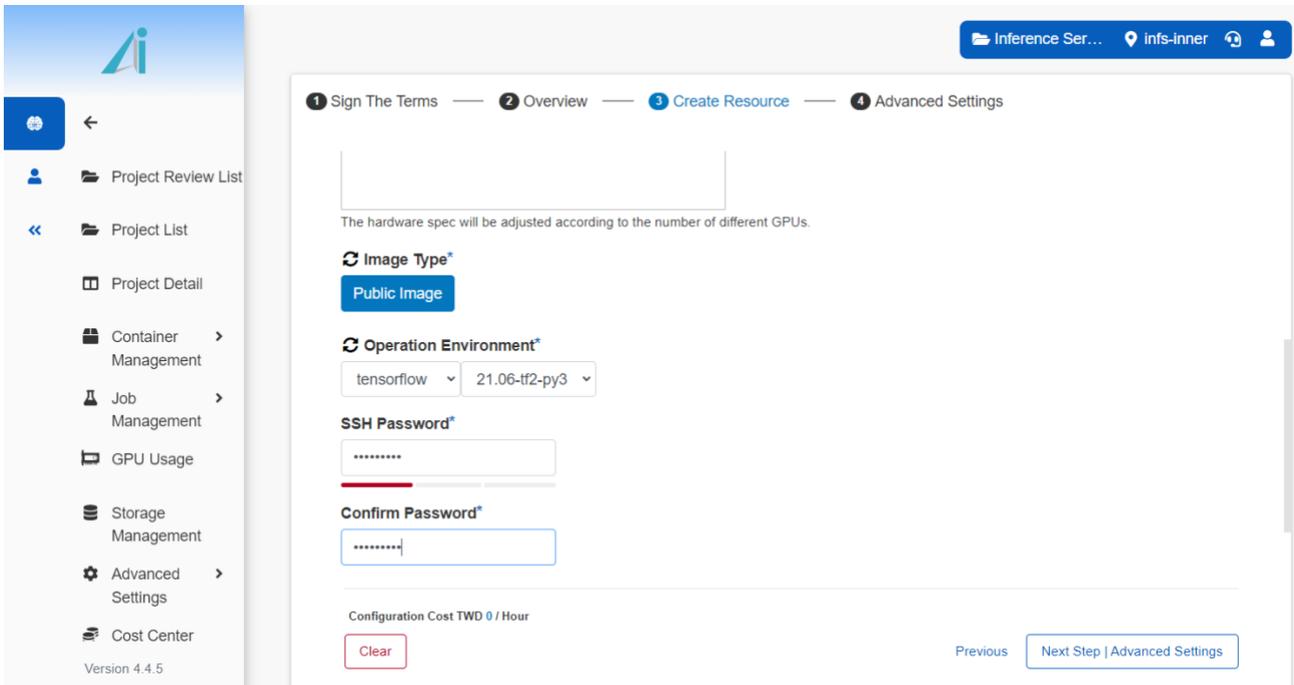
The red boxed area to the right of Hardware specifications shows the resource availability information. It will change based on the selected specification.



(6) Select to use either a public or a custom image.

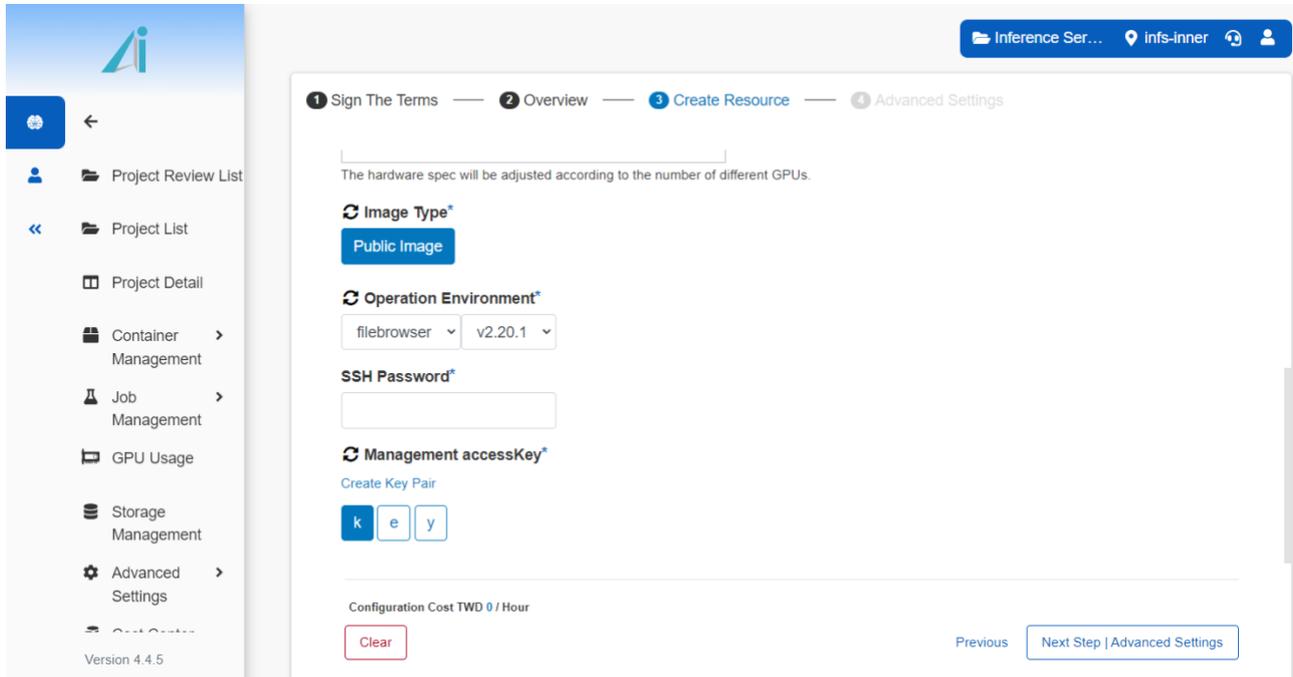
You have to create at least 1 custom image for the option to appear.

(7) Select the AI framework to use under the operation environment.



(8) Key pair management

When a user has previously created a key on the platform and has enabled the SSH log in method with the selected image template, they can choose which key pair to use for the connection.



Click "Next Step | Advance Settings" to continue

(9) Enable the option for "Shared Memory."

Shared Memory Function Explanation: This feature allows two or more processes to access the same block of memory. It is used for distributed computing in multi-GPU containers, where data is placed in shared memory to enhance GPU parallel computing efficiency.

Inference Ser... | info-inner

1 Sign The Terms — 2 Overview — 3 Create Resource — 4 Advanced Settings

If you do not have the following functional requirements, you can skip and execute the creation.

Shared Memory

Enabled

64 MB

A minimum of 64 MB is required. This setting will split the specified capacity from the selected specification memory for use as shared memory. The maximum limit is up to 70% of the selected memory capacity.

13% Shared Memory 64 MB | 88% System Memory 448 MB

Mount Volume

Enabled

Batch

Enabled

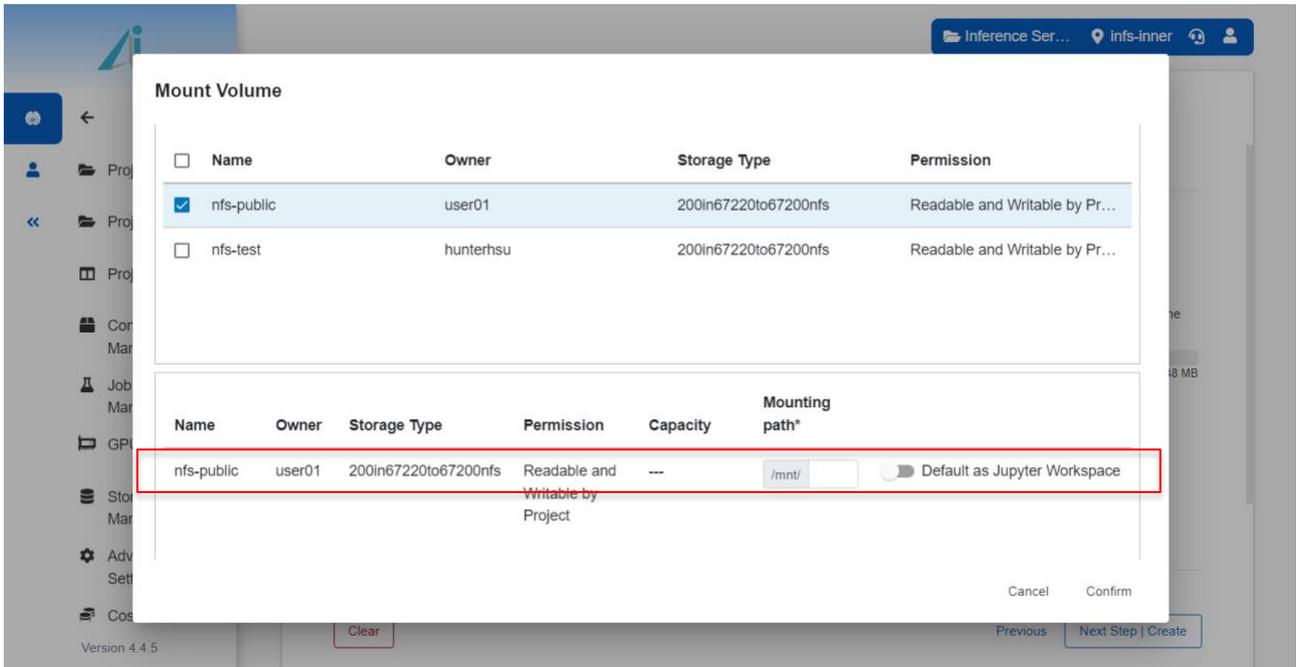
Configuration Cost TWD 10 / Hour

Clear | Previous | Next Step | Create

Version 4.4.5

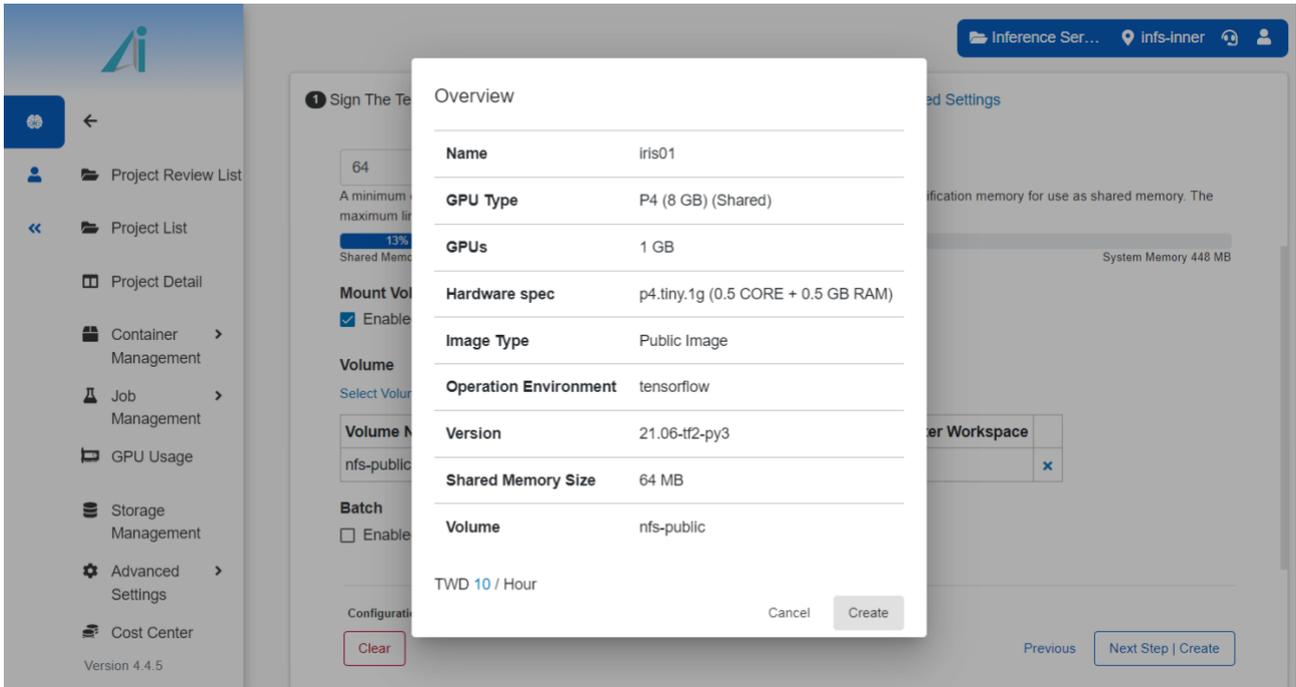
(10) Mounting Storage Device

Select a device to mount and assign a mount path, you can also set it as a working directory for your jupyter notebook.

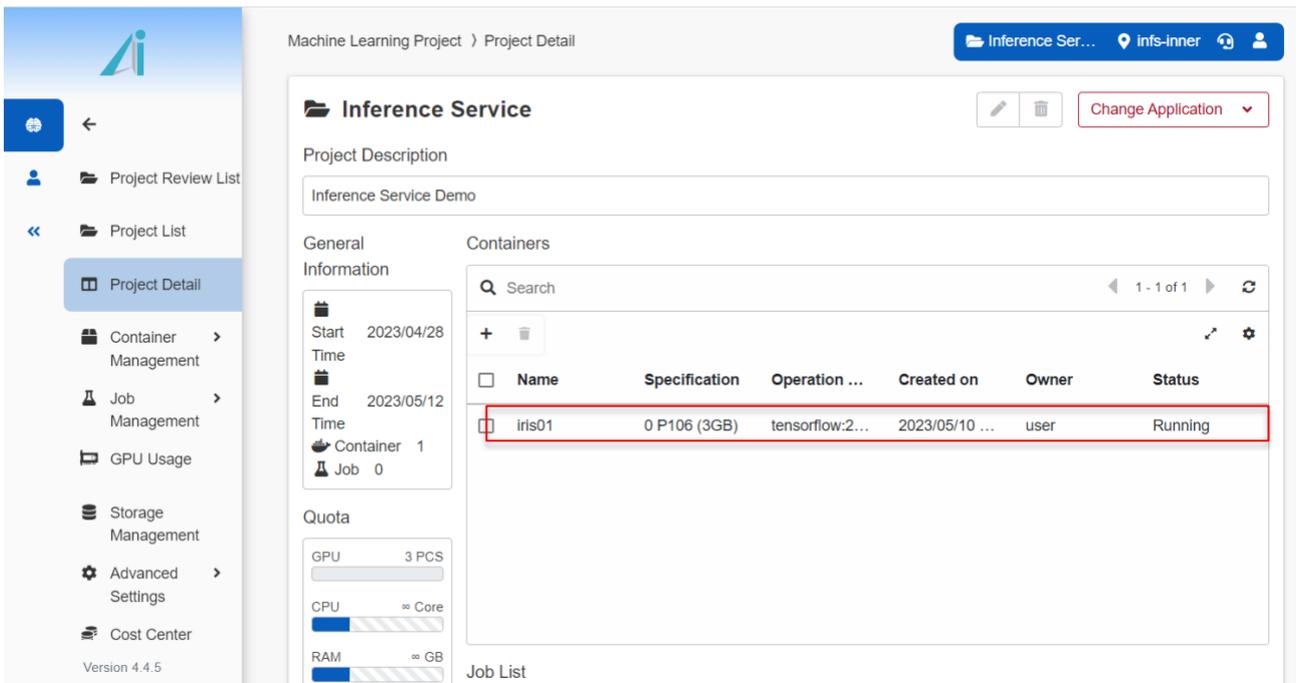


* Remark: you will have to create a storage to mount in storage management

An overview of your container will display once you click on create. Confirm to proceed.



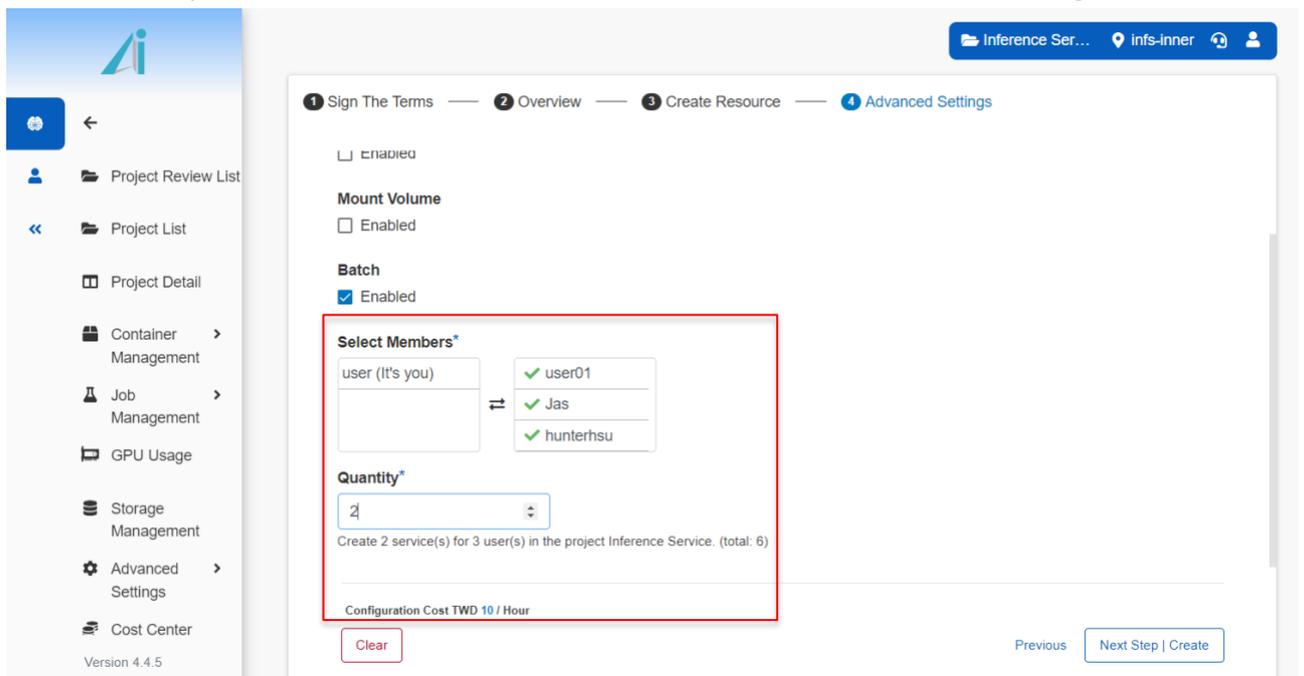
- (11) You will be directed to “project details” page while you wait for the container to initiate. When the status of the container is “Running” you can start using the service.

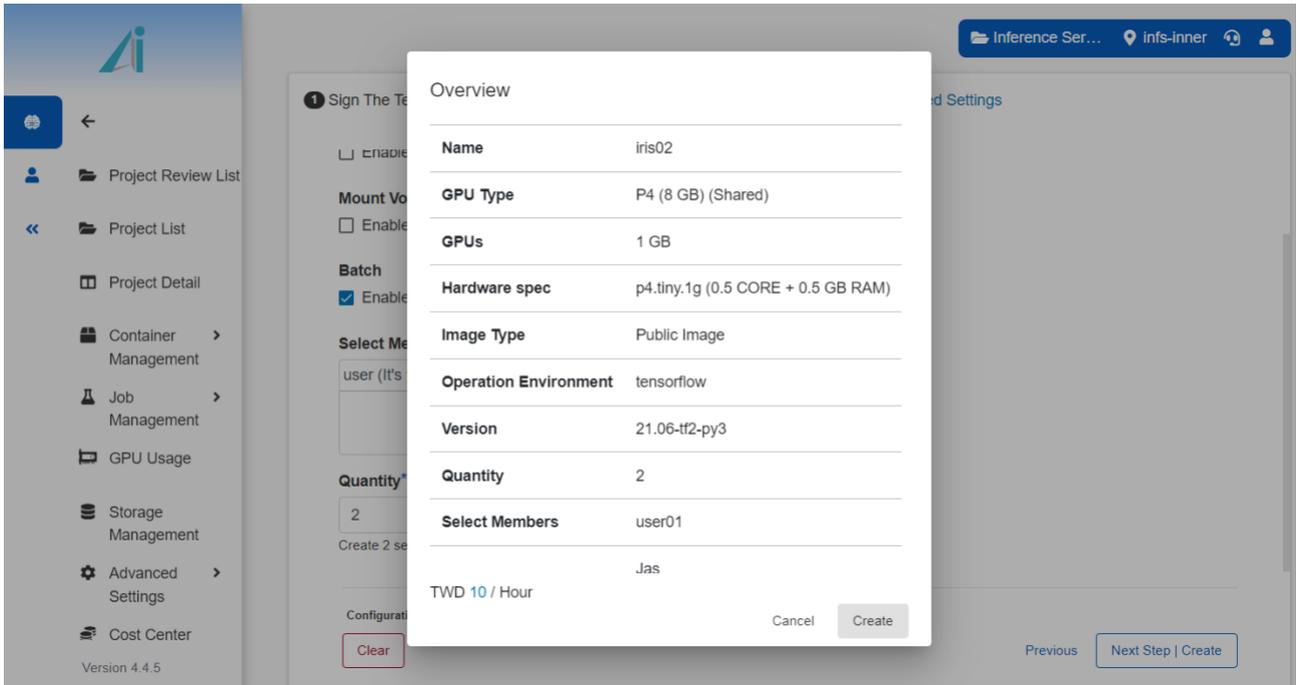


Batch Creation

Only project managers can perform this operation. This assists the project members to pre build the container by batch. The procedure is as follows:

- (1) Perform general container creation steps.
- (2) Refer to the red-boxed area in the below figure, enable batch creation in advanced settings by ticking the box.
- (3) Selecting the members to the right indicates that the container will be created for him.
- (4) Enter the quantity of the containers you would like to create for each member.
- (5) Click "Next | Create" , confirm the details in the overview, confirm the setting.





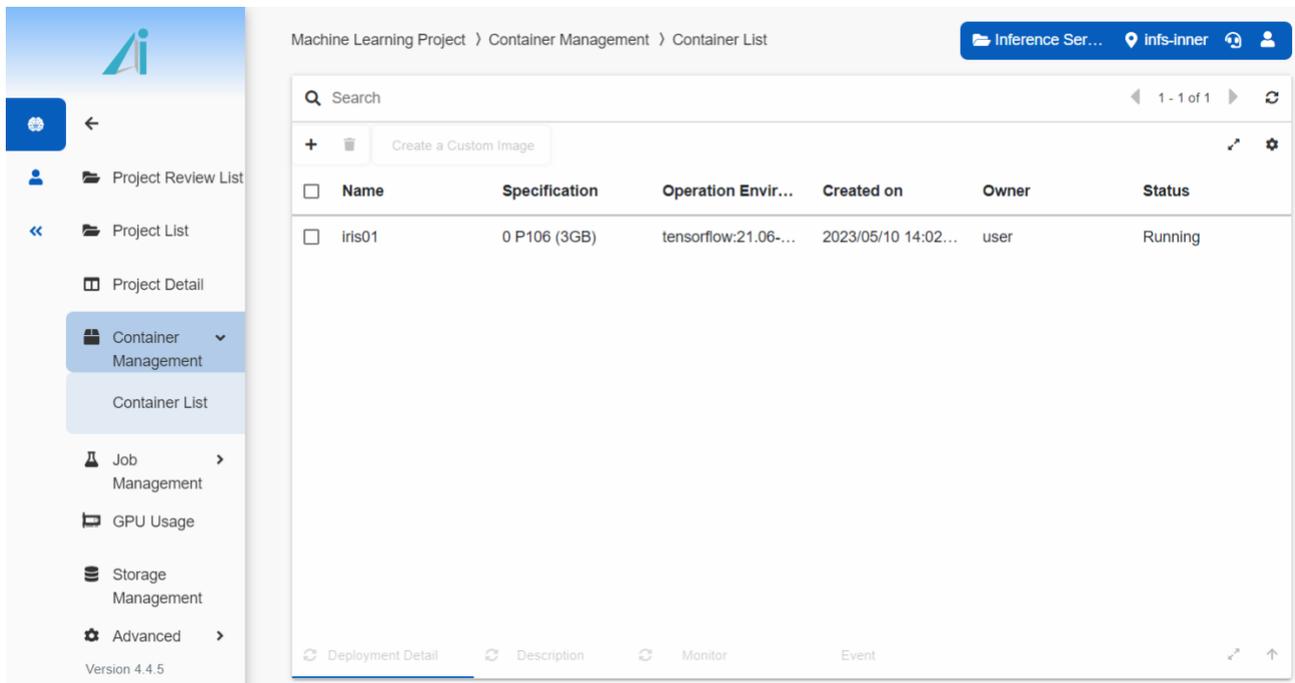
The screenshot shows a web application interface with a modal window titled "Overview" for configuring an inference service. The modal contains the following details:

Name	iris02
GPU Type	P4 (8 GB) (Shared)
GPUs	1 GB
Hardware spec	p4.tiny.1g (0.5 CORE + 0.5 GB RAM)
Image Type	Public Image
Operation Environment	tensorflow
Version	21.06-tf2-py3
Quantity	2
Select Members	user01
	Jas

At the bottom of the modal, it shows a price of "TWD 10 / Hour" and buttons for "Cancel" and "Create".

4.2 Container List

After the container creation is complete, you can go to the container list to view the list of created containers. The information on the list includes container names, specification (only displaying GPU information), operating environment, creation time, owner, and container status. When the status is not "Running," it will be displayed in gray, indicating that no operations can be performed on it.



By clicking on a container with the status "Running," you can view more detailed information below the screen. As shown in the figure, it includes "deployment detail", "descriptions", "monitor" and "events".

User Guide



- Project Review List
- Project List
- Project Detail
- Container Management**
 - Container List
- Job Management
- GPU Usage
- Storage Management
- Advanced

Version 4.4.5

Machine Learning Project > Container Management > Container List

Inference Ser...
info-inner
🔍
👤

🔍 Search 1 - 1 of 1

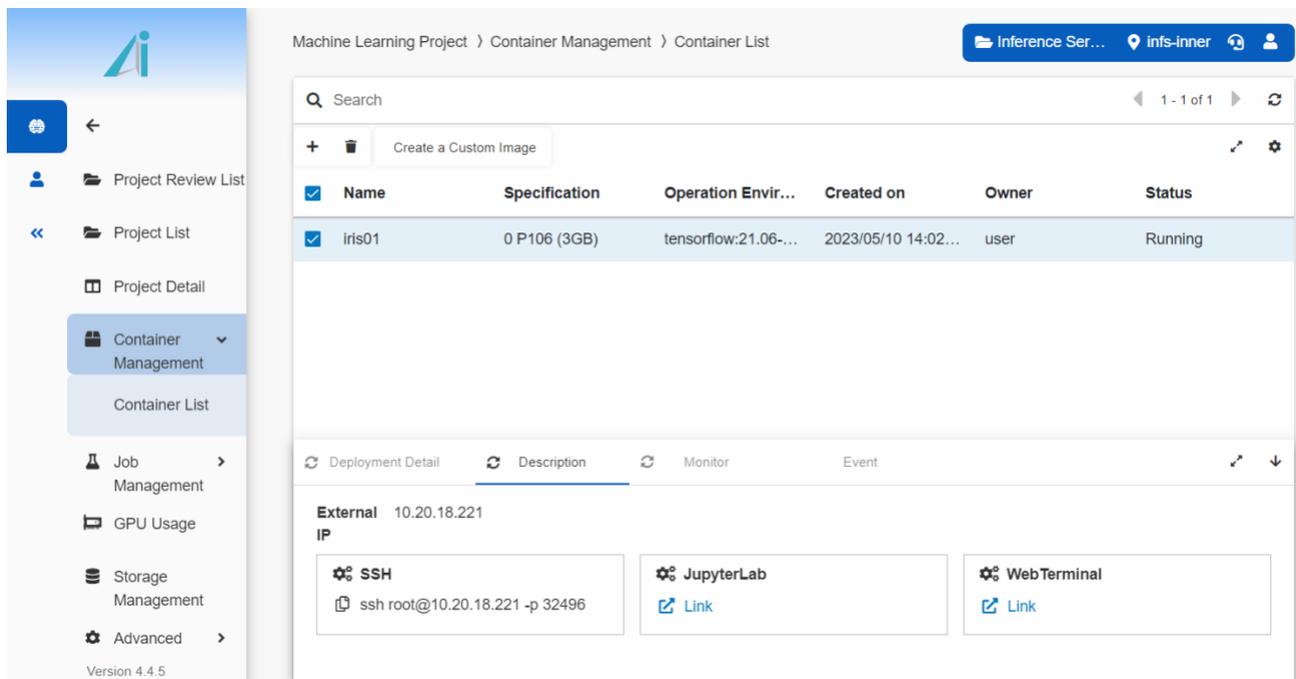
+ 🗑️ Create a Custom Image 🔗 ⚙️

<input checked="" type="checkbox"/>	Name	Specification	Operation Envir...	Created on	Owner	Status
<input checked="" type="checkbox"/>	iris01	0 P106 (3GB)	tensorflow:21.06-...	2023/05/10 14:02...	user	Running

🔄 Deployment Detail
🔄 Description
🔄 Monitor
Event

Name	iris01	Status	Running
Deployment ID	53bfbef6-7388-4d55-a7ad-caa33e39d850:iris01	Created on	2023/05/10 14:02:21
Operation En...	tensorflow:21.06-tf2-py3	End Time	2023/05/15 14:03:40
Specification	0 P106 (3GB) + 1 CORE + 1 GB RAM		
Shared Mem...	1024 MB		
Firewall Rule	Allow 0.0.0.0/0 segments to connect.		

The service information tab includes all the services the container provides. It is set up at the admin portal at MLS image settings. Supporting SSH, Jupyter, JupyterLab, TensorBoard, WebTerminal, Code Server and other self-defined services.



Using container services

(1) SSH

Using an SSH tool (such as PuTTY) in conjunction with the IP address, service port number, and the password entered during creation or the selected key, you can SSH into the container remotely for operations. Taking the example of viewing detailed information of the "dev" container, the SSH command would be "ssh root@10.20.67.221 -p 31348."

```
root@dev-5b9559b4b4-tplvq x + v
C:\>ssh root@10.20.67.221 -p 31348
The authenticity of host '[10.20.67.221]:31348 ([10.20.67.221]:31348)' can't be established.
ED25519 key fingerprint is SHA256:7VymLTZst+omGLpI6YUPHLv4z7xY9avDfw0AI0idCLA.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added '[10.20.67.221]:31348' (ED25519) to the list of known hosts.
root@10.20.67.221's password:
Welcome to Ubuntu 20.04.2 LTS (GNU/Linux 5.4.0-146-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

This system has been minimized by removing packages and content that are
not required on a system that users do not log into.

To restore this content, you can run the 'unminimize' command.

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

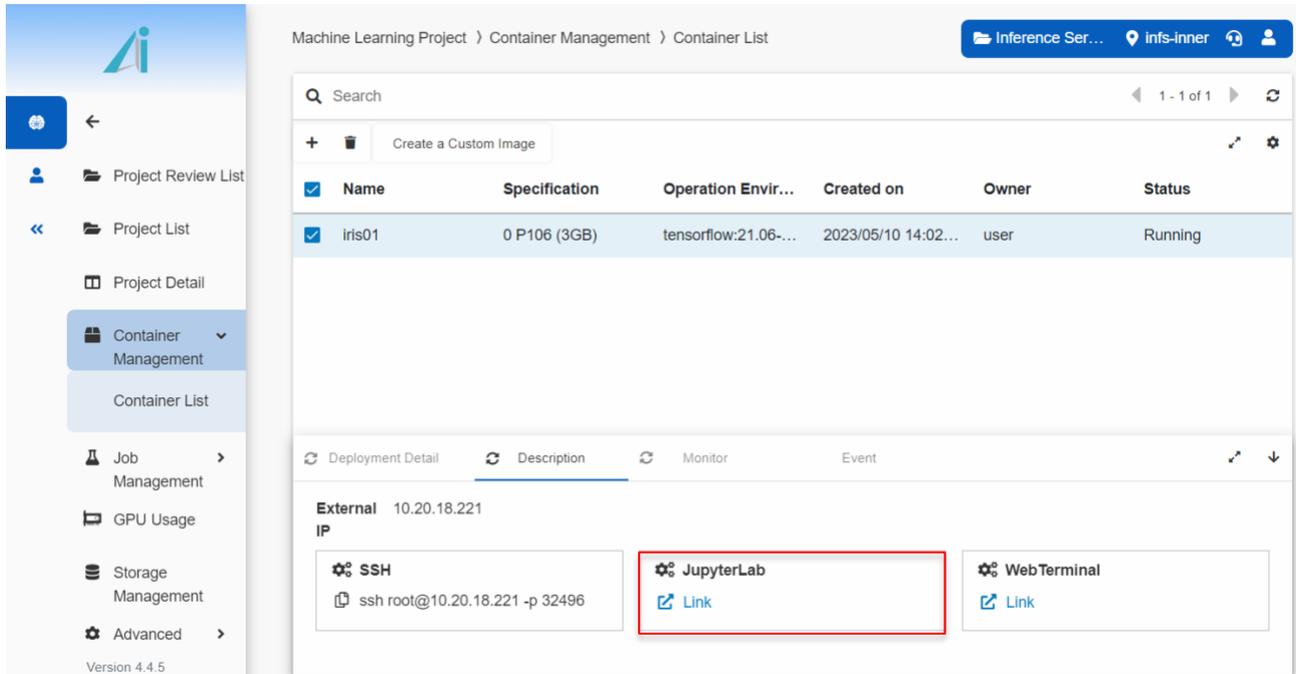
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

root@dev-5b9559b4b4-tplvq:~# ll
total 52
```

User Guide

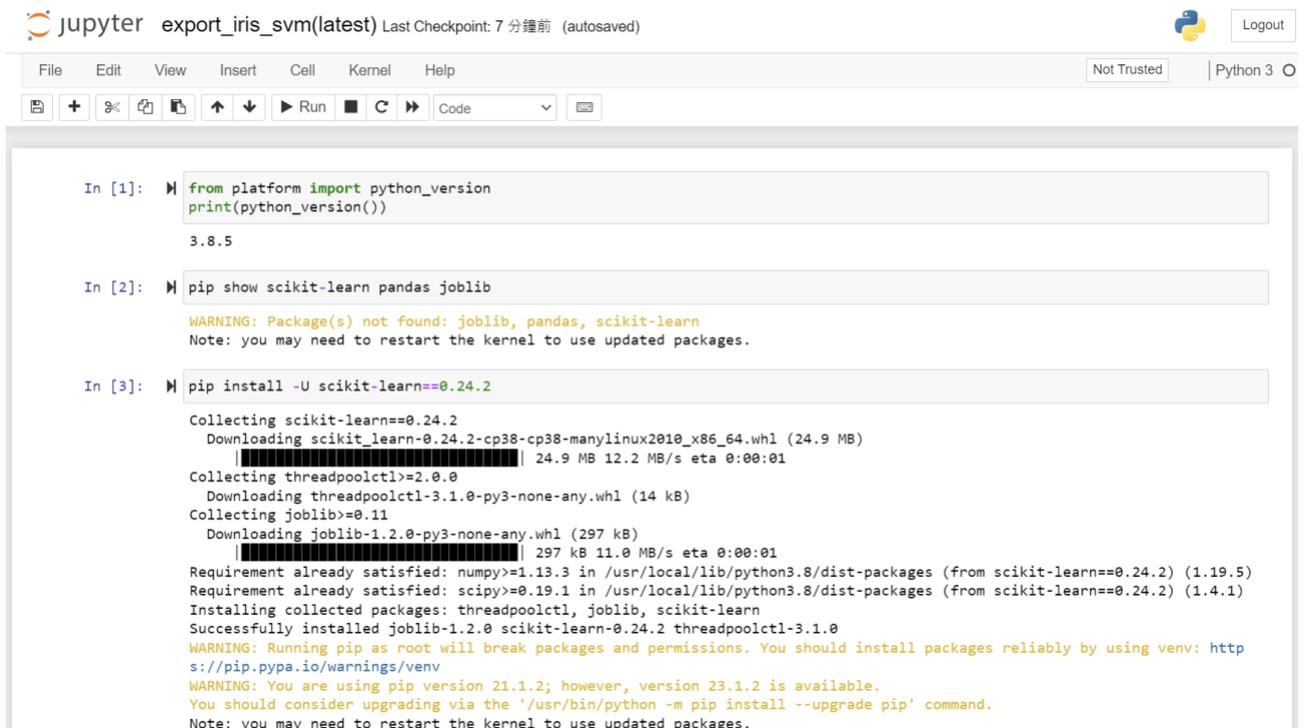
(2) Jupyter

On the "Service Information" tab, clicking the "Jupyter Link" will directly go to Jupyter Notebook. Users can use it to write Python programs.



The screenshot shows the 'Container List' page in the 'Container Management' section. A table lists containers, with 'iris01' selected. Below the table, the 'Description' tab is active, showing service information for 'External 10.20.18.221 IP'. A red box highlights the 'JupyterLab' service, which includes a 'Link' button.

Name	Specification	Operation Envir...	Created on	Owner	Status
iris01	0 P106 (3GB)	tensorflow:21.06...	2023/05/10 14:02...	user	Running



The screenshot shows a Jupyter Notebook interface with the following code and output:

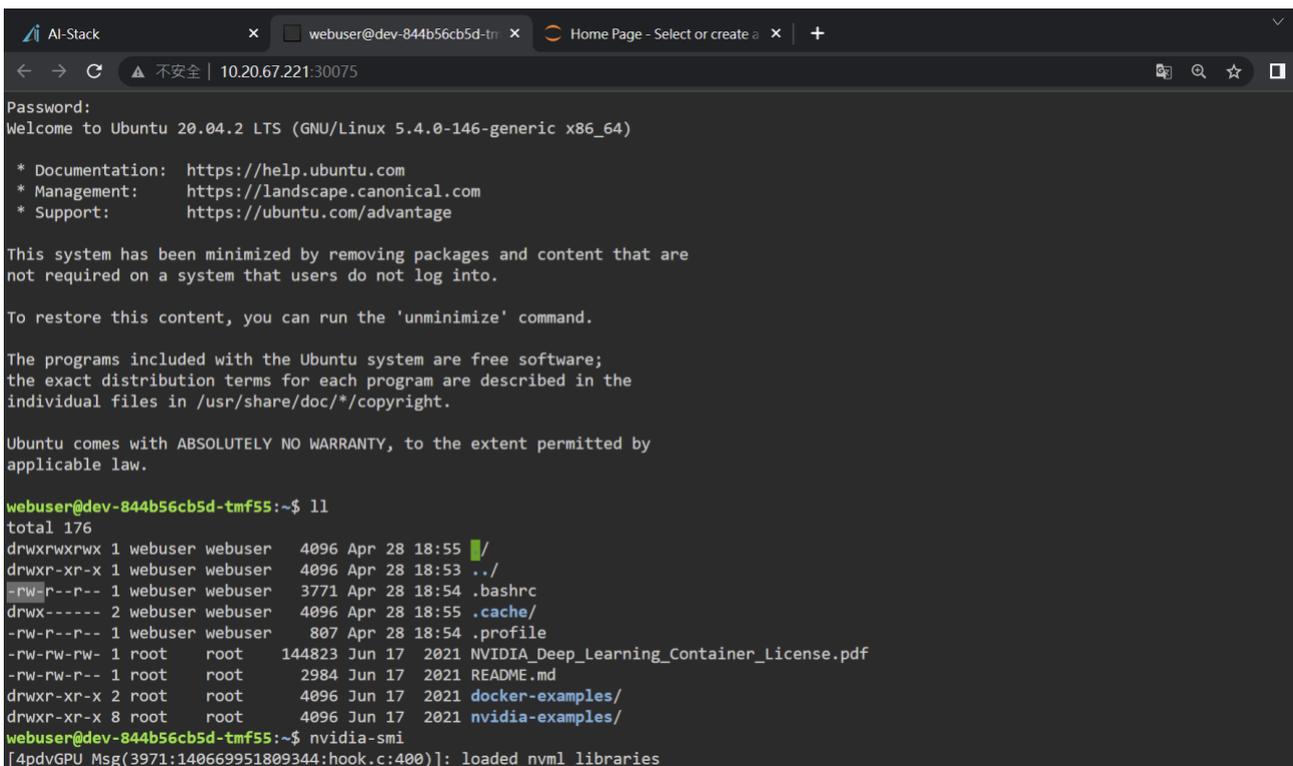
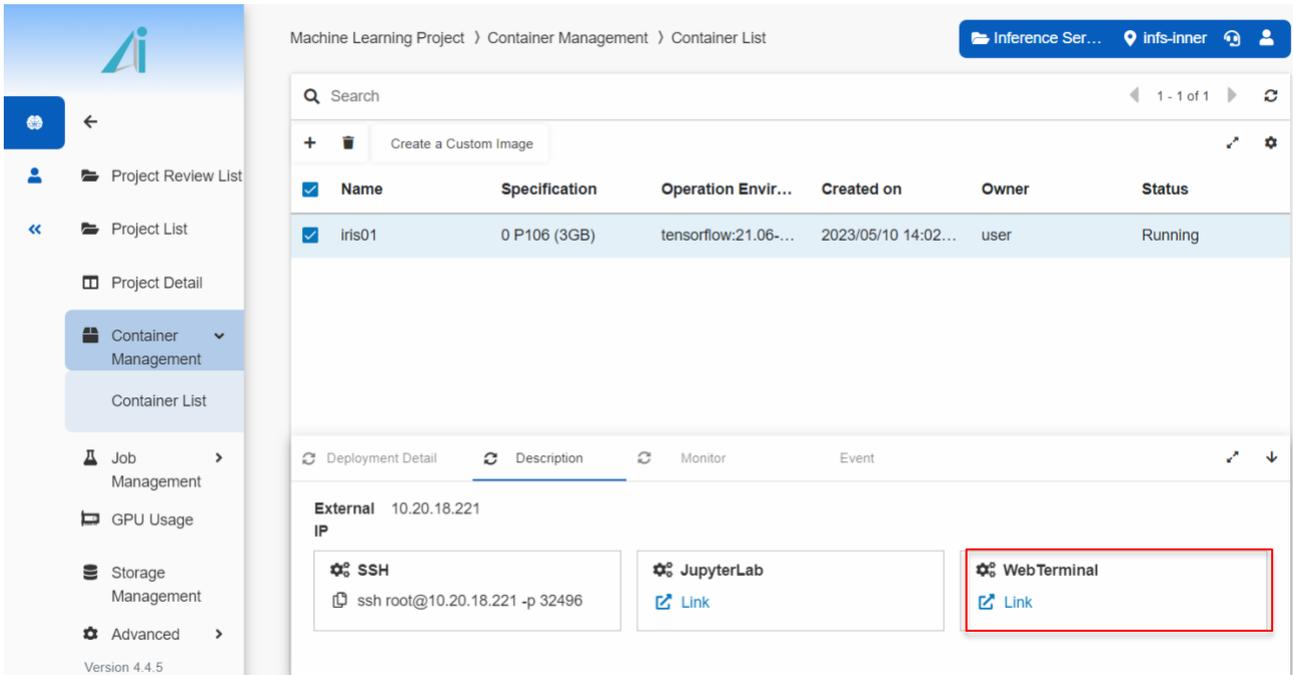
```
In [1]: from platform import python_version
print(python_version())
3.8.5
```

```
In [2]: pip show scikit-learn pandas joblib
WARNING: Package(s) not found: joblib, pandas, scikit-learn
Note: you may need to restart the kernel to use updated packages.
```

```
In [3]: pip install -U scikit-learn==0.24.2
Collecting scikit-learn==0.24.2
  Downloading scikit_learn-0.24.2-cp38-cp38-manylinux2010_x86_64.whl (24.9 MB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 24.9 MB 12.2 MB/s eta 0:00:01
Collecting threadpoolctl>=2.0.0
  Downloading threadpoolctl-3.1.0-py3-none-any.whl (14 kB)
Collecting joblib>=0.11
  Downloading joblib-1.2.0-py3-none-any.whl (297 kB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 297 kB 11.0 MB/s eta 0:00:01
Requirement already satisfied: numpy>=1.13.3 in /usr/local/lib/python3.8/dist-packages (from scikit-learn==0.24.2) (1.19.5)
Requirement already satisfied: scipy>=0.19.1 in /usr/local/lib/python3.8/dist-packages (from scikit-learn==0.24.2) (1.4.1)
Installing collected packages: threadpoolctl, joblib, scikit-learn
Successfully installed joblib-1.2.0 scikit-learn-0.24.2 threadpoolctl-3.1.0
WARNING: Running pip as root will break packages and permissions. You should install packages reliably by using venv: http
s://pip.pypa.io/warnings/venv
WARNING: You are using pip version 21.1.2; however, version 23.1.2 is available.
You should consider upgrading via the '/usr/bin/python -m pip install --upgrade pip' command.
Note: you may need to restart the kernel to use updated packages.
```

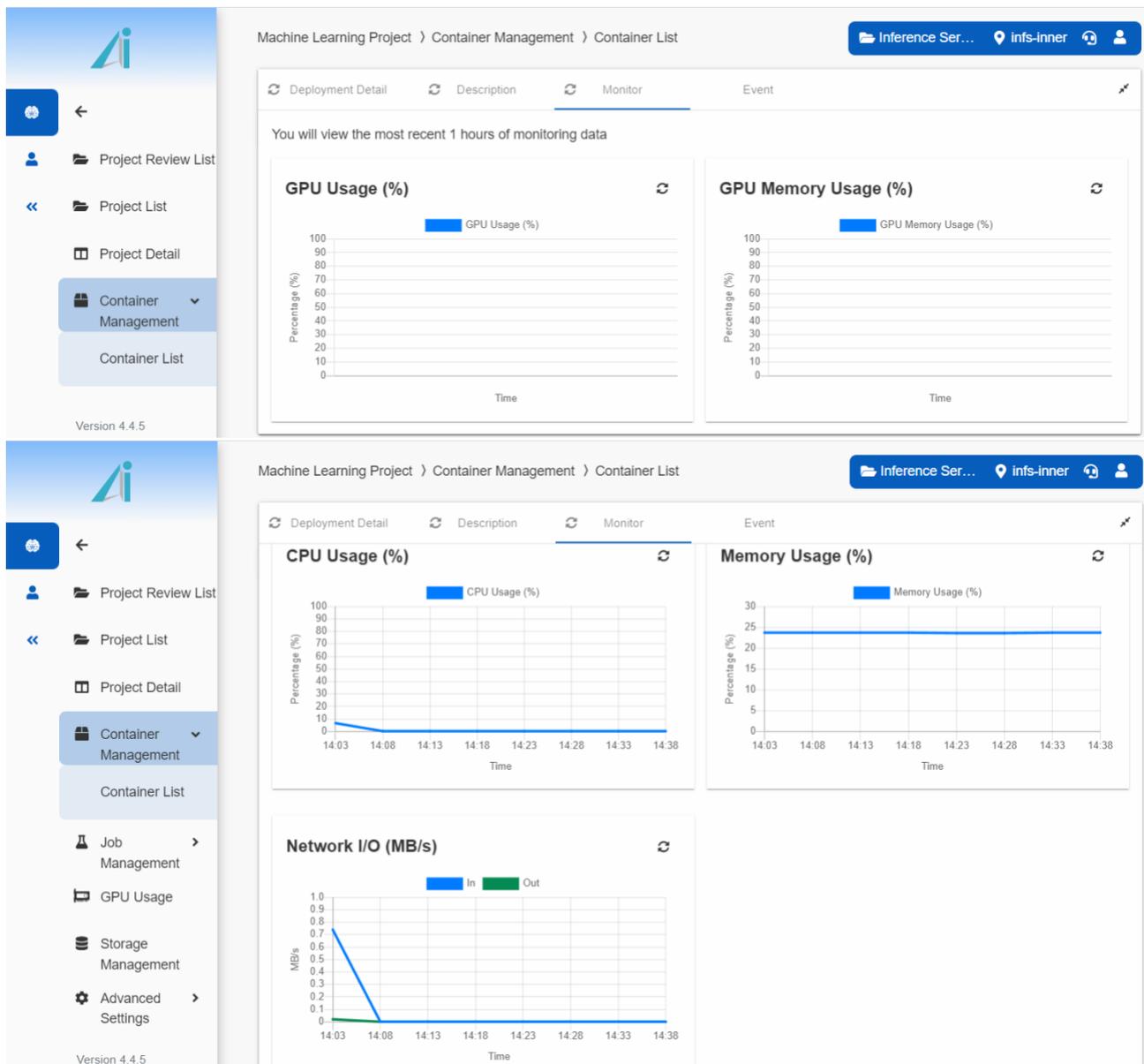
(3) WebTerminal

On the "Service Information" tab, clicking the " WebTerminal " will directly go to WebTerminal.



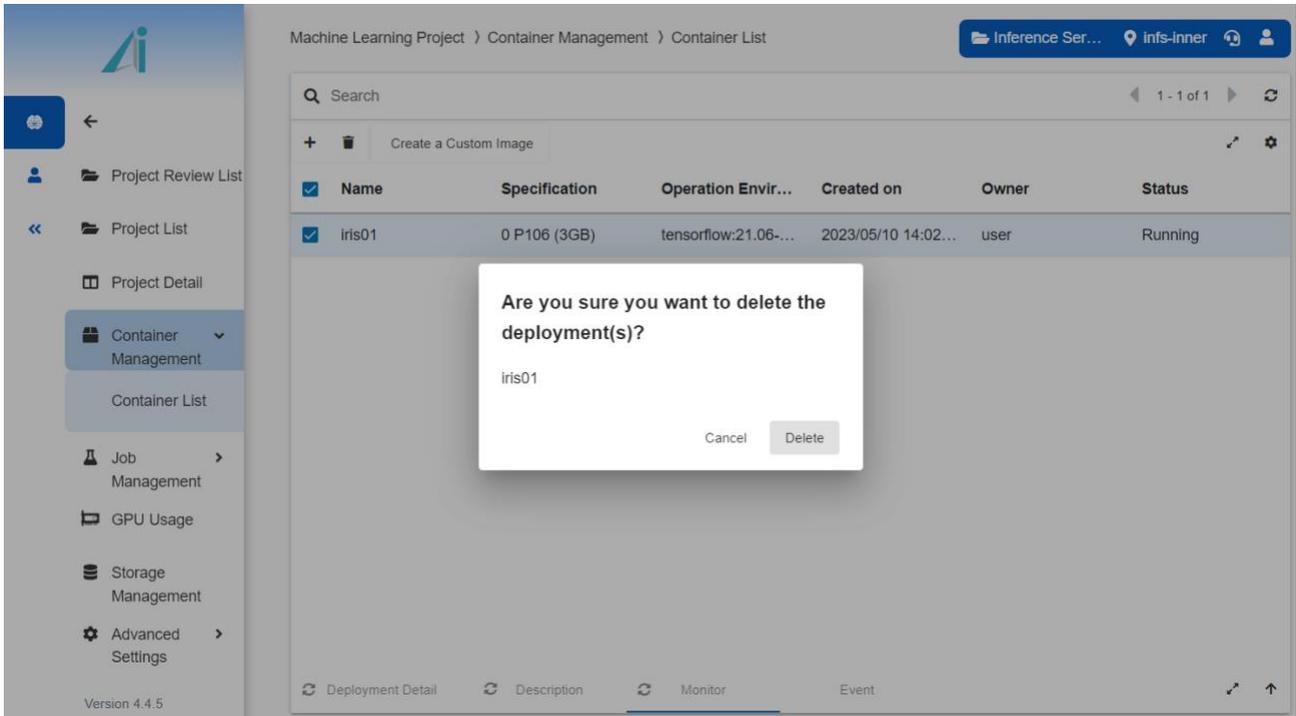
Provisioning

Users can view charts for GPU usage, GPU memory usage, CPU and memory usage, Network I/O, and more through the "Provisioning" tab. By dragging the slider above, users can switch between different intervals (ranging from 1 to 24 hours) to view the corresponding data.



4.3 Delete Container

To delete a container, you can select the target container from the list and click . A confirmation screen will appear, as shown in the figure below. After confirming that you want to delete the container, click [Delete].

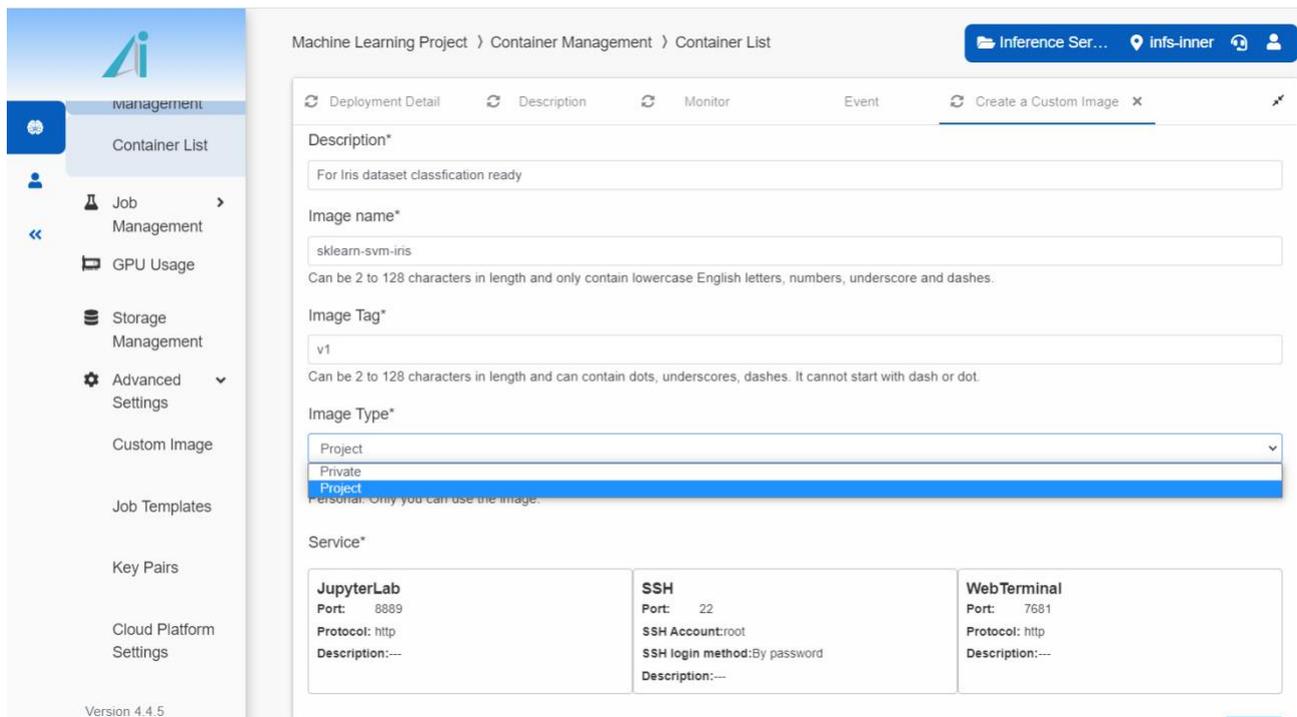


Note: Once a container is deleted, it cannot be recovered. Please ensure that you have backed up the files and data or have stored the data that needs to be retained in a separate location under the mounted device path to ensure that the data can be retrieved later.

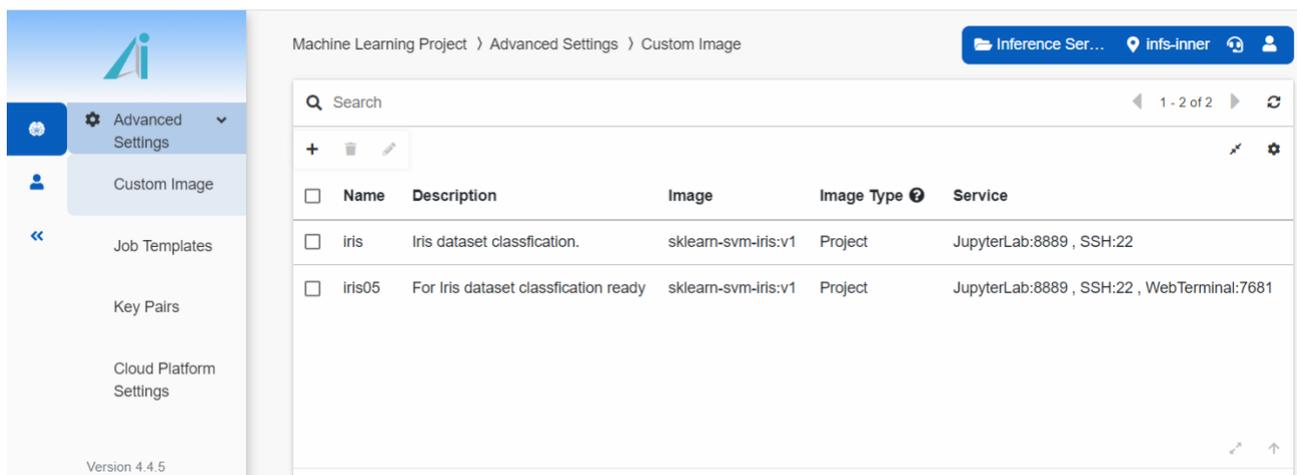
4.4 Create Custom Image

The purpose of this feature is to allow the user to make adjustments to the images. For example, importing additional packages to the original public images without altering the original public image. You can go to the settings page by clicking on [create custom image] after you selected a target container in the container List.

A user can define the name, description, image name, image tag, and the environment to build the image. One can also set the custom image for private uses or share image with project members. Refer to Section 8.1 for further details,



You can view your custom images under "Advanced Settings > Custom Image" page.

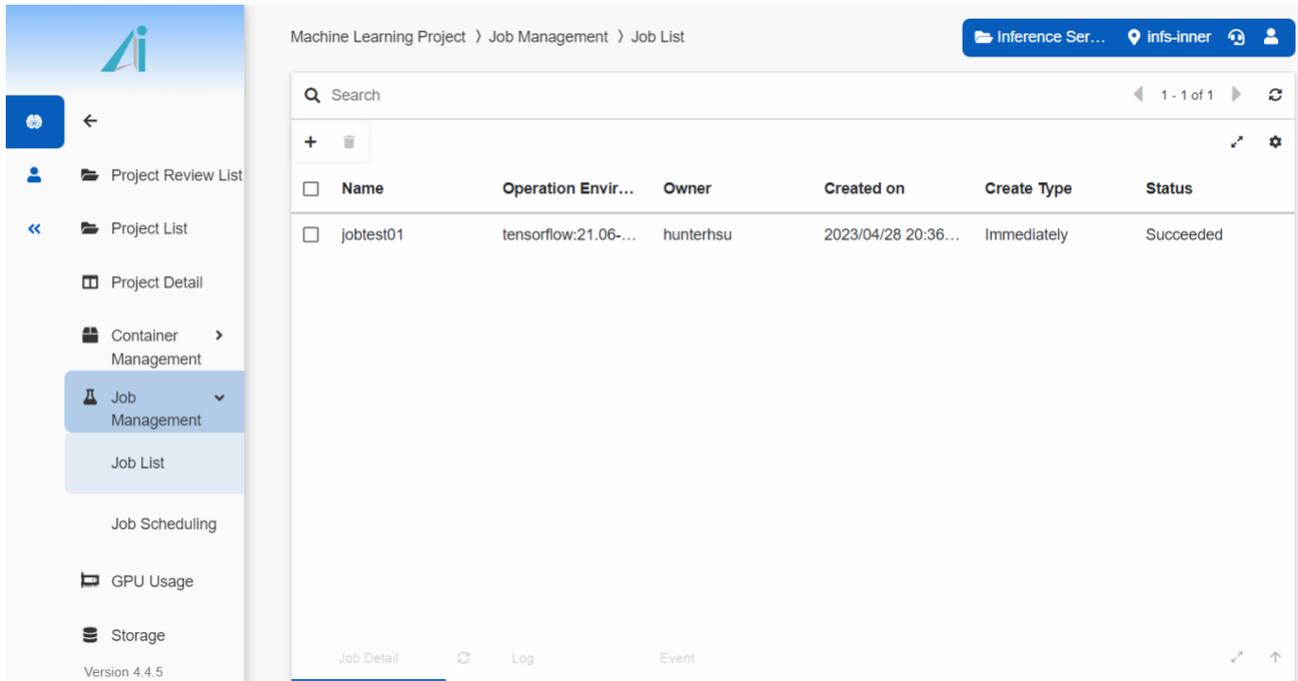


5. Job Management

Job management feature allows the user to create a job that runs a shell script at a specific time. You can choose to execute immediately or arrange to execute by batch. The platform will automatically create a container to execute the commands and delete the container once the job is completed. This increase efficiency as you do not need to terminate the job manually.

5.1 Job List

You can view all jobs in Job Management > Job List by clicking on this tab from the sidebar.



The screenshot shows the 'Job List' page in the 'Job Management' section of the 'Machine Learning Project'. The interface includes a sidebar with navigation options like 'Project Review List', 'Project List', 'Project Detail', 'Container Management', 'Job Management', 'Job List', 'Job Scheduling', 'GPU Usage', and 'Storage'. The main content area displays a table of jobs with the following columns: Name, Operation Envir..., Owner, Created on, Create Type, and Status. A single job, 'jobtest01', is listed with the status 'Succeeded'.

Name	Operation Envir...	Owner	Created on	Create Type	Status
jobtest01	tensorflow:21.06-...	hunterhsu	2023/04/28 20:36...	Immediately	Succeeded

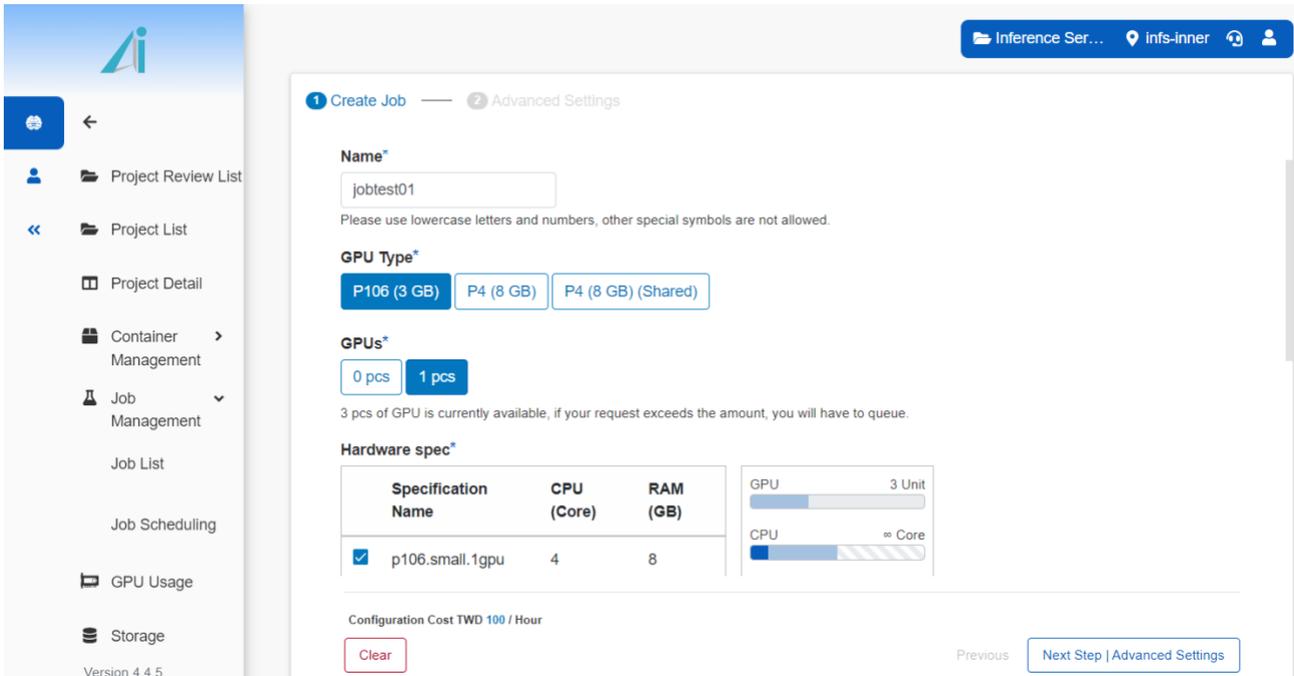
Create Jobs

Steps for job creation:

- (1) Go to the job management page and click the [+] on the top left

(2) Enter the name of your job.

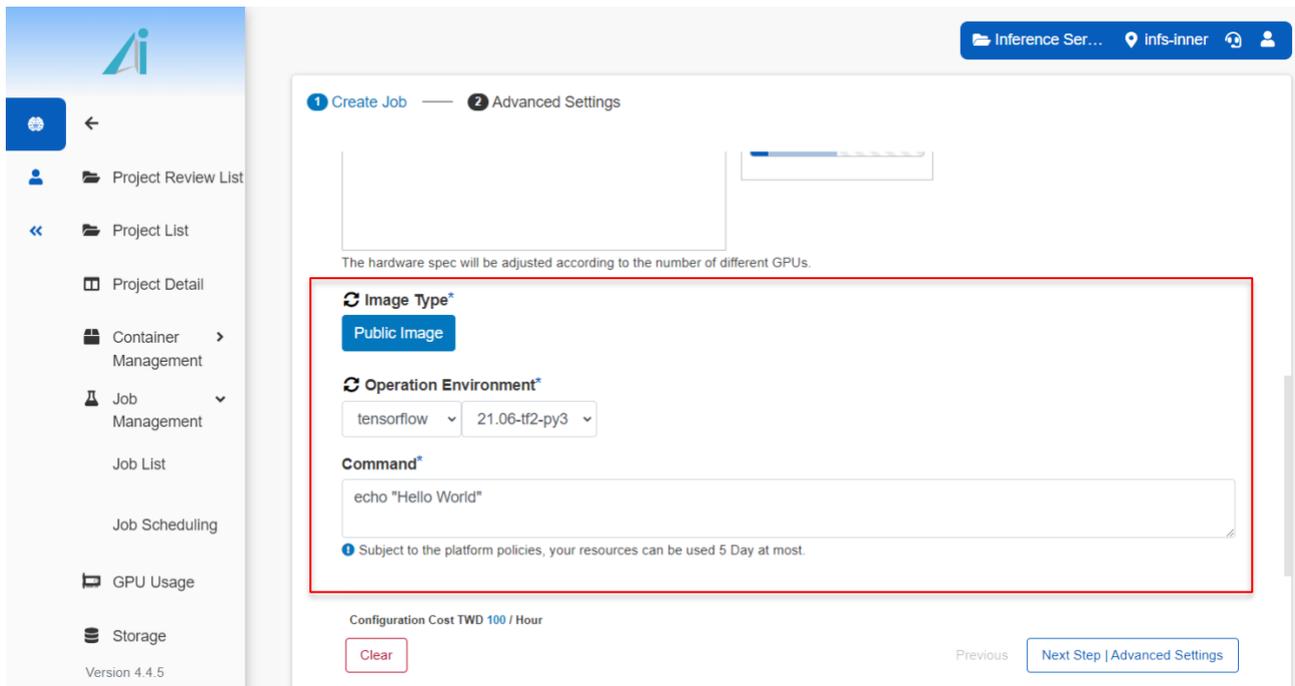
(3) Select the preferred GPU type, GPU quantity and the hardware specifications (including CPU cores and memory)



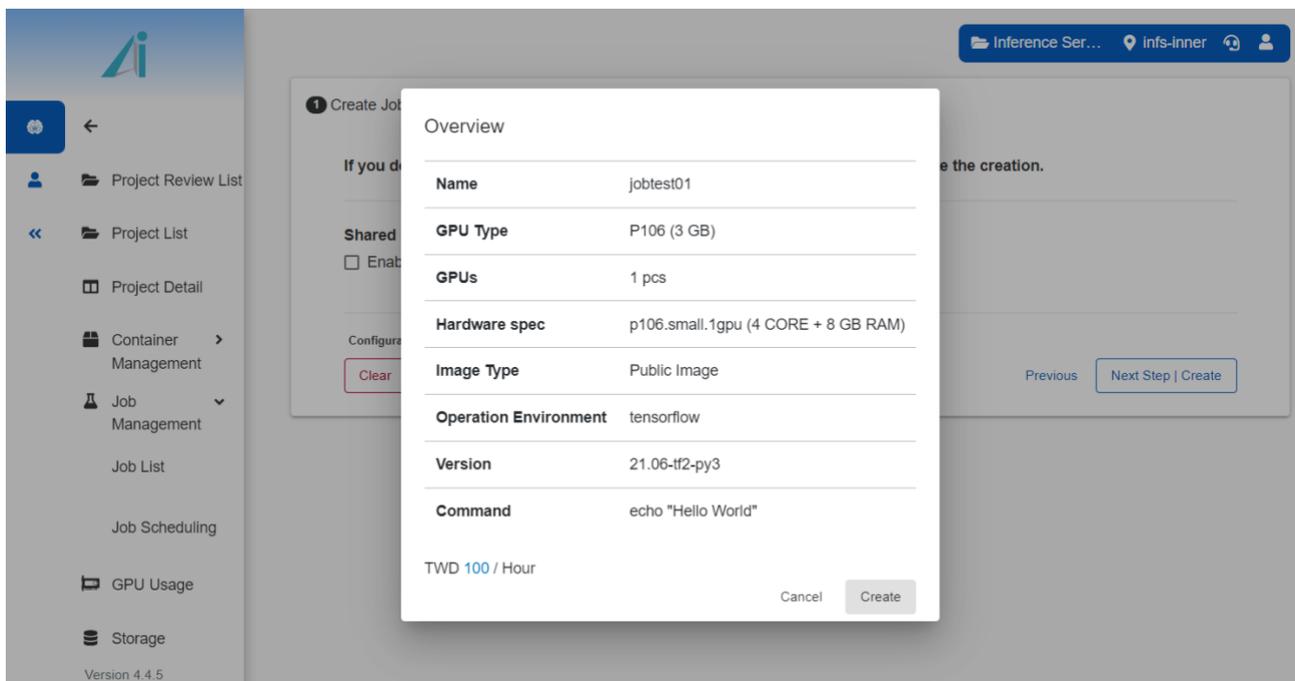
(4) Select either a public image or a customer image (if a customer image is prebuilt)

(5) Select the framework under operation environment.

(6) Enter the command that you wish to run.



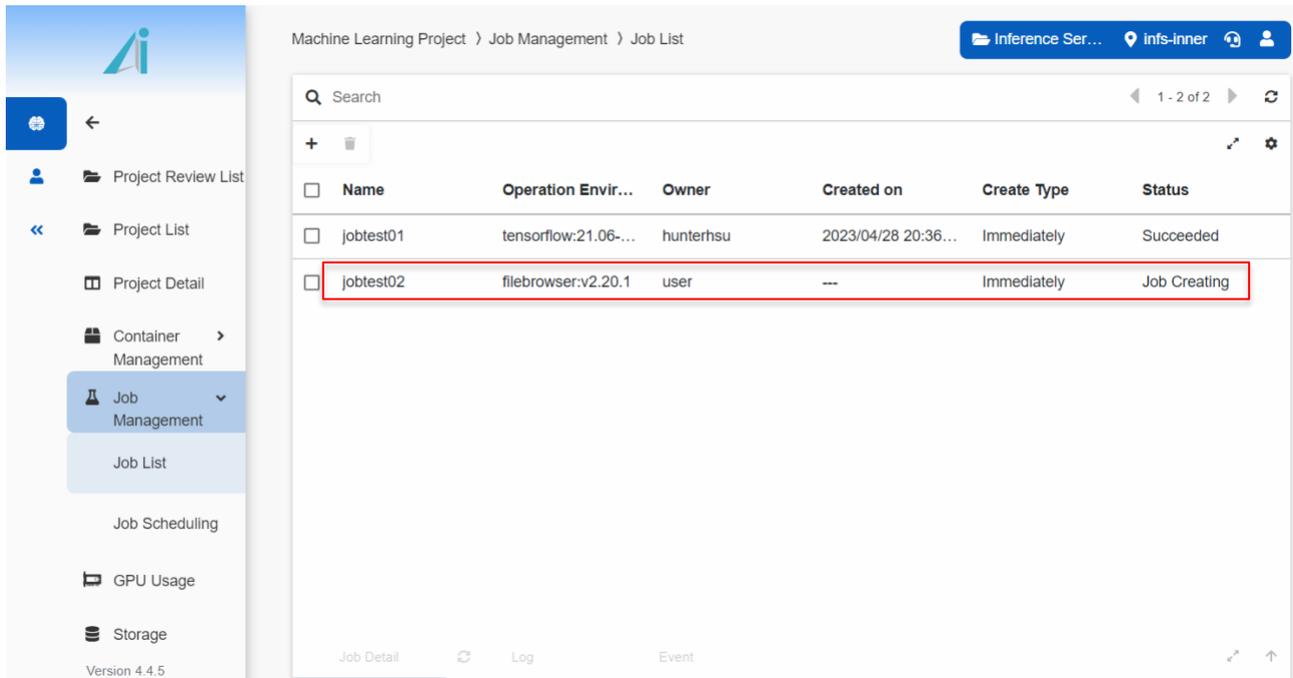
(7) Create a job by clicking on next.



User Guide

(8) Confirm the details in the overview.

(9) You can view the progress in the job list



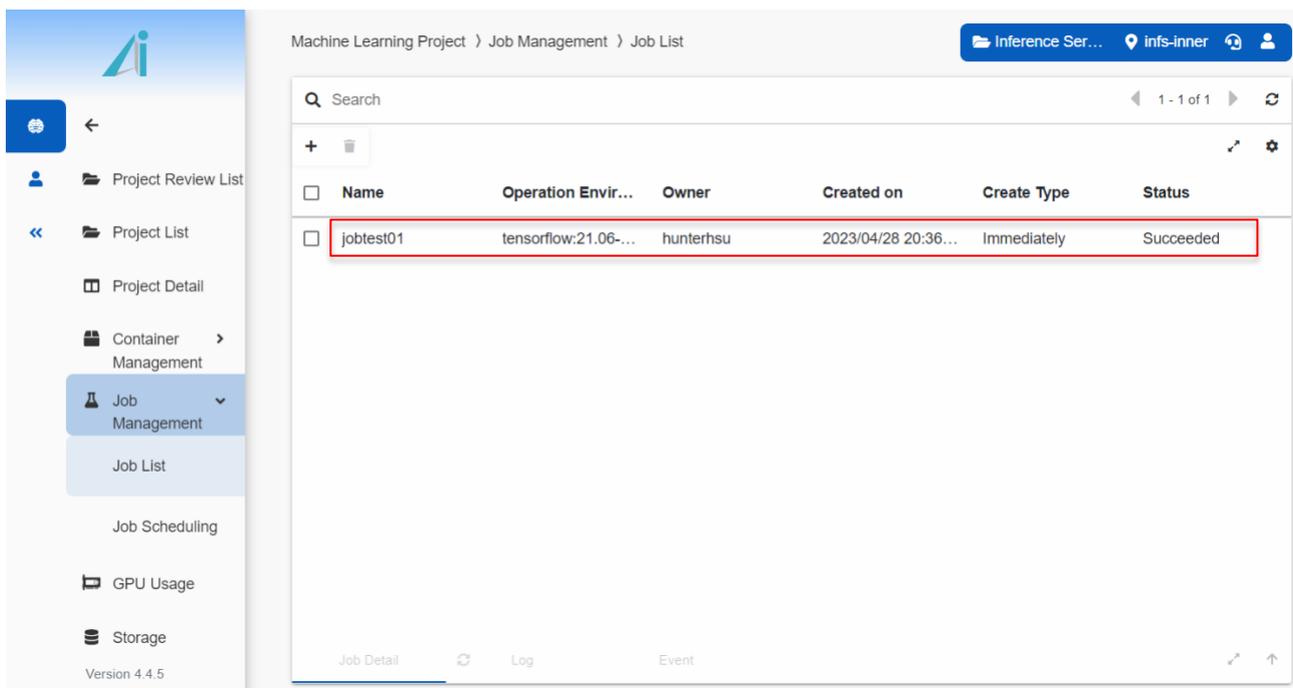
Machine Learning Project > Job Management > Job List

Inference Ser... | Infs-inner

Search 1 - 2 of 2

<input type="checkbox"/>	Name	Operation Envir...	Owner	Created on	Create Type	Status
<input type="checkbox"/>	jobtest01	tensorflow:21.06...	hunterhsu	2023/04/28 20:36...	Immediately	Succeeded
<input type="checkbox"/>	jobtest02	filebrowser:v2.20.1	user	--	Immediately	Job Creating

Job Detail | Log | Event



Machine Learning Project > Job Management > Job List

Inference Ser... | Infs-inner

Search 1 - 1 of 1

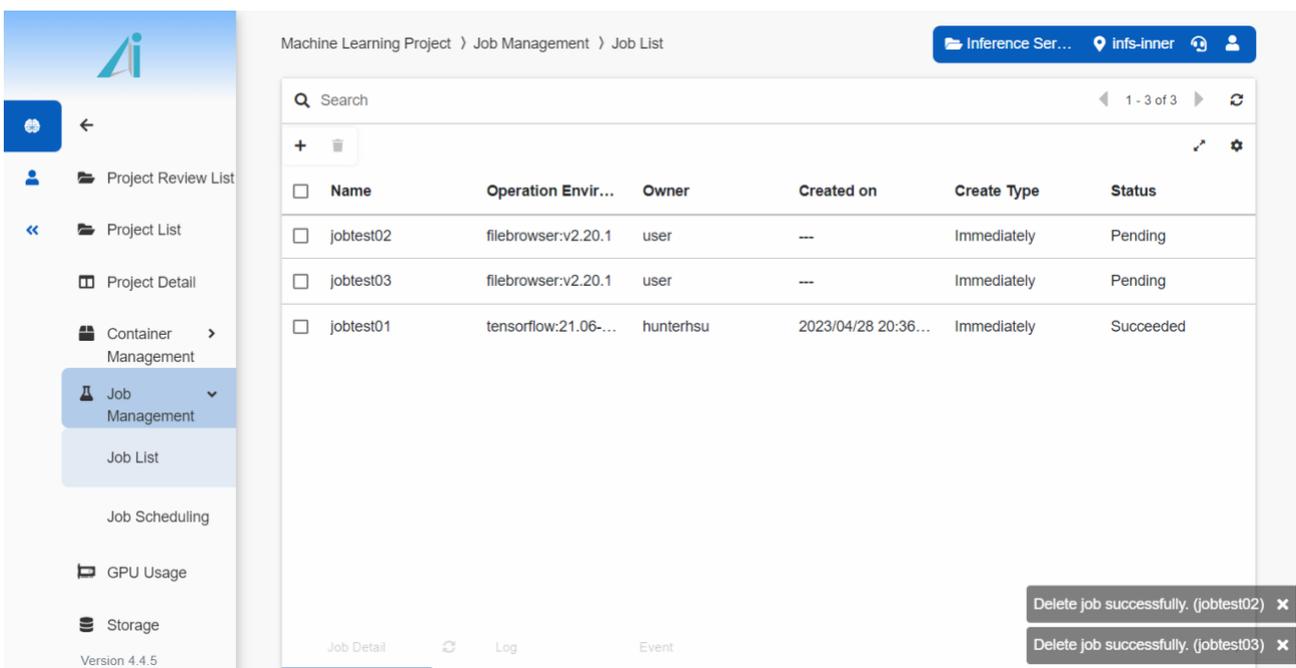
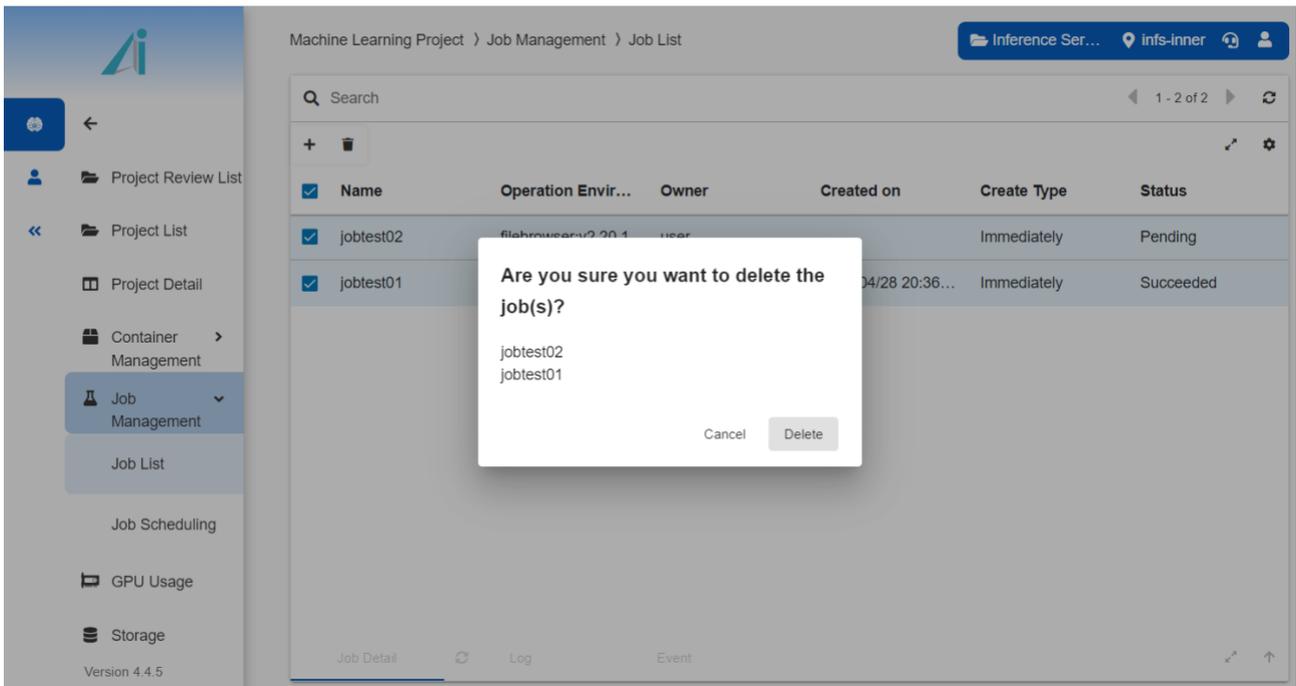
<input type="checkbox"/>	Name	Operation Envir...	Owner	Created on	Create Type	Status
<input type="checkbox"/>	jobtest01	tensorflow:21.06...	hunterhsu	2023/04/28 20:36...	Immediately	Succeeded

Job Detail | Log | Event

User Guide

Delete Jobs

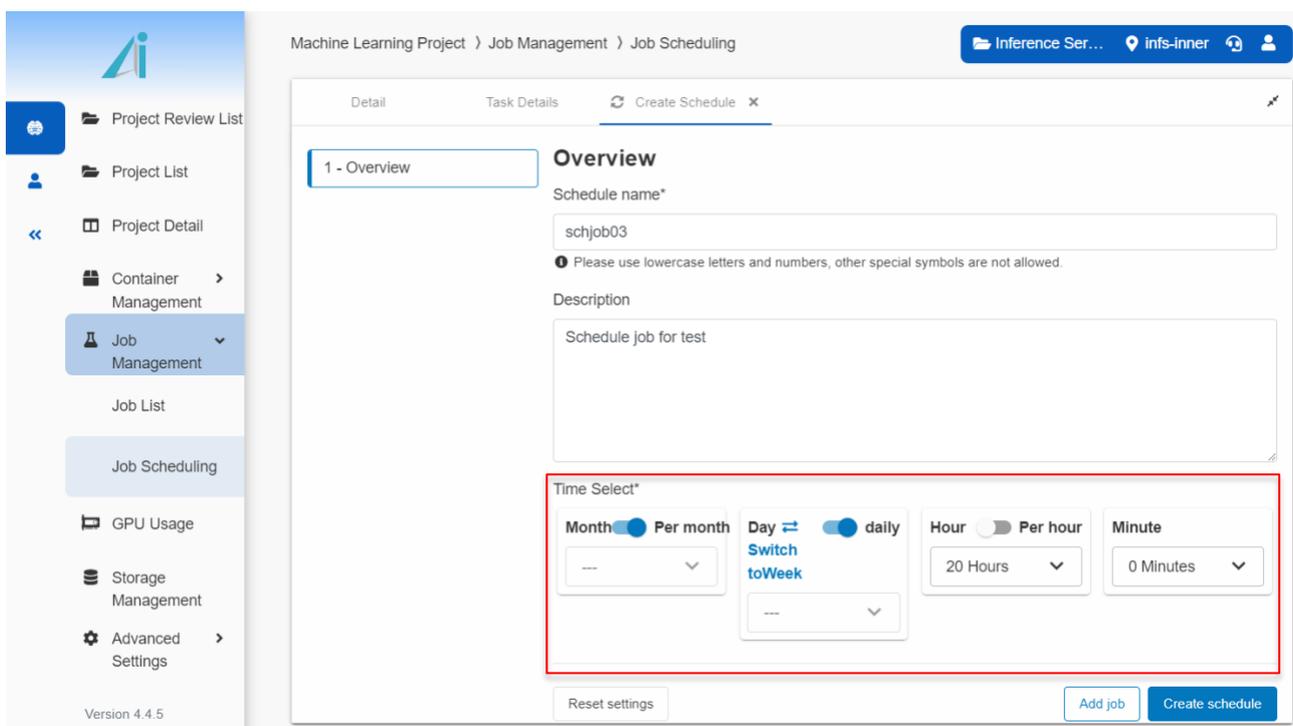
Delete the jobs by clicking  on the top left side of job list



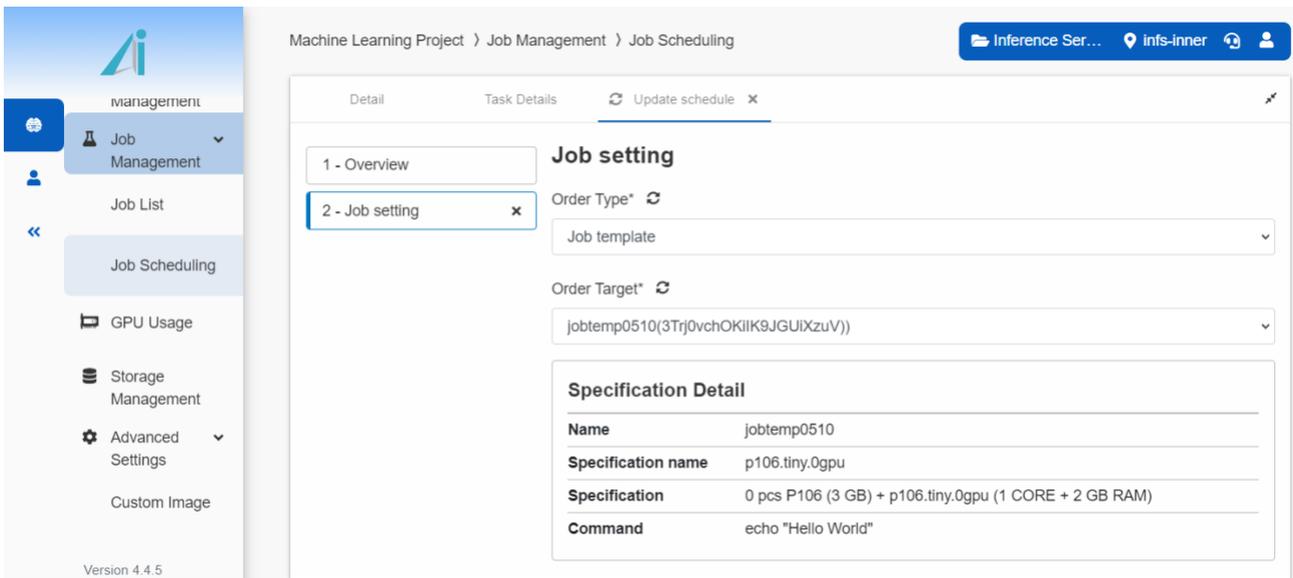
5.2 Job Scheduling

There are 2 ways of setting up job schedules:

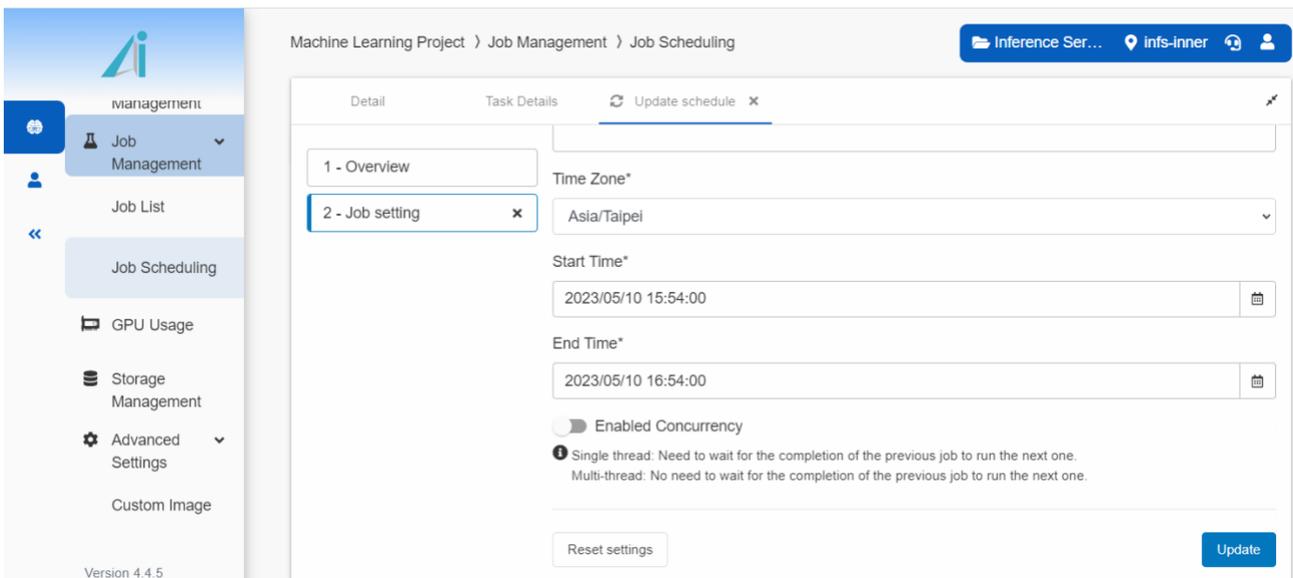
- (1) Set up job schedule in “Job Management > Job scheduling” and linked this to a job template .
 - a. Go to Job Management > job schedule
 - b. Click [+] and you will be directed to the page to create new schedule
 - c. Enter the name and the description of the schedule
 - d. Enter the execution time and frequency of your job schedule.
 - e. Click on [Create Schedule] to further set up your job.



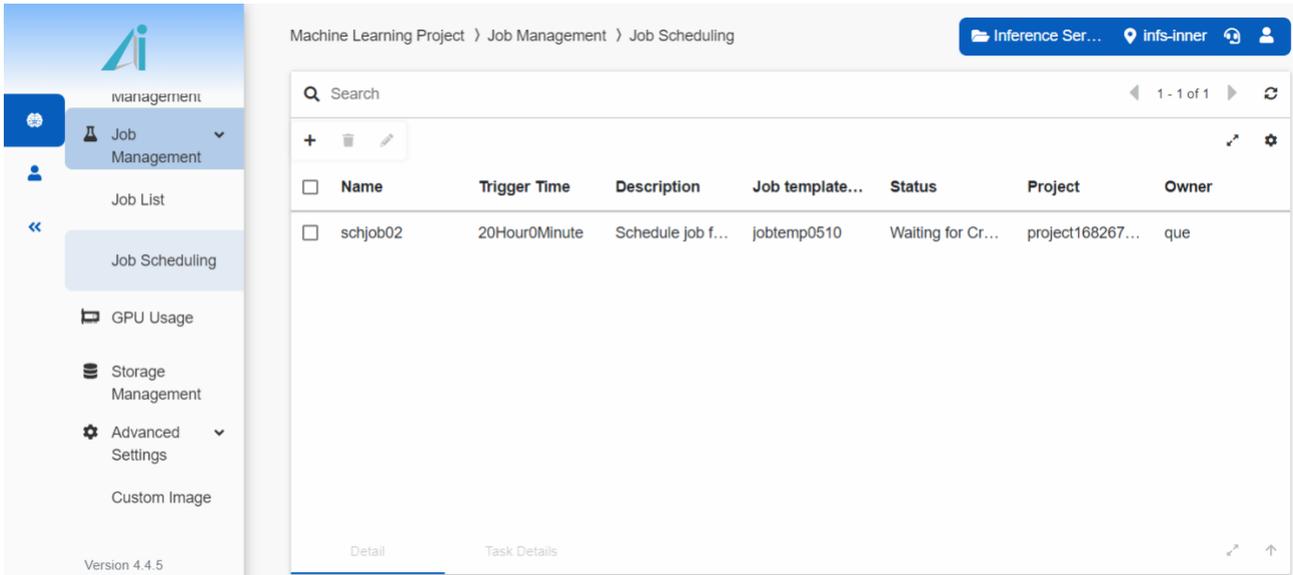
- f. Under Job target, select an existing job template* for the details to be filled. (*You will have to create a job target under Advanced Settings > Job Template. Refer for Section 8.2 for further details).



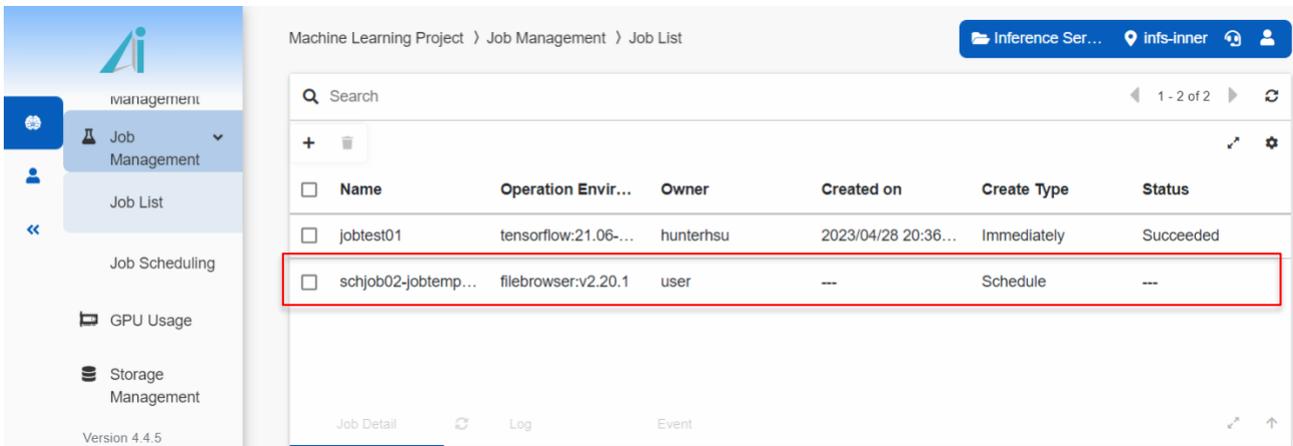
- g. Select the time zone, start time and end time
- h. Confirm whether to activate multi-thread.
 - single thread: Jobs are executed by sequence
 - multi thread: Jobs are executed in parallel.
- i. Click on the create button to confirm job creation.



- j. You can see your job in the job list when it is successfully created.

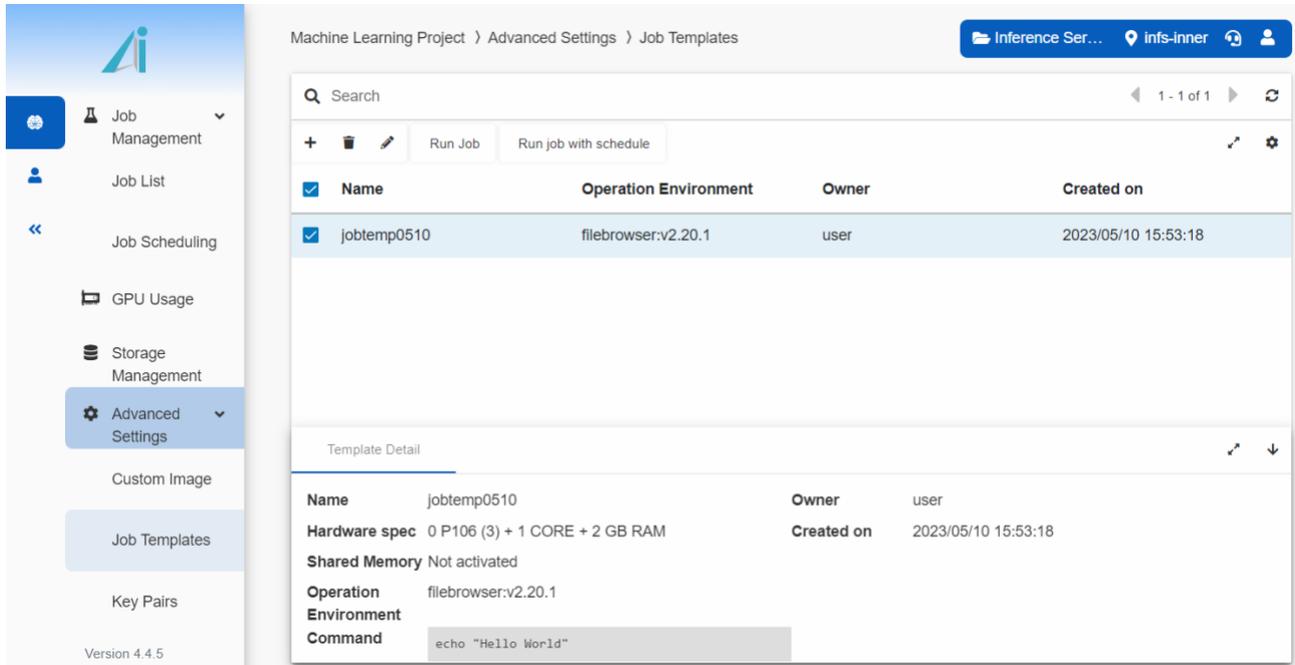


k. The job list will also show the job that is about to be initiated.

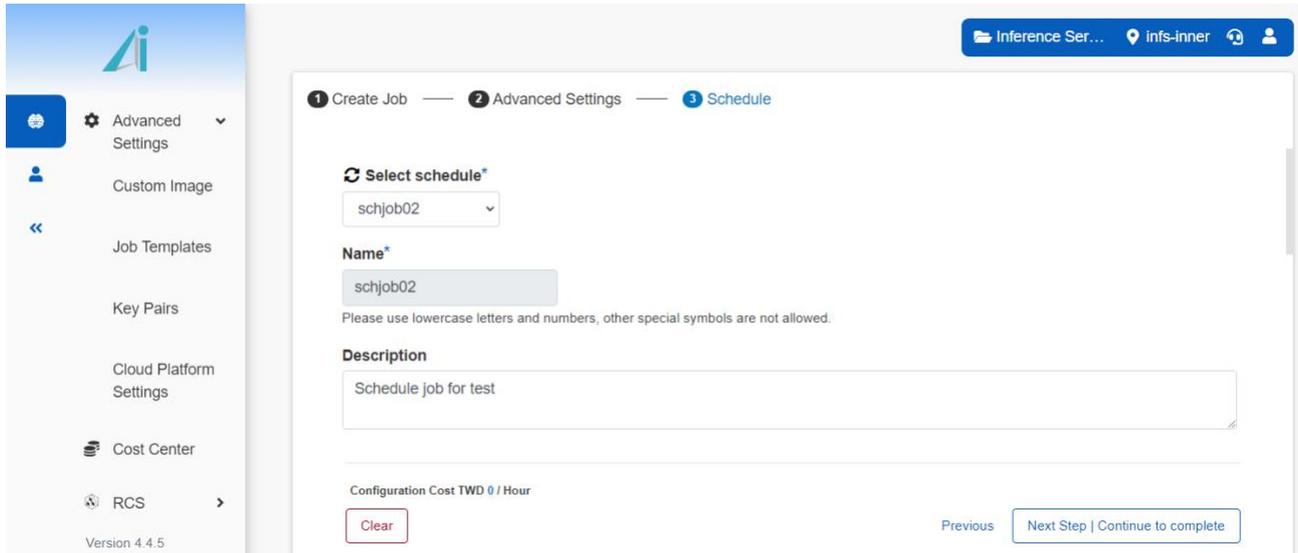


Note: You will have to delete the job schedule to delete it.

- (2) From the “Advanced Settings > Job Template” page, select the for the job template to be created or associated with a job schedule.
 - a. Go to “Advanced Settings > Job Template from the sidebar.
 - b. Select the template and click [Run job with schedule].



- c. Go to job schedule settings, and you can create a schedule, or edit existing schedules.
- d. If you choose to use existing schedules, the information will be carried out, and you will not be able to change the name of the schedule.

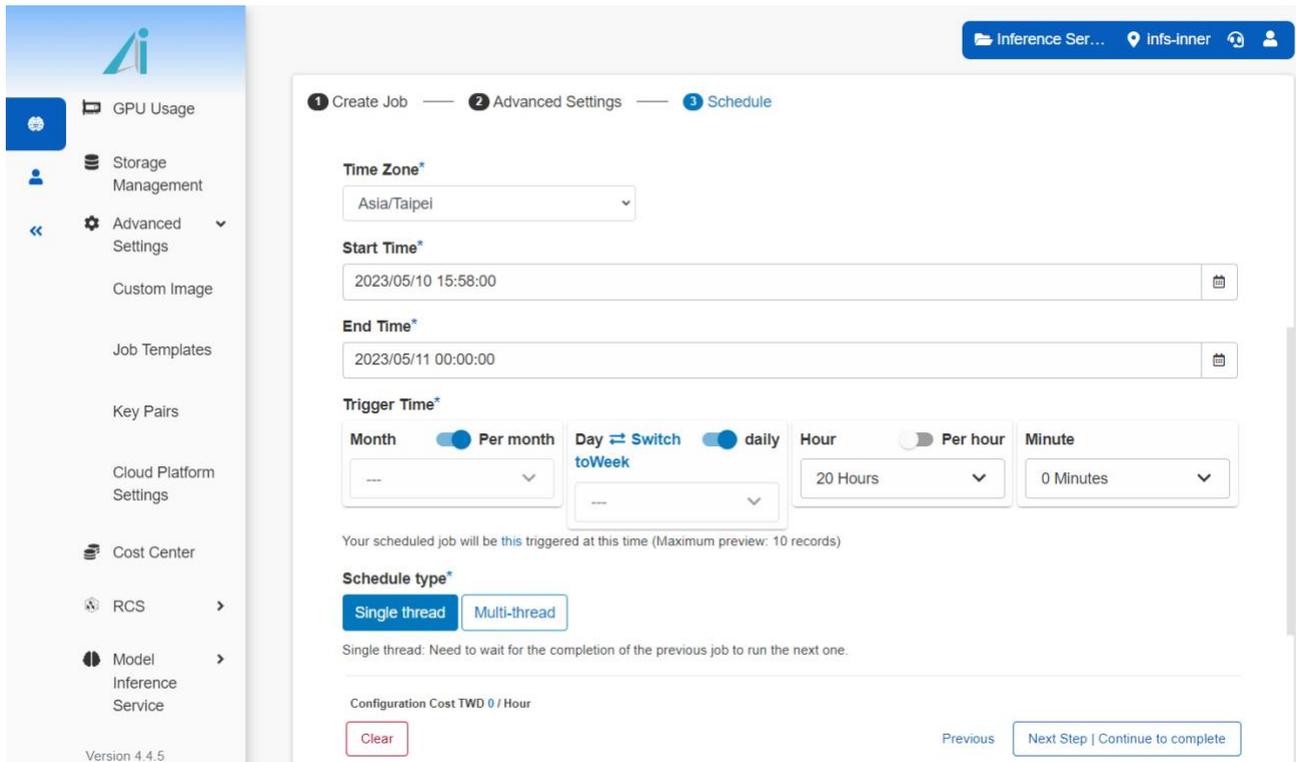


The screenshot shows the 'Schedule' step of a job creation process. The breadcrumb trail at the top indicates the steps: 1 Create Job, 2 Advanced Settings, and 3 Schedule. The main content area includes:

- Select schedule***: A dropdown menu with 'schjob02' selected.
- Name***: A text input field containing 'schjob02'. Below it, a note states: 'Please use lowercase letters and numbers, other special symbols are not allowed.'
- Description**: A text area containing 'Schedule job for test'.
- Configuration Cost TWD 0 / Hour**: A label above a 'Clear' button.
- Navigation**: 'Previous' and 'Next Step | Continue to complete' buttons.

The left sidebar contains navigation options: Advanced Settings, Custom Image, Job Templates, Key Pairs, Cloud Platform Settings, Cost Center, and RCS. The version number 'Version 4.4.5' is displayed at the bottom of the sidebar.

- e. After confirming the time zone, start date, end date, initiation time, and schedule type, click [Next Step | Continue to complete].



1 Create Job — 2 Advanced Settings — 3 Schedule

Time Zone*
Asia/Taipei

Start Time*
2023/05/10 15:58:00

End Time*
2023/05/11 00:00:00

Trigger Time*
 Month Per month Day Switch toWeek daily
 Hour Per hour Minute
 --- Day --- 20 Hours 0 Minutes

Your scheduled job will be [this](#) triggered at this time (Maximum preview: 10 records)

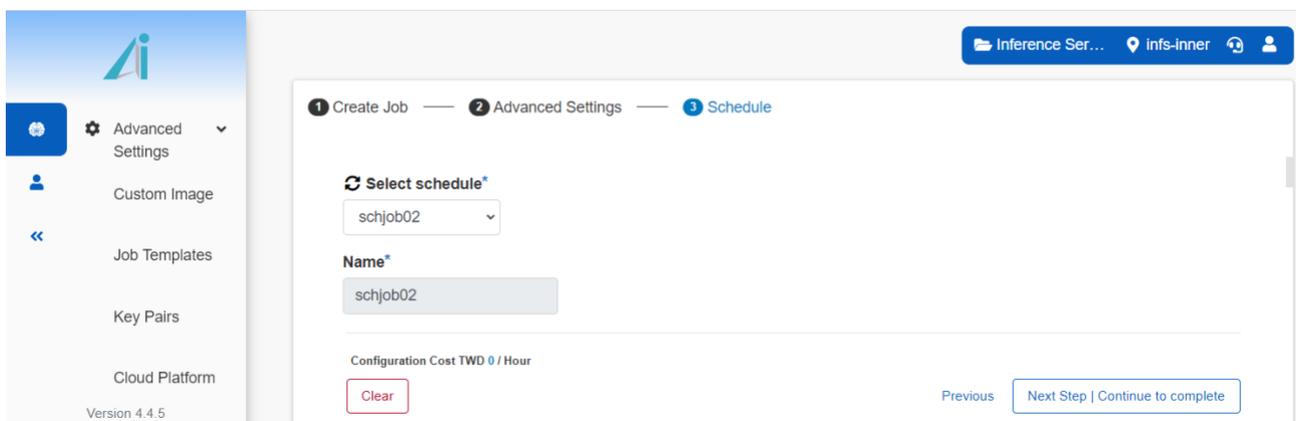
Schedule type*
 Single thread Multi-thread

Single thread: Need to wait for the completion of the previous job to run the next one.

Configuration Cost TWD 0 / Hour

Clear Previous Next Step | Continue to complete

- f. If you want to set up a new schedule, you will have to create another one from the start entering the name, start time, end time, initiation time and schedule type. Click [Next Step | Continue to Complete] to proceed.



1 Create Job — 2 Advanced Settings — 3 Schedule

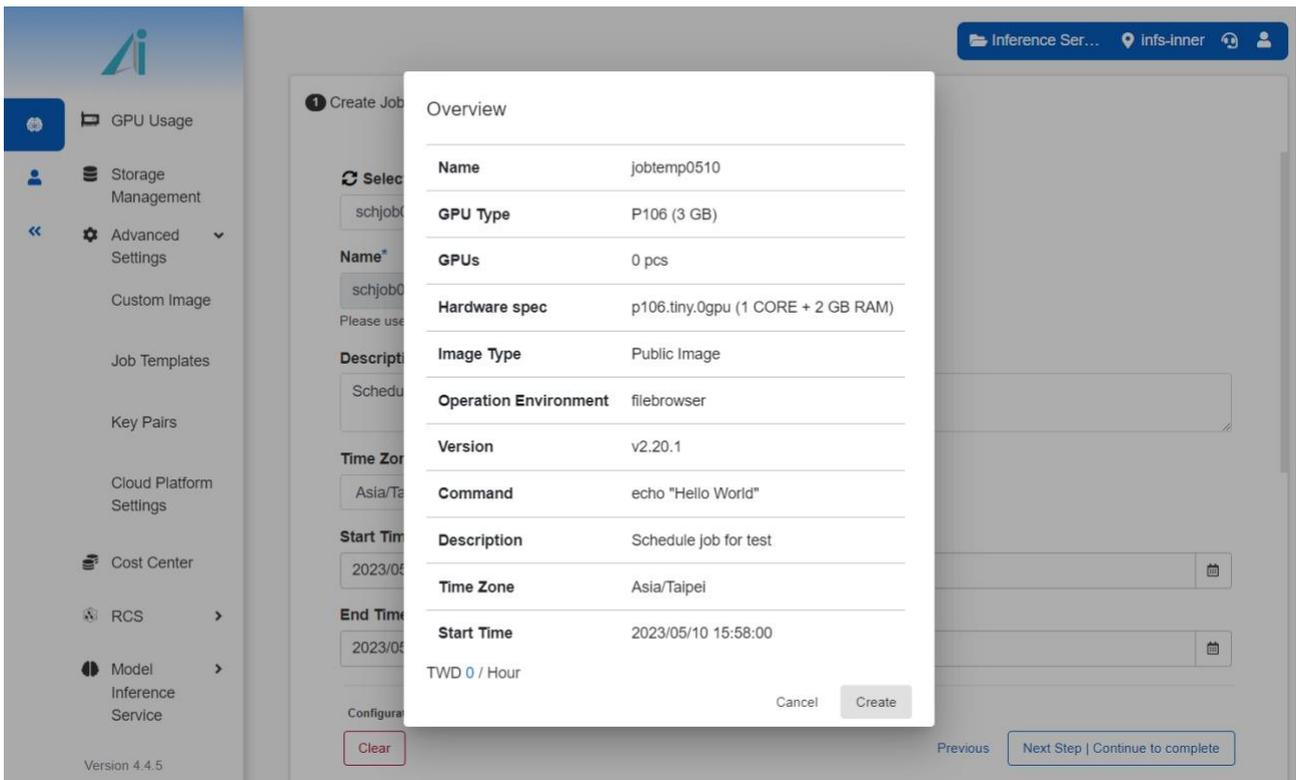
Select schedule*
schjob02

Name*
schjob02

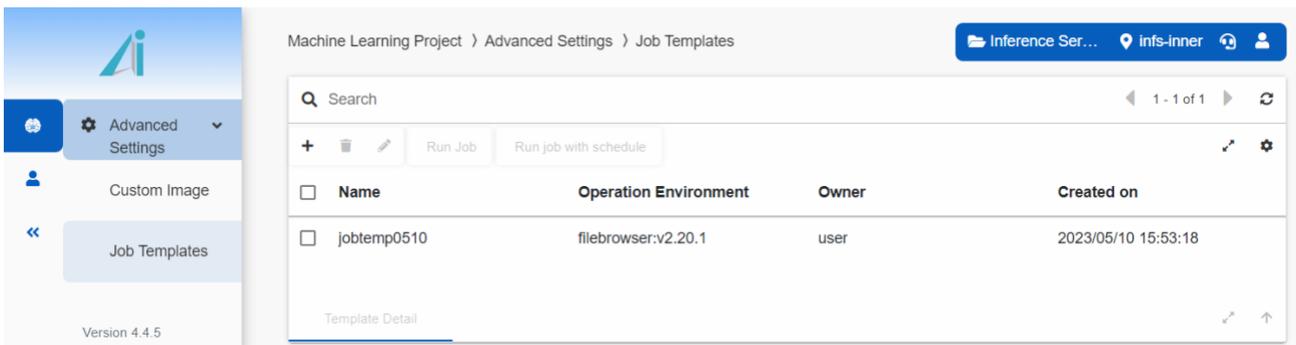
Configuration Cost TWD 0 / Hour

Clear Previous Next Step | Continue to complete

g. Confirm and click Create if after confirming all information in the overview.



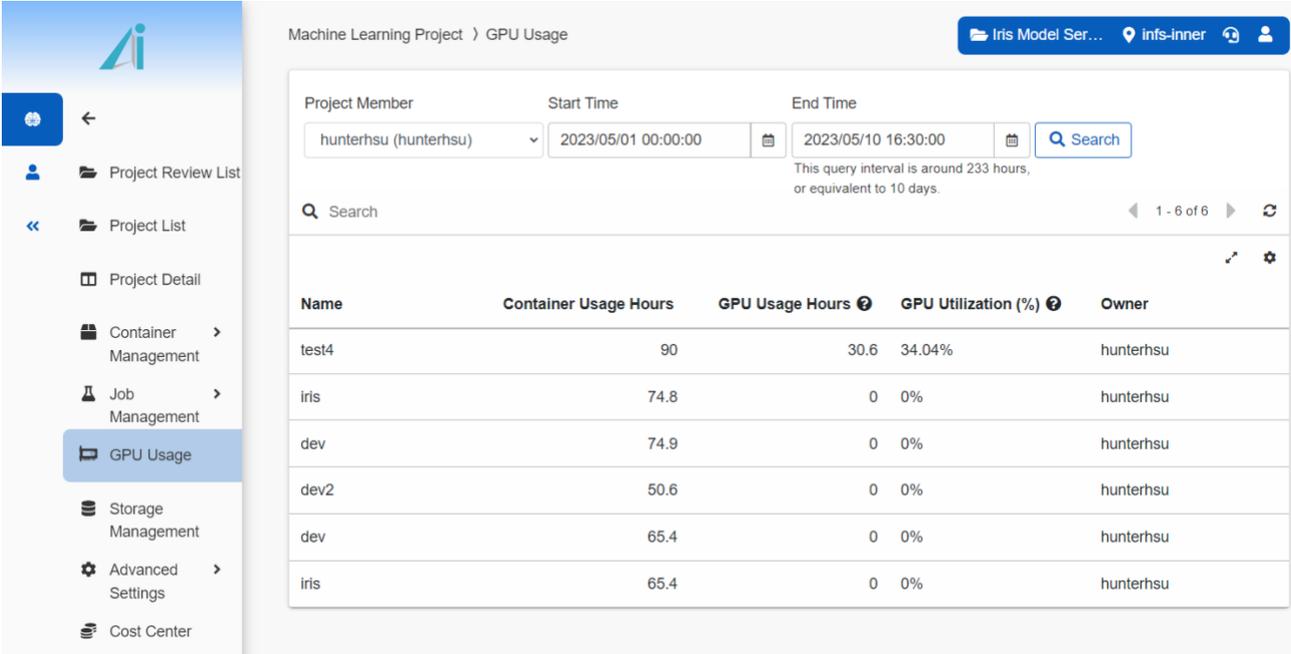
h. The job will be shown on the job list once it is successfully created.



6. GPU Utilization Rate

GPU utilization rate feature allows users to inspect the GPU utilization trends in each project. Information includes container usage times, GPU utilization times and rate.

can access this by clicking on the “GPU utilization Rate” tab in the sidebar.



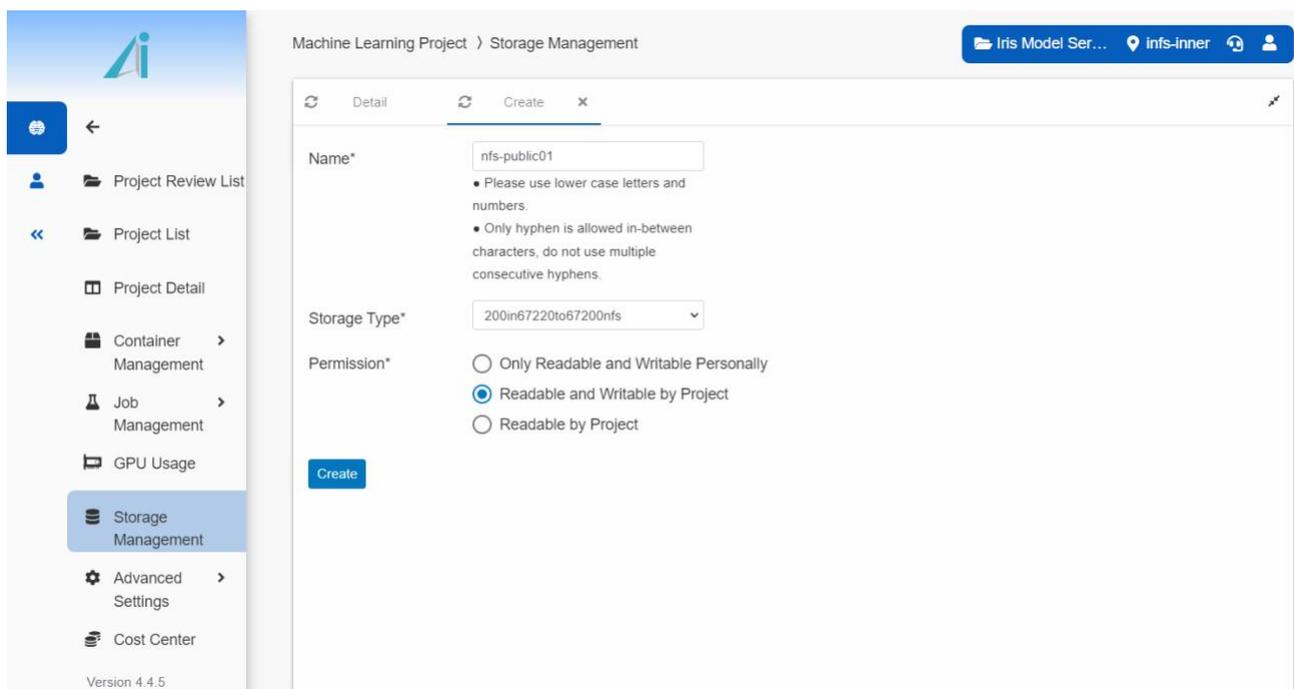
Name	Container Usage Hours	GPU Usage Hours	GPU Utilization (%)	Owner
test4	90	30.6	34.04%	hunterhsu
iris	74.8	0	0%	hunterhsu
dev	74.9	0	0%	hunterhsu
dev2	50.6	0	0%	hunterhsu
dev	65.4	0	0%	hunterhsu
iris	65.4	0	0%	hunterhsu

7. Storage Management

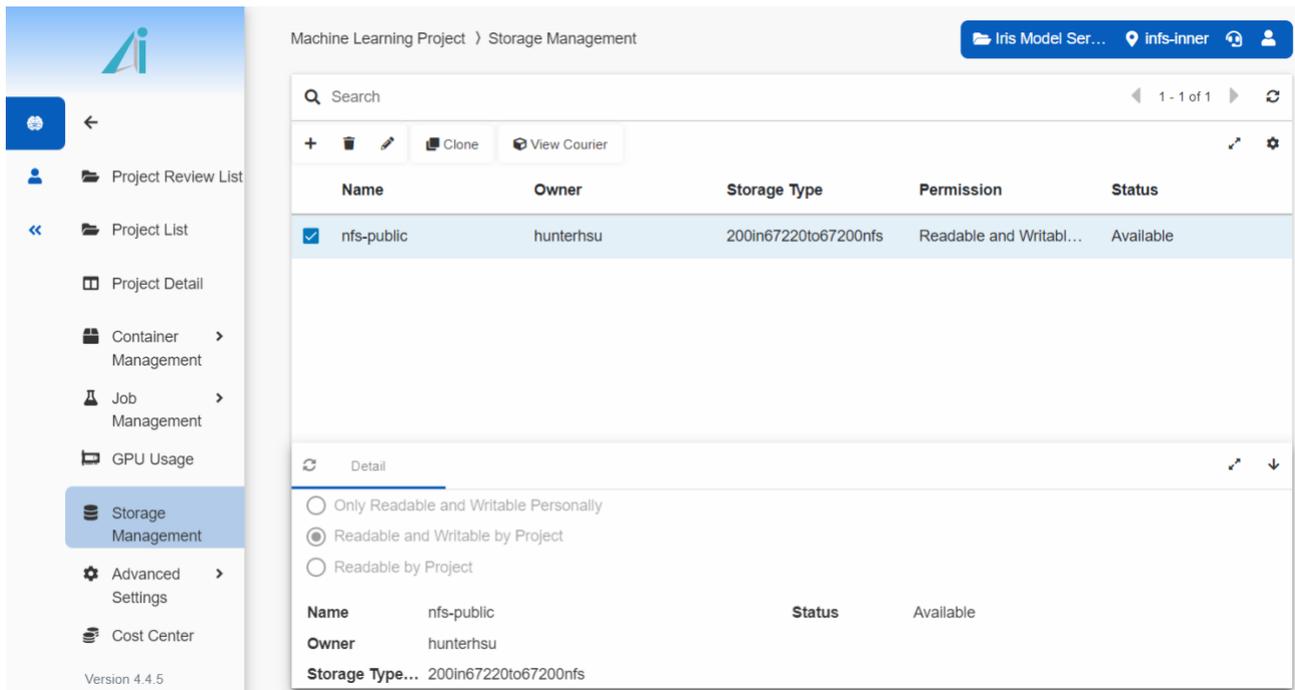
Users can create their own data storage, which can be selected as storage devices and specified mount paths when creating containers (please refer to the Section 4.1 Container Creation General Creation step 10 for further details). This provides users with additional storage space for training purposes. Furthermore, using externally mounted storage space ensures that data remains intact even if the container is deleted.

Here are the steps to create a storage:

- (1) Go to Storage Management from the sidebar.
- (2) Click [+]
- (3) Enter the name and the storage type (storage type is set up by the platform administrator from the admin portal).
- (4) Set the Permission for each storage, including "Only readable and writable personally", "Readable and Writable by project" and "Readable by project".
- (5) Click [Create] to proceed.



(6) The new storage created will appear on the storage management page.



Machine Learning Project > Storage Management

Iris Model Ser... Infs-Inner

Search 1 - 1 of 1

+ [trash] [edit] Clone View Courier [refresh] [settings]

Name	Owner	Storage Type	Permission	Status
<input checked="" type="checkbox"/> nfs-public	hunterhsu	200in67220to67200nfs	Readable and Writabl...	Available

Detail

Only Readable and Writable Personally
 Readable and Writable by Project
 Readable by Project

Name nfs-public **Status** Available
Owner hunterhsu
Storage Type... 200in67220to67200nfs

Version 4.4.5

8. Advanced Settings

8.1 Custom Image

When users create containers, they can choose from a selection of images that are pre-uploaded by the platform administrators through the admin portal. These images are commonly referred to as public images. However, there are cases where additional package installations or environment parameter adjustments are required for the containers created from these public images due to development needs. If a container is deleted, it would require time to reinstall the packages and make adjustments to rebuild the environment, which can impact work efficiency and cause inconvenience.

To address this situation, our platform provides a custom image feature. It allows users to preserve the modified container as a template for future use. These custom images can be set for personal use or shared among project members.

Create Custom Image

- (1) Go to "Advanced Setting > Custom Image" from the sidebar.
- (2) Click the [+] icon on the top left.
- (3) Select the container you wish to set as custom image, and click [Create a Custom Image]

Machine Learning Project > Container Management > Container List

Inference Ser... | info-inner

Search

1 - 1 of 1

+

Create a Custom Image

<input checked="" type="checkbox"/>	Name	Specification	Operation Envir...	Created on	Owner	Status
<input checked="" type="checkbox"/>	iris01	0 P106 (3GB)	tensorflow:21.06-...	2023/05/10 14:02...	user	Running

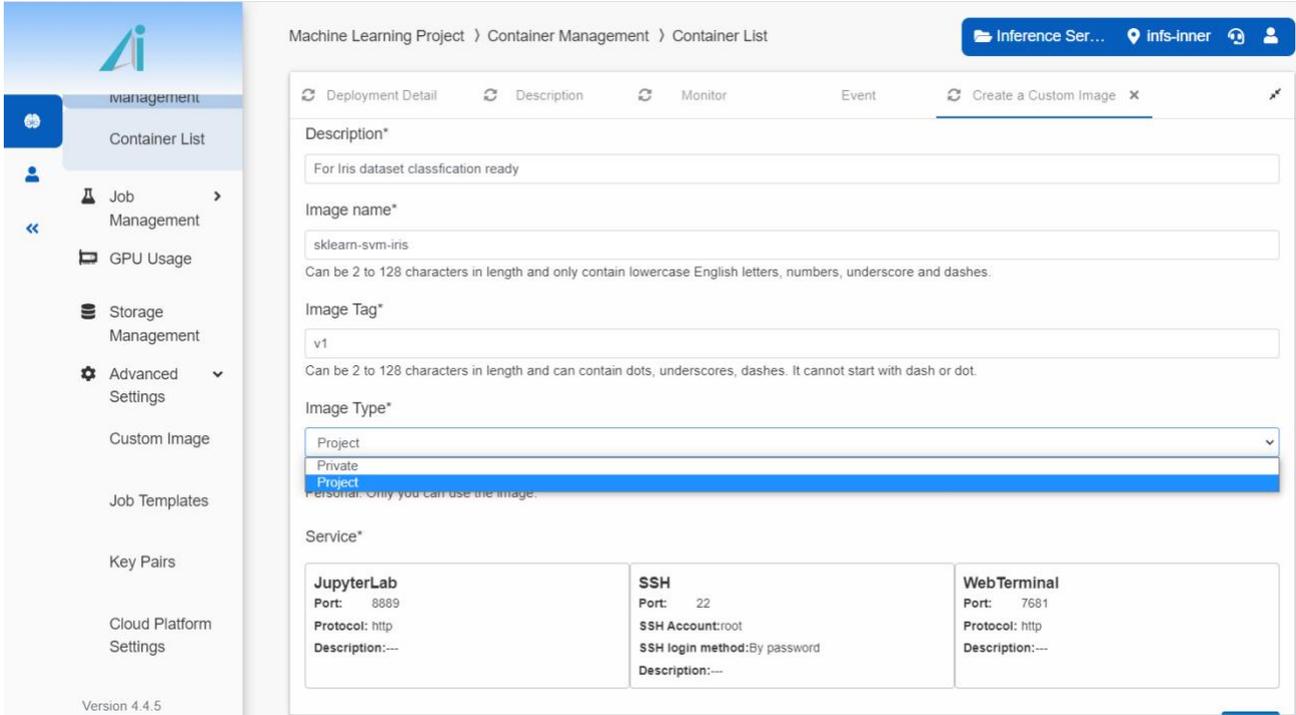
Deployment Detail | Description | Monitor | Event

Name	iris01	Status	Running
Deployment ID	53bfbef6-7388-4d55-a7ad-caa33e39d850:iris01	Created on	2023/05/10 14:02:21
Operation En...	tensorflow:21.06-tf2-py3	End Time	2023/05/15 14:03:40
Specification	0 P106 (3GB) + 1 CORE + 1 GB RAM		
Shared Mem...	1024 MB		
Firewall Rule	Allow 0.0.0.0/0 segments to connect.		

Version 4.4.5

User Guide

- (4) Enter the name and description.
- (5) Enter the image name and image tag. (These will sync to the image repository)
- (6) Select the image type to set the image for personal use or share with the project.



Machine Learning Project > Container Management > Container List

Inference Ser... | info-inner | [User Icon]

Deployment Detail | Description | Monitor | Event | Create a Custom Image x

Description*

For Iris dataset classification ready

Image name*

sklearn-svm-iris

Can be 2 to 128 characters in length and only contain lowercase English letters, numbers, underscore and dashes.

Image Tag*

v1

Can be 2 to 128 characters in length and can contain dots, underscores, dashes. It cannot start with dash or dot.

Image Type*

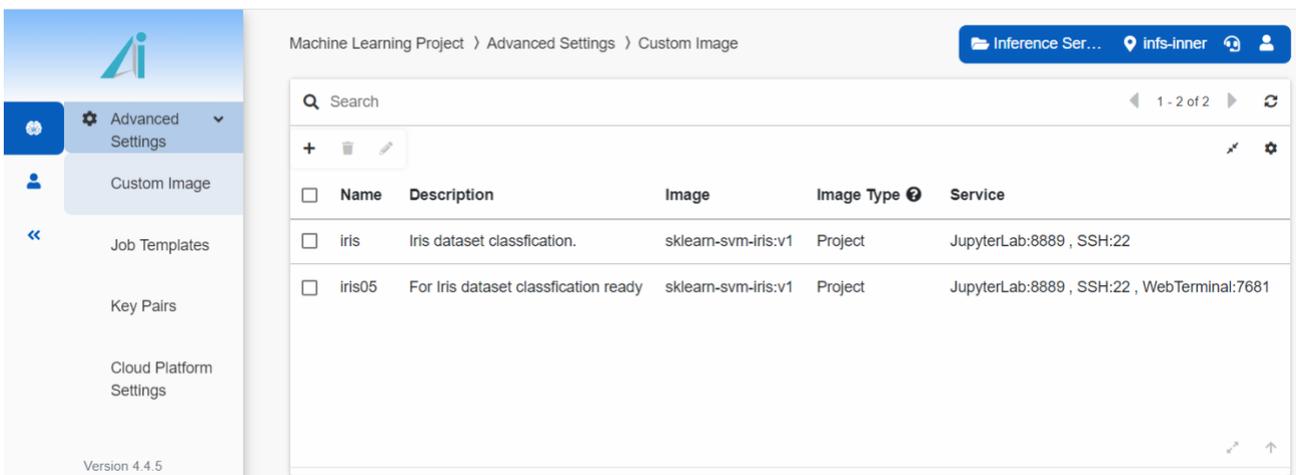
Project
Private
Project
Personal: Only you can use the image.

Service*

JupyterLab Port: 8889 Protocol: http Description:---	SSH Port: 22 SSH Account:root SSH login method:By password Description:---	WebTerminal Port: 7681 Protocol: http Description:---
--	---	---

Version 4.4.5

- (7) Confirm the information and click Create.
- (8) New image will be displayed in the custom image list.



Machine Learning Project > Advanced Settings > Custom Image

Inference Ser... | info-inner | [User Icon]

Search

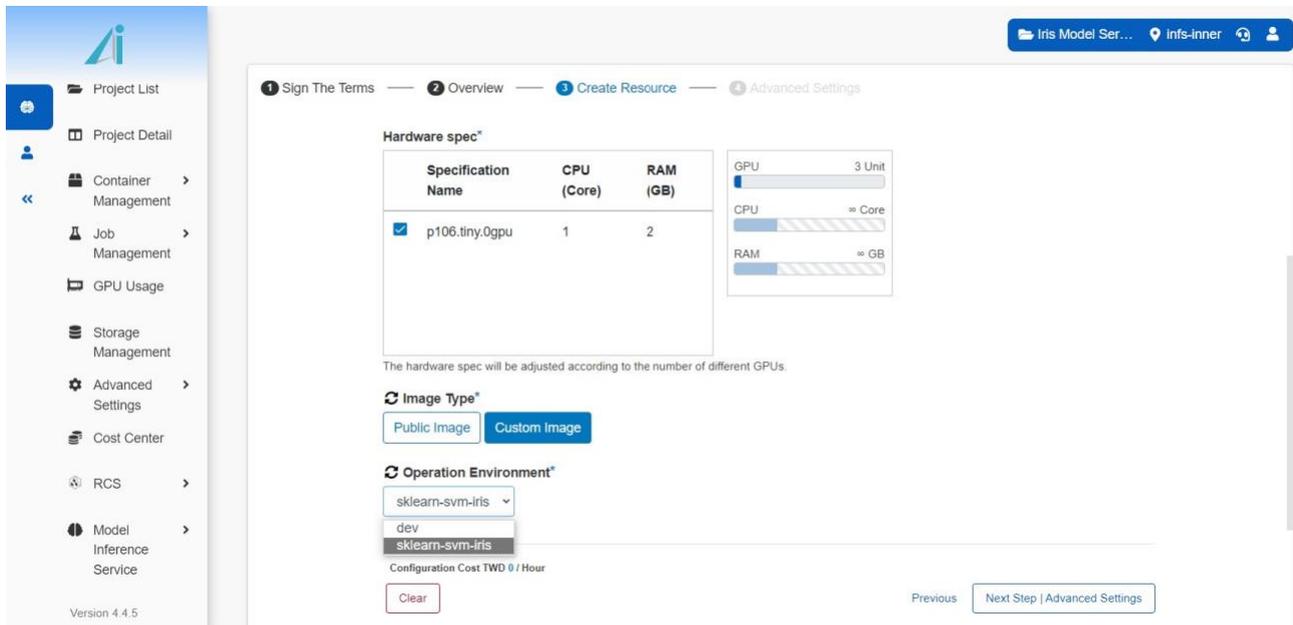
1 - 2 of 2

<input type="checkbox"/>	Name	Description	Image	Image Type ⓘ	Service
<input type="checkbox"/>	iris	Iris dataset classification.	sklearn-svm-iris:v1	Project	JupyterLab:8889 , SSH:22
<input type="checkbox"/>	iris05	For Iris dataset classification ready	sklearn-svm-iris:v1	Project	JupyterLab:8889 , SSH:22 , WebTerminal:7681

Version 4.4.5

Use Custom Image

- (1) Go to create container page. (Ref. Section 4.1 Create Containers for further details)
- (2) For Image Type, Select custom image.
- (3) All the custom images are shown in the dropdown for the operation environment for you to choose from.



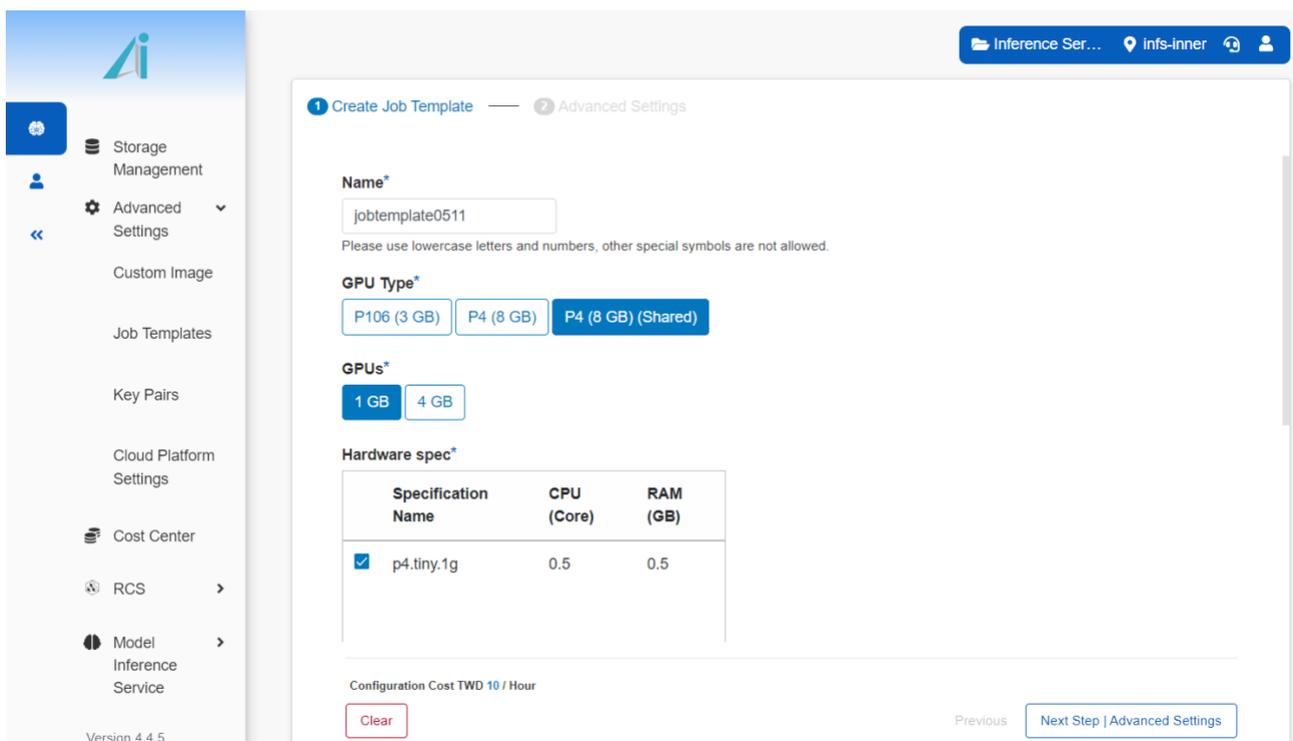
For further steps, please refer to Section 4.1 Container Creation.

8.2 Job Template

One of the ways to schedule tasks is by pre-defining job templates from scheduling settings (*refer to Section 5.4.2 Job Scheduling for more details). By utilizing these pre-defined job templates, jobs can be directly launched and executed.

Create Job Template

- (1) Go to “Advanced Setting > Job templates” from the sidebar
- (2) Click [+] on the top left corner.
- (3) Enter the name, GPU type, GPU pcs, Hardware Spec.



1 Create Job Template — **2 Advanced Settings**

Name*

Please use lowercase letters and numbers, other special symbols are not allowed.

GPU Type*

GPUs*

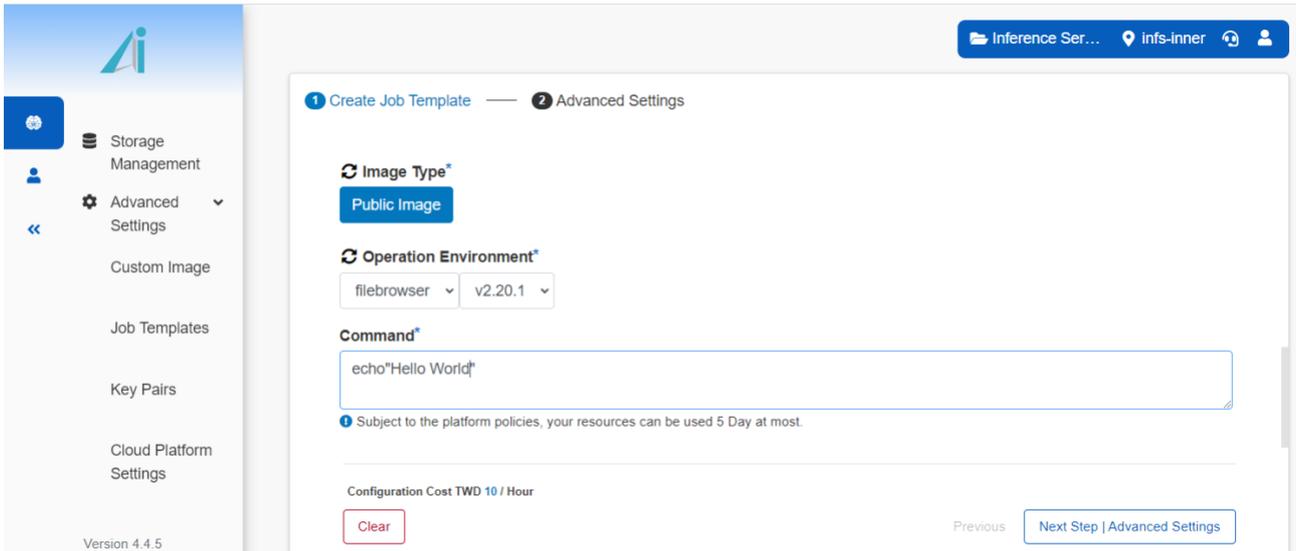
Hardware spec*

Specification Name	CPU (Core)	RAM (GB)
<input checked="" type="checkbox"/> p4.tiny.1g	0.5	0.5

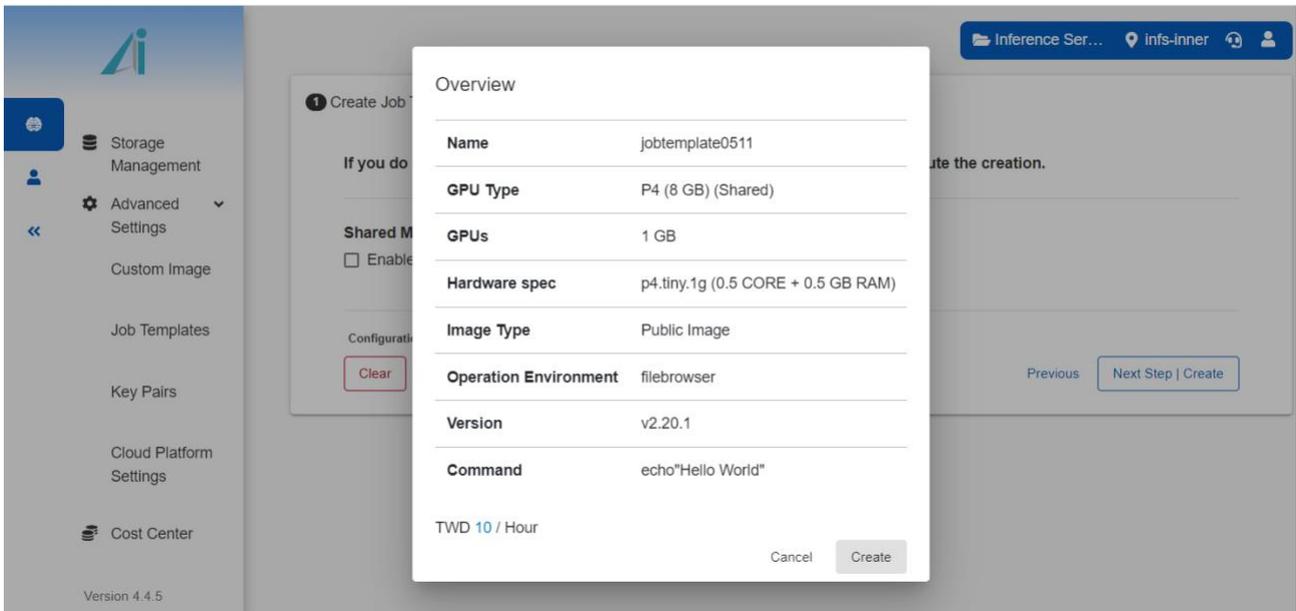
Configuration Cost TWD 10 / Hour

Previous

- (4) Select image type, operation environment and commands.

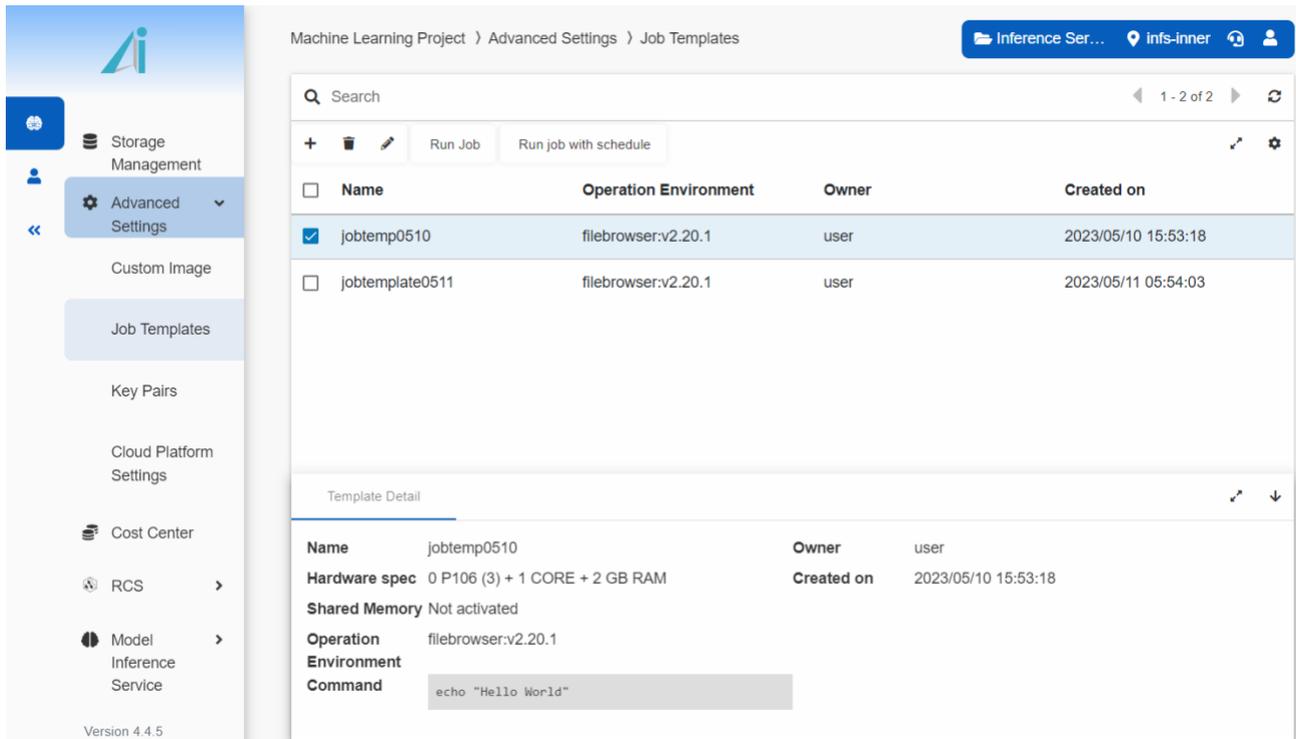


(5) Confirm the entered information on the overview page, click [create] to proceed.



Use Job Template

(1) Select a job template you wish to execute, and click [Run job]



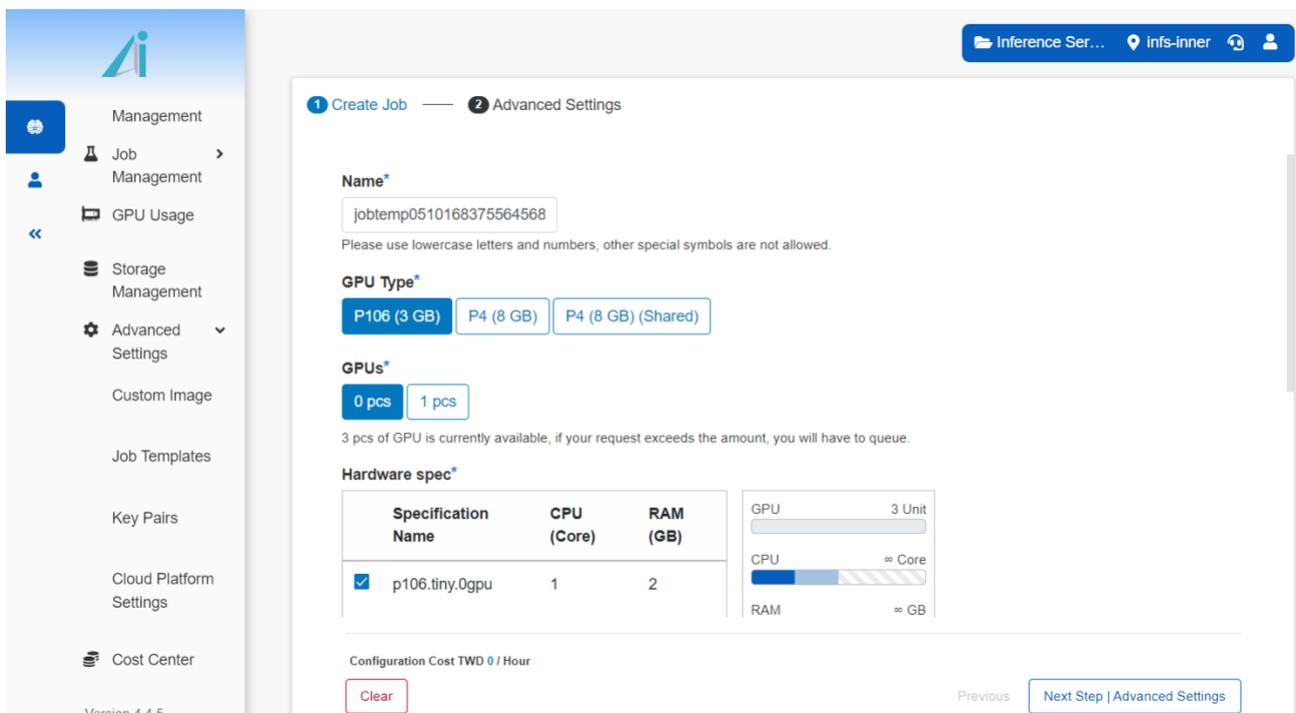
The screenshot shows the 'Job Templates' page in the Inference Service console. The left sidebar contains navigation options like 'Storage Management', 'Advanced Settings', 'Custom Image', 'Job Templates', 'Key Pairs', 'Cloud Platform Settings', 'Cost Center', 'RCS', and 'Model Inference Service'. The main content area displays a table of job templates:

<input type="checkbox"/>	Name	Operation Environment	Owner	Created on
<input checked="" type="checkbox"/>	jobtemp0510	filebrowser:v2.20.1	user	2023/05/10 15:53:18
<input type="checkbox"/>	jobtemplate0511	filebrowser:v2.20.1	user	2023/05/11 05:54:03

Below the table is a 'Template Detail' section for 'jobtemp0510':

- Name:** jobtemp0510
- Owner:** user
- Hardware spec:** 0 P106 (3) + 1 CORE + 2 GB RAM
- Created on:** 2023/05/10 15:53:18
- Shared Memory:** Not activated
- Operation Environment:** filebrowser:v2.20.1
- Command:** echo "Hello World"

(2) The information is automatically filled into the job creation page, where the job name is default to the template name followed by a serial number. You can edit the name at your will.



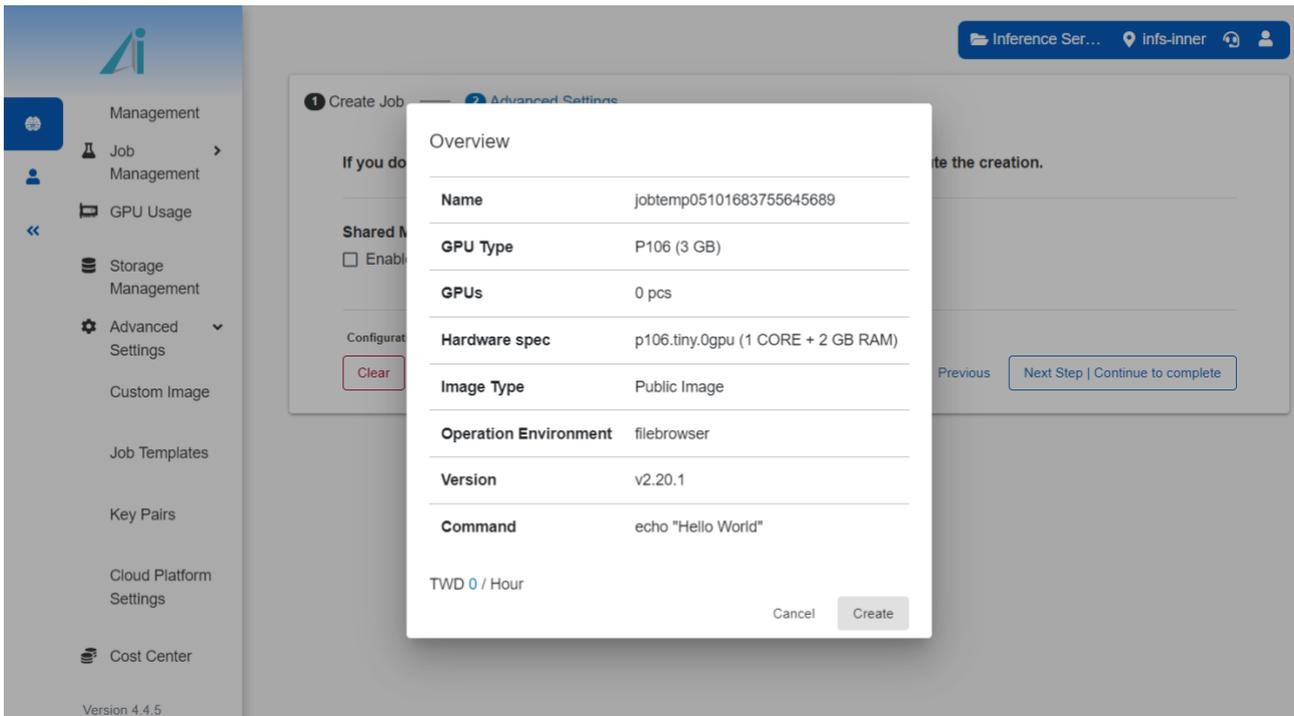
The screenshot shows the 'Advanced Settings' page for creating a job. The left sidebar is the same as in the previous screenshot. The main content area shows the 'Create Job' process with the following settings:

- Name:** jobtemp0510168375564568
- GPU Type:** P106 (3 GB) (selected), P4 (8 GB), P4 (8 GB) (Shared)
- GPUs:** 0 pcs, 1 pcs (selected)
- Hardware spec:**

Specification Name	CPU (Core)	RAM (GB)
<input checked="" type="checkbox"/> p106.tiny.0gpu	1	2
- GPU Resource Usage:** GPU: 3 Unit, CPU: ∞ Core, RAM: ∞ GB
- Configuration Cost:** TWD 0 / Hour
- Buttons:** Clear, Previous, Next Step | Advanced Settings

(3) Click [Next | Advanced Setting]

(4) Confirm the information in the overview and click Create to proceed.



(5) You will be re-directed to the job list after the creation, and the job you just created will be under the "Job creating" status.

Machine Learning Project > Job Management > Job List
Inference Ser... | info-inner



- [Project Review List](#)
- [Project List](#)
- [Project Detail](#)
- [Container Management](#)
- [Job Management](#)
- [Job List](#)
- [Job Scheduling](#)
- [GPU Usage](#)
- [Storage Management](#)
- [Advanced Settings](#)
- [Cost Center](#)

Version 4.4.5

1 - 3 of 3

<input type="checkbox"/>	Name	Operation Envir...	Owner	Created on	Create Type	Status
<input type="checkbox"/>	jobtest01	tensorflow:21.06-...	hunterhsu	2023/04/28 20:36...	Immediately	Succeeded
<input type="checkbox"/>	schjob02-jobtemp...	filebrowser:v2.20.1	user	---	Schedule	Fail to execute
<input checked="" type="checkbox"/>	jobtemp0510168...	filebrowser:v2.20.1	user	---	Immediately	Job Creating

Job Detail
Log
Event

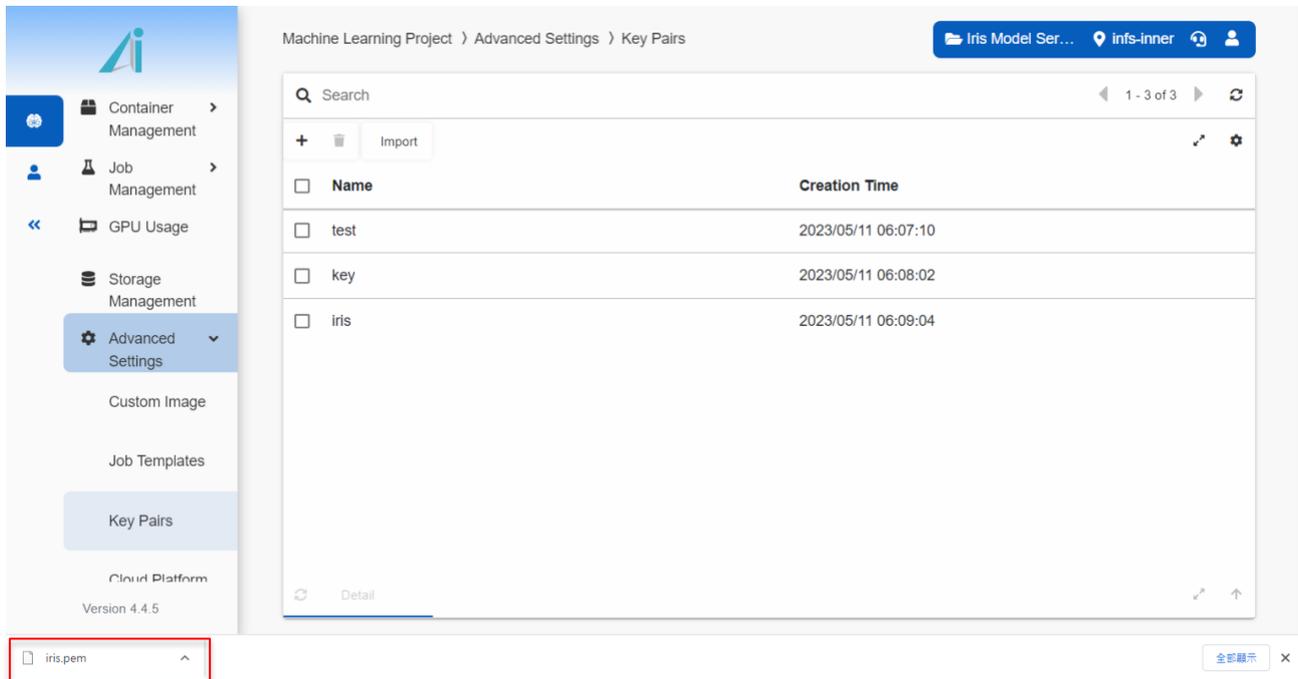
Name	jobtemp05101683755645689	Owner	user
Hardware spec	0 P106 (3) + 1 CORE + 2 GB RAM	Created on	---
Shared Memory	Not activated	Create Type	Immediately
Operation Environment	filebrowser:v2.20.1	Status	Job Creating
Command	echo "Hello World"		

8.3 Key Pair

This feature allows the users a place to manage all his key pairs.

Create Key Pair

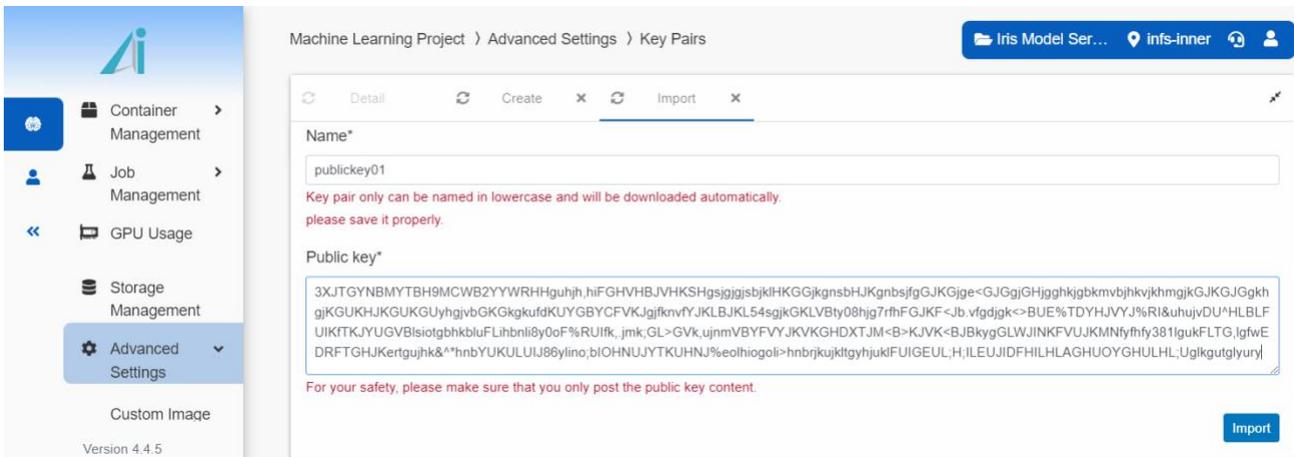
The key pair created by the users are for SSH connection for MLS services. Go to “Advanced Settings > Key Pair” and create a new key pair by clicking on the [+] on the top left and download the private key to the computer of the users.



※ Private keys need to be securely stored, as losing them will result in the inability to access services that require the corresponding public key. Once lost, the private key cannot be recovered, and the only option is to create a new key.

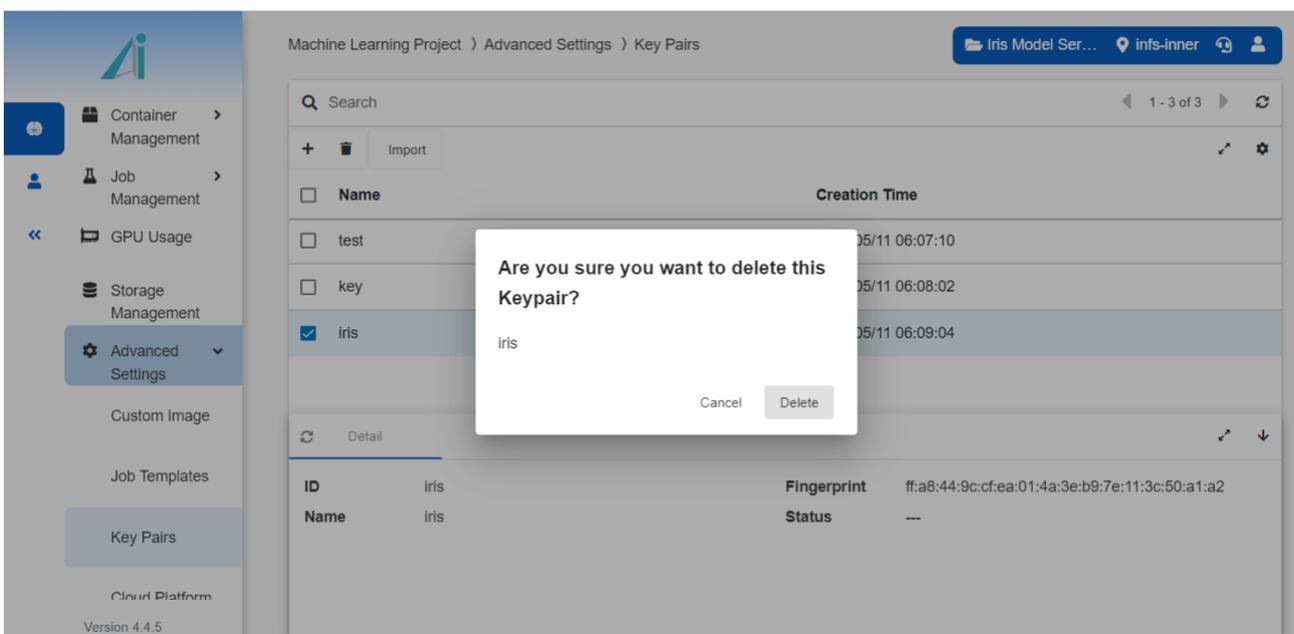
Import Keys

If you have a preferred key that you regularly use, you can click on [Import] to import the key into the system. Fill in the name and the corresponding public key of the key. For security reasons, please ensure the imported key is a public key. Once imported, you can use the key within this system.



Delete Keys

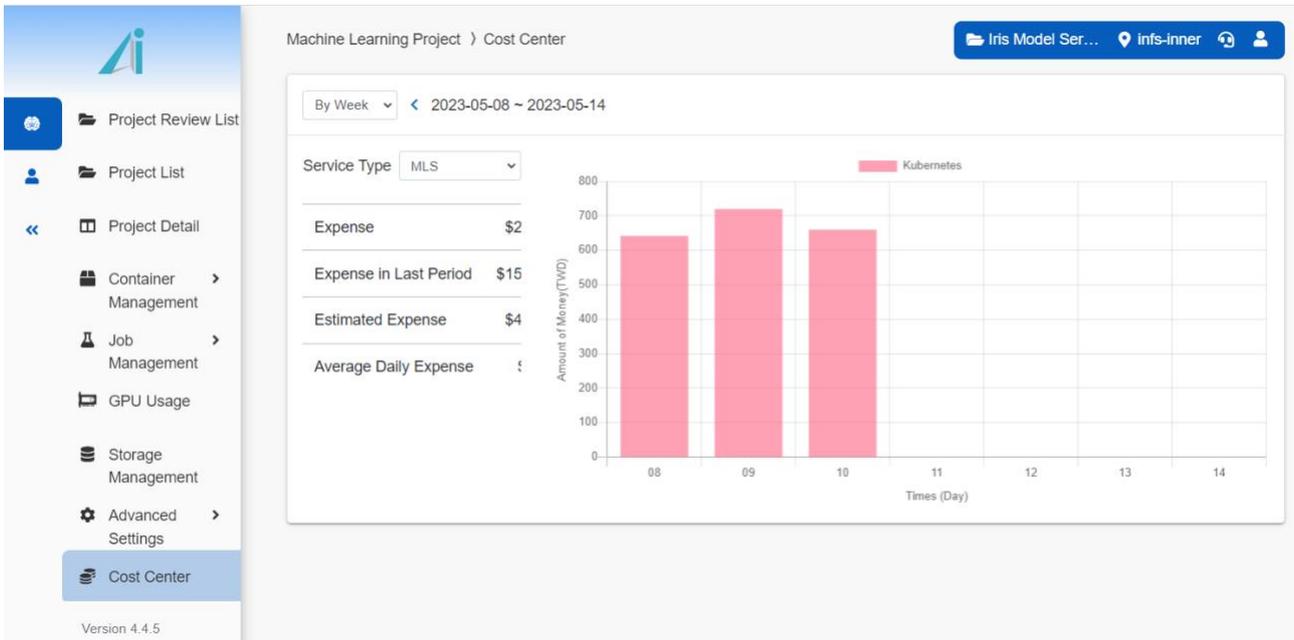
To delete a key, select the key you wish to delete from the list and click on the icon . A confirmation will appear, as shown in the following figure. If everything is correct, click [Delete] to proceed with the deletion.



9. Cost Center

The platform provides project-based cost statistics, which are calculated based on the usage duration of each container (measured by hours) multiplied by the hourly cost of the ML S specifications of each container

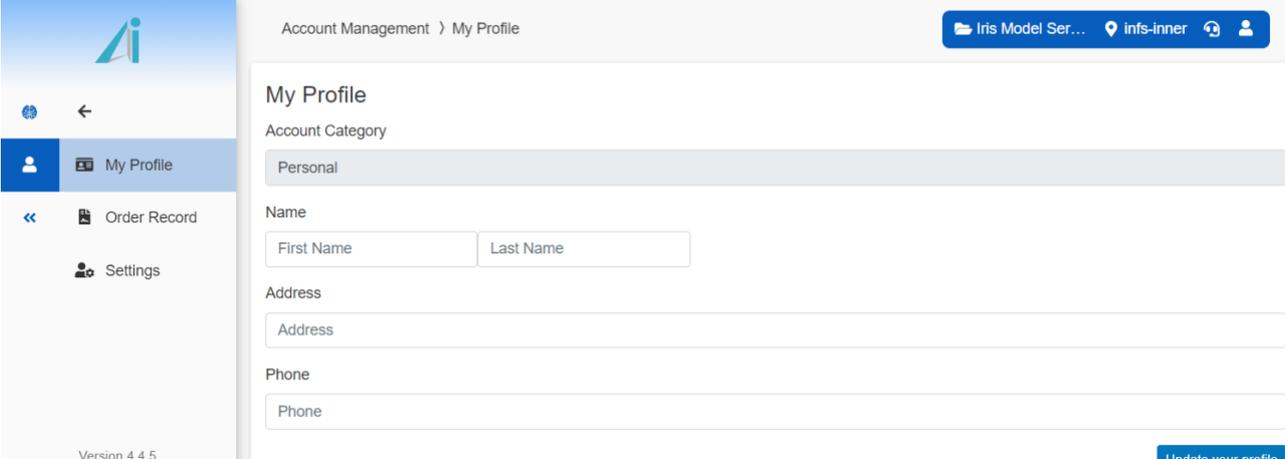
*The hourly cost of each container is set up by the platform administrators from the admin portal.



6. Account Management

1. General Information

Information of the user account is maintained in this feature. This includes name, contact information.



Account Management > My Profile

Iris Model Ser... infs-inner

My Profile

Account Category

Personal

Name

First Name Last Name

Address

Address

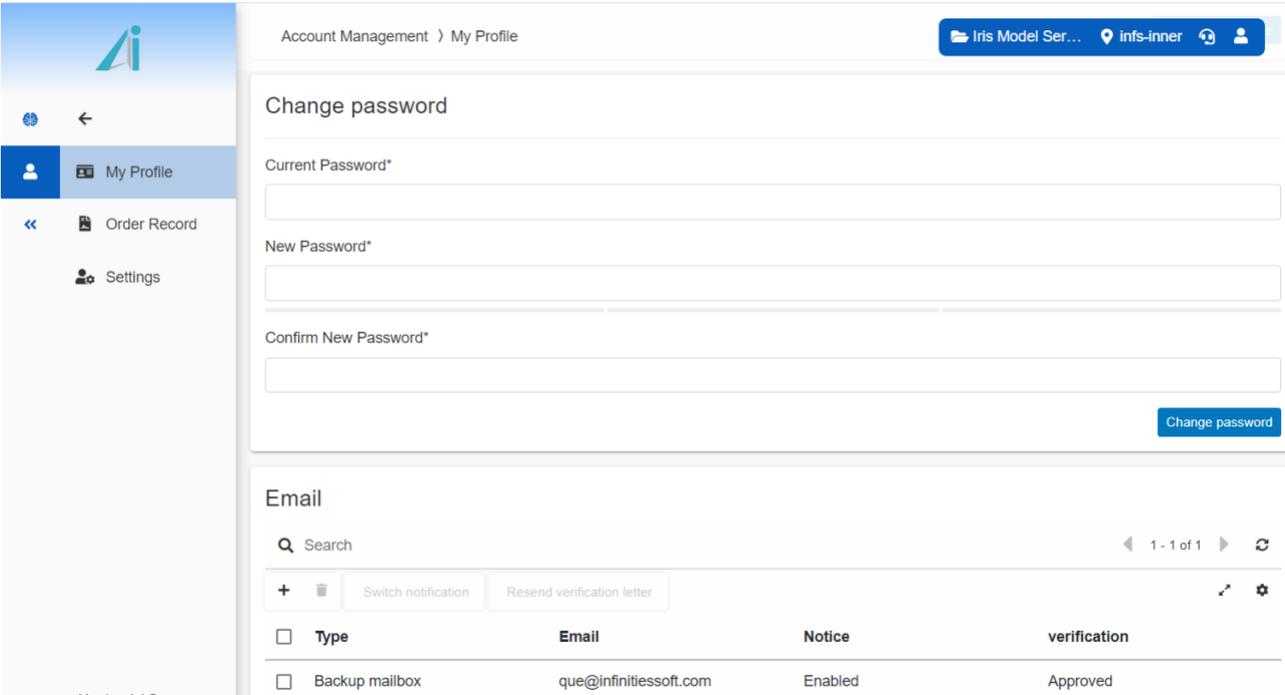
Phone

Phone

Update your profile

Version 4.4.5

Other features include changing your password and email address, turn on/ off notifications and resending confirmation emails.



Account Management > My Profile

Iris Model Ser... infs-inner

Change password

Current Password*

New Password*

Confirm New Password*

Change password

Email

Search

1 - 1 of 1

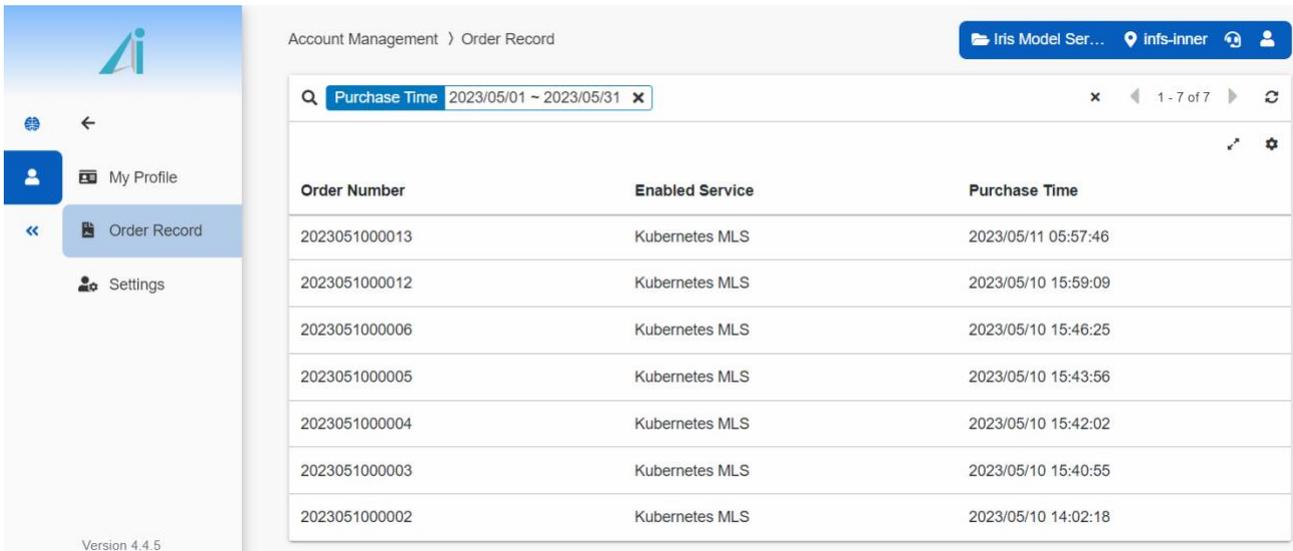
Switch notification Resend verification letter

<input type="checkbox"/>	Type	Email	Notice	verification
<input type="checkbox"/>	Backup mailbox	que@infinitiessoft.com	Enabled	Approved

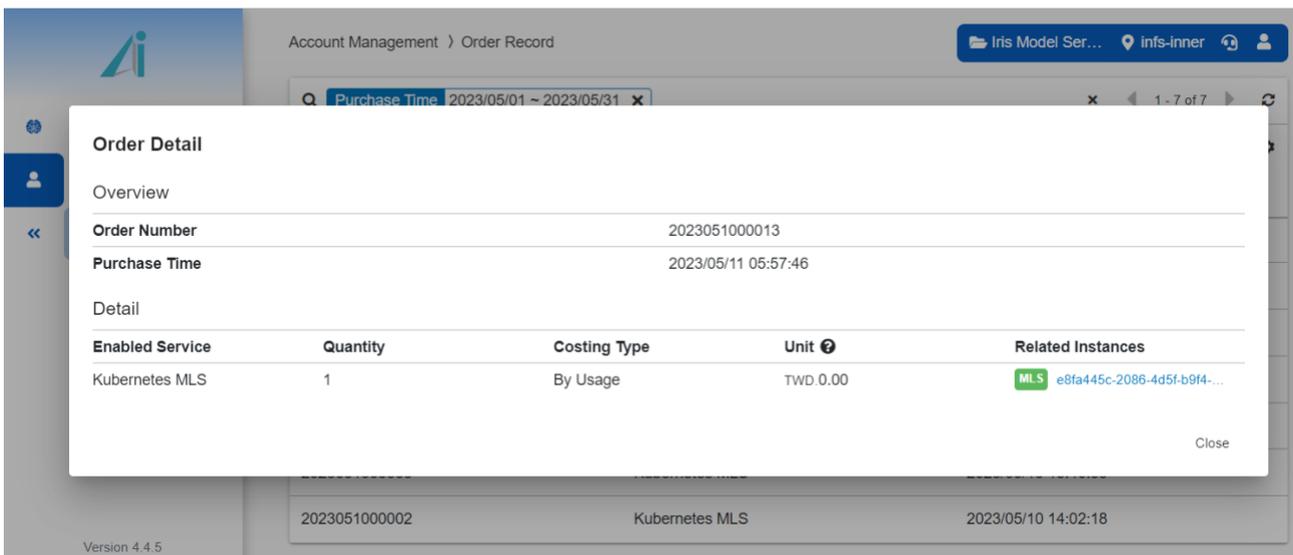
Version 4.4.5

2. Order Record

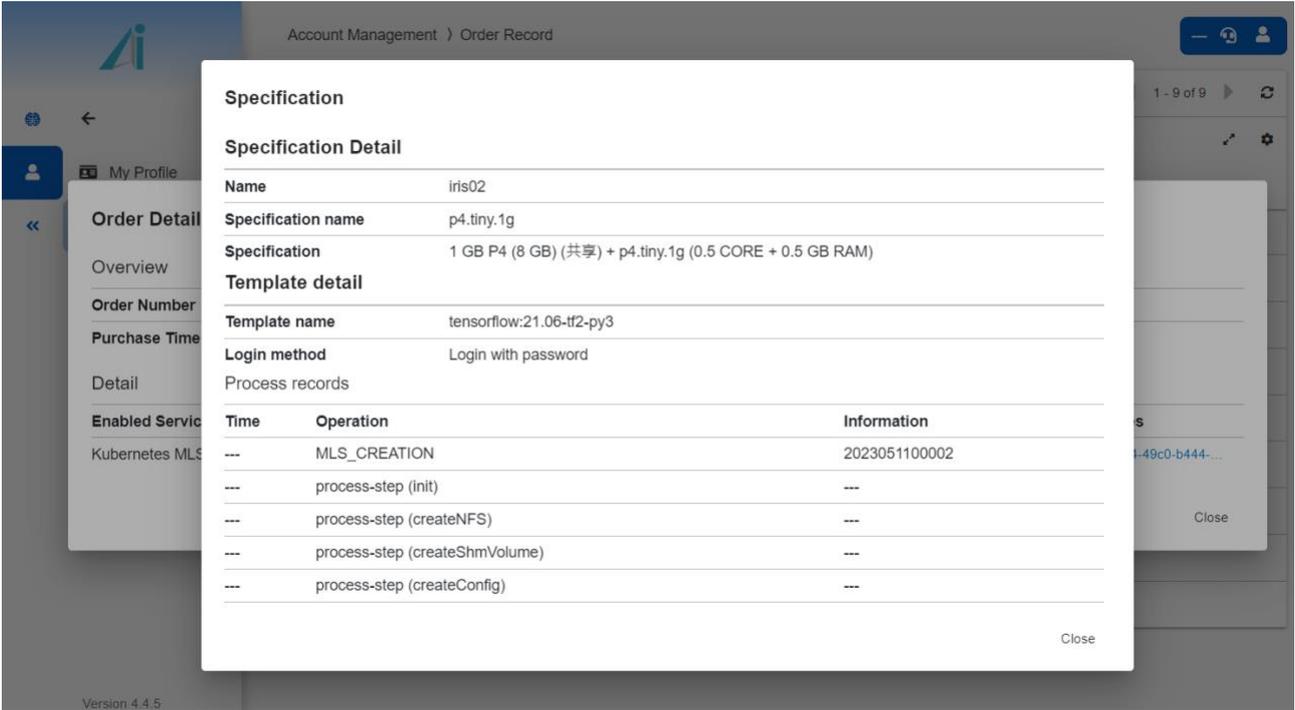
This feature allows each account to view the services created during a period of time.



Detail of an order is available through selecting a period of the [Purchase time] and clicking on a specific order.



The process records during the container creation are available through clicking on the link provided for each instance.



3. Settings

Switch between Light or dark mode.

